

# **Mathematics, Information Technologies and Applied Sciences 2017**

**post-conference proceedings of extended versions  
of selected papers**

**Editors:**

**Jaromír Baštinec and Miroslav Hrubý**

**Brno, Czech Republic, 2017**



**© University of Defence, Brno, 2017**  
**ISBN 978-80-7582-026-6**

## **Aims and target group of the conference:**

The conference **MITAV 2017** should attract in particular teachers of all types of schools and is devoted to the most recent discoveries in mathematics, informatics, and other sciences, as well as to the teaching of these branches at all kinds of schools for any age group, including e-learning and other applications of information technologies in education. The organizers wish to pay attention especially to the education in the areas that are indispensable and highly demanded in contemporary society. The goal of the conference is to create space for the presentation of results achieved in various branches of science and at the same time provide the possibility for meeting and mutual discussions among teachers from different kinds of schools and focus. We also welcome presentations by (diploma and doctoral) students and teachers who are just beginning their careers, as their novel views and approaches are often interesting and stimulating for other participants.

## **Organizers:**

Union of Czech Mathematicians and Physicists, Brno branch (JČMF),  
in co-operation with  
Faculty of Military Technology, University of Defence in Brno,  
Faculty of Science, Faculty of Education and Faculty of Economics and Administration,  
Masaryk University in Brno,  
Faculty of Electrical Engineering and Communication, Brno University of Technology.

## **Venue:**

Club of the University of Defence in Brno, Šumavská 4, Brno, Czech Republic  
June 15 and 16, 2017.

## **Conference languages:**

Czech, Slovak, English

## **Scientific committee:**

Prof. RNDr. Zuzana Došlá, DSc.	Czech Republic Faculty of Science, Masaryk University, Brno
Prof. Irada Ahaievna Dzhalladova, DrSc.	Ukraine Kyiv National Economic Vadym Getman University
Assoc. Prof. Cristina Flaut	Romania Faculty of Mathematics and Computer Science, Ovidius University, Constanta
Assoc. Prof. PaedDr. Tomáš Lengyelfalussy, Ph.D.	Slovakia Dubnica Institute of Technology in Dubnica nad Váhom
Prof. Antonio Maturo	Italy Faculty of Social Sciences of the University of Chieti – Pescara

## **Programme and organizational committee:**

Jaromír Baštinec	Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Mathematics
Luboš Bauer	Masaryk University in Brno, Faculty of Economics and Administration, Department of Applied Mathematics and Informatics
Jaroslav Beránek	Masaryk University in Brno, Faculty of Education, Department of Mathematics
Šárka Hošková-Mayerová	University of Defence in Brno, Faculty of Military Technology, Department of Mathematics and Physics
Miroslav Hrubý	University of Defence in Brno, Faculty of Military Technology, Department of Communication and Information Systems
Milan Jirsa	University of Defence in Brno, Faculty of Military Technology, Department of Communication and Information Systems
Karel Lepka	Masaryk University in Brno, Faculty of Education, Department of Mathematics
Jan Vondra	Masaryk University in Brno, Faculty of Science, Department of Mathematics and Statistics



## Programme of the conference:

*Thursday, June 15, 2017*

- 12:00-13:45 Registration of the participants
- 13:45-14:00 Opening of the conference
- 14:00-14:50 Keynote lecture No. 1 (Dag Hrubý, Czech Republic)
- 14:50-15:10 Break
- 15:10-16:00 Keynote lecture No. 2 (Ľubica Stuchlíková, Slovakia)
- 16:00-16:30 Break
- 16:30-18:00 Presentations of papers
- 18:00-19:00 Conference dinner
- 19:30-21:30 Social event (evening on the boat)

*Friday, June 16, 2017*

- 09:00-09:45 Keynote lecture No. 3 (Vladimír Baláž, Slovakia)
- 09:45-10:00 Break
- 10:00-11:30 Presentations of papers
- 11:30-11:45 Break
- 11:45-13:40 Presentations of papers
- 13:40 Closing

Each MITAV 2017 participant received printed collection of abstracts **MITAV 2017** with ISBN 978-80-7231-417-1. CD supplement of this printed volume contains all the accepted contributions of the conference.

Now, in autumn 2017, this **post-conference CD** was published, containing extended versions of selected MITAV 2017 contributions. The proceedings are published in English and contain extended versions of 28 selected conference papers. Published articles have been chosen from 53 conference papers and every article was once more reviewed.

## Webpage of the MITAV conference:

<http://mitav.unob.cz>

## **Content:**

<b>On Generalized Notion of Convergence by Means of Ideal and Its Applications</b> Vladimír Baláž .....	<b>9-20</b>
<b>Two Classes of Positive Solutions of a Discrete Equation</b> Jaromír Baštinec and Josef Diblík .....	<b>21-32</b>
<b>Metric Spaces and Continuity of Quadratic Function's Iterative Roots</b> Jaroslav Beránek .....	<b>33-42</b>
<b>Geodesic and Almost Geodesic Mappings onto Ricci Symmetric Spaces</b> Volodimir Berezovskii, Patrik Peška and Josef Mikeš .....	<b>43-49</b>
<b>Modifications of Iterative Aggregation – Disaggregation Methods</b> František Bubeník and Petr Mayer .....	<b>50-54</b>
<b>A Problem of Functional Minimizing for Single Delayed Differential System</b> Hanna Demchenko and Josef Diblík .....	<b>55-62</b>
<b>General Solution of Weakly Delayed Linear Systems with Variable Coefficients</b> Josef Diblík and Hana Halfarová .....	<b>63-76</b>
<b>Solving a Higher-Order Linear Discrete Systems</b> Josef Diblík and Kristýna Mencáková .....	<b>77-91</b>
<b>On a Quasilinear PDE Model of Population Dynamics with Random Parameters</b> Irada Dzhalladova and Michael Pokojový .....	<b>92-97</b>
<b>Some Properties of Compositions of Conformal and Geodesic Mappings</b> Irena Hinterleitner .....	<b>98-104</b>
<b>Finding the Spectral Sensitivity of Photodiode with Help of Orthogonal Projection</b> Irena Hlavičková, Martin Motyčka and Jan Škoda .....	<b>105-109</b>
<b>Sensitivity Assessment and Comparison of Maxima Methods in the Estimation of Extremal Index</b> Jan Holešovský .....	<b>110-120</b>
<b>Risk Assessment of Emergency Occurrence at Railway Cargo Transport due to Hazardous Substance Leakage</b> Šárka Hošková-Mayerová .....	<b>121-130</b>
<b>Proposal Mathematical Model for Calculation of Modal and Spectral Properties</b> Petr Hrubý, Tomáš Náhlík and Dana Smetanová .....	<b>131-140</b>
<b>The Intransitive Lie Group Actions with Variable Structure Constants</b> Veronika Chrastinová .....	<b>141-146</b>

<b>An Application of Stochastic Partial Differential Equations to Transmission Line Modelling</b>	
Edita Kolářová and Lubomír Brančík .....	147-150
<b>Priestley-Chao Estimator of Conditional Density</b>	
Kateřina Konečná .....	151-163
<b>3D Printing – Learning and Mastering</b>	
Martin Kopeček, Petr Voda, Pravoslav Stránský and Josef Hanuš .....	164-170
<b>The LMS Moodle and the Moodle Mobile Application in Educational Process of Biophysics</b>	
David Kordek, Martin Kopeček, Kristýna Čáňová, Klára Habartová and Monika Pospíšilová .....	171-176
<b>On the Theorem by Estrada and Kanwal</b>	
Ladislav Mišík .....	177-182
<b>EL-Semihypergroups in which the Quasi-Ordering is not Antisymmetric</b>	
Michal Novák .....	183-192
<b>Comparison of Two Polynomial Calibration Methods</b>	
Petra Ráboňová .....	193-207
<b>Rotary Mappings of Surfaces of Revolution</b>	
Lenka Rýparová and Josef Mikeš .....	208-216
<b>Affine Lagrangians in Second Order Field Theory</b>	
Dana Smetanová .....	217-225
<b>Engineering Education and Science &amp; Technology Popularization among Youngsters Supported by IT</b>	
Ľubica Stuchlíková, Peter Benko, František Janiček, Ondrej Pohorelec and Jiří Hrbáček .....	226-234
<b>Linear Difference Weakly Delayed Systems, the Case of Complex Conjugate Eigenvalues of the Matrix of Non-Delayed Terms</b>	
Jan Šafařík and Josef Diblík .....	235-247
<b>Homothety Curvature Homogeneity</b>	
Alena Vanžurová .....	248-255
<b>Limitation of Sequences of Banach Space through Infinite Matrix</b>	
Tomáš Visnyai .....	256-262

*List of reviewers:*

doc. RNDr. Jaromír Baštinec, CSc.  
doc. RNDr. Jaroslav Beránek, Ph.D.  
prof. RNDr. Josef Diblík, DrSc.  
Ing. Michal Fusek, Ph.D.  
RNDr. Miroslav Hrdý, Ph.D.  
Ing. Miroslav Hrubý, CSc.  
prof. RNDr. Jan Chvalina, DrSc.  
prof. Denys Khusainov, DrSc.  
doc. RNDr. Edita Kolářová, Ph.D.  
RNDr. Karel Lepka, Dr.  
doc. RNDr. Šárka Hošková-Mayerová, Ph.D.  
prof. RNDr. Miroslava Růžičková, Ph.D.  
doc. RNDr. Zdeněk Šmarda, CSc.  
doc. RNDr. Jiří Tomáš, Ph.D.

# On Generalized Notion of Convergence by Means of Ideal and Its Applications

Vladimír Baláž

Faculty of Chemical and Food Technology,  
Slovak University of Technology in Bratislava,  
Radlinského 9, 812 37 Bratislava, Slovak Republic. Email:  
vladimir.balaz@stuba.sk

*Dedicated to the memory of Professor Tibor Šalát (\*1926 – †2005)*

**Abstract:** The starting point of this paper is the notion of  $\mathcal{I}$ –convergence, introduced in this way by the paper [23].  $\mathcal{I}$ –convergence is the natural generalization of the notion of statistical convergence (see [13], [35]) which generalized the notion of classical convergence and has been developed in [4], [5], [11], [16], [21], [22], [24], [37] and [41]. This paper points out the usefulness of  $\mathcal{I}$ –convergence mainly in number theory.

**Keywords:**  $\mathcal{I}$ –convergence, density, sequence, arithmetical function.

## Introduction

The notion of statistical convergence was independently introduced by H. Fast (1951) [13] and I.J. Schoenberg (1959) [35]. The notion of  $\mathcal{I}$ –convergence from the paper [23] corresponds to the natural generalization of statistical convergence (see also [10] where  $\mathcal{I}$ –convergence is defined by means of filter—the dual notion to ideal). These notions have been developed in several directions in [4], [5], [11], [16], [21], [22], [24], [25], [37], [41] and have been used in various parts of mathematics, in particular in number theory, mathematical analysis and ergodic theory, for example [3], [7], [12], [14], [17], [18], [22], [33], [34], [38], [40], [39]. This paper points out the usefulness of  $\mathcal{I}$ –convergence mainly in number theory.

## Statistical convergence and $\mathcal{I}$ -convergence

Both notions, statistical convergence and  $\mathcal{I}$ -convergence are a very natural generalization of the classical convergence, which can be defined as follows.

**Definition 1.** We say that a sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  of real numbers *classical converges* to a real number  $L \in \mathbb{R}$  and we write  $\lim_{n \rightarrow \infty} x_n = L$ , if for each  $\varepsilon > 0$  the set  $A(\varepsilon) = \{n : |x_n - L| \geq \varepsilon\}$  is finite.

This definition is saying that for every  $\varepsilon > 0$  sets  $A(\varepsilon)$  are *small* in some sense, it is clear that it is from cardinality point of view (final sets are small and infinite are big). The notion of statistical convergence is based on the notion of asymptotic density of sets of positive integers  $\mathbb{N}$ .

**Definition 2.** Let  $A \subseteq \mathbb{N}$ . If  $m, n \in \mathbb{N}$ ,  $m \leq n$ , we denote by  $A(m, n)$  the cardinality of the set  $A \cap [m, n]$ . Limits

$$\underline{d}(A) = \liminf_{n \rightarrow \infty} \frac{A(1, n)}{n}, \quad \bar{d}(A) = \limsup_{n \rightarrow \infty} \frac{A(1, n)}{n}$$

are called the *lower* and *upper asymptotic density* of the set  $A$ , respectively. If there exists limit  $\lim_{n \rightarrow \infty} \frac{A(1, n)}{n}$ , then  $d(A) = \underline{d}(A) = \bar{d}(A)$  is said to be the *asymptotic density* of  $A$ .

We recall the definition of the notion of statistical convergence.

**Definition 3.** We say that a sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  of real numbers *statistical converges* to  $L \in \mathbb{R}$  and we write  $\lim\text{-stat } x_n = L$ , if for each  $\varepsilon > 0$  we have  $d(A(\varepsilon)) = 0$ .

Sets  $A(\varepsilon)$  have asymptotic density zero, thus they are also *small* from a different point of view than it was in the Definition 1. Mathematics has several tools how to express that a set is small, e.g. cardinality (finite sets), measure (sets having measure zero) and topology (nowhere dense or sets of the first category). The unifying principle how to express that a subset of  $\mathbb{N}$  is *small* is the notion of an *ideal*  $\mathcal{I}$  of  $\mathbb{N}$ . We say that a set  $A \subseteq \mathbb{N}$  is small set if and only if  $A \in \mathcal{I}$ . Recall the notion of an ideal  $\mathcal{I}$  of subsets of  $\mathbb{N}$ . Let  $\mathcal{I} \subseteq 2^{\mathbb{N}}$ .

$\mathcal{I}$  is called an ideal of subsets of  $\mathbb{N}$ , if  $\mathcal{I}$  is *additive* (if  $A, B \in \mathcal{I}$  then  $A \cup B \in \mathcal{I}$ ) and *hereditary* (if  $A \in \mathcal{I}$  and  $B \subset A$  then  $B \in \mathcal{I}$ ).

An ideal  $\mathcal{I}$  is said to be *non-trivial ideal* if  $\mathcal{I} \neq \emptyset$  and  $\mathbb{N} \notin \mathcal{I}$ . A non-trivial ideal  $\mathcal{I}$  is said to be *admissible ideal* if it contains all finite subsets of  $\mathbb{N}$ .

**Definition 4.** We say that a sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  of real numbers  $\mathcal{I}$ -converges to  $L \in \mathbb{R}$  and we write  $\mathcal{I} - \lim_{n \rightarrow \infty} x_n = L$ , if for each  $\varepsilon > 0$  we have  $A(\varepsilon) \in \mathcal{I}$ .

We shall say that the  $\mathcal{I}$ -convergence is generated by the ideal  $\mathcal{I}$ . In the paper [23], the concept of  $\mathcal{I}$ -convergence is transported in a metric space and there is observed that the basic properties of convergence are preserved also in a metric space. In [24], the concept of  $\mathcal{I}$ -convergence is extended to a topological space. For our purposes, sequences of real numbers are sufficient.

We recall the notion of uniform density.

**Definition 5.** Let  $A(m, n)$  denote the same as in the Definition 2. Put

$$\alpha_s = \min_{n \geq 0} A(n+1, n+s), \quad \alpha^s = \max_{n \geq 0} A(n+1, n+s).$$

The following limits exist  $\underline{u}(A) = \lim_{s \rightarrow \infty} \frac{\alpha_s}{s}$ ,  $\bar{u}(A) = \lim_{s \rightarrow \infty} \frac{\alpha^s}{s}$  and they are called *lower* and *upper uniform density* of the set  $A$ , respectively. If  $\bar{u}(A) = \underline{u}(A)$  then we denote it by  $u(A)$  and it is called the *uniform density* of  $A$ .

It is clear that for each  $A \subseteq \mathbb{N}$  we have

$$\underline{u}(A) \leq \underline{d}(A) \leq \bar{d}(A) \leq \bar{u}(A). \quad (1)$$

If  $\mathcal{I}$  is an admissible ideal then for every sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  of real numbers we have immediately that  $\lim_{n \rightarrow \infty} x_n = L$  implies that  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  also  $\mathcal{I}$ -converges to  $L$ , thus  $\mathcal{I} - \lim_{n \rightarrow \infty} x_n = L$ . The following example shows that the opposite is not true.

**Example 1.**

Let  $\mathbb{P}$  be the set of all primes. Define  $x_n = 1$  for  $n \in \mathbb{P}$  and  $x_n = 0$  otherwise. For the reason that  $u(\mathbb{P}) = 0$  (see [8]), we have that  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  is  $\mathcal{I}_u$ -convergent and also  $\mathcal{I}_d$ -convergent to 0 but it is not convergent. On the basis (1) we have  $\mathcal{I}_u \subsetneq \mathcal{I}_d$  where  $\mathcal{I}_u = \{A \subset \mathbb{N} : u(A) = 0\}$  and  $\mathcal{I}_d = \{A \subset \mathbb{N} : d(A) = 0\}$ . To prove that  $\mathcal{I}_u \neq \mathcal{I}_d$  is enough to take the set  $A = \bigcup_{k=1}^{\infty} \{10^k + 1, 10^k + 2, \dots, 10^k + k\}$  and we have  $d(A) = 0$ ,  $\underline{u}(A) = 0$  and  $\bar{u}(A) = 1$ .

## Examples of $\mathcal{I}$ -convergence

### Example 2.

- a) The class of all finite subsets of  $\mathbb{N}$  forms an admissible ideal usually denoted by  $\mathcal{I}_f$ . Then  $\mathcal{I}_f$ -convergence coincides with the classical convergence.
- b) Let  $\varrho$  be a density function on  $\mathbb{N}$ , the set  $\mathcal{I}_\varrho = \{A \subseteq \mathbb{N} : \varrho(A) = 0\}$  is an admissible ideal. The most commonly used ideals are  $\mathcal{I}_d$ ,  $\mathcal{I}_\delta$ ,  $\mathcal{I}_u$  and  $\mathcal{I}_h$  related to asymptotic, logarithmic, uniform and Alexander density respectively. These ideals generate  $\mathcal{I}_d$ -,  $\mathcal{I}_\delta$ -,  $\mathcal{I}_u$ -, and  $\mathcal{I}_h$ -convergence respectively. The definitions for those densities can be found in [2], [3], [15], [18], [21], [31] and [36].
- c) For an  $q \in (0, 1)$  the set  $\mathcal{I}_c^{(q)} = \{A \subseteq \mathbb{N} : \sum_{a \in A} a^{-q} < +\infty\}$  is an admissible ideal, which generates  $\mathcal{I}_c^{(q)}$ -convergence (see [18], [23]). The ideal  $\mathcal{I}_c^{(1)} = \{A \subseteq \mathbb{N} : \sum_{a \in A} a^{-1} < +\infty\}$  is usually denoted by  $\mathcal{I}_c$ . It is easy to see, that for any  $q_1, q_2 \in (0, 1)$ ,  $q_1 < q_2$  we have

$$\mathcal{I}_f \subsetneq \mathcal{I}_c^{(q_1)} \subsetneq \mathcal{I}_c^{(q_2)} \subsetneq \mathcal{I}_c \subsetneq \mathcal{I}_d \subsetneq \mathcal{I}_\delta. \quad (2)$$

- d) A wide class of  $\mathcal{I}$ -convergence can be obtained by means of regular non negative matrixes  $\mathbf{T} = \{t_{n,k}\}_{n,k \in \mathbb{N}}$ . For  $A \subset \mathbb{N}$  we put  $d_{\mathbf{T}}^{(n)}(A) = \sum_{k=1}^{\infty} t_{n,k} \chi_A(k)$  for  $n \in \mathbb{N}$  where  $\chi_A$  is the characteristic function of  $A$ . If  $\lim_{n \rightarrow \infty} d_{\mathbf{T}}^{(n)}(A) = d_{\mathbf{T}}(A)$  exists, then  $d_{\mathbf{T}}(A)$  is called  $\mathbf{T}$ -density of  $A$  (see [27], [23]). Put  $\mathcal{I}_{d_{\mathbf{T}}} = \{A \subset \mathbb{N} : d_{\mathbf{T}}(A) = 0\}$ . Then  $\mathcal{I}_{d_{\mathbf{T}}}$  is a non-trivial ideal and  $\mathcal{I}_{d_{\mathbf{T}}}$ -convergence contains both  $\mathcal{I}_d$ - and  $\mathcal{I}_\delta$ -convergence. For the matrix  $\mathbf{T} = \{t_{n,k}\}_{n,k \in \mathbb{N}}$  where  $t_{n,k} = \frac{\varphi(k)}{n}$  for  $k \leq n$ ,  $k \mid n$  and  $t_{n,k} = 0$  otherwise we obtain  $\varphi$ -convergence of Schoenberg (see [35]), where  $\varphi$  is Euler function.
- e) Let  $\mu$  be a finitely additive measure on  $\mathbb{N}$ . The family  $\mathcal{I}_\mu = \{A \subseteq \mathbb{N} : \mu(A) = 0\}$  is a non-trivial ideal generates  $\mathcal{I}_\mu$ -convergence. In the case if  $\mu$  is the Buck measure density ((see [9], [31]),  $\mathcal{I}_\mu$  is an admissible ideal and  $\mathcal{I}_\mu \subsetneq \mathcal{I}_d$ .
- f) Let  $\mathbb{N} = \bigcup_{j=1}^{\infty} D_j$  be a decomposition on  $\mathbb{N}$  (i.e.  $D_k \cap D_l = \emptyset$  for  $k \neq l$ ). Assume that  $D_j$  ( $j = 1, 2, \dots$ ) are infinite sets (e.g. we can choose  $D_j = \{2^{j-1} \cdot (2s-1) : s \in \mathbb{N}\}$  for  $j = 1, 2, \dots$ ). Denote  $\mathcal{I}_{\mathbb{N}}$  the class of all



$A \subseteq \mathbb{N}$  such that  $A$  intersects only a finite number of  $D_j$ . Then  $\mathcal{I}_{\mathbb{N}}$  is an admissible ideal, which generates  $\mathcal{I}_{\mathbb{N}}$ -convergence.

The fact  $\mathcal{I}_c \subsetneq \mathcal{I}_d$  follows from the following result in the paper [32]. Let  $A \subseteq \mathbb{N}$  and  $\sum_{a \in A} \frac{1}{a} < \infty$  then  $d(A) = 0$ . The opposite is not true as it shows Example 1. For the set of primes  $\mathbb{P}$ , we have  $d(\mathbb{P}) = 0$  but  $\sum_{p \in \mathbb{P}} \frac{1}{p} = \infty$ . Thus  $\mathcal{I}_c \neq \mathcal{I}_d$ .

The following example shows that for any  $q_1, q_2 \in (0, 1]$ ,  $q_1 < q_2$  we have  $\mathcal{I}_c^{(q_1)} \subsetneq \mathcal{I}_c^{(q_2)}$ .

**Example 3.**

Define the sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  as follows:  $x_n = 1$  for  $n = k^{q_1}$  and  $x_n = 0$  otherwise. Then  $\mathcal{I}_{q_2} - \lim_{n \rightarrow \infty} x_n = 0$  but  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  is not  $\mathcal{I}_{q_1}$ -convergent.

It is easy to prove the following lemma.

**Lemma 1.** If  $\mathcal{I}_1 \subseteq \mathcal{I}_2$  then the statement  $\mathcal{I}_1 - \lim_{n \rightarrow \infty} x_n = L$  implies  $\mathcal{I}_2 - \lim_{n \rightarrow \infty} x_n = L$ .

On the basis of Lema 1. if we examine the generalized convergence of the sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$ , it is interesting to find the smallest element (if such exists) in the class of all ideals  $\mathcal{I}$  (partially ordered by inclusion) for which the sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  is  $\mathcal{I}$ -convergent.

## Applications

a) **Normal order**

Recall the notion of normal order

**Definition 6.** The sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  has the *normal order*  $\mathbf{y} = (y_n)_{n=1}^{\infty}$  if for every  $\varepsilon > 0$  and almost all (almost all in the sense of asymptotic density) values  $n$  we have  $(1 - \varepsilon)y_n < x_n < (1 + \varepsilon)y_n$ .

Let  $n > 1$  be an integer with its canonical representation

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k}.$$

Put

$\omega(n)$  – the number of distinct prime factors of  $n$ , thus  $\omega(n) = k$   
 $\Omega(n)$  – the total number of prime factors of  $n$ , thus  $\Omega(n) = \alpha_1 + \alpha_2 + \dots + \alpha_k$   
 $d(n)$  – the number of divisors of  $n$ , thus  $d(n) = \sum_{d|n, d>0} 1$ .

In the paper [20] we can find that the normal order of  $\omega(n)$  is  $\ln \ln n$ . Further the normal order of  $\Omega(n)$  is also  $\ln \ln n$  and the normal order of  $\ln d(n)$  is  $\ln 2 \ln \ln n$ . for more examples of normal orders see [20], [26] and [36].

Authors of the paper [34] pointed out that one of the equivalent definition of the notion of normal order is as follows (for equivalent definitions see [36]): The sequence  $\mathbf{x} = (x_n)_{n=1}^{\infty}$  has the normal order  $\mathbf{y} = (y_n)_{n=1}^{\infty}$  if and only if  $\mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{x_n}{y_n} = 1$ .

Directly from this definition we have that sequences  $(\frac{\omega(n)}{\ln \ln n})_{n=2}^{\infty}$ ,  $(\frac{\Omega(n)}{\ln \ln n})_{n=2}^{\infty}$  and  $(\frac{\ln d(n)}{\ln 2 \ln \ln n})_{n=2}^{\infty}$  are statistically convergent to 1. Thus

$$\mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{\omega(n)}{\ln \ln n} = \mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{\Omega(n)}{\ln \ln n} = \mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{\ln d(n)}{\ln 2 \ln \ln n} = 1.$$

In [6] it is proved that these sequences are not  $\mathcal{I}_c^{(q)}$ –convergent for all  $q \in (0, 1]$ .

### b) Pascal's triangle

The  $n$ -th row of Pascal's triangle consists of the numbers  $\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n-1}, \binom{n}{n}$ . Their sum equals to  $2^n = (1 + 1)^n = \sum_{k=0}^n \binom{n}{k}$ . Let  $\Gamma(t)$  denote the number of times the positive integer  $t$ ,  $t > 2$  occurs in Pascal's triangle. That is,  $\Gamma(t)$  is the number of binomial coefficients  $\binom{n}{k}$  satisfying  $\binom{n}{k} = t$ .  $\Gamma(t) \geq 1$ . Consider that every binomial coefficient  $t = \binom{n}{2}$ ,  $n \geq 4$  occurs in Pascal's triangle at least 4 times  $(\binom{n}{2}, \binom{n}{n-2}, \binom{t}{1}, \binom{t}{t-1})$ . In [1] it is proved that the average value (recall that *average value* of  $\Gamma(t)$  is defined as  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=2}^{n-1} \Gamma(t)$ ) and normal order of  $\Gamma(t)$  is 2. Thus we have that  $\mathcal{I}_d - \lim_{t \rightarrow \infty} \Gamma(t) = 2$ . The paper [19] shows that for every  $1 \geq q > \frac{1}{2}$  we have also  $\mathcal{I}_c^{(q)} - \lim_{t \rightarrow \infty} \Gamma(t) = 2$  but it does not hold for any  $q$ ,  $0 < q \leq \frac{1}{2}$ .

### c) Niven's functions

Let  $n > 1$  be an integer with its canonical representation

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k}.$$

Put  $H(n) = \max\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ ,  $h(n) = \min\{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and  $H(1) = 1$ ,  $h(1) = 1$  (so called Niven's functions). In [29] there is proved that the average value of  $h(n)$  is 1 thus  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n h(t) = 1$  and  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n H(t) = C$ , where  $C$  is Niven's constant given by the formula  $C = 1 + \sum_{j=2}^{\infty} (1 - \frac{1}{\zeta(j)}) = 1.705211\dots$ , where  $\zeta(k) = \sum_{n=1}^{\infty} \frac{1}{n^k}$  is the Riemann zeta function. In [34] it is proved that the sequences  $(\frac{h(n)}{\ln n})_{n=2}^{\infty}$  and  $(\frac{H(n)}{\ln n})_{n=2}^{\infty}$  are dense on interval  $(0, \frac{1}{\ln 2})$  and both are  $\mathcal{I}_d$ -convergent to zero. In [6] functions  $H(n)$  and  $h(n)$  were studied again and authors proved that the situation is a bit different for  $\mathcal{I}_c^{(q)}$ -convergence of the sequences  $(\frac{h(n)}{\ln n})_{n=2}^{\infty}$ ,  $(\frac{H(n)}{\ln n})_{n=2}^{\infty}$ . The sequence  $(\frac{h(n)}{\ln n})_{n=2}^{\infty}$  is  $\mathcal{I}_c^{(q)}$ -convergent for every  $q \in (0, 1]$  and  $(\frac{H(n)}{\ln n})_{n=2}^{\infty}$  is  $\mathcal{I}_c^{(q)}$ -convergent only for  $q = 1$ , thus it is  $\mathcal{I}_c$ -convergent and it is not  $\mathcal{I}_c^{(q)}$ -convergent for any  $q \in (0, 1)$ .

d) **Functions  $f(n)$  and  $f^*(n)$**

Let  $n > 1$  be an integer with its canonical representation

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k}.$$

Put  $f(n)$  – the product over the divisors  $d$  of  $n$ , thus  $f(n) = \prod_{d|n, d>0} d$  and  $f^*(n) = \frac{f(n)}{n}$ , these functions were introduced by J. Mycielski in [28]. In the paper [40] it is proved the following equality:

$$\mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{\ln \ln f(n)}{\ln \ln n} = \mathcal{I}_d - \lim_{n \rightarrow \infty} \frac{\ln \ln f^*(n)}{\ln \ln n} = 1 + \ln 2.$$

The sequences  $(\frac{\ln \ln f(n)}{\ln \ln n})_{n=2}^{\infty}$  and  $(\frac{\ln \ln f^*(n)}{\ln \ln n})_{n=2}^{\infty}$  are not  $\mathcal{I}_c^{(q)}$ -convergent for all  $q \in (0, 1]$  (see [6]).

e) **Function  $a_p(n)$**

The function  $a_p(n)$  is defined as follows:  $a_p(1) = 0$  and if  $n > 0$ , then  $a_p(n)$  is a unique integer  $j \geq 0$  satisfying  $p^j \mid n$ , but  $p^{j+1} \nmid n$ , thus  $a_p(n) = \max\{j \in \mathbb{N} : p^j \mid n\}$  i. e.,  $p^{a_p(n)} \parallel n$ .

Let us recall the result from [38] there was proved that the sequence  $(\ln p^{\frac{a_p(n)}{\ln n}})_{n=2}^{\infty}$  is  $\mathcal{I}_d$ -convergent to 0. Moreover the sequence  $(\ln p^{\frac{a_p(n)}{\ln n}})_{n=2}^{\infty}$  is  $\mathcal{I}_c^{(q)}$ -convergent to 0 for  $q = 1$  and it is not  $\mathcal{I}_c^{(q)}$ -convergent for all  $q \in (0, 1)$ , this was shown

in [14]. In [3] it was proved that this sequence is also  $\mathcal{I}_u$ -convergent to 0. It is known that  $\mathcal{I}_u \subsetneq \mathcal{I}_d$  but the ideals  $\mathcal{I}_c$  and  $\mathcal{I}_u$  are not disjoint and moreover  $\mathcal{I}_u \not\subseteq \mathcal{I}_c$  and  $\mathcal{I}_c \not\subseteq \mathcal{I}_u$ . For example  $\mathbb{P}$ , the set of all prime numbers belongs to  $\mathcal{I}_u$  but does not belong to  $\mathcal{I}_c$ . On the other hand the set  $A$  (see Example 1.) does not belong to  $\mathcal{I}_u$  but it belongs to  $\mathcal{I}_c$ .

f) **Functions  $\gamma(n)$  and  $\tau(n)$**

In the paper [28] new arithmetical functions were introduced in connection with representation of natural numbers of the form  $n = a^b$ , where  $a, b$  are positive integers. Let

$$n = a_1^{b_1} = a_2^{b_2} = \dots = a_{\gamma(n)}^{b_{\gamma(n)}}$$

be all such representations of given natural number  $n$ , where  $a_i, b_i \in \mathbb{N}$ .

Denote by  $\tau(n) = b_1 + b_2 + \dots + b_{\gamma(n)}$ , ( $n > 1$ ).

It is clear that  $\gamma(n) \geq 1$ , because for any  $n > 1$  there exist representation in the form  $n^1$ . In [14] there is shown that the sequences  $(\gamma(n))_{n=2}^{\infty}$  and  $(\tau(n))_{n=2}^{\infty}$  are  $\mathcal{I}_c^{(q)}$ -convergent to 1 for all  $q \in (\frac{1}{2}, 1]$  and they are not  $\mathcal{I}_c^{(q)}$ -convergent for all  $q \in (0, \frac{1}{2}]$ . As a consequence of the Lemma 1. we have that both sequences are  $\mathcal{I}_d$ -convergent to 1 and so they both have normal order 1.

g) **Olivier's theorem**

The following well-known result was published by L. Olivier in 1827 (see [30]).

If  $\sum_{n=1}^{\infty} a_n < \infty$ ,  $a_n > 0$  and  $a_n \geq a_{n+1}$ , ( $n = 1, 2, \dots$ ) then  $\lim_{n \rightarrow \infty} n a_n = 0$ . Simple example show that without the monotonicity condition  $a_n \geq a_{n+1}$ , ( $n = 1, 2, \dots$ ), the sequence  $(n a_n)_{n=1}^{\infty}$  need not converge to zero. Let us consider  $a_n = \frac{1}{n}$  if  $n = k^2$ , ( $k = 1, 2, \dots$ ) and  $a_n = \frac{1}{n^2}$  otherwise. Then  $a_n > 0$  for all  $n$ ,  $\sum_{n=1}^{\infty} a_n < \infty$  but  $\lim_{n \rightarrow \infty} n a_n \neq 0$ . The authors in [39] dealt with the question as it would be the case if the sequence would converge to zero in some weaker sense. They studied the ideals  $\mathcal{I}$  with the following property:

If  $\sum_{n=1}^{\infty} a_n < \infty$ ,  $a_n > 0$ , ( $n = 1, 2, \dots$ ) then  $\mathcal{I} - \lim_{n \rightarrow \infty} n a_n = 0$ .

They proved that  $\mathcal{I}_c$  is the smallest such ideal.

## References

- [1] Abbot H. L., Erdős P., Hanson D.: On the number of times an integer occurs as a binomial coefficient. *Amer. Math. Monthly*, No-81, 1974. p. 256 - 260.
- [2] Alexander R.: Density and multiplicative structure of sets of integers. *Acta Arithm.*, No-12, 1967. p. 321-332.
- [3] Baláž V.: Remarks on uniform density  $u$ . *Proceedings IAM Workshop on Institute of Information Engineering, Automation and Mathematics*, Slovak University of Technology in Bratislava, 2007. p.43-48.
- [4] Baláž V., Strauch O., Šalát T.: Remarks on several types of convergence of bounded sequences. *Acta Math. Univ. Ostraviensis*, No-14, 2006. p. 3-12.
- [5] Baláž V., Šalát T.: Uniform density  $u$  and corresponding  $\mathcal{I}_u$ -convergence. *Math. Communications*, No-11, 2006. p. 1-7.
- [6] Baláž V., Gogola J., Visnyai T.:  $\mathcal{I}_c^{(q)}$ -convergence of arithmetical functions. *J. Number Theory*, 2017. (in press)
- [7] Baláž V., Červeňanský J., Kostyrko P., Šalát T.:  $\mathcal{I}$ -convergence and  $\mathcal{I}$ -continuity of real functions. *Acta Mathematica*, (Nitra), No-5, 2002. p. 43-50.
- [8] Brown T. C., Freedman A. R.: Arithmetic progresion in lacunary sets. *Mountain J. Math.*, No-17, 1987. p. 587-596.
- [9] Buck R. C.: The measure theoretic approach to density. *Amer. J. Math.*, No-68, 1946. p. 560-580.
- [10] Burbaki N.: *Éléments de Mathématique, Topologie Générale Livre III. (Russian translation) Obščaja topologija Osnovnye struktury*. Nauka, Moskow 1968.
- [11] Connor J. S.: The statistical and strong  $p$ -Cesaro convergence of sequences. *Analysis*, No-8, 1988. p. 47-63.
- [12] Červeňanský J., Masárová R.: Statistical convergence of sequences of real numbers and sequences of real functions. *Proceedings of 10th International Symposium on Mechatronics 2007, Trenčianska univerzita Alexandra Dubčeka*, 2007. p. 253-256.

- [13] Fast H.: Sur la convergence statistique, *Colloquium Mathematicae*, No-2, 1951. p. 241-244.
- [14] Fehér Z., László B., Mačaj M., Šalát T.: Remarks on arithmetical functions  $a_p(n)$ ,  $\gamma(n)$ ,  $\tau(n)$ . *Annales Math. et Informaticae*, No-33, 2006. p. 35-43.
- [15] Freedman A. R., Sember J.J.: Densities and summability. *Pacific Journal of Mathematics*, No-95, 1981. p. 293-305.
- [16] Fridy J. A.: On statistical convergence. *Analysis*, No-5, 1985. p. 301-313.
- [17] Furstenberg H.: *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Princeton 1981.
- [18] Gogola J., Mačaj M., Visnyai T.: On  $\mathcal{I}_c^{(q)}$ -convergence. *Annales Mathematicae et Informaticae*, No-38, 2011. p. 27-36.
- [19] Gubo Š., Mačaj M., Šalát T., Tomanová J.: On binomial coefficients. *Acta Math. (Nitra)*, No-6, 2003. p. 33-42.
- [20] Hardy G. H., Wright E. M.: *An Introduction to the Theory of Numbers*. Clarendon Press, Oxford 1954.
- [21] Kostyrko P., Mačaj M., Šalát T., Sleziak M.:  $\mathcal{I}$ -convergence and extremal  $\mathcal{I}$ -limit poits. *Mathematica Slovaca*, No-55, 2005. p. 443-464.
- [22] Kostyrko P., Mačaj M., Šalát T., Strauch O.: On Statistical limit points. *Proc. of the Amer. Math. Soc.*, No-129, 2001. p. 2647-2654.
- [23] Kostyrko P., Šalát T., Wilczyński W.:  $\mathcal{I}$ -Convergence. *Real Anal. Exchange*, No-26, 2000. p. 669-686.
- [24] Lahiri B. K., Das P.:  $\mathcal{I}$  and  $\mathcal{I}^*$ -Convergence. *Math. Bohem.*, No-2, 2005. p. 153-160.
- [25] Masárová R.: On statistical convergence of functions. *The 1st International Conference on Applied Mathematics and Informatics at Universities 2001, STU Bratislava*, 2001. p. 93-97.
- [26] Mitrinović D. S., Sándor J., Crstici B.: *Handbook of Number Theory, Mathematics and Its Applications*. vol. 351, Kluwer Academic Publishers Group, Dordrecht, Boston, London 1996.

- [27] Miller H.I.: A measure theoretic subsequence characterization of statistical convergence. *Trans. Amer. Math. Soc.*, No-347, 1945. p. 1811-1819.
- [28] Mycielski J.: Sur les représentations des nombres natural par des puissances a base et exposant naturales. *Coll. Math.*, II, 1951. p. 245-260.
- [29] Niven I.: Averages of Exponents in Factoring Integers. *Proc. Amer. Math. Soc.*, No-22, 1969. p. 356-360.
- [30] Olivier L.: Remarques sur les séries infinies et leur convergence. *J. reine angew. Math.*, No-2, 1827. p. 31-44.
- [31] Paštéka M., Šalát T., Visnyai T.: Remarks on Buck's measure density and a generalization of asymptotic density, *Tatra Mountains Mathematical Publications*, No-31, 2005. p. 87-101.
- [32] Powell B. J., Šalát T.: Convergence of subseries of the harmonic series and asymptotic densities of sets of positive integers. *Publ. Inst. Math.(Beograd)*, No-50, 1991. p. 60-70.
- [33] Renling J.: Applications of nonstandard analysis in additive number theory. *Bulletin of Symbolic Logic* , No-6, 2000. p. 331-341.
- [34] Schinzel A., Šalát T.: Remarks on maximum and minimum exponents in factoring. *Math. Slovaca*, No-44, 1994. p. 505-514.
- [35] Schoenberg I. J.: The Integrability of Certain Functions and Related Summability Methods. *The American Mathematical Monthly*, No-66, 1959. p.361-375.
- [36] Strauch O., Porubský Š.: *Distribution of Sequences : A Sampler*. Band 1 Peter Lang, Frankfurt am Main 2005.
- [37] Šalát T.: On statistically convergent sequences of real numbers. *Mathematica Slovaca*, No-30, 1980. p. 139-150.
- [38] Šalát T.: On the function  $a_p, p^{a_p(n)} \parallel n(n > 1)$ . *Mathematica Slovaca*, No-44, 1994. p. 143-151.
- [39] Šalát T., Toma V.: A classical Olivier's theorem and statistical convergence. *Annales Mathematiques Blaise Pascal*, No-10, 2003. p. 305-313.

- [40] Šalát T., Tomanová J.: On the product of divisors of a positive integer. *Math. Slovaca*, No-52, 2002. p. 271-287.
- [41] Šalát T., Visnyai T.: Subadditive measures on  $\mathbb{N}$  and the convergence of series with positive Terms. *Acta Mathematica*, No-6, 2003. p. 43-52.



# TWO CLASSES OF POSITIVE SOLUTIONS OF A DISCRETE EQUATION

**Jaromír Baštinec, Josef Diblík**

Faculty of Electrical Engineering and Communication, Brno University of Technology,  
Technická 10, 616 00 Brno, Czech Republic,  
bastinec@feec.vutbr.cz, diblik@feec.vutbr.cz

**Abstract:** *In the paper we study a class of linear discrete delayed equations with perturbations. Boundaries of perturbations guaranteeing the existence of a positive solution or a bounded vanishing solution of perturbed linear discrete delayed equation are given. In proofs of main results the discrete variant of Ważewski's topological method and method of asymptotic decompositions are utilized.*

**Keywords:** positive solution, discrete delayed equation, perturbation.

## INTRODUCTION

Discrete delayed equations are studied by many authors see e.g. books [1], [16],[17], [21] and papers [2] - [15], [18] - [20], [22].

Denote  $Z_s^q := \{s, s+1, \dots, q\}$  where  $s$  and  $q$  are integers such that  $s \leq q$ . A set  $Z_s^\infty$  is defined similarly.

In the paper the scalar linear discrete equation with delay

$$\Delta x(n) = - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} x(n-k) + \omega(n) \quad (1)$$

is considered where function  $\omega: Z_a^\infty \rightarrow R$  will be more precisely defined later,  $k \geq 1$  is fixed integer,  $n \in Z_a^\infty$ , and  $a$  is a whole number.

We will find below and upper boundaries of perturbation function  $\omega$  in order to give sufficient conditions for (1) to have positive solutions or bounded vanishing solutions.

We prove that there exist two positive solutions  $x = x_1(n)$ ,  $x = x_2(n)$  of the equation (1) defined for  $n \rightarrow \infty$  such that  $x_1(n)$  does not depend linearly on  $x_2(n)$  and

$$\lim_{n \rightarrow +\infty} \frac{x_2(n)}{x_1(n)} = 0. \quad (2)$$

Together with the equation (1) we consider equation without perturbation  $\omega$ :

$$\Delta y(n) = - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} y(n-k). \quad (3)$$

Its well-know, that the equation (3) has two linearly independent positive solutions

$$y_1(n) = \varphi_1(n) = n \left( \frac{k}{k+1} \right)^n \quad (4)$$

and

$$y_2(n) = \varphi_2(n) = \left( \frac{k}{k+1} \right)^n, \quad (5)$$

defined on  $n \in Z_a^\infty$ ,  $a \geq 0$  and satisfying

$$\lim_{n \rightarrow +\infty} \frac{y_2(n)}{y_1(n)} = 0.$$

The paper is organized as follows. In Part 1 necessary auxiliary notions and results are collected. Main results are given in Part 2.

## 1 PRELIMINARY

Let  $b, c: Z_{a-k}^\infty \rightarrow R$  be given functions such that  $b(n) < c(n)$ ,  $n \in Z_{a-k}^\infty$  and

$$\Omega(n) := \{x(n), n \in Z_{a-k}^\infty : b(n) < x < c(n)\}.$$

Let us consider the scalar discrete equation

$$\Delta u(n) = f(n, u(n), u(n-1), \dots, u(n-k)), \quad (6)$$

with  $f: Z_a^\infty \times R^{k+1} \rightarrow R$ .

Below we utilize a nonlinear result concerning the existence of a solution of (6) with the graph remaining in a prescribed set.

**Lemma 1** *Let  $f: Z_a^\infty \times R^{k+1} \rightarrow R$  be continuous with respect to last  $(k+1)$  coordinates. If, moreover, the inequalities*

$$f(n, b(n), u_1, \dots, u_k) - b(n+1) + b(n) < 0, \quad (7)$$

$$f(n, c(n), u_1, \dots, u_k) - c(n+1) + c(n) > 0 \quad (8)$$

*hold for every  $n \in Z_a^\infty$  and every*

$$u_1 \in \Omega(n-1), \dots, u_k \in \Omega(n-k),$$

*then there exists an initial problem*

$$u^*(a-m) = u_m^* \in \Omega(a-m), \quad m = 0, 1, \dots, k$$

*such that the corresponding solution  $u = u^*(n)$  of equation (6) satisfies for every  $n \in Z_{a-k}^\infty$  the inequalities*

$$b(n) < u^*(n) < c(n).$$

This result is proved in [2, 6] by a discrete variant of Ważewski's topological method.

Due to the form of the right-hand side of the linear discrete equation (1) the Lemma 1 implies the following statement.

**Lemma 2** *If inequalities*

$$- \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - b(n+1) + b(n) + \omega(n) < 0, \quad (9)$$

$$- \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - c(n+1) + c(n) + \omega(n) > 0 \quad (10)$$

where

$$b(n-k) < u_k < c(n-k)$$

hold for every  $n \in Z_a^\infty$ , then there exists an initial problem

$$u^*(a-m) = u_m^*, \quad m = 0, 1, \dots, k$$

where

$$b(a-m) < u_m^* < c(a-m)$$

such that the corresponding solution  $u = u^*(n)$  of equation (1) satisfies for every  $n \in Z_{a-k}^\infty$  the inequalities

$$b(n) < u^*(n) < c(n).$$

In the following, the symbols  $O$  and  $o$  mean the Landau order symbols.

**Lemma 3** [15] *Let  $\sigma \in \mathbb{R}$  and  $d \in \mathbb{R}$  be fixed. Then the asymptotic decomposition*

$$\left( 1 + \frac{d}{n} \right)^\sigma = 1 + \frac{\sigma d}{n} + \frac{\sigma(\sigma-1)d^2}{2n^2} + O\left( \frac{1}{n^3} \right) \quad (11)$$

holds for  $n \rightarrow \infty$ .

## 2 MAIN RESULTS

### 2.1 Existence of a positive solution asymptotically equivalent with $\varphi_1(n)$

Let  $p \in (0, 1)$  be fixed and  $\varepsilon, \delta$  are a positive constants. Define

$$b(n) := (n - \varepsilon n^p) \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty, \quad (12)$$

$$c(n) := (n + \delta n^p) \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty. \quad (13)$$

**Theorem 1** *Let  $p \in (0, 1)$ ,  $\varepsilon > 0$  and  $\delta > 0$ . If inequalities (9), (10) with functions  $b$  and  $c$ , defined by formulas (12), (13) hold for every fixed  $n \in Z_a^\infty$  and*

$$\omega(n) = o\left( \frac{1}{n^{2-p}} \left( \frac{k}{k+1} \right)^n \right),$$

then there is an initial problem

$$x_1(a_1 - m) = x_1(-m) \in \Omega(a_1 - m), \quad m = 0, 1, \dots, k$$

where  $a_1 \geq a$  exists, such that the corresponding solution  $x = x_1(n)$  of equation (1) satisfies for every  $n \in Z_{a_1-k}^\infty$  the inequalities

$$(n - \varepsilon n^p) \left( \frac{k}{k+1} \right)^n < x_1(n) < (n + \delta n^p) \left( \frac{k}{k+1} \right)^n$$

and

$$\lim_{n \rightarrow +\infty} \frac{x_1(n)}{\varphi_1(n)} = 1. \quad (14)$$

**Proof:** Without loss of generality assume that the number  $a_1 \geq a$  is sufficiently large and that for  $n \geq a_1 - k$ ,  $b(n) < c(n)$  is valid.

We prove inequality (10). Let

$$\begin{aligned} H &:= - \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - c(n+1) + c(n) + \omega(n) \\ &> - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} c(n-k) - c(n+1) + c(n) + \omega(n) \\ &= - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} [(n-k) + \delta(n-k)^p] \left( \frac{k}{k+1} \right)^{n-k} \\ &\quad - [(n+1) + \delta(n+1)^p] \left( \frac{k}{k+1} \right)^{n+1} + (n + \delta n^p) \left( \frac{k}{k+1} \right)^n + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n \left[ -\frac{1}{k+1} [(n-k) + \delta(n-k)^p] \right. \\ &\quad \left. - [(n+1) + \delta(n+1)^p] \left( \frac{k}{k+1} \right) + (n + \delta n^p) \right] + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n \left( \left[ -\frac{1}{k+1} (n-k) - (n+1) \frac{k}{k+1} + n \right] \right. \\ &\quad \left. - \frac{1}{k+1} \delta(n-k)^p - \frac{k}{k+1} \delta(n+1)^p + \delta n^p \right) + \omega(n) = (*). \end{aligned}$$

The term in square brackets is equal to zero since

$$\left[ -\frac{1}{k+1} (n-k) - (n+1) \frac{k}{k+1} + n \right] = \frac{-(n-k) - k(n+1) + n(k+1)}{k+1} = 0.$$

So we have

$$(*) = \left( \frac{k}{k+1} \right)^n \left( -\frac{\delta(n-k)^p}{k+1} - \frac{k}{k+1} \delta(n+1)^p + \delta n^p \right) + \omega(n) =$$

$$\frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \left( -\frac{(n-k)^p}{n^p} - k \frac{(n+1)^p}{n^p} + k+1 \right) + \omega(n) = (**).$$

Now we use asymptotical decomposition (see Lemma 3), for  $d = -k$  and  $\sigma = p$ :

$$\left( 1 - \frac{k}{n} \right)^p = 1 - \frac{pk}{n} + \frac{p(p-1)k^2}{2n^2} + O\left(\frac{1}{n^3}\right), \quad (15)$$

and for  $d = 1$  and  $\sigma = p$ :

$$\left( 1 + \frac{1}{n} \right)^p = 1 + \frac{p}{n} + \frac{p(p-1)}{2n^2} + O\left(\frac{1}{n^3}\right). \quad (16)$$

After substitution we have

$$\begin{aligned} (**) &= \frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \left[ -\left( 1 - \frac{pk}{n} + \frac{p(p-1)k^2}{2n^2} + O\left(\frac{1}{n^3}\right) \right) \right. \\ &\quad \left. - k \left( 1 + \frac{p}{n} + \frac{p(p-1)}{2n^2} + O\left(\frac{1}{n^3}\right) \right) + k+1 \right] + \omega(n). \\ &= \frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \left[ -1 + \frac{pk}{n} - \frac{p(p-1)k^2}{2n^2} \right. \\ &\quad \left. - k - \frac{pk}{n} - \frac{p(p-1)k}{2n^2} + O\left(\frac{1}{n^3}\right) + k+1 \right] + \omega(n). \\ &= \frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \left[ -\frac{p(p-1)k^2}{2n^2} - \frac{p(p-1)k}{2n^2} + O\left(\frac{1}{n^3}\right) \right] + \omega(n). \\ &= \frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \left[ \frac{p(p-1)}{2n^2} (-k)(k+1) + O\left(\frac{1}{n^3}\right) \right] + \omega(n) = (***) . \end{aligned}$$

The term in square brackets is positive, because  $p \in (0, 1)$  and the difference  $(p-1)$  is negative. Since

$$\omega(n) = o\left(\frac{1}{n^{2-p}} \left( \frac{k}{k+1} \right)^n\right),$$

then

$$(***) \sim \frac{\delta n^p}{k+1} \left( \frac{k}{k+1} \right)^n \frac{p(p-1)}{2n^2} (-k)(k+1) = \delta n^{p-2} \left( \frac{k}{k+1} \right)^n \frac{p(p-1)}{2} (-k) > 0$$

and  $H$  is positive too.

Similarly we can prove that inequality (9) holds for sufficiently large  $n$ .

The limit (14) is obvious, because

$$0 < (n - \varepsilon n^p) \left( \frac{k}{k+1} \right)^n < x_1(n) < (n + \delta n^p) \left( \frac{k}{k+1} \right)^n$$

and

$$\lim_{n \rightarrow +\infty} \frac{x_1(n)}{\varphi_1(n)} \leq \lim_{n \rightarrow +\infty} \frac{(n + \delta n^p) \left( \frac{k}{k+1} \right)^n}{n \left( \frac{k}{k+1} \right)^n} = 1.$$

□

## 2.2 Existence of a bounded solution

New, let  $p \in (0, 1)$  be fixed and  $\varepsilon, \delta$  are a positive constant. Define

$$b(n) := -\varepsilon n^p \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty, \quad (17)$$

$$c(n) := \delta n^p \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty. \quad (18)$$

**Theorem 2** *Let  $p \in (0, 1)$  and  $\varepsilon > 0, \delta > 0$ . If inequalities (9), (10) with functions  $b$  and  $c$ , defined by formulas (17), (18) hold for every fixed  $n \in Z_a^\infty$ ,*

$$\omega(n) = o \left( \frac{1}{n^{2-p}} \left( \frac{k}{k+1} \right)^n \right),$$

*then there is an initial problem*

$$x^*(a_1 - m) = x_{-m}^* \in \Omega(a_1 - m), \quad m = 0, 1, \dots, k$$

*where  $a_1 \geq a$  exists, such that the corresponding solution  $x = x^*(n)$  of equation (1) satisfies for every  $n \in Z_{a_1-k}^\infty$  the inequalities*

$$-\varepsilon n^p \left( \frac{k}{k+1} \right)^n < x^*(n) < \delta n^p \left( \frac{k}{k+1} \right)^n.$$

**Proof:** Without loss of generality assume that the number  $a_1 \geq a$  is sufficiently large and that for  $n \geq a_1 - k, b(n) < c(n)$  is valid.

We prove inequality (10). Let

$$\begin{aligned} G &:= - \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - c(n+1) + c(n) + \omega(n) \\ &> - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} c(n-k) - c(n+1) + c(n) + \omega(n) \\ &= - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} \delta (n-k)^p \left( \frac{k}{k+1} \right)^{n-k} \\ &\quad - \delta (n+1)^p \left( \frac{k}{k+1} \right)^{n+1} + \delta n^p \left( \frac{k}{k+1} \right)^n + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n \left[ -\frac{1}{k+1} \delta (n-k)^p - \delta (n+1)^p \left( \frac{k}{k+1} \right) + \delta n^p \right] + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n n^p \left[ -\frac{\delta}{k+1} \frac{(n-k)^p}{n^p} - \delta \frac{(n+1)^p}{n^p} \frac{k}{k+1} + \delta \right] + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n \delta n^p \left[ -\frac{1}{k+1} \frac{(n-k)^p}{n^p} - \frac{(n+1)^p}{n^p} \frac{k}{k+1} + 1 \right] + \omega(n) \end{aligned}$$

$$= \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \delta n^p \left[ - \left( 1 - \frac{k}{n} \right)^p - k \left( 1 + \frac{1}{n} \right)^p + (k+1) \right] + \omega(n) = (\star).$$

Now we use formulas (15), (16). After substitution into  $(\star)$  we have

$$\begin{aligned} (\star) &= \delta n^p \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \left[ -1 + \frac{pk}{n} - \frac{p(p-1)k^2}{2n^2} - O\left(\frac{1}{n^3}\right) \right. \\ &\quad \left. - k - \frac{pk}{n} - \frac{p(p-1)k}{2n^2} - O\left(\frac{1}{n^3}\right) + (k+1) \right] + \omega(n) \\ &= \delta n^p \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \left[ -\frac{p(p-1)k^2}{2n^2} - \frac{p(p-1)k}{2n^2} + O\left(\frac{1}{n^3}\right) \right] + \omega(n) \\ &= \delta n^p \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \left[ \frac{p(p-1)}{2n^2} (-k^2 - k) + O\left(\frac{1}{n^3}\right) \right] + \omega(n) \\ &= \delta n^p \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \left[ \frac{p(p-1)}{2n^2} (-k)(k+1) + O\left(\frac{1}{n^3}\right) \right] + \omega(n) = (\star\star). \end{aligned}$$

Since for sufficiently large  $n$

$$\omega(n) = o\left(\frac{1}{n^{2-p}} \left(\frac{k}{k+1}\right)^n\right),$$

then

$$(\star\star) \sim \delta n^p \left( \frac{k}{k+1} \right)^n \frac{1}{k+1} \frac{p(p-1)}{2n^2} (-k)(k+1) = \delta n^{p-2} \left( \frac{k}{k+1} \right)^n \frac{p(p-1)}{2} (-k).$$

Because  $p \in (0; 1)$ , then  $(p-1)$  is negative and  $(\star\star)$  is positive. So  $G$  is positive too.

Now we prove the inequality (9) for

$$b(n) := -\varepsilon n^p \left( \frac{k}{k+1} \right)^n.$$

We get

$$\begin{aligned} G_2 &:= - \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - b(n+1) + b(n) + \omega(n) \\ &< - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} b(n-k) - b(n+1) + b(n) + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^k \frac{\varepsilon}{k+1} (n-k)^p \left( \frac{k}{k+1} \right)^{n-k} + \varepsilon (n+1)^p \left( \frac{k}{k+1} \right)^{n+1} \\ &\quad - \varepsilon n^p \left( \frac{k}{k+1} \right)^n + \omega(n) \\ &= \left( \frac{k}{k+1} \right)^n \left[ \frac{\varepsilon}{k+1} (n-k)^p + \frac{\varepsilon k}{k+1} (n+1)^p - \varepsilon n^p \right] + \omega(n) \end{aligned}$$

$$= \left( \frac{k}{k+1} \right)^n \frac{\varepsilon n^p}{k+1} \left[ \frac{(n-k)^p}{n^p} + k \frac{(n+1)^p}{n^p} - k - 1 \right] + \omega(n) = (\star \star \star).$$

Now we use formulas (15), (16). After substitution into  $(\star \star \star)$  we have

$$\begin{aligned} (\star \star \star) &= \left( \frac{k}{k+1} \right)^n \frac{\varepsilon n^p}{k+1} \left[ 1 - \frac{pk}{n} + \frac{p(p-1)k^2}{2n^2} + O\left(\frac{1}{n^3}\right) \right. \\ &\quad \left. + k + \frac{pk}{n} + \frac{p(p-1)k}{2n^2} + O\left(\frac{1}{n^3}\right) - k - 1 \right] + \omega(n) = \\ &= \left( \frac{k}{k+1} \right)^n \frac{\varepsilon n^p}{k+1} \left[ \frac{p(p-1)}{2n^2} k(k+1) + O\left(\frac{1}{n^3}\right) \right] + \omega(n) = (\nabla). \end{aligned}$$

For sufficiently large  $n$  we get

$$(\nabla) \sim \left( \frac{k}{k+1} \right)^n \frac{\varepsilon n^p}{k+1} \frac{p(p-1)}{2n^2} k(k+1) = \left( \frac{k}{k+1} \right)^n \varepsilon n^{p-2} \frac{p(p-1)}{2} k.$$

Because  $p \in (0; 1)$ ,  $(p-1)$  is negative,  $\nabla$  is negative, and  $G_2$  is negative, too.

Since

$$\lim_{n \rightarrow +\infty} n^p \left( \frac{k}{k+1} \right)^n = 0$$

we get

$$\lim_{n \rightarrow +\infty} x^*(n) = 0.$$

□

### 2.3 Existence of a positive solution asymptotically non-comparable with $\varphi_1(n)$

Let  $p \in (0, 1)$  be fixed and let  $\delta > p$  be a positive constant. Define

$$b(n) := 0, \quad n \in Z_{a-k}^\infty, \tag{19}$$

$$c(n) := \delta n^p \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty. \tag{20}$$

**Theorem 3** *Let  $p \in (0, 1)$ ,  $\delta > 0$ . If inequalities (9), (10) with functions  $b$  and  $c$ , defined by formulas (19), (20) hold for every fixed  $n \in Z_a^\infty$ ,  $\omega(n) < 0$ ,*

$$\omega(n) = o\left(\frac{1}{n^{2-p}} \left(\frac{k}{k+1}\right)^n\right),$$

*then there is an initial problem*

$$x_2(a_1 - m) = x_2(-m) \in \Omega(a_1 - m), \quad m = 0, 1, \dots, k$$

*where  $a_1 \geq a$  exists, such that the corresponding solution  $x = x_2(n)$  of equation (1) satisfies for every  $n \in Z_{a_1-k}^\infty$  the inequalities*

$$0 < x_2(n) < \delta n^p \left( \frac{k}{k+1} \right)^n.$$



**Proof:** Verification of inequality (10) can be done by the same scheme as the proof of Theorem 2. Now we prove the inequality (9) for  $b(n) := 0$ . Then

$$\begin{aligned} G_3 &:= - \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - b(n+1) + b(n) + \omega(n) \\ &< - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} b(n-k) - b(n+1) + b(n) + \omega(n) \\ &= \omega(n) < 0. \end{aligned}$$

□

## 2.4 Existence of a positive solution of equation (1) asymptotically non-comparable with $\varphi_1(n)$

New, let  $p, q \in (0, 1)$  be fixed and  $\delta > p$  is a positive constant. Define

$$b(n) := q \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty, \quad (21)$$

$$c(n) := \delta n^p \left( \frac{k}{k+1} \right)^n, \quad n \in Z_{a-k}^\infty. \quad (22)$$

**Theorem 4** *Let  $p, q \in (0, 1)$  and  $\delta > 0$ . If inequalities (9), (10) with functions  $b$  and  $c$ , defined by formulas (21), (22) hold for every fixed  $n \in Z_a^\infty$ ,  $\omega(n) < 0$ ,*

$$\omega(n) = o \left( \frac{1}{n^{2-p}} \left( \frac{k}{k+1} \right)^n \right),$$

*then there is an initial problem*

$$z_2(a_2 - m) = z_2(-m) \in \Omega(a_2 - m), \quad m = 0, 1, \dots, k$$

*where  $a_2 \geq a$  exists, such that the corresponding solution  $z = z_2(n)$  of equation (1) satisfies for every  $n \in Z_{a_2-k}^\infty$  the inequalities*

$$q \left( \frac{k}{k+1} \right)^n < z_2(n) < \delta n^p \left( \frac{k}{k+1} \right)^n.$$

**Proof:** Verification of inequality (10) can be done by the same scheme as the proof of Theorem 2. Now we prove the inequality (9) for  $b(n) := q(k/k+1)^n$ . Then

$$\begin{aligned} G_5 &:= - \left( \frac{k}{k+1} \right)^k \frac{u_k}{k+1} - b(n+1) + b(n) + \omega(n) \\ &< - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} b(n-k) - b(n+1) + b(n) + \omega(n) \end{aligned}$$

$$\begin{aligned}
&= - \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} q \left( \frac{k}{k+1} \right)^{n-k} - q \left( \frac{k}{k+1} \right)^{n+1} \\
&\quad + q \left( \frac{k}{k+1} \right)^n + \omega(n) \\
&= \left( \frac{k}{k+1} \right)^n q \left[ -\frac{1}{k+1} - \frac{k}{k+1} + 1 \right] + \omega(n) \\
&= \left( \frac{k}{k+1} \right)^n q \left[ \frac{-1 - k + k + 1}{k+1} \right] + \omega(n) = \omega(n) < 0.
\end{aligned}$$

□

**Remark 1** It is well-known (we refer, e.g., to [17]) that the equation with constant and positive coefficient  $p$ :

$$\Delta x(n) = -px(n-k)$$

has a positive solution if

$$p \leq c_k$$

and every solution is oscillating for  $n \rightarrow \infty$  if

$$p > c_k$$

where

$$c_k = \left( \frac{k}{k+1} \right)^k \frac{1}{k+1}.$$

Therefore, Theorem 1 gives sufficient conditions for the existence of a positive solution when  $\lim_{n \rightarrow \infty} p(n)$  exists and equals to a critical constant, i.e.,

$$\lim_{n \rightarrow \infty} p(n) = \lim_{n \rightarrow \infty} \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} [1 + \omega(n)] = \left( \frac{k}{k+1} \right)^k \frac{1}{k+1} = c_k.$$

**Remark 2** If Theorem 1 and Theorem 3 hold simultaneously then for solutions  $x_1(n)$  and  $x_2(n)$  relation (2) holds since

$$\lim_{n \rightarrow +\infty} \frac{x_2(n)}{x_1(n)} \leq \lim_{n \rightarrow \infty} \frac{\delta n^p \left( \frac{k}{k+1} \right)^n}{(n - \varepsilon n^p) \left( \frac{k}{k+1} \right)^n} = \lim_{n \rightarrow +\infty} \frac{\delta}{(n^{1-p} - \varepsilon)} = 0.$$

## CONCLUSION

In the paper, sufficient conditions for existence of two classes of eventually positive and asymptotically different solutions or for the existence of bounded solutions of equation (1) are given provided that the perturbation function  $\omega$  satisfies prescribed asymptotic behavior.

## References

- [1] R. P. Agarwal: *Difference Equations and Inequalities. Theory, Methods and Applications*, Second edition, Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker, Inc. New York, 2000.
- [2] Baštinec, J., Diblík, J., Zhang, B.: *Existence of bounded solutions of discrete delayed equations*, Proceedings of the Sixth International Conference on Difference Equations, CRC, Boca Raton, FL, 359–366, 2004.
- [3] J. Čermák, J. Janský: *Stability switches in linear delay difference equations*, Appl. Math. Comput. **243** (2014) 755–766.
- [4] J. Čermák, J. Janský, P. Kunderát: *On necessary and sufficient conditions for the asymptotic stability of higher order linear difference equations*, J. Difference Equ. Appl. **18** (2012) 1781–1800.
- [5] J. Čermák, P. Tomášek: *On delay-dependent stability conditions for a three-term linear difference equation*, Funkcial. Ekvac. **57** (2014) 91–106.
- [6] Diblík, J.: *Asymptotic behavior of solutions of discrete equations*, Functional Differential Equations, **11** (2004), 37–48.
- [7] Diblík, J., Baštinec, J., Morávková, B.: *Oscillation of solutions of a linear second order discrete delayed equation*. In *6. konference o matematice a fyzice na vysokých školách technických s mezinárodní účastí*. Brno, UNOB Brno. 2009. p. 39 - 50. ISBN 978-80-7231-667-0.
- [8] Diblík, J., Baštinec, J., Morávková, B.: *Oscillation of Solutions of a Linear Second Order Discrete Delayed Equation*. In *XXVII International Colloquium on the Management of Educational Process*. Brno, FEM UNOB. 2009. p. 28 - 35. ISBN 978-80-7231-650-2.
- [9] Diblík, J., Hlavičková, I.: *Asymptotic upper and lower estimates of a class of positive solutions of a discrete linear equation with a single delay*. Abstract and Applied Analysis, 2012, ArticleID 764351, 1–12. ISSN: 1085– 3375.
- [10] Diblík, J., Hlavičková, I.: *Asymptotic behavior of solutions of delayed difference equations*. Abstract and Applied Analysis, 2011, Article ID 67196, 1–24. ISSN: 1085–3375.
- [11] Diblík, J., Hlavičková, I.: *Combination of Liapunov and retract methods in the investigation of the asymptotic behavior of solutions of systems of discrete equations*. Dynamic systems and applications, 2009, 18, 3– 4, 507-538. ISSN: 1056–2176.
- [12] Diblík, J., Khusainov, D., Baštinec, J., Sirenko, A.: *Exponential stability of perturbed linear discrete systems*. Advances in Difference Equations, 2016, vol. 2016, no. 2, p. 1-20. ISSN: 1687-1847.
- [13] Diblík, J., Khusainov, D., Baštinec, J., Sirenko, A.: *Exponential stability of linear discrete systems with constant coefficients and single delay*. Applied Mathematics Letters, 2016, vol. 2016, no. 51, p. 68-73. ISSN: 0893-9659.
- [14] J. Diblík, D. Ya. Khusainov: *Representation of solutions of discrete delayed system  $x(k+1) = Ax(k) + Bx(k-m) + f(k)$* , J. Math. Anal. Appl. **318**, 2006, 63–76.
- [15] Diblík, J., Kocsch, N.: *Positive Solutions of the Equation  $x'(t) = -c(t)x(t-\tau)$  in the Critical Case*, *Journal of Mathematical Analysis and Applications*, **250**, 635 - 659 (2000). doi:10.1006/jmaa.2000.7008,
- [16] S. N. Elaydi: *An Introduction to Difference Equations*, Undergraduate Texts in Mathematics, Springer, Third Edition, 2005.
- [17] Györi, I., Ladas, G.: *Oscillation Theory of Delay Differential Equations*, Clarendon Press (1991).

- [18] E. Kaslik: *Stability results for a class of difference systems with delay*, Adv. Difference Equ. **2009** (2009), Article ID 938492, 1–13.
- [19] M.M. Kipnis, R.M. Nigmatullin: *Stability of the trinomial linear difference equations with two delays*, Autom. Remote Control **65** (11) (2004) 1710–1723.
- [20] S.A. Kuruklis: *The asymptotic stability of  $x_{n+1} - ax_n + bx_{n-k} = 0$* , J. Math. Anal. Appl. **188** (1994) 719–731.
- [21] V. Lakshmikantham, Donato Trigiante: *Theory of Difference Equations (Numerical Methods and Applications)*, Marcel Dekker, Second Edition, 2002.
- [22] M. Medved', L. Škripková: *Sufficient conditions for the exponential stability of delay difference equations with linear parts defined by permutable matrices*, Electron. J. Qual. Theory Differ. Equ. **22**, 2012, 1–13.

## Acknowledgement

The work was supported by the project FEKT-S-17-4225 of Brno University of Technology.

# METRIC SPACES AND CONTINUITY OF QUADRATIC FUNCTION'S ITERATIVE ROOTS

Jaroslav Beránek

Faculty of Education, Masaryk University

Poříčí 7, 603 00 Brno, Czech Republic

beranek@ped.muni.cz

**Abstract:** *The article includes one interesting and atypical approach to the continuity of second iterative roots of quadratic function  $q(x) = x^2$ . In the first part of the paper there is mentioned the description of second iterative roots of this quadratic function and the proposition that the set of discontinuous second iterative roots of quadratic function  $q$  is uncountable. In the second part there is constructed quasi-metric  $d$ , so that each second iterative root of quadratic function  $q$  is a continuous mapping of space  $(R, d)$  into itself.*

**Keywords:** Quadratic function, iteration, iterative root, uncountable set, metric space.

## INTRODUCTION

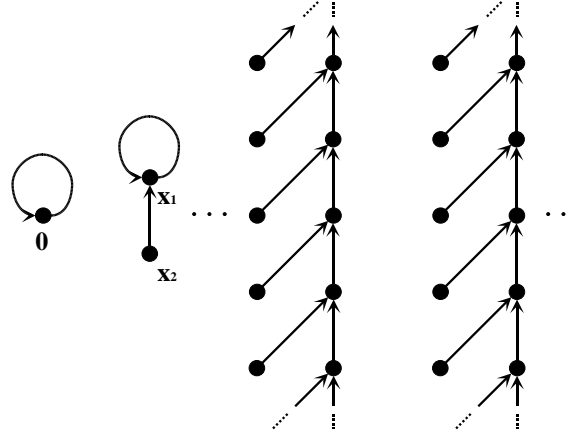
This article is devoted to iterative roots of the simplest real quadratic function  $q(x) = x^2$  and it follows up in the topic the previous author's article published in the conference proceedings MITAV 2016. (See [4]). While examining the existence and properties of real functions' iterative roots, the essential issue is the discrete point of view approach to these functions when we treat these functions as mono-unary algebras and represent them with the help of vertex graphs (See e.g. [3]). Such approach occurs quite rarely while teaching mathematics both at secondary schools and universities. However, it facilitates efficient solutions to many problems, not only from the theory of functional equations and mono-unary algebras, but also from the theory of metric spaces, continuous mappings and others. First, let us mention basic notions from the theory of mono-unary algebras, then introduce the definition and description (the intuitive and formally precise ones) of the second iterative roots of real function  $q(x) = x^2$ . It is also necessary to give some statements which were proved in articles [2], [4] and [5].

## 1 BASIC NOTIONS AND VERTEX GRAPHS

A *mono-unary algebra* is an ordered pair  $(A, f)$ , where  $A$  is a non-empty set and  $f$  is the mapping of set  $A$  into itself. Such algebra will be shortly called a *unar* according to [9]. Unar  $(A, f)$  is called *continuous* if for each pair of its elements  $a, b \in A$  there exists a pair of non-negative integers  $m, n$  with the property  $f^n(a) = f^m(b)$ ; otherwise the unar is called *non-continuous*. Symbol  $f^n$  denotes the  $n$ -th iteration of mapping  $f$  defined as follows:  $f^1 = f$ ,  $f^n = f \circ f^{n-1}$ , where the symbol  $\circ$  means the composition of mappings. Unar  $(B, g)$  is a subunar of unar  $(A, f)$ , where  $B$  is a non-empty subset of set  $A$  with the property  $f(B) \subset B$  and  $g$  is the contraction of mapping  $f$  on set  $B$ . The maximal subunar (with respect to the set inclusion of the carriers) of unar  $(A, f)$  is called a *component* of unar  $(A, f)$ . The smallest subunar (if it exists) is called a *cycle* of this component. Let us remark that in the iteration theory the carrier

of the component of  $\text{unar}(A, f)$  is called an *orbit* of transformation  $f$ . On every  $\text{unar}(A, f)$  there can be defined a pre-ordering  $\leq_f$  (a reflexive and transitive binary relation) as follows: For  $a, b \in A$  holds  $a \leq_f b$  if and only if there exists non-negative integer  $n$  with the property  $f^n(a) = b$ . This relation  $\leq_f$  is an ordering if and only if transformation  $f$  has no multiple-element cycles, but only the one-element ones ([6], [12]). Let  $n \in \mathbb{N}$ ,  $n > 1$ . Mapping  $g: A \rightarrow A$  is called an *iterative root* of order  $n$  of mapping  $f$  if  $g^n = f$ . The set of all second iterative roots of function  $q(x) = x^2$  will be denoted  $\sqrt{q}^*$ .

Now let us mention the description of quadratic real function  $q(x) = x^2$  using its vertex graph.  $\text{Unar}(\mathbf{R}, q)$  has just two finite components with carriers  $K_0 = \{0\}$ ,  $K_1 = \{-1, 1\}$  with one-element cycles  $\{0\}$ ,  $\{1\}$  respectively. Further, it consists of innumerably many countable components  $K_t$ ,  $t \in (0, 1)$ . These infinite components  $K_t$  are isomorphic to each other, i.e. they have the same vertex graph. This vertex graph is the same as for  $\text{unar}(\mathbf{Z}, \mu)$ , where  $\mu: \mathbf{Z} \rightarrow \mathbf{Z}$ ,  $\mu(z) = z + 2$  for odd  $z$ ,  $\mu(z) = z + 1$  for even  $z$ ; according to [9] these components are called two-sidedly infinite chains with short chains. The vertex graph of function  $q$  is shown in Figure 1 (here  $x_1 = 1$ ,  $x_2 = -1$ ):



**Fig. 1.** Vertex graph of quadratic function  $q(x) = x^2$ .  
Source: own

Now let us describe formally the structure of functions  $f \in \sqrt{q}^*$ . We will show that there can only be two possibilities for such structure (the proof is given in [2] and [5]). In the next part of the article we will demonstrate how favourable and useful such description is. Let us further note that in this description the term component represents the infinite component because for finite components  $K_0, K_1$  and function  $f \in \sqrt{q}^*$  there holds  $f(0) = 0$ ,  $f(-1) = f(1) = 1$ .

*Definition:* Let for two components  $K_i, K_j$  of  $\text{unar}(\mathbf{R}, q)$  and for any function  $f \in \sqrt{q}^*$  there hold:  $f(x) \in K_j$  and  $f(y) \in K_i$  for every pair  $(x, y) \in K_i \times K_j$ . Then this pair of components  $(K_i, K_j)$  will be called an  $f$ -pair of components of  $\text{unar}(\mathbf{R}, q)$ .

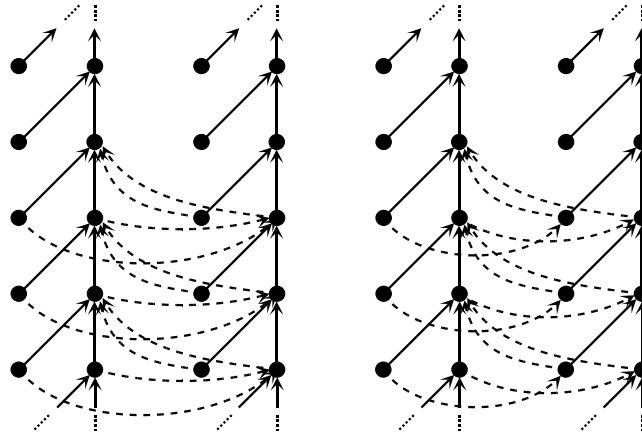
*Definition:* Let  $(K_i, K_j)$  be an  $f$ -pair of components of  $\text{unar}(\mathbf{R}, q)$ ,  $f \in \sqrt{q}^*$ . Then function  $f|_{K_i} = f_{(i,j)}: K_i \rightarrow K_j$  and function  $f|_{K_j} = f_{(j,i)}: K_j \rightarrow K_i$  will be called connective functions with respect to the  $f$ -pair of components  $(K_i, K_j)$ .

*Note:* It is obvious that any component  $(K, f_K)$  of unar  $(\mathbf{R}, f)$  can be created as follows:

$K = K_i \cup K_j$ , where  $(K_i, K_j)$  is an  $f$ -pair of components,  $f_K = f_{(j,i)} \cup f_{(i,j)}$ , where  $f_{(i,j)}, f_{(j,i)}$  are connective functions with respect to an  $f$ -pair of components  $(K_i, K_j)$ .

*Lemma 1:* Let us denote  $P = \mathbf{R} - \{-1, 0, 1\}$ . Let  $f \in \sqrt{q}^*$ . Let  $(K, f_K)$  be any component of unar  $(\mathbf{R}, f)$ ,  $K \subset P$ . Then there exists an  $f$ -pair  $(K_i, K_j)$  of components of unar  $(\mathbf{R}, q)$ , for which  $f_K$  is either an even non-negative function or  $f_K = f_{(i,j)} \cup f_{(j,i)}$  (with a suitable choice of indices  $i, j$ ), where  $f_{(i,j)}$  is an odd connective function and  $f_{(j,i)}$  is an even connective function with respect to an  $f$ -pair of components  $(K_i, K_j)$ .

*Note:* Both possible cases of the construction of the second iterative roots of function  $q$  for infinite components  $K_i$  are demonstrated in the following Figure 2. In the left part there is depicted the case when  $f_K$  is an odd non-negative function, in the right part there is the case  $f_K = f_{(i,j)} \cup f_{(j,i)}$ , where  $f_{(i,j)}$  is an odd connective function and  $f_{(j,i)}$  is an even connective function.



**Fig. 2.** Second iterative roots of function  $q$  - vertex graphs.  
Source: own

## 2 CONTINUITY OF SECOND ITERATIVE ROOTS OF QUADRATIC FUNCTIONS

Now let us mention some statements which deal with iterative roots of function  $q$  (See e.g. [2], [4], [5]). Let us further note that the monoid of endomorphisms  $End(\mathbf{R}, q)$  is the set of all real functions commuting to (interchangeable with) function  $q$ , with operation function composition.

*Lemma 2:* For every function  $f: \mathbf{R} \rightarrow \mathbf{R}$ , which is the solution of equation  $f^2 = q$ , there holds  $f \circ q = q \circ f$  when  $\sqrt{q}^* \subset End(\mathbf{R}, q)$ .

*Corollary:* Every solution  $f: \mathbf{R} \rightarrow \mathbf{R}$  of functional equation  $f^2(x) = x^2$  is the solution of functional equation  $f(x^2) = [f(x)]^2$ .

Now we will consider the continuity of the solution of equation  $f^2 = q$ , first in the classical meaning of the real function connectivity. In [7] there is given one continuous solution, which is  $f(x) = |x|^{\sqrt{2}}$ . This continuous solution of the given equation is a rare case. The following Lemma 3 (See also [2]) can provide a certain answer to the problem of the continuity of the solution of equation  $f^2 = q$ .

*Lemma 3:* There exists set  $F \subset \{f: \mathbf{R} \rightarrow \mathbf{R}; f^2 = q\}$  with the following properties:

1<sup>0</sup>  $\text{card } F = 2^{\aleph_0} (= c)$ ;

2<sup>0</sup> every function  $f \in F$  has an infinite number of the points of discontinuity.

*Proof:* Let  $\{a_n\}_{n \in \mathbf{N}}, \{b_n\}_{n \in \mathbf{N}}$  be sequences for which  $a_1 = 2, b_1 = \frac{5}{2} = 2 + \frac{1}{2}, a_n = b_{n-1}, b_n = a_n + \frac{1}{2^n}$  for every  $n \in \mathbf{N}, n \geq 2$ . It is obvious that the union of all intervals  $\langle a_n, b_n \rangle$  for  $n \in \mathbf{N}$  is the interval  $\langle 2, 3 \rangle$ . Let us denote  $P$  as the set of all increasing sequences  $p = \{r_n\}_{n \in N_0}, p(n) = r_n$ , of real numbers from interval  $\langle 2, 3 \rangle$  defined as follows:  $p(0) = r_0 = 2$  for every  $p \in P, p(n) = r_n \in (a_n, b_n)$ , while  $\{p(n); p \in P\} = (a_n, b_n)$  for every  $n \in \mathbf{N}$ . Then  $\text{card } P = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0^2} = 2^{\aleph_0}$ . As for every sequence  $p \in P$  there holds  $3 - \frac{1}{2^{n-1}} < p(n) < 3$  for every  $n \in \mathbf{N}$ , then  $\lim_{n \rightarrow \infty} p(n) = 3$  for every sequence  $p \in P$ . It is obvious that there are not any two members of any sequence  $p \in P$  which would belong to the same component of  $\text{unar}(\mathbf{R}, q) = \sum_{t \in (0, 1)} (K_t, q_t)$ , and likewise there are not any two members of sequence  $\frac{1}{p}$  (tj.  $\left\{ \frac{1}{r_n} \right\}_{n \in N_0}$ ), where  $r_n = p(n)$ , which would belong to the same component of  $\text{unar}(\mathbf{R}, q)$  for

any  $p \in P$ . We will further demonstrate that for  $s, t \in (0, 1)$  with the property  $r_m \in K_s, \frac{1}{r_n} \in K_t$ ,

where  $r_m, r_n$  are members of one sequence or different ones from  $P$ , there holds  $s \neq t$ . Let us assume to the contrary that there exist sequences  $\{r_n\}_{n \in N_0}, \{s_m\}_{m \in N_0} \in P$  and number  $t \in (0, 1)$

such that  $r_n \in K_p, \frac{1}{s_m} \in K_t$ . Then for suitable  $k \in \mathbf{N}$  there applies either  $r_n = (\frac{1}{s_m})^{2^k}$  or  $\frac{1}{s_m} = r_n^{2^k}$ .

From there we can get  $s_m^{2^k} \cdot r_n = 1$  or  $s_m \cdot r_n^{2^k} = 1$ , which is the contradiction to the assumption  $r_n \geq 2, s_m \geq 2$ .

Now we will assign to each sequence  $p \in P$  the function  $f_p \in \sqrt{q}^*$  as follows: Let  $p = \{r_n\}_{n \in N_0}$ .

For every  $n \in N_0$  and every  $k \in N_0$  let us set  $f_p(r_n^{2^k}) = r_n^{-2^k}, f_p(r_n^{-2^k}) = r_n^{2^{k+1}}$ . Let  $K_s, K_t$  be the pair of different infinite components of  $\text{unar}(\mathbf{R}, q)$ , for which  $r_n \in K_s, r_n^{-1} \in K_t$  for some  $n \in N_0$ . Function  $f_p$  will be extended to  $K_s$  so that the restriction  $f_p|_{K_s}$  would be the homomorphism of  $\text{unar}(K_s, q_s)$  to  $(K_p, q_t)$ , while  $f_p(-x) = f_p(x) > 0$  (it is a simple special case of the construction of all homomorphisms of one unar to another from [8]). Similarly,  $f_p|_{K_t}$  will



be homomorphism  $(K_r, q_t)$  into  $(K_s, q_s)$  such that  $f_p(-x) = f_p(x) > 0$  for every  $x \in K_t$ , and  $f_p|_{K_t^+}, f_p|_{K_s^+}$  (where  $K^+ = \{x \in K; x > 0\}$ ) project isomorphic corresponding two-sidedly infinite chains to each other.

For every  $p \in P$  let us denote  $K^{(p)}$  as the set of all infinite components of unar  $(\mathbf{R}, q)$  defined as follows:  $K \in K^{(p)}$  if and only if  $p(n) \notin K$ ,  $\frac{1}{p(n)} \notin K$  for every  $n \in N_0$ . Let  $K(3)$ , or  $K(\frac{1}{3})$  denote the component of unar  $(\mathbf{R}, q)$  containing number 3, or  $\frac{1}{3}$ . Because  $3 < 3^{2^k}$  and  $1 < 3^{2^{-k}} < 1,8$  for every  $k \in \mathbf{N}$ , then  $K(3) \cap (\frac{1}{3}, \frac{1}{2}) = \emptyset = K(3) \cap (2, 3)$ , so  $K(3) \in K^{(p)}$  for every  $p \in P$ . Further,  $3 < 3^{2^k}$  implies  $(\frac{1}{3})^{2^k} < \frac{1}{3}$  and also  $3 < 2^{2^k}$  for every  $k \in \mathbf{N}$ , from where there follows the inequality  $\frac{1}{2} < (\frac{1}{3})^{2^{-k}} < 1$  (the last inequality is evident), so also  $K(\frac{1}{3}) \cap (\frac{1}{3}, \frac{1}{2}) = \emptyset = K(\frac{1}{3}) \cap (2, 3)$ . Then there holds  $K(3) \in K^{(p)}$ ,  $K(\frac{1}{3}) \in K^{(p)}$  for every sequence  $p \in P$ . There holds  $\text{card } K^{(p)} = 2^{\aleph_0}$ . Let  $\{K_1^{(p)}, K_2^{(p)}\}$  be a two-element decomposition of set  $K^{(p)}$  such that  $K(3), K(\frac{1}{3}) \in K_1^{(p)}$ ,  $\text{card } K_1^{(p)} = \text{card } K_2^{(p)}$ . Let  $\varphi^{(p)}, \varphi^{(p)}: K_1^{(p)} \rightarrow K_2^{(p)}$ , be a bijection (firmly selected for every  $p \in P$  and the given decomposition of set  $K^{(p)}$ ). For every pair of components  $K \in K_1^{(p)}$ ,  $\varphi^{(p)}(K) \in K_2^{(p)}$  let us extend  $f_p$  to  $K \cup \varphi^{(p)}(K)$  so that  $f_p(K) \subset \varphi^{(p)}(K)$ ,  $f_p(\varphi^{(p)}(K)) \subset K$ , further  $f_p|_K$  is an odd function,  $f_p|_{\varphi^{(p)}(K)}$  is an even one and there holds  $f_p^2(x) = x^2$  for every number  $x \in K \cup \varphi^{(p)}(K)$ . Next, let us set  $f_p(0) = 0$ ,  $f_p(-1) = f_p(1) = 1$ . Thus the function  $f_p$  is defined on set  $\mathbf{R}$ . Now, we will show that  $f_p \in \sqrt{q}^*$ .

If  $x$  is an element of the union of sets  $(K \cup K_0 \cup K_1)$  for  $K \in K^{(p)}$ , then  $f_p^2(x) = x^2$  (with respect to the extension  $f_p$  on the union of the given components). Let  $x = r_n^{2^k}$ , where  $p(n) = r_n$  and  $k, n \in N_0$ . Then  $f_p^2(x) = f_p(r_n^{-2^k}) = r_n^{2^{k+1}} = x^2$ , and similarly if  $x = r_n^{-2^k}$ , then  $f_p^2(x) = f_p(r_n^{2^{k+1}}) = r_n^{-2^{k+1}} = x^2$ . If  $x \neq r_n^{2^k}$ , then for every  $n \in N_0$  and every  $k \in N_0$  let us denote  $K_t$  as the component of unar  $(\mathbf{R}, q)$  with the property  $x \in K_r$ ,  $r_n \in K_t$  for some  $n \in N_0$ . In view of the definition of function  $f_p$  on set  $K_t \cup K_s$ , where  $K_s$  is a component of unar  $(\mathbf{R}, q)$  containing  $\frac{1}{r_n}$ , we will get that  $f_p^2(x) = x^2$ .

Such assignment  $p \rightarrow f_p$  for every  $p \in P$  defines the mapping of set  $P$  to  $\sqrt{q}^*$ . We will show that the mapping is an injection. Let  $p_1, p_2 \in P$ ,  $p_1 \neq p_2$ ,  $p_1(n) = r_n$ ,  $p_2(n) = s_n$  for every  $n \in N_0$ ,  $r_n, s_n \in (2, 3)$ . According to the definition of function  $f_{p_1}$  there holds  $f_{p_1}(-r_1) = f_{p_1}(r_1) = \frac{1}{r_1}$ ,  $f_{p_1}(-\frac{1}{r_1}) = f_{p_1}(\frac{1}{r_1}) = r_1^2$ . Let  $r_1 \in K_t$ ,  $\frac{1}{r_1} \in K_s$ ,  $t, s \in (0, 1)$ .

Then  $K_t, K_s \in K^{(p_2)}$ . If the components  $K_t, K_s$  belong to one block of decomposition  $\{K_1^{(p_2)}, K_2^{(p_2)}\}$  of the set of components  $K^{(p_2)}$ , then  $f_{p_1}(r_1) \neq f_{p_2}(r_1)$ , because  $f_{p_2}(r_1) \notin K_s$ . Let then  $K_t \in K_1^{(p_2)}, K_s \in K_2^{(p_2)}$  and let then assume that  $K_s \neq \varphi^{(p_2)}(K_t)$ , (where  $\varphi^{(p_2)}$  is a bijection corresponding to decomposition  $\{K_1^{(p_2)}, K_2^{(p_2)}\}$ ). Then there again holds  $f_{p_2}(r_1) \notin K_s$ , so  $f_{p_1}(r_1) \neq f_{p_2}(r_1)$ . To the contrary, let us assume that  $K_s = \varphi^{(p_2)}(K_t)$  (the case  $\varphi^{(p_2)}(K_s) \neq K_t$  is analogical to the case  $\varphi^{(p_2)}(K_t) \neq K_s$ ). Then  $f_{p_2}(-\frac{1}{r_1}) = -f_{p_2}(\frac{1}{r_1}) < 0$ ,  $f_{p_1}(-\frac{1}{r_1}) = r_1^2 > 0$ , and there holds again  $f_{p_1}(-\frac{1}{r_1}) \neq f_{p_2}(-\frac{1}{r_1})$ . Thus we proved that in all possible cases there holds  $f_{p_1} \neq f_{p_2}$ . If we set  $F = \{f_p; p \in P\}$ , then  $\text{card } F = \text{card } P = 2^{\aleph_0}$ , and at the same time  $F \subset \sqrt{q}^*$ .

In order to conclude the proof there remains to show that every function  $f \in F$  has an infinite set of discontinuity points. Let us denote set  $M$ ,  $M = \{3^{2^k}; k \in \mathbf{Z}\} \cup \{(\frac{1}{3})^{2^k}; k \in \mathbf{Z}\}$ . Let  $f_p \in F$  be any function, let  $p$  be the corresponding sequence from set  $P$ , where  $r_n = p(n)$ . Let  $k \in \mathbf{Z}$  be random. There holds

$$\lim_{n \rightarrow \infty} r_n^{2^k} = 3^{2^k}, \quad \lim_{n \rightarrow \infty} f_p(r_n^{2^k}) = \lim_{n \rightarrow \infty} r_n^{-2^k} = 3^{-2^k}, \quad \lim_{n \rightarrow \infty} f_p(r_n^{-2^k}) = \lim_{n \rightarrow \infty} r_n^{2^{k+1}} = 3^{2^{k+1}}.$$

According to the definition of function  $f_p$  there holds  $f_p(3) \notin K(\frac{1}{3})$ , so  $f_p(3^{2^k}) \neq 3^{-2^k}$ ,  $f_p(3^{-2^k}) \neq 3^{2^{k+1}}$  for every  $k \in \mathbf{Z}$ , therefore function  $f_p$  is discontinuous in every point of set  $M$ . Thus the proof is finished.

In the following part of the paper we will deal with the problem of the continuity of the second iterative roots of function  $q$  generally, i.e. in the non-Euclidean metric. We will present the construction of non-symmetrical non-discrete quasi-metric  $d$  (in the definition of the metric there is omitted the requirement for the symmetry) which gains infinitely many values on  $\mathbf{R} \times \mathbf{R}$  and for which every solution of equation  $f^2 = q$  represents the continuous mapping of space  $(\mathbf{R}, d)$  into itself. Let us note that this required quasi-metric  $d$  is function  $d: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}_0^+$ , for which there holds:

- (i)  $(\forall x, y \in \mathbf{R}) d(x, y) = 0 \Leftrightarrow x = y$ ,
- (ii)  $(\forall x, y, z \in \mathbf{R}) d(x, y) + d(y, z) \geq d(x, z)$ .

Let us now define quasi-pseudo-metrics  $d_0, \dots, d_k$  (there does not hold (i)),  $k \in N_0$ , as follows:

$$d_0(x, y) = \begin{cases} 0 & \exists m \in N_0 : x = y^{2^m} \\ 1 & \text{otherwise} \end{cases},$$

$$d_k(x, y) = \begin{cases} 0 & x = y \vee \exists m \geq k+1 : x = y^{2^m} \\ 1 & \text{otherwise} \end{cases}.$$

Let us further define function  $d: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}_0^+$  with the following formula:

$$d(x, y) = \sum_{k=0}^{\infty} \frac{1}{2^k} d_k(x, y).$$

We can easily prove that this non-negative real function  $d$  already satisfies (i) and (ii), so it is a quasi-metric on  $\mathbf{R}$ . The condition (i) follows from the evident fact that  $d(x, y) = 0$  if and only if  $d_k(x, y) = 0$  for every non-negative integer  $k$ , which according to the definition of quasi-pseudo-metrics  $d_k$  is valid only in the case  $x = y$ . The condition (ii) can be proved as follows:

$$\begin{aligned} d(x, y) + d(y, z) &= \sum_{k=0}^{\infty} \frac{1}{2^k} d_k(x, y) + \sum_{k=0}^{\infty} \frac{1}{2^k} d_k(y, z) = \sum_{k=0}^{\infty} \frac{1}{2^k} [d_k(x, y) + d_k(y, z)] \geq \\ &\sum_{k=0}^{\infty} \frac{1}{2^k} d_k(x, z) = d(x, z). \end{aligned}$$

Therefore function  $d$  is a real quasi-metric. The range of this quasi-metric are elements of the set union  $\{0, 2\} \cup \{\frac{1}{2^n}; n \in \mathbf{N}_0\}$ , i.e. this quasi-metric assumes infinitely many values on  $\mathbf{R}$ .

The only problem represents the value  $d(1, -1)$  due to the shape of the vertex graph of function  $q(x) = x^2$ . So it is necessary to define  $d(1, -1) = 1$ . Now the quasi-metric  $d$  is defined on the whole set  $\mathbf{R} \times \mathbf{R}$ .

In the following Theorem and its proof we will denote  $P = \mathbf{R} - \{-1\}$ .  $q_p$  is the contraction of function  $q$  on set  $P$ ,  $d_p$  is the contraction of function  $d$  on set  $P \times P$ . Finally, let us mention the denotation  $C(\mathbf{R}, d)$  for the monoid of all real functions of one variable continuous in metric  $d$ .

*Theorem:* There exists quasi-metric  $d$  on set  $\mathbf{R}$  for which holds:

- (a) Every function  $f \in \sqrt{q}^*$  is the continuous mapping of space  $(\mathbf{R}, d)$  into itself.
- (b) The bijective mapping  $f$  of subspace  $(P, d_p)$  of space  $(\mathbf{R}, d)$  into itself is an isometric mapping if and only if there holds  $f(x^2) = [f(x)]^2$  for every  $x \in P$  (i.e.  $f \in \text{End}(P, q_p)$ ).

*Proof:* (a) Let us consider the above defined quasi-metric  $d: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}_0^+$ . We will prove that there holds  $\sqrt{q}^* \subset C(\mathbf{R}, d)$ . Such inclusion can be obtained as the result of a stronger statement, namely from the inclusion  $\text{End}(\mathbf{R}, q) \subset C(\mathbf{R}, d)$ , which together with inclusion  $\sqrt{q}^* \subset \text{End}(\mathbf{R}, q)$  proved in Lemma 2 (See [2], Theorem 1) leads to the inclusion in (a).

Let then  $f \in \text{End}(\mathbf{R}, q)$ , i.e.  $f(x^2) = [f(x)]^2$  for every  $x \in \mathbf{R}$ . Let  $x_0 \in \{0, 1, -1\}$ , then  $f(x_0) \in \{0, 1\}$ . If  $\varepsilon > 0$  is random, for  $\delta < 1$  (with respect to the definition of  $d$ ) there holds  $K_d(x_0, \delta) = \{x_0\}$ , and thus  $f(K_d(x_0, \delta)) \subset K_d(f(x_0), \varepsilon)$ , which means the continuity of function  $f$  in point  $x_0$ .

Now let  $x_0$  be an element of an infinite component of  $\text{unar}(\mathbf{R}, q)$ . Let  $\varepsilon > 0$  be random. If  $\varepsilon > 2$ , then  $K_d(f(x_0), \varepsilon) = \mathbf{R}$ , so for any  $\delta > 0$  there holds  $f(K_d(x_0, \delta)) \subset K_d(f(x_0), \varepsilon)$ . Let us assume that  $0 < \varepsilon \leq 2$ . Let us set  $\delta = \varepsilon$ . If  $\varepsilon > 1$ , then  $K_d(x, \varepsilon) = \{x^{2^n}, n \in \mathbf{N}_0\}$ . We will show that  $f(K_d(x_0, \delta)) = K_d(f(x_0), \varepsilon)$ . Let  $t \in f(K_d(x_0, \delta))$ ,  $x \in K_d(x_0, \delta)$  with the property  $f(x) = t$ . Then  $x = x_0^{2^n}$  for suitable  $n \in \mathbf{N}_0$ . We will obtain

$t = f(x) = f(x_0^{2^n}) = [f(x_0)]^{2^n} \in K_d(f(x_0), \varepsilon)$ , so  $f(K_d(x_0, \delta)) \subset K_d(f(x_0), \varepsilon)$ . Now let  $t \in K_d(f(x_0), \varepsilon)$ . Then again there exists a number  $n \in \mathbf{N}_0$  such that  $t = [f(x_0)]^{2^n} = f(x_0^{2^n}) \in f(K_d(x_0, \delta))$ . Thereby we get the desired equality  $f(K_d(x_0, \delta)) = K_d(f(x_0), \varepsilon)$ . If  $\varepsilon \leq 1$ , let us denote  $p$  as the smallest natural number with the property  $2^{1-p} < \varepsilon$ . Then  $K_d(x_0, \delta) = K_d(x_0, \varepsilon) = \{x_0^{2^n}; n = 0, p+1, p+2, \dots\}$ ;  $K_d(f(x_0), \varepsilon) = \{f(x_0^{2^n}); n = 0, p+1, p+2, \dots\}$ . Applying the similar method as described above, we will show that there holds  $f(K_d(x_0, \delta)) \subset K_d(f(x_0), \varepsilon)$ . Thus we will obtain the continuity of function  $f$  in every point  $x_0 \in \mathbf{R}$ . According to Lemma 2 there holds  $\sqrt{q}^* \subset \text{End}(\mathbf{R}, q)$ . The just proven statement (a) means that there holds  $\text{End}(\mathbf{R}, q) \subset C(\mathbf{R}, d)$ ; thus altogether  $\sqrt{q}^* \subset C(\mathbf{R}, d)$ .

(b) Let us consider subspace  $(P, d_p)$  of space  $(\mathbf{R}, d)$ . Let  $f: P \rightarrow P$  be a bijective mapping with the property  $f(x^2) = [f(x)]^2$  for every  $x \in P$ . For any pair of points  $x_1, x_2 \in P$ ,  $x_1 \neq x_2$ , there will hold one of these cases:

- (i)  $x_1 = x_2^{2^n}$  for suitable  $n \in \mathbf{N}$ ,
- (ii)  $x_2 = x_1^{2^m}$  for suitable  $m \in \mathbf{N}$ ,
- (iii)  $x_1 \parallel_q x_2$ .

In the case (i) there holds  $d_p(f(x_1), f(x_2)) = d_p(f(x_1), f(x_1^{2^m})) = d_p(f(x_1), [f(x_1)]^{2^m}) = \frac{1}{2^{m+1}} =$

$$d_p(x_1, x_1^{2^m}) = d_p(x_1, x_2).$$

In the case (ii)  $d_p(x_1, x_2) = 2 = d_p(f(x_1), f(x_2))$ ; let us note that  $f$  is an automorphism of  $\text{unar}(P, q_p)$ .

In the case (iii) there holds  $d_p(x_1, x_2) = 2$ . If there was  $d_p(f(x_1), f(x_2)) \leq 1$ , then according to the definition of quasi-metric  $d_p$  we would find out that there exists  $n \in \mathbf{N}$  with the property  $f(x_1) = [f(x_2)]^{2^n}$ . Then  $f(x_2^{2^n}) = f(x_1)$  (with respect to formula  $f \circ q = q \circ f$ ). At the same time there applies  $x_1 \neq x_2^{2^n}$ , which is the contradiction to the bijectivity of mapping  $f$ . Then there must hold  $d_p(f(x_1), f(x_2)) = 2$ , so  $f$  is an isometric mapping of space  $(P, d_p)$  into itself.

On the contrary, let  $f$  be an isometric mapping of space  $(P, d_p)$  into itself. Let  $x_0 \in P$  be any point,  $x_0 \notin \{0, 1\}$ . There holds  $d_p(f(x_0^2), f(x_0)) = d_p(x_0^2, x_0) = 1$ , from where  $f(x_0^2) = [f(x_0)]^2$  with respect to the definition of quasi-metric  $d$ . Let  $x_0 \in \{0, 1\}$ . Then set  $\{x_0\}$  is ambiguous (i.e. open and closed) in space  $(P, d_p)$ . As the isometry retains the set ambiguity and the isometry is a homeomorphism, there holds  $f(x_0) \in \{0, 1\}$ . There is no one-point subset of

space  $(P, d_p)$  different from  $\{0\}, \{1\}$  which is ambiguous. But then with respect to the definition of quasi-metric  $d_p$  and the bijectivity of mapping  $f$  there holds:  $d_p(f(0), f(1)) = d_p(f(1), f(0)) = 2 = d_p(0, 1) = d_p(1, 0)$ . If we consider the definition of quasi-metric  $d_p$  and the form of the vertex graph of function  $q$ , we will get the formula  $f \circ q = q \circ f$ . Thus the proof is finished.

## CONCLUSION

The paper follows up the author's article [4] which was published in the Proceedings from the international conference MITAV in 2016. It is devoted to the discrete description of real functions (unlike the commonly used continuous representation) and to the possible application of their discrete description. The representation of real functions in the form of vertex graphs makes possible the decision about the existence of the given real function's iterative roots and it enables the effective formal mathematical description of these iterative roots. Here is given the mathematical description of the second iterative roots of the simplest quadratic function  $y = x^2$ ; further there follows the proof of the Theorem dealing with the continuity of these second iterative roots in standard Euclidean metric. The paper is concluded with the Theorem according to which there exists real quasi-metric  $d$  such that every second iterative root of function  $f(x) = x^2$  is a continuous mapping of space  $(\mathbf{R}, d)$  into itself. Particularly, the author points out the fact that the application of the discrete approach to functions allows us to solve problems which at standard continuous approach would be solved with great difficulties (e.g. some types of functional equations of one variable). The given theory presents a number of open problems and topics for further creative mathematical exploration. In some of the statements, the second iterative root can be replaced generally by the iterative root of order  $n$ ; similarly, the existence of a metric (not only of a quasi-metric) satisfying analogous properties as the ones in the main Theorem of this paper is still open for investigation. A lot of interesting and open problems can be found in [1] [6], [10], [11] and [12].

## References

- [1] Beránek, J., Chvalina, J. *On Tabor's problem concerning a certain quasi-ordering of iterative roots of functions*. Aequ. Math. No. 39, 1990, p. 1–5.
- [2] Beránek, J. *O spojitosti iterativních kořenů nejjednodušší kvadratické funkce*. In: *Proceeding of Contributions of Žilina Didactic Conference*. Žilina: University of Žilina, 2004. ISBN 80-8070-270-5.
- [3] Beránek, J. *Funkcionální rovnice*. Brno: Masaryk University, 2004, 74 pp. ISBN 80-210-3422-X.
- [4] Beránek, J. *Hyperbolic sine and cosine from the iteration theory point of view*. In Baštinec, J., Hrubý, M. *Mathematics, Information Technologies and Applied Sciences 2016, Post-Conference Proceedings of Extended Versions of Selected Papers*. Brno: University of Defence in Brno, 2016. p. 31-41. ISBN 978-80-7231-400-3.
- [5] Chvalina, J., Beránek, J. *O iteračních odmocninách kvadratické funkce*. In: *Proceeding of Contributions of Faculty of Education UJEP Brno*. Brno: UJEP Brno, 1990, p.7–19.
- [6] Chvalina, J. *Funkcionální grafy, kvaziuspořádané množiny a komutativní hypergrupy*. Brno: Masaryk University, 1995, 205 pp. ISBN 80-210-1148-3.
- [7] Isaacs, R. *Iterates of fractional order*. Canad. J. Math. No. 2, 1950, p. 409-416.

- [8] Novotný, M. *O jednom problému z theorie zobrazení*. Publ. Fac. Sci. Univ. Masaryk, No. 344, 1953, p. 53-64.
- [9] Skornjakov, L. A. *Unars*. In: *Colloq. Math. Soc. János Bolyai 29*, Esztergom: Univ. Algebra, 1977, p. 735-743.
- [10] Smítal, J. *O funkciách a funkcionálnych rovniciach*. Bratislava: Alfa, 1984, 143 pp. ISBN 63-146-84.
- [11] Snowden, M., Howie, J. M. *Square roots in finite full transformation semigroups*. Glasgow Math. J. 23, no. 2, 1982, p. 137–149.
- [12] Targonski, G. *Topics in Iteration Theory*. Göttingen and Zürich: Vandenhoeck et Ruprecht, 1981.

# Geodesic and almost geodesic mappings onto Ricci symmetric spaces

V. Berezovskii, P. Peška and J. Mikeš

Uman National University of Horticulture, Dept. of Mathematics

Instytutska 1, Uman, Ukraine

Department of Algebra and Geometry,

Faculty of Science, Palacký University Olomouc,

17. listopadu 1192/12, 774 16 Olomouc

Email: berez.volod@rambler.ru, patrik\_peska@seznam.cz,  
josef.mikes@upol.cz

**Abstract:** This paper is devoted to study of geodesic and almost geodesic mappings of special spaces with affine connection. In the first section, we mention the basic definition of geodesic and almost geodesic mappings. The next section is devoted to geodesic mappings onto Ricci symmetric manifolds and its fundamental differential equation in Cauchy type form in covariant derivatives. We also study almost geodesic mappings of the first type onto symmetric space.

**Keywords:** geodesic mapping, almost geodesic mapping, spaces with affine connection, (pseudo-) Riemannian space

## INTRODUCTION

This paper is dedicated to further development of theories of geodesic and almost geodesic mappings of spaces with affine connection on some special spaces, especially symmetric and Ricci symmetric spaces.

T. Levi-Civita [14] set and solved a special equation for the problem of finding Riemannian spaces with common geodesics. It is worth to note that it was connected with studying the equations of mechanical systems.

The theory of geodesic mappings has been developed later in the works by Thomas, Weyl, Shirokov, Solodovnikov, Sinyukov, Mikeš and others. Studying geodesic mappings was followed up in the works by Kagan, Vrceanu, Shapiro, Vedenyapin and others. The mentioned authors identified special classes  $(n - 1)$ -projective spaces, see [15, 17, 18, 19, 26].

The quasi geodesic mappings introduced A.Z. Petrov [23]. Special quasi geodesic mappings, in particular, are holomorphically projective mappings of Kähler spaces, which first had been studied by Otsuki and Tashiro [22], Prvanovich [25] and others, see [16, 17, 18, 19, 26].

The natural generalizations of these classes are almost geodesic mappings, which were introduced by Sinyukov. He also defined three types of almost geodesic mappings  $\pi_1$ ,  $\pi_2$  and  $\pi_3$ , see [26]. The almost geodesic mappings have been developed later in the works by Sobchuk, Yablonskaya,

Berezovskii, Mikeš, see [1, 2, 3, 4, 5, 6, 16, 18], [17], pp. 455–480.

Recently, some questions related to geodesic and almost geodesic mappings, and their generalizations have also been studied in [9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 24, 27].

In this paper there were obtained the main equations of geodesic mappings of spaces with affine connection onto Ricci symmetric manifolds and almost geodesic mappings of the first type of spaces with affine connection onto symmetric spaces in a form of closed systems of Cauchy type in covariant derivatives. In this paper was also set the number of essential real parameters for the general solution of that system.

We assume that the studied spaces are simply connected with dimension  $n > 2$ , and we consider that geometric objects are continuous and smooth enough.

## 1 GEODESIC AND ALMOST GEODESIC MAPPINGS THEORY

Let us mention the following definition of geodesic and almost geodesic curves and mappings, see [17], pp. 88, 257, 455–458, [26], pp. 43, 70, 156–162.

### 1.1 Geodesics and geodesic mappings

It is known, the curve  $\ell$  defined in space with affine connection  $A_n$  is called *geodesic* if there exists a parallel tangent vector field along it.

The diffeomorphism  $f: A_n \rightarrow \bar{A}_n$  between spaces with affine connection is called *geodesic mapping* if any geodesic of  $A_n$  is mapped onto geodesic on  $\bar{A}_n$ .

The diffeomorphism  $f$  is geodesic mapping if and only if in a common coordinate system  $x = (x^1, x^2, \dots, x^n)$  respective mapping  $f$  yields the Levi-Civita equation

$$\bar{\Gamma}_{ij}^h(x) = \Gamma_{ij}^h(x) + \psi_i \delta_j^h + \psi_j \delta_i^h, \quad (1)$$

where  $\bar{\Gamma}_{ij}^h(x)$  and  $\Gamma_{ij}^h(x)$  are components of affine connection of manifolds  $A_n$  and  $\bar{A}_n$ , respectively,  $\psi_i$  are components of covector, and  $\delta_i^h$  is the Kronecker delta.

### 1.2 Almost geodesics curves and mappings

The curve  $\ell$  defined in space with affine connection  $A_n$  is called *almost geodesic* if there exists two dimensional parallel plane along it, which contains its tangent vector.

The diffeomorphism  $f: A_n \rightarrow \bar{A}_n$  is called *almost geodesic mapping (AGM)* if any geodesic of  $A_n$  is mapped onto almost geodesic in  $\bar{A}_n$ .

The diffeomorphism  $f$  is almost geodesic mapping if and only if in a common coordinate system  $x = (x^1, x^2, \dots, x^n)$  respective mapping  $f$  the deformation tensor of connections

$$P_{ij}^h(x) = \bar{\Gamma}_{ij}^h(x) - \Gamma_{ij}^h(x) \quad (2)$$



yields the Sinyukov's equation

$$A_{\alpha\beta\gamma}^h \lambda^\alpha \lambda^\beta \lambda^\gamma = a \lambda^h + \lambda P_{\alpha\beta}^h \lambda^\alpha \lambda^\beta,$$

where  $A_{ijk}^h = P_{ij,k}^h + P_{ij}^\alpha P_{\alpha k}^h$ ,  $\lambda^h$  is a vector,  $a$  and  $b$  are functions of variables  $x^h$  and  $\lambda^h$ . Here and after a symbol “ $\cdot$ ” denotes a covariant derivative respective connection of  $A_n$ .

Sinyukov defined three types of almost geodesic mappings  $\pi_1$ ,  $\pi_2$  and  $\pi_3$ . We proved [1], that if dimension  $n > 5$  no other types exist.

Almost geodesic mappings of type  $\pi_1$  are characterized by the following conditions on deformation tensor

$$A_{(ijk)}^h = a_{(ij}\delta_{j)}^h + b_{(i}P_{jk)}^h, \quad (3)$$

where  $a_{ij}$  is a symmetric tensor,  $b_i$  is covector, and  $(i, j, k)$  is symmetrization of mentioned indices without division.

If in an equation (3) the covector  $b_i$  is vanishing, then the mapping is called *canonical*. It is known [26], p. 171, [17], p. 463, that any almost geodesic mappings of type  $\pi_1$  can be regarded as a composition of canonical almost geodesic mapping of type  $\pi_1$  and geodesic mapping.

## 2 THE GEODESIC MAPPINGS ONTO RICCI SYMMETRIC MANIFOLDS

In this section we will study the geodesic mappings of  $A_n$  onto Ricci symmetric spaces  $\bar{A}_n$ . The manifold with affine connection is called *Ricci symmetric* if the Ricci tensor in it is absolutely parallel, see [17], p. 87.

Therefore, Ricci symmetric manifold  $\bar{A}_n$  is characterized by the condition

$$\bar{R}_{ij|k} \equiv 0, \quad (4)$$

where  $\bar{R}_{ij}$  is the Ricci tensor of  $\bar{A}_n$ , the symbol “ $|$ ” denotes a covariant derivative of connection on  $\bar{A}_n$ .

Because for Riemannian tensor (or curvature tensor)  $\bar{R}$  of  $\bar{A}_n$  holds

$$\bar{R}_{ijk|m}^h = \frac{\partial \bar{R}_{ijk}^h}{\partial x^m} + \bar{\Gamma}_{m\alpha}^h \bar{R}_{ijk}^\alpha - \bar{\Gamma}_{mi}^\alpha \bar{R}_{\alpha jk}^h - \bar{\Gamma}_{mj}^\alpha \bar{R}_{i\alpha k}^h - \bar{\Gamma}_{mk}^\alpha \bar{R}_{i\alpha j}^h,$$

then from the formula (2) we obtain

$$\bar{R}_{ijk|m}^h = \bar{R}_{ijk,m}^h + P_{m\alpha}^h \bar{R}_{ijk}^\alpha - P_{mi}^\alpha \bar{R}_{\alpha jk}^h - P_{mj}^\alpha \bar{R}_{i\alpha k}^h - P_{mk}^\alpha \bar{R}_{i\alpha j}^h, \quad (5)$$

where  $\bar{R}_{ijk}^h$  are components of Riemannian tensor  $\bar{R}$ .

Contracting formula (5) with respect to indices  $h$  and  $k$ , we obtain

$$\bar{R}_{ij|m} = \bar{R}_{ij,m} - P_{mi}^\alpha \bar{R}_{\alpha j} - P_{mj}^\alpha \bar{R}_{i\alpha}. \quad (6)$$

To follow, let us suppose the manifolds  $\bar{A}_n$  is Ricci symmetric. Using the formula (4), we get

$$\bar{R}_{ij,m} = P_{mi}^\alpha \bar{R}_{\alpha j} + P_{mj}^\alpha \bar{R}_{i\alpha}. \quad (7)$$

Taking into consideration the Levi-Civita equation (1) and basing on the formula (7), we can write

$$\bar{R}_{ij,m} = 2\psi_m \bar{R}_{ij} + \psi_i \bar{R}_{mj} + \psi_j \bar{R}_{im}. \quad (8)$$

It is known, that between Riemannian tensors on  $A_n$  and  $\bar{A}_n$  there is a dependence

$$\bar{R}_{ijk}^h = R_{ijk}^h + P_{ik,j}^h - P_{ij,k}^h + P_{ik}^\alpha P_{j\alpha}^h - P_{ij}^\alpha P_{k\alpha}^h, \quad (9)$$

where  $R_{ijk}^h$  are components of Riemannian tensor  $R$  on  $A_n$ .

Because deformation tensor  $P$  has the structure (1) from the formula (9) after computation we obtain

$$\bar{R}_{ijk}^h = R_{ijk}^h - \delta_j^h \psi_{i,k} + \delta_k^h \psi_{i,j} - \delta_i^h \psi_{j,k} + \delta_i^h \psi_{k,j} + \delta_j^h \psi_i \psi_k - \delta_k^h \psi_i \psi_j. \quad (10)$$

Let us contract (10) with respect to indices  $h$  and  $k$ . As the result we get

$$\bar{R}_{ij} = R_{ij} + n\psi_{i,j} - \psi_{j,i} + (1 - n)\psi_i \psi_j. \quad (11)$$

From the equation (11), we obtain the following

$$\psi_{i,j} = \frac{1}{n^2 - 1} [n\bar{R}_{ij} + \bar{R}_{ji} - (nR_{ij} + R_{ji})] + \psi_i \psi_j. \quad (12)$$

The following theorem holds.

**Theorem 1** *The manifold  $A_n$  admits geodesic mapping onto Ricci symmetric manifold  $\bar{A}_n$  if and only if it consists a solution of closed system of Cauchy type equations in covariant derivative (8) and (12) with respect to unknown functions  $\bar{R}_{ij}(x)$  and  $\psi_i(x)$ .*

*General solution of the above system depends on at most than  $n(n+1)$  essential real parameters.*

Because the systems (8) and (12) have only one solution for the initial conditions in point  $x_0$

$$\bar{R}_{ij}(x_0) \text{ and } \psi_i(x_0),$$

from this follows the above number of essential real parameter.

### 3 AGM OF THE FIRST TYPE ONTO SYMMETRIC SPACES

Now, let us consider canonical almost geodesic mappings of spaces with affine connection  $A_n$  onto symmetric space  $\bar{A}_n$ .

A space  $\bar{A}_n$  with affine connection  $\bar{\nabla}$  is called (locally) *symmetric* if Riemannian tensor in it is absolutely parallel (P.A. Shirokov, É. Cartan [7], S. Helgason [8], see [17], p.286, [26], p.42. Therefore, symmetric manifolds  $\bar{A}_n$  are characterized by the condition

$$\bar{R}_{ijk|m}^h \equiv 0. \quad (13)$$

Further, let us suppose that manifold  $\bar{A}_n$  is symmetric. Basing on the formula (13), from (5) we obtain

$$\bar{R}_{ijk,m}^h = P_{mi}^\alpha \bar{R}_{\alpha jk}^h + P_{mj}^\alpha \bar{R}_{i\alpha k}^h + P_{mk}^\alpha \bar{R}_{ij\alpha}^h - P_{m\alpha}^h \bar{R}_{ijk}^\alpha. \quad (14)$$

Formula (14) can be applied to general mappings of  $A_n$  onto symmetric spaces  $\bar{A}_n$ .

It is known that the equation (3) has the following form

$$3(P_{ij,k}^h + P_{ij}^\alpha P_{\alpha k}^h) = R_{(ij)k}^h - \bar{R}_{(ij)k}^h + P_{mk}^\alpha \bar{R}_{ij\alpha}^h + \delta_{(k}^h a_{ij}) + b_{(i} P_{jk)}^h. \quad (15)$$

From formula (15) of canonical almost geodesic mappings of the first type, we obtain the equation

$$P_{ij,k}^h = \frac{1}{3} (R_{(ij)k}^h - \bar{R}_{(ij)k}^h + \delta_{(k}^h a_{ij})) - P_{ij}^\alpha P_{\alpha k}^h. \quad (16)$$

From the integrability condition of the equation (16) considering formulas (15) and (16), after computation, we get the following

$$\begin{aligned} \delta_{(m}^h a_{ij),k} - \delta_{(k}^h a_{ij),m} = & -3P_{\alpha j}^h R_{ikm}^\alpha - 3P_{i\alpha}^h R_{jkm}^\alpha - R_{(ij)m}^\alpha P_{\alpha k}^h + \\ & R_{(ij)k}^\alpha P_{\alpha m}^h + R_{(ij)k,m}^h - R_{(ij)m,k}^h + 3\bar{R}_{\alpha km}^h P_{ij}^\alpha - P_{mi}^\alpha \bar{R}_{(j\alpha)k}^h + \\ & P_{ki}^\alpha \bar{R}_{(j\alpha)m}^h - P_{mj}^\alpha \bar{R}_{(i\alpha)k}^h + P_{kj}^\alpha \bar{R}_{(i\alpha)m}^h - \delta_{(m}^\alpha a_{ij}) P_{\alpha k}^h + \delta_{(k}^\alpha a_{ij}) P_{\alpha m}^h. \end{aligned} \quad (17)$$

After contracting the equation (17) with respect to indices  $h$  and  $m$ , we obtain

$$\begin{aligned} (n+1)a_{ij,k} - a_{ik,j} - a_{jk,i} = & -3P_{\alpha(j}^\beta R_{i)k\beta}^\alpha - P_{\alpha k}^\beta R_{(ij)\beta}^\alpha + P_{\alpha\beta}^\beta R_{(ij)k}^\alpha + \\ & R_{(ij)k,\beta}^\beta - R_{(ij),k}^\beta + 3P_{ij}^\alpha \bar{R}_{\alpha k}^\beta - P_{\beta i}^\alpha \bar{R}_{(j\alpha)k}^\beta + P_{ki}^\alpha \bar{R}_{j\alpha}^\beta - \\ & P_{\beta j}^\alpha \bar{R}_{(i\alpha)k}^\beta + P_{kj}^\alpha \bar{R}_{(i\alpha)}^\beta - \delta_{(\beta}^\alpha a_{ij}) P_{\alpha k}^\beta + \delta_{(k}^\alpha a_{ij}) P_{\alpha\beta}^\beta. \end{aligned} \quad (18)$$

Let us alternate the equation (18) with respect to indices  $j$  and  $k$ . Then, we can write (18) in the following form

$$\begin{aligned} (n-1)a_{ij,k} = & -3P_{\alpha(j}^\beta R_{i)k\beta}^\alpha - P_{\alpha k}^\beta R_{(ij)\beta}^\alpha + P_{\alpha\beta}^\beta R_{(ij)k}^\alpha + R_{(ij)k,\beta}^\beta - \\ & R_{(ij),k}^\beta + 3P_{ij}^\alpha \bar{R}_{\alpha k}^\beta - P_{\alpha i}^\beta \bar{R}_{(j\beta)k}^\alpha + P_{ki}^\alpha \bar{R}_{\alpha j}^\beta - P_{\alpha j}^\beta \bar{R}_{(i\beta)k}^\alpha + \\ & P_{kj}^\alpha \bar{R}_{(i\alpha)}^\beta - \delta_{(\beta}^\alpha a_{ij}) P_{\alpha k}^\beta + \delta_{(k}^\alpha a_{ij}) P_{\alpha\beta}^\beta - \frac{1}{n+2} B_{(ij)k}, \end{aligned} \quad (19)$$

where

$$\begin{aligned} B_{ijk} = & P_{\alpha k}^\beta (R_{ij\beta}^\alpha + R_{\beta ji}^\alpha) - P_{\alpha k}^\beta (R_{ik\beta}^\alpha + R_{\beta ki}^\alpha) + 3P_{\alpha\beta}^\beta R_{ijk}^\alpha + 3R_{ijk,\beta}^\beta - \\ & R_{(ij),k}^\beta + R_{(ik),j}^\beta + 2P_{ij}^\alpha \bar{R}_{\alpha k}^\beta - 2P_{ik}^\alpha \bar{R}_{\alpha j}^\beta + P_{ki}^\alpha \bar{R}_{j\alpha}^\beta - \\ & P_{ij}^\alpha \bar{R}_{k\alpha}^\beta - P_{\beta j}^\alpha \bar{R}_{(i\alpha)k}^\beta + P_{\beta k}^\alpha \bar{R}_{(i\alpha)j}^\beta - a_{ij} P_{\alpha k}^\alpha - \\ & a_{\alpha j} P_{ik}^\alpha + a_{ik} P_{i\alpha}^\alpha + a_{\alpha k} P_{ij}^\alpha. \end{aligned} \quad (20)$$

It is evident, the equations (14), (16) and (19) in the given manifold  $A_n$  have a form of closed system of Cauchy type equation regarding unknown function  $\bar{R}_{ijk}^h(x)$ ,  $P_{ij}^h(x)$  and  $a_{ij}(x)$  which also satisfies the algebraic conditions

$$\bar{R}_{i(jk)}^h = 0, \quad \bar{R}_{(ij)k}^h = 0, \quad P_{ij}^h = P_{ji}^h, \quad a_{ij} = a_{ji}. \quad (21)$$

The following theorem hold.

**Theorem 2** *The manifold  $A_n$  admits canonic almost geodesic mapping of type  $\pi_1$  onto symmetric manifold  $\bar{A}_n$  if and only if it contains a solution of a closed mixed system of Cauchy type equations in covariant derivative (14), (16), (19) and (21) in respect to unknown functions  $\bar{R}_{ijk}^h(x)$ ,  $P_{ij}^h(x)$  and  $a_{ij}(x)$ .*

*General solution of the above system depends on no more than  $1/2 n(n^3 + 2n + 1)$  essential real parameters.*

The systems (14), (16), (19) have only one solution for the initial conditions in point  $x_0$

$$\bar{R}_{i(jk)}^h(x_0) = 0, \quad P_{ij}^h(x_0), \quad a_{ij}(x_0),$$

which has to satisfy the condition (21). From this follows the above number of essential real parameters.

## CONCLUSION

In this paper we obtained the fundamental equations of geodesic mappings of spaces with affine connection onto Ricci symmetric manifolds and almost geodesic mappings of the first type of spaces with affine connection onto symmetric spaces. The fundamental equations have a closed Cauchy type form in covariant derivatives. We also set the number of essential real parameters for the general solution of such system.

## References

- [1] Berezovski, V.E., Mikeš, J. On the classification of almost geodesic mappings of affine-connected spaces. IN: *DGA, Proc. Conf.*, Dubrovnik/Yugosl. 1988, p. 41-48.
- [2] Berezovski, V.E., Mikeš, J. On canonical almost geodesic mappings of the first type of affinely connected spaces. Translation of *Izv. Vyssh. Uchebn. Zaved. Mat.* 2014, no. 2, 38. Russian Math. (Iz. VUZ) 58 2014, 2, p. 1-5.
- [3] Berezovski, V.E.; Guseva, N.I., Mikeš, J. On special first-type almost geodesic mappings of affine connection spaces preserving a certain tensor. (Russian); translated from *Mat. Zametki* 98. *Math. Not.*, 98:3, 2015, p. 515-518.
- [4] Berezovski, V.E., Mikeš, J., Peška, P. Geodesic mappings of manifolds with affine connection onto symmetric manifolds. In: *Geometry, integrability and quantization XVIII.*, Bulgar. Acad. Sci., Sofia, 2017, p. 99-104.
- [5] Berezovski, V.E., Bácsó, S., Mikeš, J. Diffeomorphism of affine connected spaces which preserved Riemannian and Ricci curvature tensors. *Miskolc Math. Notes*, 18:1, 2017, p. 117-124.
- [6] Berezovski, V.E., Mikeš, J., Vanžová, A. Fundamental PDE's of the canonical almost geodesic mappings of type  $\tilde{\pi}_1$ . *Bull. Malays. Math. Sci. Soc.*, 37:3, 2014, p. 647-659.
- [7] Cartan, É. Sur une classe remarquable d'espaces de Riemann. I, II. *Bull. S.M.F.* 54, 214-264, 1926; 55, 114-134, 1927.
- [8] Helgason, S. *Differential geometry, Lie groups, and symmetric spaces.* AMS, 1978.
- [9] Hinterleitner, I. Geodesic mappings on compact Riemannian manifolds with conditions on sectional curvature. *Publ. Inst. Math.*, 94:108, 2013, p. 125-130.

- [10] Hinterleitner, I., Mikeš, J. Fundamental equations of geodesic mappings and their generalizations. *J. Math. Sci.*, 174, 2011, p. 537–554.
- [11] Hinterleitner, I., Mikeš, J. Geodesic mappings and differentiability of metrics, affine and projective connections. *Filomat*, 29, 2015, p. 1245–1249.
- [12] Hinterleitner, I., Mikeš, J., Peška, P. On  $F_2^\varepsilon$ -planar mappings of (pseudo-) Riemannian manifolds. *Arch. Math.*, 50, 2014, p. 287–295.
- [13] Hinterleitner, I., Mikeš, J., Peška, P. Fundamental equations of F-planar mappings. *Lobachevskii J. Math.* 38:4, 2017, p. 653–659.
- [14] Levi-Civita, T. Sulle transformation delle dinamiche. *Ann. Mat. Milano*, Ser. 2, 24, 1896, p. 255–300.
- [15] Mikeš, J. Geodesic mappings of affine-connected and Riemannian spaces. *J. Math. Sci.*, 78, 1996, p. 311–333.
- [16] Mikeš, J. Holomorphically Projective mappings and their generalizations. *J. Math. Sci.*, 89, 1998, p. 1334–1353.
- [17] Mikeš, J. et al. *Differential geometry of special mappings*. Olomouc: Palacky Univ. Press, 2015, 566 pp.
- [18] Mikeš, J., Berezovski, V.E., Stepanova, E., Chudá, H. Geodesic mappings and their generalizations. *J. Math. Sci.*, 217:5, 2016, p. 607–623.
- [19] Mikeš, J., Vanžurová, A., Hinterleitner, I. *Geodesic mappings and some generalizations*. Olomouc: Palacky Univ. Press, 2009, 304 pp.
- [20] Najdanović, M.S., Velimirović, L.S. On the Willmore energy of curves under second order infinitesimal bending. *Miskolc Math. Not.*, 17:2, 2016, p. 979–987.
- [21] Najdanović, M.S., Zlatanović, M., Hinterleitner, I. Conformal and geodesic mappings of generalized equidistant spaces. *Publ. Inst. Math*, 98:112, 2015. p. 71–84.
- [22] Otsuki, T., Tashiro, Y. On curves in Kaehlerian spaces, *Math. J. Okayama Univ.*, 4, 1954, p. 57–78.
- [23] Petrov, A. Modeling of the paths of test particles in gravitation theory. *Gravit. and the Theory of Relativity*, 4:5, 1968, p. 7–21.
- [24] Peška, P., Mikeš, J., Chudá, H., Shiha, M. On Holomorphically Projective Mappings of Parabolic Kähler Spaces. *Miskolc Math. Not.*, 17:2, 2016, p. 1011–1019.
- [25] Prvanović, M. A note on holomorphically projective transformations of the Kähler spaces. *Tensor*, 35:1, 1981, p. 99–104.
- [26] Sinyukov, N.S. Geodesic mappings of Riemannian spaces. Nauka, Moscow, 1979, pp. 256.
- [27] Zlatanović, M., Velimirović, L., Stanković, M. Necessary and sufficient conditions for equitortion geodesic mapping. *J. Math. Anal. Appl.*, 435, 2016, p. 578–592.

## Acknowledgement

The paper was supported by the project IGA PrF 2017012 Palacký University Olomouc.

# MODIFICATIONS OF ITERATIVE AGGREGATION – DISAGGREGATION METHODS

**František Bubeník, Petr Mayer**

Faculty of Civil Engineering, Czech Technical University in Prague

Thákurova 7, 166 29 Praha 6, Czech Republic

Frantisek.Bubenik@cvut.cz, Petr.Mayer@cvut.cz

**Abstract:** *This paper deals with iterative aggregation – disaggregation methods (IAD Methods) which are a class of important numerical methods. The algorithm of the classical method and some modifications are introduced and convergence is investigated. An always - convergent iterative aggregation - disaggregation method is introduced and this is a new significant asset of this paper. Some properties of the method are derived.*

**Keywords:** iterative aggregation - disaggregation methods, numerical methods, Markov chains, stationary distributions, always – convergent algorithms.

## INTRODUCTION

We will deal with the search for the stationary probability distribution of a homogeneous Markovian chain. For a description of the chain we use the so called transition matrix, which is a column stochastic.

**Definition 1** *A matrix  $\mathbf{T} \in \mathfrak{R}^{n \times n}$  is a column stochastic matrix if its elements are non negative and  $\mathbf{e}^T \mathbf{T} = \mathbf{e}^T$ , where  $\mathbf{e} = (1, \dots, 1)^T \in \mathfrak{R}^n$ .*

We consider the problem

$$\mathbf{T}\pi = \pi, \quad \mathbf{e}^T \pi = \mathbf{1},$$

where  $\mathbf{T}$  is a column stochastic matrix and  $\pi$  is a stationary probability vector.

Such problems can be encountered in various important branches, for example, in particular in queueing theory and performance analysis or when investigating a quantitative risk and reliability analysis for signaling systems. Practical problems can be very large and a possibility how to eliminate the size is to apply the IAD Methods.

## 1 THE ITERATIVE AGGREGATION – DISAGGREGATION ALGORITHM

In this section we describe the classic iterative aggregation - disaggregation algorithm (IAD algorithm), see for example [1], [2] or [3].

We introduce an aggregation mapping

$$g: \{1, \dots, N\} \rightarrow \{1, \dots, n\}, n \ll N,$$

where  $n$  is the size of the coarse space.

The indices which are mapped to the same values of  $g$  define one aggregation group. The optimal choice of mapping  $g$  is difficult and often depends on further information about the solved problem. Distinctions between two choices of  $g$  for the same matrix  $\mathbf{T}$  can be substantial.

By means of aggregation mappings we define the restriction and prolongation matrices.

The **restriction matrix**  $\mathbf{R} \in \mathfrak{R}^{n \times N}$  is defined by nonzero elements  $r_{g(i),i} = 1$ , this is

$$(\mathbf{R} \mathbf{x})_j = \sum_{i=1, g(i)=j}^N x_i.$$

The **prolongation matrix**  $\mathbf{S}(\mathbf{x}) \in \mathfrak{R}^{N \times n}$  is parameterized by a vector  $\mathbf{x} \in \mathfrak{R}^N$ ; the nonzero elements of the matrix are

$$(\mathbf{S}(\mathbf{x}))_{i,g(i)} = \frac{x_i}{(\mathbf{R} \mathbf{x})_{g(i)}},$$

it means that  $(\mathbf{S}(\mathbf{x}) \mathbf{z})_i = z_{g(i)} x_i / (\mathbf{R} \mathbf{x})_{g(i)}$ .

As an illustrative example of an aggregation mapping we can introduce, for example, the following:

$$g : \{1, 2, 3\} \rightarrow 1, \quad g : \{4, 5, 6\} \rightarrow 2, \quad g : \{7, 8, 9\} \rightarrow 3.$$

Then the restriction and prolongation matrices are

$$\mathbf{R} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{S}(\mathbf{x}) = \left( \begin{array}{c|c|c} \frac{x_1 \ x_2 \ x_3}{(\mathbf{R} \mathbf{x})_1} & 0 & 0 \\ \hline 0 & \frac{x_4 \ x_5 \ x_6}{(\mathbf{R} \mathbf{x})_2} & 0 \\ \hline 0 & 0 & \frac{x_7 \ x_8 \ x_9}{(\mathbf{R} \mathbf{x})_3} \end{array} \right),$$

(where the symbol  $\frac{x_1 \ x_2 \ x_3}{(\mathbf{R} \mathbf{x})_1}$  stands for column  $\left( \frac{x_1}{(\mathbf{R} \mathbf{x})_1}, \frac{x_2}{(\mathbf{R} \mathbf{x})_1}, \frac{x_3}{(\mathbf{R} \mathbf{x})_1} \right)^T$ , the other symbols analogously).

One of the most important properties of these matrices is expressed in the following lemma.

**Lemma 1** *If  $\mathbf{x}$  is a positive vector then  $\mathbf{R} \mathbf{S}(\mathbf{x}) = \mathbf{I}$ .*

**Proof.** The proof follows from the definitions of  $\mathbf{R}$  and  $\mathbf{S}(\mathbf{x})$ .

Let us denote by  $\mathbf{A}(\mathbf{x}) = \mathbf{R} \mathbf{T} \mathbf{S}(\mathbf{x})$  the **aggregated matrix** defined by a vector  $\mathbf{x}$  and by an aggregation mapping  $g$ . Some properties of the matrix  $\mathbf{A}(\mathbf{x})$  are introduced in the following lemma.

**Lemma 2** *Let  $\mathbf{T}$  be a column stochastic matrix, let  $g$  be an aggregation mapping and  $\mathbf{x} \in \mathfrak{R}^N$  such that  $\mathbf{x} \geq \mathbf{0}$  and  $\mathbf{R} \mathbf{x} > \mathbf{0}$ . Then the aggregated matrix  $\mathbf{A}(\mathbf{x})$  is a column stochastic matrix. If the matrix  $\mathbf{T}$  is irreducible and the vector  $\mathbf{x}$  is strictly positive, then  $\mathbf{A}(\mathbf{x})$  is irreducible.*

With the previous knowledge we can define the following algorithm for an irreducible stochastic matrix  $\mathbf{T}$  and for a positive initial approximation  $\mathbf{x}_{\text{init}}$ .

Suppose that matrices  $\mathbf{W}_1$  and  $\mathbf{W}_2$  form the regular splitting of the matrix  $\mathbf{I} - \mathbf{T}$ . It means that  $\mathbf{I} - \mathbf{T} = \mathbf{W}_1 - \mathbf{W}_2$ , where  $\mathbf{W}_1$  is a  $M$ -matrix and where  $\mathbf{W}_2$  is a nonnegative matrix.

**Algorithm IAD 1** (input:  $\mathbf{T}$ ,  $\mathbf{W}_1$ ,  $\mathbf{W}_2$ ,  $\mathbf{x}_{\text{init}}$ ,  $\varepsilon$ ,  $g$ ,  $s$ ; output:  $\mathbf{x}$ )

1.  $k := 1$ ,  $\mathbf{x}_1 := \mathbf{x}_{\text{init}}$
2. while  $\|\mathbf{T} \mathbf{x}_k - \mathbf{x}_k\| > \varepsilon$  do
3.    $\tilde{\mathbf{x}} := (\mathbf{W}_1^{-1} \mathbf{W}_2)^s \mathbf{x}_k$
4.    $\mathbf{A}(\tilde{\mathbf{x}}) := \mathbf{R} \mathbf{T} \mathbf{S}(\tilde{\mathbf{x}})$
5.   solve  $\mathbf{A}(\tilde{\mathbf{x}}) \mathbf{z} = \mathbf{z}$  and  $\mathbf{e}^T \mathbf{z} = 1$
6.    $k := k + 1$
7.    $\mathbf{x}_k = \mathbf{S}(\tilde{\mathbf{x}}) \mathbf{z}$
8. end while

Convergence theory for the Algorithm IAD 1 is not quite clear generally. There is a significant theorem which is based on Theorem 3.2. proved in paper [2]. Using our notation it can be reformulated as follows

**Theorem 1** *Let  $\mathbf{T}$  be an irreducible column stochastic matrix. Let  $\mathbf{W}_1$  be the identity matrix and  $s = 1$  in the step 3 of the Algorithm. Then the Algorithm IAD 1 is locally convergent.*

Next convergence theories are, for example, in [1] or [3], but they all require some additional assumptions.

## 2 AN ALWAYS – CONVERGENT VARIANT OF THE IAD METHOD

We now introduce a significant alternative IAD method. Let again

$$\mathbf{I} - \mathbf{T} = \mathbf{W}_1 - \mathbf{W}_2.$$

Then

$$\mathbf{0} = \mathbf{e}^T (\mathbf{I} - \mathbf{T}) = \mathbf{e}^T (\mathbf{W}_1 - \mathbf{W}_2) = \mathbf{e}^T \mathbf{W}_1 - \mathbf{e}^T \mathbf{W}_2.$$

It implies that

$$\mathbf{e}^T \mathbf{W}_1 = \mathbf{e}^T \mathbf{W}_2.$$

Since  $\mathbf{W}_1$  is a  $M$ -matrix then it has a non-negative inverse, we can then write

$$\mathbf{e}^T = \mathbf{e}^T \mathbf{W}_2 \mathbf{W}_1^{-1}.$$

Denote

$$\tilde{\mathbf{T}} = \mathbf{W}_2 \mathbf{W}_1^{-1}.$$

We can see that

$$\mathbf{e}^T = \mathbf{e}^T \tilde{\mathbf{T}}$$

and because  $\mathbf{W}_2$  and  $\mathbf{W}_1^{-1}$  are non-negative, then

$$\tilde{\mathbf{T}} \geq \mathbf{0}.$$



Thus  $\tilde{\mathbf{T}}$  is a stochastic matrix. Then the matrix  $\hat{\mathbf{T}}$  created as

$$\hat{\mathbf{T}} = \frac{1}{2}(\mathbf{I} + \tilde{\mathbf{T}})$$

is also stochastic and  $\text{diag}(\tilde{\mathbf{T}}) > 0$ .

**Lemma 3** *Let  $\hat{\mathbf{T}}\rho = \rho$ . Then  $\pi = \mathbf{W}_1^{-1}\rho$  is a solution of the problem*

$$\mathbf{T}\pi = \pi.$$

**Proof.** It is easy to see that  $\hat{\mathbf{T}}$  and  $\tilde{\mathbf{T}}$  have the same eigenvectors and therefore it is sufficient to prove the Lemma for  $\tilde{\mathbf{T}}$ . From the assumption we have

$$\tilde{\mathbf{T}}\rho = \rho$$

this is

$$\mathbf{W}_2\mathbf{W}_1^{-1}\rho = \rho.$$

If we denote  $\sigma = \mathbf{W}_1^{-1}\rho$ , we get

$$\mathbf{W}_2\sigma = \mathbf{W}_1\sigma$$

and then successively

$$\mathbf{W}_1\sigma - \mathbf{W}_2\sigma = \mathbf{0}$$

$$(\mathbf{W}_1 - \mathbf{W}_2)\sigma = \mathbf{0}$$

$$(\mathbf{I} - \mathbf{T})\sigma = \mathbf{0}$$

and thus we get

$$\mathbf{T}\sigma = \sigma.$$

As well, it is seen that the relation between the eigenvector of  $\tilde{\mathbf{T}}$ , which is  $\rho$ , and the eigenvector of  $\mathbf{T}$  is that  $\pi = \mathbf{W}_1^{-1}\rho$ . Then  $\sigma = \pi$  and the Lemma 3 is proved.

**Remark 1**  $\tilde{\mathbf{T}}$  and  $\hat{\mathbf{T}}$  have the same eigenvectors for eigenvalue 1 (In fact, they have the same all eigenvectors).

We now consider IAD 1 with  $\hat{\mathbf{T}}$  instead of  $\mathbf{T}$ . New decomposition is that  $\mathbf{I} - \hat{\mathbf{T}} = \mathbf{W}_1 - \mathbf{W}_2$ , where  $\mathbf{W}_1 = \mathbf{I}$  and then  $\mathbf{W}_2 = \hat{\mathbf{T}}$ .

Then the modified steps 3. and 4. from the algorithm IAD 1 are as follows:

3.  $\tilde{\mathbf{x}} := \hat{\mathbf{T}} \mathbf{x}_k$ ,
4.  $\mathbf{A}(\tilde{\mathbf{x}}) := \mathbf{R} \hat{\mathbf{T}} \mathbf{S}(\tilde{\mathbf{x}})$ .

The approximations in 3. can be expressed as

$$\hat{\mathbf{T}} \mathbf{x}_k = \frac{1}{2}(\mathbf{I} + \tilde{\mathbf{T}})\mathbf{x}_k = \frac{1}{2}\mathbf{x}_k + \frac{1}{2}\tilde{\mathbf{T}}\mathbf{x}_k = \frac{1}{2}\mathbf{x}_k + \frac{1}{2}\mathbf{W}_2\mathbf{W}_1^{-1}\mathbf{x}_k = \frac{1}{2}\mathbf{x}_k + \frac{1}{2}\mathbf{W}_2\pi_k,$$

where  $\pi_k = \mathbf{W}_1^{-1} \mathbf{x}_k$  is the  $k$ -th approximation of  $\pi$ .

Then, using Lemma 1, we can write in step 4.

$$\mathbf{R} \hat{\mathbf{T}} \mathbf{S}(\tilde{\mathbf{x}}) = \mathbf{R} \frac{1}{2}(\mathbf{I} + \tilde{\mathbf{T}}) \mathbf{S}(\tilde{\mathbf{x}}) = \frac{1}{2} \left[ \mathbf{R} \mathbf{S}(\tilde{\mathbf{x}}) + \mathbf{R} \tilde{\mathbf{T}} \mathbf{S}(\tilde{\mathbf{x}}) \right] = \frac{1}{2} \left[ \mathbf{I} + \mathbf{R} \tilde{\mathbf{T}} \mathbf{S}(\tilde{\mathbf{x}}) \right] = \frac{1}{2}(\mathbf{I} + \mathbf{B}(\tilde{\mathbf{x}})),$$

where  $\mathbf{B}(\tilde{\mathbf{x}}) = \mathbf{R} \tilde{\mathbf{T}} \mathbf{S}(\tilde{\mathbf{x}})$ . We can also use the form  $\mathbf{B}(\tilde{\mathbf{x}}) = \mathbf{R} \mathbf{W}_2 \mathbf{W}_1^{-1} \mathbf{S}(\tilde{\mathbf{x}})$ .

We have got a new modified algorithm:

**Algorithm IAD 2** (input:  $\mathbf{T}$ ,  $W_1$ ,  $W_2$ ,  $\pi_{init}$ ,  $\varepsilon$ ,  $g$ ; output:  $\pi$ )

1.  $k := 1$ ,  $\pi_1 := \pi_{init}$ ,  $\mathbf{x}_1 := \mathbf{W}_1 \pi_1$
2. while  $\|\mathbf{T} \pi_k - \pi_k\| > \varepsilon \cdot \mathbf{e}^T \pi_k$  do
3.    $\tilde{\mathbf{x}} := \frac{1}{2} \mathbf{x}_k + \frac{1}{2} \mathbf{W}_2 \pi_k$ , where  $\pi_k := \mathbf{W}_1^{-1} \mathbf{x}_k$
4.    $\mathbf{A}(\tilde{\mathbf{x}}) := \frac{1}{2}(\mathbf{I} + \mathbf{B}(\tilde{\mathbf{x}}))$ , where  $\mathbf{B}(\tilde{\mathbf{x}}) := \mathbf{R} \mathbf{W}_2 \mathbf{W}_1^{-1} \mathbf{S}(\tilde{\mathbf{x}})$
5.   solve  $\mathbf{A}(\tilde{\mathbf{x}}) \mathbf{z} = \mathbf{z}$  and  $\mathbf{e}^T \mathbf{z} = 1$
6.    $k := k + 1$
7.    $\mathbf{x}_k = \mathbf{S}(\tilde{\mathbf{x}}) \mathbf{z}$
8. end while
9.  $\pi := \pi_k / (\mathbf{e}^T \pi_k)$

**Theorem 2** *The Algorithm IAD 2 converges locally for every irreducible column stochastic matrix, arbitrary choice of the aggregation mapping and arbitrary regular splitting of the matrix  $\mathbf{I} - \mathbf{T}$ .*

The proof is clear from Theorem 1.

## CONCLUSION

In this work there is introduced a new algorithm for the iterative aggregation – disaggregation method. This algorithm is locally convergent for every irreducible column stochastic transition matrix, for any aggregation mapping and for arbitrary regular splitting of the matrix  $\mathbf{I} - \mathbf{T}$ . It is a new asset and added value of the authors. There are no other requirements and the convergence is without any additional conditions.

## References

- [1] Marek I., Mayer P.: Iterative aggregation – disaggregation methods for computing some characteristics of Markov chains, *Large Scale Scientific Computing, Third International Conference, LSSC 2001*, pp. 68–82, Sozopol, Bulgaria, 2001.
- [2] Pultarová I.: Local convergence analysis of iterative aggregation – disaggregation methods with polynomial correction, *Linear Algebra and its Applications*, 421 (2007), pp. 122–137, 2007.
- [3] Stewart W. J.: *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, 1994.

# A PROBLEM OF FUNCTIONAL MINIMIZING FOR SINGLE DELAYED DIFFERENTIAL SYSTEM

Hanna Demchenko, Josef Diblík

Faculty of Electrical Engineering and Communication, Brno University of Technology

Technická 3058/10, 61600, Brno, Czech Republic

xdemch02@stud.feec.vutbr.cz, diblik@feec.vutbr.cz

**Abstract:** *In the contribution, a linear differential system with a single delay*

$$\frac{dx(t)}{dt} = A_0x(t) + A_1x(t - \tau) + bu(t), \quad t \geq t_0$$

where  $A_0, A_1$  are  $n \times n$  constant matrices,  $x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^n$ ,  $\tau > 0$ ,  $t_0 \in \mathbb{R}$ ,  $u \in \mathbb{R}$ , is considered. A problem of minimizing (by a suitable control function  $u(t)$ ) a functional

$$I = \int_{t_0}^{\infty} (x^T(t)C_{11}x(t) + x^T(t)C_{12}x(t - \tau) + x^T(t - \tau)C_{21}x(t) + x^T(t - \tau)C_{22}x(t - \tau) + du^2(t))dt,$$

where  $C_{11}, C_{12}, C_{21}, C_{22}$  are  $n \times n$  constant matrices,  $d > 0$ , and the integrand is a positive-definite quadratic form, is considered. To solve the problem, Malkin's approach and Lyapunov's second method are utilized.

**Keywords:** delayed differential system, Lyapunov-Krasovskii functional, integral quality criterion, optimal control.

## INTRODUCTION

Assume that a system of differential equations of delayed type

$$x'(t) = f(t, x_t, u(t)), \quad t \geq t_0, \quad (1)$$

describes a process controlled by a vector-function  $u: [t_0, \infty) \rightarrow \mathbb{R}^m$ . Let, in (1),  $t_0 \in \mathbb{R}$ ,  $f: D \rightarrow \mathbb{R}^n$ ,

$$D := \{(t, x, u) \in [t_0, \infty) \times C_\tau^n \times \mathbb{R}^m, \|x\|_\tau < M, \|u\| < M\},$$

$M > 0$ ,  $n, m \in \mathbb{N}$ ,  $C_\tau^n = C([- \tau, 0], \mathbb{R}^n)$  is the space of continuous mappings from the interval  $[- \tau, 0]$  into  $\mathbb{R}^n$ ,

$$\|x(t)\|_\tau := \max_{\theta \in [- \tau, 0]} (\|x(t + \theta)\|), \quad t \geq t_0,$$

$$\|x(s)\| := \max_{i=1, \dots, n} \{|x_i(s)|\}, \quad s \in [t_0 - \tau, \infty)$$

and  $x_t \in C_\tau^n$  is defined by  $x_t(\theta) := x(t + \theta)$ ,  $\theta \in [- \tau, 0]$ .

Below we assume that

1.  $f(t, \theta_n^*, \theta_m) = \theta_n, t \geq t_0$ , where  $\theta_n, \theta_m$  are  $n$  and  $m$  dimensional zero vectors and  $\theta_n^* \in C_\tau^n$  is a zero vector-function.
2. The functional  $f$  is locally Lipschitzian in every bounded neighborhood of each point  $(t^*, x, u) \in D$ .

Let us formulate the following problem for system (1): Determine a control function  $u: [t_0, \infty) \rightarrow \mathbb{R}^m$  such that zero solution  $x(t) = \theta_n, t \geq t_0$  of system (1) will be asymptotically stable and for an arbitrary solution  $x = x(t), t \geq t_0 - \tau$  of system (1) satisfying  $\|x\| < M$ , the integral

$$\int_{t_0}^{\infty} \omega(t, x_t, u(t)) dt \quad (2)$$

exists and attains minimum value (provided that  $\omega: D \rightarrow \mathbb{R}$  is a positive-definite functional and that the indefinite integral exists).

To formulate a result related to this problem we need to define an auxiliary functional

$$V: [t_0, \infty) \times C_\tau^n \rightarrow \mathbb{R}.$$

Below we assume that there exists the derivative  $dV(t, x_t)/dt$  of functional  $V(t, x_t)$  along trajectories of system (1).

Repeating well-known definitions ([3], see also [2]), we say that functional  $V$  is positive definite if there exists a continuous nondecreasing function  $w$  on  $[0, \infty)$  which is zero at 0 and positive on  $(0, \infty)$  such that

$$V(t, x_t) \geq w(\|x(t)\|), \quad t \geq t_0$$

where  $x$  is assumed to be defined on  $[t_0 - r, \infty)$ .

Functional  $V$  has an infinitesimal upper bound if there exists a continuous nondecreasing function  $W$  on  $[0, \infty)$  which is zero at 0 and positive on  $(0, \infty)$  such that

$$V(t, x_t) \leq W(\|x_t\|_r), \quad t \geq t_0.$$

A positive-definite functional  $V: (\alpha, \infty) \times C_\tau^n(D) \rightarrow \mathbb{R}$  having an infinitesimal upper bound is called a Lyapunov-Krasovskii functional.

Define an auxiliary functional  $B: D_1 \rightarrow \mathbb{R}$  where

$$D_1 := \{(v, t, x, u) \in \mathbb{R} \times [t_0, \infty) \times C_\tau^n \times \mathbb{R}^m, \|x\|_\tau < M, \|u\| < M\},$$

by formula

$$B(V, t, x_t, u) := \frac{dV(t, x_t)}{dt} + \omega(t, x_t, u(t)) \quad (3)$$

The following theorem, motivated by optimality results for non-delayed systems in the book by Malkin [4, Theorem IV] (wherein Lyapunov's second method is utilized for the proof), is true.

**Theorem 1** Assume that, for the system of differential equations (1), there exists a positive definite functional  $V$  having an infinitesimal upper bound and a vector-function  $u_0: [t_0, \infty) \rightarrow \mathbb{R}^m$ ,  $\|u_0(t)\| \leq M$ ,  $t \geq t_0$  such that

i) Functional  $\omega(t, x_t, u_0(t))$  is positive-definite for every  $t \geq t_0$ ,  $\|x_t\|_\tau < M$ .

ii) Identity  $B(V, t, x_t, u_0(t)) \equiv 0$  is true on  $[t_0, \infty)$  for every solution

$$x: [t_0 - \tau, \infty) \rightarrow \mathbb{R}^n$$

of system (1) where  $u = u_0$ .

iii) Inequality  $B(V, t, x_t, u(t)) \geq 0$  holds on  $[t_0, \infty)$  for every solution

$$x: [t_0 - \tau, \infty) \rightarrow \mathbb{R}^n$$

of system (1) and for every vector-function  $u: [t_0, \infty) \rightarrow \mathbb{R}^m$  with

$$\|u(t)\| < M, \quad t \in [t_0, \infty).$$

Then, the function  $u_0$  is a solution of the problem (1), (2) and

$$\int_{t_0}^{\infty} \omega(t, x_t, u_0(t)) dt = \min_u \left[ \int_{t_0}^{\infty} \omega(t, x_t, u(t)) dt \right] = V(t_0, x_{t_0}).$$

In the following part we apply Theorem 1 to a linear differential system with a single delay.

## 1 SYSTEM WITH A SINGLE DELAY AND A SCALAR CONTROL FUNCTION

Consider linear systems with constant coefficients, a single delay and a scalar control function

$$\frac{dx(t)}{dt} = A_0 x(t) + A_1 x(t - \tau) + bu(t), \quad t \geq 0, \quad (4)$$

where  $A_0, A_1$  are  $n \times n$  constant matrices,  $x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^n$ ,  $\tau > 0$  and  $u \in \mathbb{R}$ . We need to find a control function  $u = u_0(t)$  such that the system is asymptotically stable and an integral quality criterion

$$I = \int_0^{\infty} (x^T(t)C_{11}x(t) + x^T(t)C_{12}x(t - \tau) + x^T(t - \tau)C_{21}x(t) + x^T(t - \tau)C_{22}x(t - \tau) + du^2(t)) dt. \quad (5)$$

takes a minimum value provided that  $d > 0$ ,  $n \times n$  constant matrices  $C_{11}, C_{22}$  and  $2n \times 2n$  matrix

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

are positive-definite symmetric matrices where  $C_{21}$ ,  $C_{12}$  are  $n \times n$  constant matrices and  $C_{21} = C_{12}^T$ . Below, by  $\Theta$ , is denoted  $n \times n$  null matrix.

In the following theorem we use a Lyapunov-Krasovskii functional

$$V(t, x_t) = x^T(t)Hx(t) + \int_{t-\tau}^t x^T(s)Gx(s)ds, \quad (6)$$

where  $n \times n$  matrices  $H$  and  $G$  are symmetric positive-definite. Such kind of functional is often used (see, e.g. [1] and the references therein).

**Theorem 2** Assume that matrices  $H$  and  $G$  satisfying the matrix equation

$$A_0^T H + HA_0 + G + C_{11} - \frac{1}{d}Hbb^T H = \Theta. \quad (7)$$

If, moreover,

$$HA_1 + C_{12} = \Theta \quad (8)$$

and

$$C_{22} = G, \quad (9)$$

then the optimal control function of problem (4), (5) exists and equals

$$u_0(t) = -\frac{1}{d}b^T Hx(t), \quad t \geq 0. \quad (10)$$

Moreover, system (4) with  $u(t) = u_0(t)$ , i.e.,

$$\frac{dx(t)}{dt} = A_0x(t) + A_1x(t - \tau) + bu_0(t), \quad t \geq 0,$$

is asymptotically stable and

$$\begin{aligned} V(t_0, x(t_0)) &= \int_0^\infty (x^T(t)C_{11}x(t) + x^T(t)C_{12}x(t - \tau) + x^T(t - \tau)C_{21}x(t) \\ &\quad + x^T(t - \tau)C_{22}x(t - \tau) + du_0^2(t))dt = \min_u \int_0^\infty (x^T(t)C_{11}x(t) + x^T(t)C_{12}x(t - \tau) \\ &\quad + x^T(t - \tau)C_{21}x(t) + x^T(t - \tau)C_{22}x(t - \tau) + du^2(t)) dt. \end{aligned}$$

**PROOF.** We utilize Theorem 1. Let  $t_0 = 0$ . In accordance with condition *ii)* of Theorem 1 we analyze the expression  $B$  given by (3), i.e.,

$$\begin{aligned} B(V, t, x_t, u_0) &= [A_0x(t) + A_1x(t - \tau) + bu_0(t)]^T Hx(t) + x^T(t)H[A_0x(t) + A_1x(t - \tau) + bu_0(t)] \\ &\quad + x^T(t)Gx(t) - x^T(t - \tau)Gx(t - \tau) + x^T(t)C_{11}x(t) + x^T(t)C_{12}x(t - \tau) \\ &\quad + x^T(t - \tau)C_{21}x(t) + x^T(t - \tau)C_{22}x(t - \tau) + du_0^2(t) \equiv 0. \end{aligned}$$

Simplifying the last expression, we get

$$\begin{aligned} B(V, t, x_t, u_0) &= x^T(t)[A_0^T H + H A_0 + G + C_{11}]x(t) + x^T(t - \tau)[A_1^T H + C_{21}]x(t) \\ &+ x^T(t)[H A_1 + C_{12}]x(t - \tau) + x^T(t - \tau)[C_{22} - G]x(t - \tau) + 2x^T(t)H b u_0(t) + d u_0^2(t) \equiv 0. \end{aligned} \quad (11)$$

Looking for an extremum of (11), we get

$$B'_{u_0}(V, t, x_t, u_0(t)) = 2b^T H x(t) + 2d u_0(t) = 0,$$

i.e.,

$$u_0(t) = -\frac{1}{d} b^T H x(t),$$

which is the minimum of the function  $B$  because

$$B''_{u_0 u_0}(V, t, x_t, u_0(t)) = 2d > 0.$$

For (11) to be valid, i.e.,

$$\begin{aligned} B(V, t, x_t, u_0) &= x^T(t)[A_0^T H + H A_0 + G + C_{11} - \frac{1}{d} H b b^T H]x(t) \\ &+ x^T(t - \tau)[A_1^T H + C_{21}]x(t) + x^T(t)[H A_1 + C_{12}]x(t - \tau) \\ &+ x^T(t - \tau)[C_{22} - G]x(t - \tau) \equiv 0. \end{aligned}$$

we obtain

$$\begin{aligned} A_0^T H + H A_0 + G + C_{11} - \frac{1}{d} H b b^T H &= \Theta, \\ H A_1 + C_{12} &= \Theta, \\ C_{22} &= G. \end{aligned}$$

Thus, for the control function defined by (10) and the Lyapunov-Krasovskii functional (6), the system (4) is asymptotically stable and the quality criterion (5) takes a minimum value.  $\square$

**Example.** Consider system (4) with

$$A_0 = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -1 & -0.1 \\ -0.5 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

i.e.,

$$\begin{aligned} \dot{x}_1(t) &= -2x_1(t) + x_2(t) - x_1(t - \tau) - 0.1x_2(t - \tau) + u(t), \\ \dot{x}_2(t) &= x_1(t) - 2x_2(t) - 0.5x_1(t - \tau) - x_2(t - \tau) + u(t) \end{aligned}$$

with a quadratic quality criterion (5) with

$$C_{11} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}, \quad C_{12} = \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix}, \quad C_{21} = \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix}, \quad C_{22} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}, \quad d = 1,$$

i.e.,

$$\begin{aligned}
I = \int_0^\infty & \left[ (x_1(t), x_2(t)) \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + (x_1(t), x_2(t)) \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} \begin{pmatrix} x_1(t-\tau) \\ x_2(t-\tau) \end{pmatrix} \right. \\
& + (x_1(t-\tau), x_2(t-\tau)) \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \\
& \left. + (x_1(t-\tau), x_2(t-\tau)) \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x_1(t-\tau) \\ x_2(t-\tau) \end{pmatrix} + u^2(t) \right] dt.
\end{aligned}$$

By formula (10) we obtain the optimal stabilization control function in the form

$$u_0(t) = -(h_1 + h_2)x_1 - (h_2 + h_3)x_2. \quad (12)$$

We need to find matrix  $H$ . In our case we can compute expression (7), using (9), i.e.,

$$\begin{aligned}
& A_0^T H + H A_0 + G + C_{11} - \frac{1}{d} H b b^T H = \\
& = \begin{pmatrix} -4h_1 + 2h_2 + 6 - (h_1 + h_2)^2 & h_1 - 4h_2 + h_3 - (h_1 + h_2)(h_2 + h_3) \\ h_1 - 4h_2 + h_3 - (h_1 + h_2)(h_2 + h_3) & 2h_2 - 4h_3 + 6 - (h_2 + h_3)^2 \end{pmatrix} = \Theta.
\end{aligned}$$

which means that

$$\begin{cases} -4h_1 + 2h_2 + 6 - (h_1 + h_2)^2 = 0, \\ h_1 - 4h_2 + h_3 - (h_1 + h_2)(h_2 + h_3) = 0, \\ 2h_2 - 4h_3 + 6 - (h_2 + h_3)^2 = 0. \end{cases} \quad (13)$$

To solve it we can, for example, add the first, the third and the second (multiplied by 2) equations. We obtain

$$-2h_1 - 4h_2 - 2h_3 + 12 - [(h_1 + h_2) + (h_2 + h_3)]^2 = -2[h_1 + 2h_2 + h_3] + 12 - [h_1 + 2h_2 + h_3]^2 = 0.$$

If put

$$h_1 + 2h_2 + h_3 = K, \quad (14)$$

then we have

$$K^2 + 2K - 12 = 0$$

and  $K = -1 \pm \sqrt{13}$ .

After subtracting the first equation from the third, we obtain

$$4h_1 - 4h_3 + (h_1 + h_2)^2 - (h_2 + h_3)^2 = 4(h_1 - h_3) + (h_1 + 2h_2 + h_3)(h_1 - h_3) = (h_1 - h_3)(4 + K) = 0$$

and

$$h_1 = h_3.$$

Using the last equation to (14) we find



$$h_1 + h_2 = \frac{K}{2}. \quad (15)$$

The second equation of system (13), turns into

$$2h_1 - 4h_2 - (h_1 + h_2)^2 = 0 \Rightarrow h_1 - 2h_2 = \frac{K^2}{8}. \quad (16)$$

From (15) and (16) we find that

$$h_1 = h_3 = \frac{K}{3} + \frac{K^2}{24},$$

$$h_2 = \frac{K}{6} - \frac{K^2}{24}.$$

Also (8) should be valid, so

$$C_{12} = -HA_1.$$

For  $K = -1 - \sqrt{13}$  matrix  $H$  is not positive definite, so by (12) the optimal stabilization control function will be

$$u_0(t) = \frac{1 - \sqrt{13}}{2}(x_1(t) + x_2(t)).$$

## CONCLUSION

In the paper we extended Malkin's approach, utilizing Lyapunov's second method, to solve optimal stabilization problem for linear delayed differential system with a single delay and a scalar control function. In spite of Malkin's approach making it possible to find optimal control functions for large classes of systems of ordinary linear differential systems and minimizing problems, the results that can be derived for the linear delayed differential systems considered are not so general and cover only narrow classes of problems.

## References

- [1] Baštinec, J., Diblík, J., Khusainov, D.Ya., Ryvolová, A. Exponential stability and estimation of solutions of linear differential systems of neutral type with constant coefficients. In: *Boundary Value Problems*. 2010. Available at: <<https://doi.org/10.1155/2010/956121>>. ISSN 1687-2770.
- [2] Baštinec, J., Klimešová, M. Stability of the zero solution of stochastic differential systems with four-dimensional Brownian motion. In: *Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 7-30. [Online]. [Cit. 2017-07-26]. Available at: <<http://mitav.unob.cz/data/MITAV2016Proceedings.pdf>>. ISBN 978-80-7231-400-3.

- [3] Elsgolc, L.E., Norkin, S.B.: *Introduction to the Theory and Application of Differential Equations with an Deviating Argument*. Elsevier, 1973.
- [4] Malkin, I.G.: *Theory of Stability of Motion*. Second revised edition, (Russian), Moscow: Nauka Publisher, 1966, 530 pp.

### **Acknowledgement**

The work presented in this paper has been supported by the Grant of Faculty of Electrical Engineering and Communication, BUT (research project No. FEKT-S-17-4225).

# GENERAL SOLUTION OF WEAKLY DELAYED LINEAR SYSTEMS WITH VARIABLE COEFFICIENTS

**J. Diblík, H. Halfarová**

Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering  
Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic  
diblik@feec.vutbr.cz, halfarova.h@fce.vutbr.cz

**Abstract:** *Weakly delayed planar linear discrete systems with variable coefficients are considered. For one of the possible cases of the roots of the matrix of linear non-delayed terms, general solution is constructed. The dimensionality of the space of solutions is discussed as well.*

**Keywords:** discrete linear system, weakly delayed system, delay.

## INTRODUCTION

Let  $s$  and  $q$  be integers such that  $s \leq q$  and define a set of integers

$$\mathcal{Z}_s^q := \{s, s+1, \dots, q\}$$

used throughout the paper. Similarly, the sets with infinite boundaries  $\mathcal{Z}_s^\infty$ ,  $\mathcal{Z}_\infty^q$  and  $\mathcal{Z}_\infty^\infty$  can be defined. In [1], linear discrete planar systems with constant coefficients

$$x(k+1) = Ax(k) + Bx(k-m) \quad (1)$$

were considered where  $m > 0$  is a fixed integer,  $k \in \mathcal{Z}_0^\infty$ ,  $A = (a_{ij})$ ,  $i, j = 1, 2$  and  $B = (b_{ij})$ ,  $i, j = 1, 2$  are constant  $2 \times 2$  matrices, and  $x: \mathcal{Z}_{-m}^\infty \rightarrow \mathbb{R}^2$ . The number  $m$  represents a delay in (1). Formulas for a general solution were found in [1], provided that the system (1) is a system with weak delay as defined below.

**Definition 1** *The system (1) is called a system with weak delay if, for every  $\lambda \in \mathbb{C} \setminus \{0\}$ ,*

$$\det(A + \lambda^{-m}B - \lambda I) = \det(A - \lambda I). \quad (2)$$

Later, systems with weak delay have been considered, e.g., in [2, 3]. Since the property of being a system with a weak delay is related, as (2) suggests, with the coefficients of the matrices  $A$ ,  $B$  rather than with the delay  $m$ , in [3] with such systems, the term *weakly delayed systems* rather than *weak delay* is used.

The Definition 1 is applicable for planar as well as higher-dimensional systems [4, 5].

Let an initial (Cauchy) problem for equation (1) be given by the relation

$$x(k) = \varphi(k) \quad (3)$$

where  $k = -m, -m+1, \dots, 0$  and  $\varphi = (\varphi_1, \varphi_2)^T: \mathcal{Z}_{-m}^0 \rightarrow \mathbb{R}^2$ . Due to the linearity of (1), the initial problem (1), (3) has a unique solution on  $\mathcal{Z}_{-m}^\infty$ . For completeness, the solution  $x: \mathcal{Z}_{-m}^\infty \rightarrow \mathbb{R}^2$  of (1), (3) is defined as an infinite sequence

$$\{x(-m) = \varphi(-m), x(-m+1) = \varphi(-m+1), \dots, x(0) = \varphi(0), x(1), x(2), \dots, x(k), \dots\}$$

if, for any  $k \in \mathcal{Z}_0^\infty$ , equality (1) holds.

## 0.1 Weakly delayed systems with variable coefficients

Consider linear discrete planar systems with variable coefficients

$$x(k+1) = A(k)x(k) + B(k)x(k-m), \quad k \in \mathcal{Z}_0^\infty \quad (4)$$

where, unlike the system (1),  $A(k) = (a_{ij}(k))$ ,  $i, j=1,2$  and  $B(k) = (b_{ij}(k))$ ,  $i, j=1,2$  are  $2 \times 2$  matrices with variable coefficients. Definition 1 can be modified to system (4) as follows:

**Definition 2** *The system (4) is called a system with weak delay if, for every  $\lambda \in \mathcal{C} \setminus \{0\}$  and every fixed  $k \in \mathcal{Z}_0^\infty$ ,*

$$\det(A(k) + \lambda^{-m}B(k) - \lambda I) = \det(A(k) - \lambda I). \quad (5)$$

Let us consider a linear transformation

$$x(k) = Sy(k), \quad k \in \mathcal{Z}_{-m}^\infty \quad (6)$$

with the nonsingular  $2 \times 2$  matrix  $S$ . Then, (4) is transformed to

$$y(k+1) = S^{-1}A(k)Sy(k) + S^{-1}B(k)Sy(k-m), \quad k \in \mathcal{Z}_0^\infty. \quad (7)$$

In [1] is showed that, if a system (1) if weakly delayed, then its arbitrary linear nonsingular transformation again leads to a weakly delayed system. The same property holds for weakly delayed systems with variable coefficients.

**Lemma 1** *If the system (4) is weakly delayed, then the system (7) is weakly delayed provided that the transformation (6) is nonsingular.*

PROOF. Assume that (5) holds for every  $\lambda \in \mathcal{C} \setminus \{0\}$  and every fixed  $k \in \mathcal{Z}_0^\infty$ . Then,

$$\begin{aligned} \det(S^{-1}A(k)S + \lambda^{-m}S^{-1}B(k)S - \lambda I) \\ &= \det[S^{-1}(A(k) + \lambda^{-m}B(k) - \lambda I)S] \\ &= \det(A(k) + \lambda^{-m}B(k) - \lambda I) \\ &= \det(A(k) - \lambda I) \\ &= \det[S^{-1}(A(k) - \lambda I)S] \\ &= \det(S^{-1}A(k)S - \lambda I). \end{aligned}$$

□

## 0.2 Coefficient criterion for determining a weakly delayed system

It is easy to find conditions that are necessary and sufficient for a system to be weakly delayed.

**Theorem 1** *System (4) is weakly delayed if and only if*

$$\text{tr } B(k) = \det B(k) = 0, \quad k \in \mathcal{Z}_0^\infty \quad (8)$$

and

$$\det(A(k) + B(k)) = \det A(k), \quad k \in \mathcal{Z}_0^\infty. \quad (9)$$

PROOF. Computing the determinant on the left-hand side of equation (5), we derive

$$\begin{aligned}
& \det(A(k) + \lambda^{-m}B(k) - \lambda I) \\
&= \begin{vmatrix} a_{11}(k) + b_{11}(k)\lambda^{-m} - \lambda & a_{12}(k) + b_{12}(k)\lambda^{-m} \\ a_{21}(k) + b_{21}(k)\lambda^{-m} & a_{22}(k) + b_{22}(k)\lambda^{-m} - \lambda \end{vmatrix} \\
&= \begin{vmatrix} a_{11}(k) - \lambda & a_{12}(k) \\ a_{21}(k) & a_{22}(k) - \lambda \end{vmatrix} - \lambda^{-m+1}(b_{11}(k) + b_{22}(k)) \\
&\quad + \lambda^{-m} \left[ \begin{vmatrix} a_{11}(k) & a_{12}(k) \\ b_{21}(k) & b_{22}(k) \end{vmatrix} + \begin{vmatrix} b_{11}(k) & b_{12}(k) \\ a_{21}(k) & a_{22}(k) \end{vmatrix} \right] \\
&\quad + \lambda^{-2m} \begin{vmatrix} b_{11}(k) & b_{12}(k) \\ b_{21}(k) & b_{22}(k) \end{vmatrix} \\
&= \begin{vmatrix} a_{11}(k) - \lambda & a_{12}(k) \\ a_{21}(k) & a_{22}(k) - \lambda \end{vmatrix} - \lambda^{-m+1} \text{tr } B(k) \\
&\quad + \text{tr } B(k) + \lambda^{-2m} \det B(k).
\end{aligned}$$

Then, (5) will hold if and only if (8) and (9) are valid since, in that case,

$$\det(A(k) + \lambda^{-m}B(k) - \lambda I) = \det(A(k) - \lambda I) = \begin{vmatrix} a_{11}(k) - \lambda & a_{12}(k) \\ a_{21}(k) & a_{22}(k) - \lambda \end{vmatrix}.$$

□

## 1 PROBLEM UNDER CONSIDERATION AND RESULTS

A complete investigation of the system (1) is carried out in [1]. The construction of a general solution is given for every case of the Jordan form of the matrix  $A$  of non-delayed linear terms except for the case of two complex conjugate roots of the characteristic equation  $\det(A - \lambda I) = 0$ . For general systems of linear equations with variable coefficients (4), it is much more difficult to derive formulas for general solutions in way similar to [1]. Therefore, we will restrict the problem to the construction of a general solution only in one of the possible cases. In particular, we will assume that only the matrix  $B(k)$  in (4) is variable while  $A(k) = A = (a_{ij})$ ,  $i, j = 1, 2$  is a constant matrix. Thus, we will consider a system

$$x(k+1) = Ax(k) + B(k)x(k-m), \quad k \in \mathcal{Z}_0^\infty. \quad (10)$$

In the proof, we need a formula for the solution of a scalar linear non-delayed difference equation of the form

$$z(k+1) = az(k) + \omega(k), \quad k \in \mathcal{Z}_{k_0}^\infty$$

with  $a \in \mathcal{C}$  and  $\omega: \mathcal{Z}_{k_0}^\infty \rightarrow \mathcal{C}$ . By, e.g., [6], the solution of the initial problem

$$z(k_0) = z_0 \quad (11)$$

is given by the formula

$$z(k) = a^{k-k_0} z_0 + \sum_{r=k_0}^{k-1} a^{k-1-r} \omega(r), \quad k \in \mathcal{Z}_{k_0+1}^\infty. \quad (12)$$

By definition, we set

$$\sum_{s=i}^j f(s) = 0$$

if, for integers  $i, j$ , inequality  $i > j$  holds.

Assume that  $S^{-1}AS = \Lambda$  for a nonsingular  $2 \times 2$  matrix  $S$  where  $\Lambda$  is the Jordan form of  $A$  depending on the roots of the characteristic equation

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0. \quad (13)$$

The equation  $y(k) = S^{-1}x(k)$  transforms (10) into

$$y(k+1) = \Lambda y(k) + B^*(k)y(k-m), \quad k \in \mathcal{Z}_0^\infty \quad (14)$$

where  $B^*(k) = S^{-1}B(k)S$ ,  $B^*(k) = (b_{ij}^*(k))$ ,  $i, j = 1, 2$ . The transformed initial problem for (14) is

$$y(k) = \varphi^*(k), \quad k \in \mathcal{Z}_{-m}^0 \quad (15)$$

with  $\varphi^* = (\varphi_1^*, \varphi_2^*)^T: \mathcal{Z}_{-m}^0 \rightarrow \mathbb{R}^2$ ,  $\varphi^*(k) = S^{-1}\varphi(k)$ .

### 1.1 Two real distinct roots of characteristic equation (13)

Assume that the characteristic equation (13) has two real distinct roots  $\lambda_1, \lambda_2$ . Then, obviously,  $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ . The necessary and sufficient conditions (8), (9) in the case of the system (14) are

$$b_{11}^*(k) + b_{22}^*(k) = 0, \quad k \in \mathcal{Z}_0^\infty, \quad (16)$$

$$\begin{vmatrix} b_{11}^*(k) & b_{12}^*(k) \\ b_{21}^*(k) & b_{22}^*(k) \end{vmatrix} = b_{11}^*(k)b_{22}^*(k) - b_{12}^*(k)b_{21}^*(k) = 0, \quad k \in \mathcal{Z}_0^\infty, \quad (17)$$

$$\begin{vmatrix} \lambda_1 & 0 \\ b_{21}^*(k) & b_{22}^*(k) \end{vmatrix} + \begin{vmatrix} b_{11}^*(k) & b_{12}^*(k) \\ 0 & \lambda_2 \end{vmatrix} = \lambda_1 b_{22}^*(k) + \lambda_2 b_{11}^*(k) = 0, \quad k \in \mathcal{Z}_0^\infty. \quad (18)$$

By (16), (18) we have  $b_{11}^*(k) = b_{22}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$  (since  $\lambda_1 \neq \lambda_2$ ) and (17) implies  $b_{12}^*(k)b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ .

**Theorem 2** *Let system (10) be weakly delayed and let the characteristic equation (13) have two real distinct roots  $\lambda_1, \lambda_2$ . Then,  $b_{11}^*(k) = b_{22}^*(k) = b_{12}^*(k)b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ . The solution of the initial problem (10), (3) is  $x(k) = Sy(k)$ ,  $k \in \mathcal{Z}_{-m}^\infty$  where, in the case  $b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ ,  $y(k) = (y_1(k), y_2(k))^T$  has the form*

$$(y_1(k), y_2(k)) = (\varphi_1^*(k), \varphi_2^*(k)), \quad k \in \mathcal{Z}_{-m}^0, \quad (19)$$

and

$$y_1(k) = \lambda_1^k \varphi_1^*(0) + \sum_{r=0}^{k-1} \lambda_1^{k-1-r} b_{12}^*(r) \varphi_2^*(r-m), \quad k \in \mathcal{Z}_1^{m+1}, \quad (20)$$

$$\begin{aligned} y_1(k) = & \lambda_1^k \varphi_1^*(0) + \sum_{r=0}^m \lambda_1^{k-1-r} b_{12}^*(r) \varphi_2^*(r-m) \\ & + \varphi_2^*(0) \sum_{r=m+1}^{k-1} \lambda_1^{k-1-r} \lambda_2^{r-m} b_{12}^*(r), \quad k \in \mathcal{Z}_{m+2}^\infty, \end{aligned} \quad (21)$$

$$y_2(k) = \lambda_2^k \varphi_2^*(0), \quad k \in \mathcal{Z}_1^\infty. \quad (22)$$

PROOF. Since  $b_{11}^*(k) = b_{22}^*(k) = b_{12}^*(k) b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ , the transformed system (14) is

$$y_1(k+1) = \lambda_1 y_1(k) + b_{12}^*(k) y_2(k-m), \quad (23)$$

$$y_2(k+1) = \lambda_2 y_2(k), \quad (24)$$

$$k \in \mathcal{Z}_0^\infty$$

if  $b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ . Consider the initial problem (23), (24), (15). Solve the equation (24). Using initial data (15), we obtain

$$y_2(k) = \begin{cases} \varphi_2^*(k) & \text{if } k \in \mathcal{Z}_{-m}^0, \\ \lambda_2^k \varphi_2^*(0) & \text{if } k \in \mathcal{Z}_1^\infty. \end{cases} \quad (25)$$

Thus, the formula (22) holds. Substituting the first formula in (25) into (23), we get

$$y_1(k+1) = \lambda_1 y_1(k) + b_{12}^*(k) \varphi_2^*(k-m) \quad \text{if } k \in \mathcal{Z}_0^m \quad (26)$$

and substituting the second formula in (25) into (23), we derive

$$y_1(k+1) = \lambda_1 y_1(k) + b_{12}^*(k) \lambda_2^{k-m} \varphi_2^*(0) \quad \text{if } k \in \mathcal{Z}_{m+1}^\infty. \quad (27)$$

Consider equation (26) together with the initial problem (selected from (15)), i.e.

$$\begin{cases} y_1(k+1) = \lambda_1 y_1(k) + b_{12}^*(k) \varphi_2^*(k-m), & k \in \mathcal{Z}_0^\infty, \\ y_1(0) = \varphi_1^*(0). \end{cases}$$

Applying formula (12) with the prescribed initial data (11) where  $k_0 = 0$  and  $z(0) = y_1(0)$ , we have

$$y_1(k) = \lambda_1^k \varphi_1^*(0) + \sum_{r=0}^{k-1} \lambda_1^{k-1-r} b_{12}^*(r) \varphi_2^*(r-m), \quad k \in \mathcal{Z}_1^{m+1}. \quad (28)$$

To solve equation (27) for  $k \in \mathcal{Z}_{m+1}^\infty$ , we need the initial data  $y_1(m+1)$ . Deriving them from formula (28) for  $k = m+1$ , consider the problem

$$\begin{cases} y_1(k+1) = \lambda_1 y_1(k) + b_{12}^*(k) \lambda_2^{k-m} \varphi_2^*(0), & k \in \mathcal{Z}_{m+1}^\infty, \\ y_1(m+1) = \lambda_1^{m+1} \varphi_1^*(0) + \sum_{r=0}^m \lambda_1^{m-r} b_{12}^*(r) \varphi_2^*(r-m). \end{cases}$$

Applying formula (12) again with the prescribed initial data (11) where  $k_0 = m+1$  and  $z(m+1) = y_1(m+1)$ , we have

$$\begin{aligned}
y_1(k) &= \lambda_1^{k-(m+1)} y_1(m+1) + \varphi_2^*(0) \sum_{r=m+1}^{k-1} \lambda_1^{k-1-r} \lambda_2^{r-m} b_{12}^*(r) \\
&= \lambda_1^{k-(m+1)} \left[ \lambda_1^{m+1} \varphi_1^*(0) + \sum_{r=0}^m \lambda_1^{m-r} b_{12}^*(r) \varphi_2^*(r-m) \right] \\
&\quad + \varphi_2^*(0) \sum_{r=m+1}^{k-1} \lambda_1^{k-1-r} \lambda_2^{r-m} b_{12}^*(r) \\
&= \lambda_1^k \varphi_1^*(0) + \sum_{r=0}^m \lambda_1^{k-1-r} b_{12}^*(r) \varphi_2^*(r-m) + \varphi_2^*(0) \sum_{r=m+1}^{k-1} \lambda_1^{k-1-r} \lambda_2^{r-m} b_{12}^*(r). \tag{29}
\end{aligned}$$

where  $k \in \mathcal{Z}_{m+2}^\infty$ . Formulas (28), (29) together with the initial data for  $y_1$  are equivalent with (19) (restricted to the co-ordinate  $y_1$ ) and (20)–(21).  $\square$

**Example 1** Let (1) is reduced to

$$x_1(k+1) = -x_2(k) + 0.5(-1)^k x_1(k-1) + 0.5(-1)^k x_2(k-1), \tag{30}$$

$$x_2(k+1) = -x_1(k) - 0.5(-1)^k x_1(k-1) - 0.5(-1)^k x_2(k-1), \tag{31}$$

$$k \in \mathcal{Z}_0^\infty$$

where  $m = 1$ ,

$$A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad B(k) = \begin{pmatrix} 0.5(-1)^k & 0.5(-1)^k \\ -0.5(-1)^k & -0.5(-1)^k \end{pmatrix}.$$

Since (8) and (9) hold, system (30), (31) is, by Theorem 1, weakly delayed. Transformation (6) where

$$\mathcal{S} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad \mathcal{S}^{-1} = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

transforms system (30), (31) to a system

$$y_1(k+1) = y_1(k) + (-1)^k y_2(k-1),$$

$$y_2(k+1) = -y_2(k),$$

$$k \in \mathcal{Z}_0^\infty$$

where (in accordance with (14), (32), (33))

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \lambda_1 = 1, \quad \lambda_2 = -1, \quad B^*(k) = \begin{pmatrix} 0 & (-1)^k \\ 0 & 0 \end{pmatrix}.$$

Compute

$$\varphi^*(k) = \begin{pmatrix} \varphi_1^*(k) \\ \varphi_2^*(k) \end{pmatrix} = \mathcal{S}^{-1} \varphi(k) = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} \varphi_1(k) \\ \varphi_2(k) \end{pmatrix} = \begin{pmatrix} 0.5\varphi_1(k) - 0.5\varphi_2(k) \\ 0.5\varphi_1(k) + 0.5\varphi_2(k) \end{pmatrix}.$$



Theorem 2 is valid and, by formulas (19), (20), (22), we derive solution of the system (30), (31):

$$(y_1(k), y_2(k)) = (0.5\varphi_1(k) - 0.5\varphi_2(k), \varphi_2^*(k)), \quad k \in \mathcal{Z}_{-1}^0,$$

and

$$\begin{aligned} y_1(k) &= 0.5\varphi_1(0) - 0.5\varphi_2(0) + \sum_{r=0}^{k-1} (-1)^r (0.5\varphi_1(r-1) + 0.5\varphi_2(r-1)), \quad k \in \mathcal{Z}_1^2, \\ y_1(k) &= 0.5\varphi_1(0) - 0.5\varphi_2(0) + \sum_{r=0}^1 (-1)^r (0.5\varphi_1(r-1) + 0.5\varphi_2(r-1)) \\ &\quad - (0.5\varphi_1(0) + 0.5\varphi_2(0))(k-2), \quad k \in \mathcal{Z}_3^\infty, \\ y_2(k) &= (-1)^k (0.5\varphi_1(0) + 0.5\varphi_2(0)), \quad k \in \mathcal{Z}_1^\infty. \end{aligned}$$

**Theorem 3** *Let system (10) be weakly delayed and let the characteristic equation (13) have two real distinct roots  $\lambda_1, \lambda_2$ . Then,  $b_{11}^*(k) = b_{22}^*(k) = b_{12}^*(k)b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ . The solution of the initial problem (10), (3) is  $x(k) = Sy(k)$ ,  $k \in \mathcal{Z}_{-m}^\infty$  where, in the case  $b_{12}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ ,  $y(k) = (y_1(k), y_2(k))^T$  has the form*

$$(y_1(k), y_2(k)) = (\varphi_1^*(k), \varphi_2^*(k)), \quad k \in \mathcal{Z}_{-m}^0,$$

and

$$\begin{aligned} y_1(k) &= \lambda_1^k \varphi_1^*(0), \quad k \in \mathcal{Z}_1^\infty, \\ y_2(k) &= \lambda_2^k \varphi_2^*(0) + \sum_{r=0}^{k-1} \lambda_2^{k-1-r} b_{21}^*(r) \varphi_1^*(r-m), \quad k \in \mathcal{Z}_1^{m+1}, \\ y_2(k) &= \lambda_2^k \varphi_2^*(0) + \sum_{r=0}^m \lambda_2^{k-1-r} b_{21}^*(r) \varphi_1^*(r-m) \\ &\quad + \varphi_1(0) \sum_{r=m+1}^{k-1} \lambda_1^{r-m} \lambda_2^{k-1-r} b_{21}^*(r), \quad k \in \mathcal{Z}_{m+2}^\infty. \end{aligned}$$

PROOF. If  $b_{12}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ , then the transformed system (14) is

$$y_1(k+1) = \lambda_1 y_1(k), \tag{32}$$

$$y_2(k+1) = \lambda_2 y_2(k) + b_{21}^*(k) y_1(k-m), \tag{33}$$

$$k \in \mathcal{Z}_0^\infty.$$

We investigate this by the same scheme as we investigated the problem (23), (24), (15). Due to the symmetry of both problems, the formulas for solutions of the problem (32), (33), (15) can be derived from the formulas describing the solutions of the problem (23), (24), (15) by replacing  $\lambda_1$  with  $\lambda_2$ ,  $b_{12}^*(k)$  with  $b_{21}^*(k)$ , and  $\varphi_1^*, \varphi_2^*$  with  $\varphi_2^*, \varphi_1^*$ .  $\square$

## 2 A GENERALIZATION

Our following considerations need an analogy of formula (12) to systems of equations. Consider a system

$$z(k+1) = Cz(k) + \Omega(k), \quad k \in \mathcal{Z}_{k_0}^\infty \quad (34)$$

where  $C$  is a  $2 \times 2$  constant matrix and  $z(k)$ ,  $\Omega(k)$ ,  $k \in \mathcal{Z}_{k_0}^\infty$  are  $2 \times 1$  vectors. By, e.g., [6], the solution of the initial problem (34), (35) where

$$z(k_0) = z_0 \quad (35)$$

is given by the formula

$$z(k) = C^{k-k_0} z_0 + \sum_{r=k_0}^{k-1} C^{k-1-r} \Omega(r), \quad k \in \mathcal{Z}_{k_0+1}^\infty. \quad (36)$$

In the above Theorem 2 and Theorem 3, two main cases were considered, namely, in the first theorem, the case  $b_{21}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$  and, in the second one, the case  $b_{12}^*(k) = 0$ ,  $k \in \mathcal{Z}_0^\infty$ . Now we will discuss the general case

$$b_{11}^*(k) = b_{22}^*(k) = b_{12}^*(k)b_{21}^*(k) = 0, \quad k \in \mathcal{Z}_0^\infty \quad (37)$$

without additional assumptions. Define a  $2 \times 2$  matrix (assuming (37))

$$G(k) := B^*(k) = \begin{pmatrix} 0 & b_{12}^*(k) \\ b_{21}^*(k) & 0 \end{pmatrix}, \quad k \in \mathcal{Z}_0^\infty$$

and write the system (14) as

$$y(k+1) = \Lambda y(k) + G(k)y(k-m), \quad k \in \mathcal{Z}_0^\infty.$$

The initial problem

$$\begin{aligned} y(k+1) &= \Lambda y(k) + G(k)y(k-m), \quad k \in \mathcal{Z}_{k_0}^\infty, \\ y(k_0) &= y_0 \end{aligned}$$

can be rewritten, by formula (36), as

$$y(k) = \Lambda^{k-k_0} y_0 + \sum_{r=k_0}^{k-1} \Lambda^{k-1-r} G(r)y(r-m), \quad k \in \mathcal{Z}_{k_0+1}^\infty. \quad (38)$$

Formula (38) can serve as a tool for solving of the initial problem

$$y(k+1) = \Lambda y(k) + G(k)y(k-m), \quad k \in \mathcal{Z}_0^\infty, \quad (39)$$

$$y(0) = \varphi^*(0), \quad (40)$$

where (by (15))

$$y(k) = \varphi^*(k), \quad k \in \mathcal{Z}_{-m}^0 \quad (41)$$

by what is called the step method. Now this method will be applied.

### 2.0.1 Step I

Denote by  $y^1(k)$ ,  $k \in \mathcal{Z}_1^{m+1}$  the solution of the problem (39), (40) where  $k \in \mathcal{Z}_0^m$ . Consider a modified problem (39), (40)

$$y^1(k+1) = \Lambda y^1(k) + G(k)\varphi^*(k-m), \quad k \in \mathcal{Z}_0^m, \quad (42)$$

$$y^1(0) = \varphi^*(0). \quad (43)$$

By (38) we derive

$$y^1(k) = \Lambda^k \varphi^*(0) + \sum_{r=0}^{k-1} \Lambda^{k-1-r} G(r) \varphi^*(r-m), \quad k \in \mathcal{Z}_1^{m+1}. \quad (44)$$

### 2.0.2 Step II

Denote by  $y^2(k)$ ,  $k \in \mathcal{Z}_{m+2}^{2(m+1)}$  the solution of the problem (39), (40) where  $k \in \mathcal{Z}_{m+1}^{2m+1}$  and consider a modified problem (39), (40)

$$y^2(k+1) = \Lambda y^2(k) + G(k)y^1(k-m), \quad k \in \mathcal{Z}_{m+1}^{2m+1},$$

$$y^2(m+1) = y^1(m+1).$$

By (38) we derive

$$y^2(k) = \Lambda^{k-m-1} y^1(m+1) + \sum_{r=m+1}^{k-1} \Lambda^{k-1-r} G(r) y^1(r-m), \quad k \in \mathcal{Z}_{m+2}^{2(m+1)}. \quad (45)$$

From (45) we get

$$\begin{aligned} y^2(k) &= \Lambda^{k-m-1} y^1(m+1) + \sum_{r=m+1}^{k-1} \Lambda^{k-1-r} G(r) y^1(r-m) \\ &= \Lambda^{k-m-1} \left[ \Lambda^{m+1} \varphi^*(0) + \sum_{r=0}^m \Lambda^{m-r} G(r) \varphi^*(r-m) \right] \\ &\quad + \sum_{r=m+1}^{k-1} \Lambda^{k-1-r} G(r) \left[ \Lambda^{r-m} \varphi^*(0) + \sum_{r_1=0}^{r-m-1} \Lambda^{r-m-1-r_1} G(r_1) \varphi^*(r_1-m) \right] \\ &= \Lambda^k \varphi^*(0) + \sum_{r=0}^m \Lambda^{k-1-r} G(r) \varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-1-r} G(r) \Lambda^{r-m} \varphi^*(0) \\ &\quad + \sum_{r=m+1}^{k-1} \sum_{r_1=0}^{r-m-1} \Lambda^{k-1-r} G(r) \Lambda^{r-m-1-r_1} G(r_1) \varphi^*(r_1-m), \quad k \in \mathcal{Z}_{m+2}^{2(m+1)}. \end{aligned}$$

### 2.0.3 Step s

Let  $s$  be a natural number,  $s \geq 1$ . Denote by  $y^s(k)$ ,  $k \in \mathcal{Z}_{(s-1)m+s}^{s(m+1)}$  the solution of the problem (39), (40) where  $k \in \mathcal{Z}_{(s-1)(m+1)}^{sm+(s-1)}$ . Consider a modified problem (39), (40)

$$\begin{aligned} y^s(k+1) &= \Lambda y^s(k) + G(k)y^{(s-1)}(k-m), \quad k \in \mathcal{Z}_{(s-1)(m+1)}^{sm+(s-1)}, \\ y^s((s-1)(m+1)) &= y^{(s-1)}((s-1)(m+1)) \end{aligned}$$

By (38) we derive

$$\begin{aligned} y^s(k) &= \Lambda^{k-(s-1)(m+1)} y^{(s-1)}((s-1)(m+1)) \\ &\quad + \sum_{r=(s-1)(m+1)}^{k-1} \Lambda^{k-1-r} G(r) y^{(s-1)}(r-m), \quad k \in \mathcal{Z}_{(s-1)m+s}^{s(m+1)}. \end{aligned} \quad (46)$$

Changing  $s$  by  $(s-1)$  in (46), we get

$$\begin{aligned} y^s(k) &= \Lambda^{k-(s-1)(m+1)} y^{(s-1)}((s-1)(m+1)) \\ &\quad + \sum_{r=(s-1)(m+1)}^{k-1} \Lambda^{k-1-r} G(r) y^{(s-1)}(r-m) \\ &= \Lambda^{k-(s-1)(m+1)} \left[ \Lambda^{(s-1)(m+1)-(s-2)(m+1)} y^{(s-2)}((s-2)(m+1)) \right. \\ &\quad \left. + \sum_{r=(s-2)(m+1)}^{(s-1)(m+1)-1} \Lambda^{(s-1)(m+1)-1-r} G(r) y^{(s-2)}(r-m) \right] \\ &\quad + \sum_{r=(s-1)(m+1)}^{k-1} \Lambda^{k-1-r} G(r) \left[ \Lambda^{r-m-(s-2)(m+1)} y^{(s-2)}((s-2)(m+1)) \right. \\ &\quad \left. + \sum_{r_1=(s-2)(m+1)}^{r-(m+1)} \Lambda^{r-m-1-r_1} G(r_1) y^{(s-2)}(r_1-m) \right] \\ &= \Lambda^{k-(s-2)(m+1)} y^{(s-2)}((s-2)(m+1)) + \sum_{r=(s-2)(m+1)}^{(s-1)(m+1)-1} \Lambda^{k-1-r} G(r) y^{(s-2)}(r-m) \\ &\quad + \sum_{r=(s-1)m+1}^{k-1} \Lambda^{k-1-r} G(r) \Lambda^{r-m-(s-2)(m+1)} y^{(s-2)}((s-2)(m+1)) \\ &\quad + \sum_{r=(s-1)(m+1)}^{k-1} \sum_{r_1=(s-2)(m+1)}^{r-m-1} \Lambda^{k-1-r} G(r) \Lambda^{r-m-r_1} G(r_1) y^{(s-2)}(r_1-m). \end{aligned}$$

Based on the above we formulate the following theorem.

**Theorem 4** *Let system (10) be weakly delayed and let its characteristic equation (13) have two real distinct roots. Assume*

$$G(k) \sum_{r=0}^{k-m-1} \Lambda^{k-m-1-r} G(r) \varphi^*(r-m) = \theta \quad (47)$$

for every  $k \in \mathcal{Z}_{m+2}^{2m+1}$  and

$$G(k) \left[ \sum_{r=0}^m \Lambda^{k-m-1-r} G(r) \varphi^*(r-m) + \sum_{r=m+1}^{k-m-1} \Lambda^{k-m-1-r} G(r) \Lambda^{r-m} \varphi^*(0) \right] = \theta \quad (48)$$

for every  $k \in \mathcal{Z}_{2(m+1)}^{\infty}$  where  $\theta$  is null vector. Then, the solution of the initial problem (10), (3) is  $x(k) = Sy(k)$ ,  $k \in \mathcal{Z}_{-m}^{\infty}$  where

$$y(k) = \varphi^*(k), \quad k \in \mathcal{Z}_{-m}^0, \quad (49)$$

and

$$y(k) = \Lambda^k \varphi^*(0) + \sum_{r=0}^{k-1} \Lambda^{k-1-r} G(r) \varphi^*(r-m), \quad k \in \mathcal{Z}_1^{m+1}, \quad (50)$$

$$\begin{aligned} y(k) &= \Lambda^k \varphi^*(0) + \sum_{r=0}^m \Lambda^{k-1-r} G(r) \varphi^*(r-m) \\ &+ \sum_{r=m+1}^{k-1} \Lambda^{k-1-r} G(r) \Lambda^{r-m} \varphi^*(0), \quad k \in \mathcal{Z}_{m+2}^{\infty}, \end{aligned} \quad (51)$$

PROOF. Formula (49) is obvious since it represents the initial condition of the original problem. Let us prove that (50) is valid. On the considered interval, this formula coincides with formula (44) which solves problem (42), (43) being equivalent with (39), (40).

Finally, we prove that formula (51) holds. Consider system (39), modified to the case considered:

$$y(k+1) = \Lambda y(k) + G(k) y(k-m), \quad k \in \mathcal{Z}_{m+1}^{\infty}. \quad (52)$$

The left-hand side of (52) for  $y$  given by (51) is equal to (for  $k \in \mathcal{Z}_{m+1}^{\infty}$ ) to

$$\begin{aligned} \mathcal{L} = y(k+1) &= \Lambda^{k+1} \varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r} G(r) \varphi^*(r-m) \\ &+ \sum_{r=m+1}^k \Lambda^{k-r} G(r) \Lambda^{r-m} \varphi^*(0) \\ &= \Lambda^{k+1} \varphi_1^*(0) + \sum_{r=0}^m \Lambda^{k-r} G(r) \varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-r} G(r) \Lambda^{r-m} \varphi^*(0) \\ &\quad + G(k) \Lambda^{k-m} \varphi^*(0). \end{aligned}$$

The right-hand side of (52) for  $y$  given by (51) is equal (for  $k \in \mathcal{Z}_{2(m+1)}^\infty$ ) to

$$\begin{aligned}
\mathcal{R} &= \Lambda y(k) + G(k)y(k-m) \\
&= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) \\
&\quad + G(k) \left[ \Lambda^{k-m}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-m-1-r}G(r)\varphi^*(r-m) \right. \\
&\quad \left. + \sum_{r=m+1}^{k-m-1} \Lambda^{k-m-1-r}G(r)\Lambda^{r-m}\varphi^*(0) \right] \\
&= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) \\
&\quad + G(k)\Lambda^{k-m}\varphi^*(0) \\
&\quad + G(k) \left[ \sum_{r=0}^m \Lambda^{k-m-1-r}G(r)\varphi^*(r-m) + \sum_{r=m+1}^{k-m-1} \Lambda^{k-m-1-r}G(r)\Lambda^{r-m}\varphi^*(0) \right]
\end{aligned}$$

and, due to (48),

$$\begin{aligned}
\mathcal{R} &= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) \\
&\quad + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) + G(k)\Lambda^{k-m}\varphi^*(0) = \mathcal{L}.
\end{aligned}$$

The right-hand side of (52) for  $y$  given by (51) is equal to (for  $k \in \mathcal{Z}_{m+2}^{2m+1}$ )

$$\begin{aligned}
\mathcal{R}_1 &= \Lambda y(k) + G(k)y(k-m) \\
&= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) \\
&\quad + G(k) \left[ \Lambda^{k-m}\varphi^*(0) + \sum_{r=0}^{k-m-1} \Lambda^{k-m-r}G(r)\varphi^*(r-m) \right] \\
&= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) \\
&\quad + G(k)\Lambda^{k-m}\varphi^*(0) + G(k) \sum_{r=0}^{k-m-1} \Lambda^{k-m-r}G(r)\varphi^*(r-m)
\end{aligned}$$

and, due to (47),

$$\begin{aligned}\mathcal{R}_1 &= \Lambda^{k+1}\varphi^*(0) + \sum_{r=0}^m \Lambda^{k-r}G(r)\varphi^*(r-m) \\ &\quad + \sum_{r=m+1}^{k-1} \Lambda^{k-r}G(r)\Lambda^{r-m}\varphi^*(0) + G(k)\Lambda^{k-m}\varphi^*(0) = \mathcal{L}.\end{aligned}$$

□

**Remark 1** *Let us remark that, for  $k = m + 1$ , the formula (51) gives the same value as the formula (50) for  $k = m + 1$ . Theorems 2, 3 are particular cases of Theorem 4.*

## 2.1 Active initial data

Analyzing carefully the formulas for solution in Theorem 2, we deduce the following. The dimension of the set of initial data, being initially  $2(m + 1)$  (see (19)), is reduced. In formulas (20)–(22), only the initial data

$$\varphi_1^*(0), \varphi_2^*(0), \varphi_2^*(-1), \dots, \varphi_2^*(-m) \quad (53)$$

play a role and the rest of the initial data

$$\varphi_1^*(-1), \dots, \varphi_1^*(-m)$$

are “hidden” and not used in the computations. Thus, parts of the solutions are identical and depend on  $m + 2$  initial data (parameters) itemized in (53) only.

Similarly, in formulas given in Theorem 3, only parameters

$$\varphi_2^*(0), \varphi_1^*(0), \varphi_1^*(-1), \dots, \varphi_1^*(-m) \quad (54)$$

determine the solutions. Consequently, the solutions are partially pasted as well and depend on  $m + 2$  initial data (parameters) described in (54).

The same analysis can be performed in the case of Theorem 4. Again, the solution described by formulas (50), (51) depends on  $m + 2$  initial data because every multiplication of the type  $G(r)\varphi^*(r - m)$ ,  $r \in \mathbb{Z}_0^m$  deletes one of the two initial pieces of information given by  $\varphi^*(r - m)$ . Therefore formula (50) contains  $m + 2$  initial items. The number of initial items in formula (51) is  $m + 2$  as well.

## CONCLUSION

Considered in the paper weakly delayed systems can be simplified and their solutions can be found in an explicit analytical form. In the case of discrete systems of two equations, to obtain the corresponding eigenvalues it is sufficient to solve only the second order polynomial characteristic equation

$$\det(A - \lambda I) = 0$$

rather than a  $2(m + 1)$ -th order polynomial equation

$$\det(A + \lambda^{-m}B(k) - \lambda I) = 0$$

where  $A$  is an  $2 \times 2$  constant matrix and  $B(k)$ ,  $k \in \mathcal{Z}_0^\infty$  is an  $2 \times 2$  variable matrix,  $A$  and  $B(k)$  satisfy (8), (9). In the paper is considered the case when characteristic equation has two real distinct roots. Moreover, in the considered case the solution of the initial problem depends on  $m + 2$  initial data only although the Cauchy is usually determined by  $2(m + 1)$  parameters, i.e.  $2(m + 1)$ -dimensional space of initial data is reduced to  $m + 2$ -dimensional space. It is an open question if is possible to investigate the remaining cases and to construct the general solution if the Jordan forms of the matrix of linear terms are different from that investigated in this paper.

## References

- [1] Diblík, J., Khusainov, D., Šmarda, Z. Construction of the general solution of planar linear discrete systems with constant coefficients and weak delay. *Advances in Difference Equations*, **2009**, Art. ID 784935, p.–18 pages, doi:10.1155/2009/784935.
- [2] Diblík, J., Halfarová, J. Explicit general solution of planar linear discrete systems with constant coefficients and weak delays. *Advances in Difference Equations*, 2013, 2013:50 doi:10.1186/1687-1847-2013-50, p. 1-29.
- [3] Diblík, J., Halfarová, H. General explicit solution of planar weakly delayed linear discrete systems and pasting its solutions. *Abstract and Applied Analysis*, vol. 2014, Article ID 627295, p. 1–37.
- [4] Šafařík, J., Diblík, J., Halfarová, H. Weakly delayed systems of linear discrete equations in  $R^3$ . *MITAV 2015, Post-Conference Proceedings of Extended Versions of Selected Papers*. Brno, Univerzita obrany v Brně, 2015, p. 105–121. Available at: <<http://mitav.unob.cz/data/MITAV%202015%20Proceedings.pdf>>. ISBN 978-80-7231-436-2.
- [5] Šafařík, J., Diblík, J. Weakly delayed difference systems in  $R^3$  and their solution. In: *Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 84-104. [Online]. [Cit. 2017-07-26]. Available at: <<http://mitav.unob.cz/data/MITAV2016Proceedings.pdf>>. ISBN 978-80-7231-400-3.
- [6] Elaydi, S. N. *An Introduction to Difference Equations*. Third Edition, Springer, 2005.

## Acknowledgement

The first author has been supported by the project No. LO1408, AdMaS UP-Advanced Materials, Structures and Technologies (supported by Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme I).



# SOLVING A HIGHER-ORDER LINEAR DISCRETE SYSTEMS

**J. Diblík, K. Mencáková**

Faculty of Electrical Engineering and Computer Science  
Brno University of Technology, Technická 8, 616 00 Brno, Czech Republic  
diblik@feec.vutbr.cz, mencakova.k@fce.vutbr.cz

**Abstract:** *In this paper there is considered a linear discrete homogenous system of the order  $(m + 2)$ :*

$$\Delta^2 x(k) + B^2 x(k - m) = f(k), \quad k \in \mathbb{N}_0,$$

*where  $B$  is a constant  $n \times n$  regular matrix,  $m \in \mathbb{N}_0$  and  $x: \{-m, -m + 1, \dots\} \rightarrow \mathbb{R}^n$ ,  $f: \mathbb{Z}_0^\infty \rightarrow \mathbb{R}^n$ .*

*Two linearly independent solutions will be found as special matrix functions called delayed discrete cosine and delayed discrete sine.*

*Formulas for solutions are gotten utilizing these matrix functions. An example illustrating results is given as well.*

**Keywords:** delayed discrete cosine, delayed discrete sine, discrete equation.

## INTRODUCTION

Below it is used following notation:  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ ,  $\mathbb{Z}$  is the set of all integers and for integers  $s$ ,  $r$ ,  $s \leq r$  define  $\mathbb{Z}_s^r := \{s, s + 1, \dots, r\}$ . Similarly symbols  $\mathbb{Z}_{-\infty}^r$ ,  $\mathbb{Z}_s^\infty$  are defined.

In this paper we consider a linear discrete homogeneous system of the order  $(m + 2)$ :

$$\Delta^2 x(k) + B^2 x(k - m) = f(k), \quad k \in \mathbb{Z}_0^\infty, \quad (1)$$

where  $B$  is a constant  $n \times n$  regular matrix,  $m \in \mathbb{N}_0$  and  $x: \mathbb{Z}_{-m}^\infty \rightarrow \mathbb{R}^n$ ,  $f: \mathbb{Z}_0^\infty \rightarrow \mathbb{R}^n$ .

Solution  $x = x(k)$  of (1) is defined as a function  $x: \mathbb{Z}_{-m}^\infty \rightarrow \mathbb{R}^n$  satisfying (1) for  $k \in \mathbb{Z}_0^\infty$ . We will find two linearly independent solutions of (1) as a special matrix functions called delayed discrete cosine and delayed discrete sine. With their aid a solution of initial Cauchy problem is given as well. Previously, a similar problem for discrete linear systems

$$\Delta x(k) = Bx(k - m) + f(k), \quad k \in \mathbb{Z}_0^\infty, \quad (2)$$

where  $m$  is a fixed integer,  $B$  is a constant  $n \times n$  matrix, was considered in [1]. A fundamental matrix was constructed as a delayed discrete matrix exponential. In [2] a particular case of (2) was investigated and a new formula for solution of initial-value problem (when  $m = 1$ ,  $x: \mathbb{Z}_{-1}^\infty \rightarrow \mathbb{R}^2$ ,  $f$  is a null vector) was derived. In the paper [3], there was considered a differential system of the second order with delay

$$\ddot{x}(t) + \Omega^2 x(t - \tau) = 0, \quad (3)$$

where  $\tau > 0$  and  $\Omega$  is a constant  $n \times n$  matrix and a generalization is given in [4]. A fundamental matrix for (3) was constructed as a special matrix functions called delayed matrix cosine and

delayed matrix sine. Such special matrix functions served as a motivation for the present investigation.

## 1 DELAYED DISCRETE COSINE AND SINE

In this part there we define auxiliary discrete functions – delayed discrete cosine, delayed discrete sine and some of their properties are proved.

Below, we use the following definition of combinative numbers:

$$\binom{p}{q} := \begin{cases} \frac{p!}{q! \cdot (p-q)!} & \text{if } p \geq q \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where  $p, q$  are whole numbers and in common usage  $0! = 1$ .

Symbols  $\Theta$ ,  $I$  and  $\theta$  stand for  $n \times n$  null matrix,  $n \times n$  unit matrix and  $n \times 1$  vector.

**Definition 1.** Delayed discrete cosine is defined as:

$$\text{Cos}_m Bk := \begin{cases} \Theta & \text{if } k \in \mathbb{Z}_{-\infty}^{-m-1}, \\ I & \text{if } k \in \mathbb{Z}_{-m}^1, \\ I - B^2 \cdot \binom{k}{2} & \text{if } k \in \mathbb{Z}_2^{(m+2)+1}, \\ I - B^2 \cdot \binom{k}{2} + B^4 \cdot \binom{k-m}{4} & \text{if } k \in \mathbb{Z}_{(m+2)+2}^{2(m+2)+1}, \\ \dots & \\ I - B^2 \cdot \binom{k}{2} + B^4 \cdot \binom{k-m}{4} - B^6 \cdot \binom{k-2m}{6} + \dots & \\ + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l} & \\ \text{if } k \in \mathbb{Z}_{(l-1)(m+2)+2}^{l(m+2)+1}, \quad \ell = 0, 1, 2, \dots, & \\ \dots & \end{cases}$$

**Definition 2.** Delayed discrete sine is defined as:

$$\text{Sin}_m Bk := \begin{cases} \Theta & \text{if } k \in \mathbb{Z}_{-\infty}^{-m-1}, \\ B \cdot \binom{k+m}{1} & \text{if } k \in \mathbb{Z}_{-m}^1, \\ B \cdot \binom{k+m}{1} - B^3 \cdot \binom{k}{3} & \text{if } k \in \mathbb{Z}_2^{(m+2)+1}, \\ B \cdot \binom{k+m}{1} - B^3 \cdot \binom{k}{3} + B^5 \cdot \binom{k-m}{5} & \text{if } k \in \mathbb{Z}_{(m+2)+2}^{2(m+2)+1}, \\ \dots & \\ B \cdot \binom{k+m}{1} - B^3 \cdot \binom{k}{3} + B^5 \cdot \binom{k-m}{5} + \dots & \\ + (-1)^l B^{2l+1} \cdot \binom{k-(l-1)m}{2l+1} & \\ \text{if } k \in \mathbb{Z}_{(l-1)(m+2)+2}^{l(m+2)+1}, \quad \ell = 0, 1, 2, \dots, & \\ \dots & \end{cases}$$

We remind of the definition of summation:

$$\sum_{j=\alpha}^{\beta} g(j) := \begin{cases} g(\alpha) + g(\alpha+1) + \dots + g(\beta-1) + g(\beta) & \text{if } \alpha \leq \beta, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where  $\alpha, \beta \in \mathbb{Z}$ .

The definitions of  $\text{Cos}_m Bk$  and  $\text{Sin}_m Bk$  can be shortly expressed as

$$\text{Cos}_m Bk := \sum_{j=0}^{\lceil (k-1)/(m+2) \rceil} (-1)^j B^{2j} \binom{k-(j-1)m}{2j} \quad \text{if } k \in \mathbb{Z}_{-\infty}^{\infty}, \quad (6)$$

$$\text{Sin}_m Bk := \sum_{j=0}^{\lceil (k-1)/(m+2) \rceil} (-1)^j B^{2j+1} \binom{k-(j-1)m}{2j+1} \quad \text{if } k \in \mathbb{Z}_{-\infty}^{\infty}. \quad (7)$$

In the following theorem, there will be given basic properties of  $\text{Cos}_m Bk$  and  $\text{Sin}_m Bk$ .

**Theorem 1.** For  $\text{Cos}_m Bk$ ,  $\text{Sin}_m Bk$  and any  $k \in \mathbb{Z}$  are hold:

$$\Delta \text{Cos}_m Bk = -B \cdot \text{Sin}_m B(k-m), \quad (8)$$

$$\Delta \text{Sin}_m Bk = B \cdot \text{Cos}_m Bk. \quad (9)$$

*Proof.* In the proof we will use well-known formula

$$\binom{p+1}{q} - \binom{p}{q} = \binom{p}{q-1}, \quad (10)$$

where  $p, q$  are whole numbers.

At the first we prove the formula (8). The proof is divided into three parts.

a) If  $k \in \mathbb{Z}_{-\infty}^{-m-2}$  :

$$\Delta \text{Cos}_m Bk = \text{Cos}_m B(k+1) - \text{Cos}_m Bk = \Theta - \Theta = \Theta = -B \cdot \text{Sin}_m B(k-m).$$

For these  $k$  formula (8) obviously holds.

b) If  $k \in \mathbb{Z}_{(l-1)(m+2)+2}^{l(m+2)}$  we get

$$\begin{aligned} \Delta \text{Cos}_m Bk &= \text{Cos}_m B(k+1) - \text{Cos}_m Bk \\ &= I - B^2 \cdot \binom{k+1}{2} + B^4 \cdot \binom{k+1-m}{4} - B^6 \cdot \binom{k+1-2m}{6} + \dots \\ &\quad + (-1)^l B^{2l} \cdot \binom{k+1-(l-1)m}{2l} - \left[ I - B^2 \cdot \binom{k}{2} + B^4 \cdot \binom{k-m}{4} \right. \\ &\quad \left. - B^6 \cdot \binom{k-2m}{6} + \dots + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l} \right] \\ &= I - I - B^2 \cdot \left[ \binom{k+1}{2} - \binom{k}{2} \right] + B^4 \cdot \left[ \binom{k-m+1}{4} - \binom{k-m}{4} \right] \\ &\quad - B^6 \cdot \left[ \binom{k-2m+1}{6} - \binom{k-2m}{6} \right] + \dots \\ &\quad + (-1)^l B^{2l} \cdot \left[ \binom{k-(l-1)m+1}{2l} - \binom{k-(l-1)m}{2l} \right]. \end{aligned}$$

Now we use formula (10).

$$\begin{aligned} \Delta \text{Cos}_m Bk &= -B^2 \cdot \binom{k}{1} + B^4 \cdot \binom{k-m}{3} - B^6 \cdot \binom{k-2m}{5} + \dots \\ &\quad + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l-1} \\ &= -B \cdot \left[ B \cdot \binom{k}{1} - B^3 \cdot \binom{k-m}{3} + B^5 \cdot \binom{k-2m}{5} + \dots \right. \\ &\quad \left. + (-1)^{l-1} B^{2l-1} \cdot \binom{k-(l-1)m}{2l-1} \right] \\ &= -B \cdot \left[ B \cdot \binom{k-m+m}{1} - B^3 \cdot \binom{k-m}{3} + B^5 \cdot \binom{k-m-m}{5} + \dots \right. \\ &\quad \left. + (-1)^{l-1} B^{2(l-1)+1} \cdot \binom{k-(l-2)m-m}{2(l-1)+1} \right] \\ &= -B \cdot \text{Sin}_m B(k-m). \end{aligned}$$

In this case formula (8) holds too.

c) Let  $k = l(m+2) + 1$ . Then

$$\Delta \text{Cos}_m Bk = \text{Cos}_m B(k+1) - \text{Cos}_m Bk$$

$$\begin{aligned}
&= I - B^2 \cdot \binom{k+1}{2} - B^4 \cdot \binom{k+1-m}{4} + \dots \\
&\quad + (-1)^l B^{2l} \cdot \binom{k+1-(l-1)m}{2l} + (-1)^{l+1} B^{2(l+1)} \cdot \binom{k+1-lm}{2(l+1)} \\
&\quad - \left[ I - B^2 \cdot \binom{k}{2} - B^4 \cdot \binom{k-m}{4} + \dots + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l} \right] \\
&= I - I - B^2 \cdot \left[ \binom{k+1}{2} - \binom{k}{2} \right] + B^4 \cdot \left[ \binom{k-m+1}{4} - \binom{k-m}{4} \right] + \\
&\quad \dots + (-1)^l B^{2l} \cdot \left[ \binom{k-(l-1)m+1}{2l} - \binom{k-(l-1)m}{2l} \right] \\
&\quad + (-1)^{l+1} B^{2(l+1)} \cdot \binom{k-lm+1}{2(l+1)}.
\end{aligned}$$

Utilizing formula (10)

$$\begin{aligned}
\Delta \text{Cos}_m Bk &= -B^2 \cdot \binom{k}{1} + B^4 \cdot \binom{k-m}{3} + \dots \\
&\quad + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l-1} + (-1)^{2l+1} B^{2(l+1)} \cdot \binom{k-lm+1}{2(l+1)} \\
&= -B \cdot \left[ B \cdot \binom{k}{1} - B^3 \cdot \binom{k-m}{3} + \dots \right. \\
&\quad \left. + (-1)^{l-1} B^{2l-1} \cdot \binom{k-(l-1)m}{2l-1} + (-1)^l B^{2l+1} \cdot \binom{k-lm+1}{2l+2} \right] \\
&= -B \cdot \left[ B \cdot \binom{k+m-m}{1} - B^3 \cdot \binom{k-m}{3} + \dots \right. \\
&\quad \left. + (-1)^{l-1} B^{2l-1} \cdot \binom{k+m-lm}{2l-1} + (-1)^l B^{2l+1} \cdot \binom{k-lm+1}{2l+2} \right] = (*).
\end{aligned}$$

The last binomial coefficient can be decomposed by (10):

$$\begin{aligned}
\binom{k-lm+1}{2l+2} &= \binom{k-lm}{2l+1} + \binom{k-lm}{2l+2} \\
&= \binom{k-m-lm+m}{2l+1} + \binom{l(m+2)+1-lm}{2l+2} \\
&= \binom{k-m-(l-1)m}{2l+1} + \binom{lm+2l+1-lm}{2l+2} \\
&= \binom{k-m-(l-1)m}{2l+1} + \binom{2l+1}{2l+2} \\
&= \binom{k-m-(l-1)m}{2l+1} + 0 = \binom{k-m-(l-1)m}{2l+1}.
\end{aligned}$$

Then

$$\begin{aligned}
(*) &= -B \cdot \left[ B \cdot \binom{k+m-m}{1} - B^3 \cdot \binom{k-m}{3} + \dots \right. \\
&\quad \left. + (-1)^{l-1} B^{2l-1} \cdot \binom{k+m-lm}{2l-1} + (-1)^l B^{2l+1} \cdot \binom{k-m-(l-1)m}{2l+1} \right] \\
&= -B \cdot \text{Sin}_m B(k-m).
\end{aligned}$$

So formula (8) holds in this case and we proved it for each  $k \in \mathbb{Z}$ .

Now we prove formula (9), again in three parts.

a) If  $k \in \mathbb{Z}_{-\infty}^{-m-2}$

$$\Delta \text{Sin}_m Bk = \text{Sin}_m B(k+1) - \text{Sin}_m Bk = \Theta - \Theta = \Theta = B \cdot \text{Cos}_m Bk$$

and the formula obviously holds.

b) If  $k \in \mathbb{Z}_{l(m+2)+2}^{(l+1)(m+2)}$  then

$$\begin{aligned}
\Delta \text{Sin}_m Bk &= \text{Sin}_m B(k+1) - \text{Sin}_m Bk \\
&= B \cdot \binom{k+m+1}{1} - B^3 \cdot \binom{k+1}{3} + B^5 \cdot \binom{k+1-m}{5} + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \binom{k+1-(l-1)m}{2l+1} - \left[ B \cdot \binom{k+m}{1} \right. \\
&\quad \left. - B^3 \cdot \binom{k}{3} + B^5 \cdot \binom{k-m}{5} + \dots + (-1)^l B^{2l+1} \cdot \binom{k-(l-1)m}{2l+1} \right] \\
&= B \cdot \left[ \binom{k+m+1}{1} - \binom{k+m}{1} \right] - B^3 \cdot \left[ \binom{k+1}{3} - \binom{k}{3} \right] \\
&\quad + B^5 \cdot \left[ \binom{k-m+1}{5} - \binom{k-m}{5} \right] + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \left[ \binom{k-(l-1)m+1}{2l+1} - \binom{k-(l-1)m}{2l+1} \right].
\end{aligned}$$

We use formula (10).

$$\begin{aligned}
\Delta \text{Sin}_m Bk &= B \cdot \binom{k+m}{0} - B^3 \cdot \binom{k}{2} + B^5 \cdot \binom{k-m}{4} + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \binom{k-(l-1)m}{2l} \\
&= B \cdot \left[ I - B^2 \cdot \binom{k}{2} + B^4 \cdot \binom{k-m}{4} + \dots \right. \\
&\quad \left. + (-1)^l B^{2l} \cdot \binom{k-(l-1)m}{2l} \right] \\
&= B \cdot \text{Cos}_m Bk.
\end{aligned}$$

In this case formula (9) holds too.

c) Let  $k = l(m + 2) + 1$  we get

$$\begin{aligned}
\Delta \text{Sin}_m Bk &= \text{Sin}_m B(k + 1) - \text{Sin}_m Bk \\
&= B \cdot \binom{k + m + 1}{1} - B^3 \cdot \binom{k + 1}{3} + B^5 \cdot \binom{k + 1 - m}{5} + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \binom{k + 1 - (l-1)m}{2l+1} + (-1)^{l+1} B^{2(l+1)+1} \cdot \binom{k + 1 - lm}{2(l+1)+1} \\
&\quad - \left[ B \cdot \binom{k + m}{1} - B^3 \cdot \binom{k}{3} + B^5 \cdot \binom{k - m}{5} + \dots \right. \\
&\quad \left. + (-1)^l B^{2l+1} \cdot \binom{k - (l-1)m}{2l+1} \right] \\
&= B \cdot \left[ \binom{k + m + 1}{1} - \binom{k + m}{1} \right] - B^3 \cdot \left[ \binom{k + 1}{3} - \binom{k}{3} \right] \\
&\quad + B^5 \cdot \left[ \binom{k - m + 1}{5} - \binom{k - m}{5} \right] + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \left[ \binom{k - (l-1)m + 1}{2l+1} - \binom{k - (l-1)m}{2l+1} \right] \\
&\quad + (-1)^{l+1} B^{2(l+1)+1} \cdot \binom{k - lm + 1}{2(l+1)+1}.
\end{aligned}$$

Utilizing formula (10)

$$\begin{aligned}
\Delta \text{Sin}_m Bk &= B \cdot \binom{k + m}{0} - B^3 \cdot \binom{k}{2} + B^5 \cdot \binom{k - m}{4} + \dots \\
&\quad + (-1)^l B^{2l+1} \cdot \binom{k - (l-1)m}{2l} + (-1)^{l+1} B^{2l+3} \cdot \binom{k - lm + 1}{2l+3} \\
&= B \cdot \left[ I - B^2 \cdot \binom{k}{2} + B^4 \cdot \binom{k - m}{4} + \dots + (-1)^l B^{2l} \cdot \binom{k - (l-1)m}{2l} \right] \\
&\quad + (-1)^{l+1} B^{2l+3} \cdot \binom{lm + 2l + 1 - lm + 1}{2l+3} \\
&= B \cdot \text{Cos}_m Bk,
\end{aligned}$$

because the last binomial coefficient

$$\binom{lm + 2l + 1 - lm + 1}{2l+3} = \binom{2l+2}{2l+3} = 0.$$

So formula (9) holds in the last case and for every  $k \in \mathbb{Z}$  too.

□

**Remark 1.** From formulas (8), (9) follows that for any  $k \in \mathbb{Z}$ :

$$\Delta^2 \text{Cos}_m Bk = -B^2 \cdot \text{Cos}_m B(k - m),$$

$$\Delta^2 \text{Sin}_m Bk = -B^2 \cdot \text{Sin}_m B(k - m),$$

i.e.  $\text{Cos}_m Bk \cdot C_1$  and  $\text{Sin}_m Bk \cdot C_2$ , where  $C_1, C_2$  are any constant vectors, are linearly independent solutions of homogenous system (1).

## 2 SOLUTION OF AN INITIAL PROBLEM FOR HOMOGENOUS SYSTEM

Obviously, the expression

$$x(k) = C_1 \cdot \text{Cos}_m Bk + C_2 \cdot \text{Sin}_m Bk, \quad k \geq 2,$$

where  $C_1, C_2$  are arbitrary constant  $1 \times n$  vectors, is a family of solutions of system (1).

**Theorem 2.** *Solution of initial problem (1), (11), where*

$$x(k) = \varphi(k) = (\varphi_1(k), \dots, \varphi_n(k))^T, \quad k = -m, \dots, 1, \quad (11)$$

*is expressed by formula*

$$x(k) = (\text{Cos}_m Bk) \varphi(-m) + B^{-1} \left[ (\text{Sin}_m Bk) \Delta \varphi(-m) + \sum_{j=-m+1}^0 \text{Sin}_m B(k - m - j) \cdot \Delta^2 \varphi(j - 1) \right], \quad (12)$$

where  $k \in \mathbb{Z}_{-m}^\infty$ .

*Proof.* According to Remark 1 it is easy to see that  $x(k)$  from formula (12) is a solution of equation (1) for any  $k \in \mathbb{Z}_2^\infty$ .

We prove that formula (12) satisfy also conditions (11). If  $k \in \mathbb{Z}_{-m}^1$ , then  $\text{Cos}_m Bk = I$ ,  $\text{Sin}_m Bk = B \cdot \binom{k+m}{1}$  and

$$x(k) = \varphi(-m) + B^{-1} \cdot B \cdot \binom{k+m}{1} \cdot \Delta \varphi(-m) + B^{-1} \cdot \sum_{j=-m+1}^0 \text{Sin}_m B(k - m - j) \cdot \Delta^2 \varphi(j - 1).$$

Now we divide the proof into three cases.

a) For  $k = -m$  we get

$$\begin{aligned} x(-m) &= \varphi(-m) + \binom{0}{1} \Delta \varphi(-m) + B^{-1} \cdot \sum_{j=-m+1}^0 \text{Sin}_m B(-2m - j) \Delta^2 \varphi(j - 1) \\ &= \varphi(-m) + \theta + \theta = \varphi(-m). \end{aligned}$$

b) If  $k = -m + 1$ , we have

$$x(-m + 1) = \varphi(-m) + \binom{1}{1} \Delta \varphi(-m) + B^{-1} \cdot \sum_{j=-m+1}^0 \text{Sin}_m B(-2m + 1 - j) \Delta^2 \varphi(j - 1)$$



$$= \varphi(-m) + \varphi(-m+1) - \varphi(-m) + \theta = \varphi(-m+1).$$

c) Finally let  $k \in \mathbb{Z}_{-m+2}^1$ . Then

$$\begin{aligned} x(k) &= \varphi(-m) + \binom{k+m}{1} \cdot [\varphi(-m+1) - \varphi(-m)] \\ &\quad + B^{-1} \left[ \sum_{j=-m+1}^{k-1} \text{Sin}_m B(k-m-j) \Delta^2 \varphi(j-1) \right. \\ &\quad \left. + \sum_{j=k}^0 \text{Sin}_m B(k-m-j) \Delta^2 \varphi(j-1) \right] = (*). \end{aligned}$$

The second sum is composed of all null terms. So we have

$$\begin{aligned} (*) &= \varphi(-m) - \binom{k+m}{1} \varphi(-m) + \binom{k+m}{1} \varphi(-m+1) \\ &\quad + B^{-1} \cdot \sum_{j=-m+1}^{k-1} \text{Sin}_m B(k-m-j) \cdot \Delta^2 \varphi(j-1) \\ &= \binom{1-k-m}{1} \varphi(-m) + \binom{k+m}{1} \varphi(-m+1) \\ &\quad + B^{-1} \cdot [\text{Sin}_m B(k-m+m-1) \Delta^2 \varphi(-m) \\ &\quad + \text{Sin}_m B(k-m+m-2) \Delta^2 \varphi(-m+1) \\ &\quad + \text{Sin}_m B(k-m+m-3) \Delta^2 \varphi(-m+2) + \dots \\ &\quad + \text{Sin}_m B(k-m-k+3) \Delta^2 \varphi(k-4) + \text{Sin}_m B(k-m-k+2) \Delta^2 \varphi(k-3) \\ &\quad + \text{Sin}_m B(k-m-k+1) \Delta^2 \varphi(k-2)] \\ &= \binom{1-k-m}{1} \varphi(-m) + \binom{k+m}{1} \varphi(-m+1) + B^{-1} \cdot [\text{Sin}_m B(k-1) \Delta^2 \varphi(-m) \\ &\quad + \text{Sin}_m B(k-2) \Delta^2 \varphi(-m+1) + \text{Sin}_m B(k-3) \Delta^2 \varphi(-m+2) + \dots \\ &\quad + \text{Sin}_m B(-m+3) \Delta^2 \varphi(k-4) + \text{Sin}_m B(-m+2) \Delta^2 \varphi(k-3) \\ &\quad + \text{Sin}_m B(-m+1) \Delta^2 \varphi(k-2)] \\ &= \binom{1-k-m}{1} \varphi(-m) + \binom{k+m}{1} \varphi(-m+1) + B^{-1} \cdot \left[ B \cdot \binom{k-1+m}{1} \Delta^2 \varphi(-m) \right. \\ &\quad + B \cdot \binom{k-2+m}{1} \Delta^2 \varphi(-m+1) + B \cdot \binom{k-3+m}{1} \Delta^2 \varphi(-m+2) + \dots \\ &\quad \left. + B \cdot \binom{3}{1} \Delta^2 \varphi(k-4) + B \cdot \binom{2}{1} \Delta^2 \varphi(k-3) + B \cdot \binom{1}{1} \Delta^2 \varphi(k-2) \right] \\ &= (1-k-m) \varphi(-m) + (k+m) \varphi(-m+1) + B^{-1} \cdot B \cdot [(k-1+m) \Delta^2 \varphi(-m) \end{aligned}$$

$$\begin{aligned}
& + (k-2+m)\Delta^2\varphi(-m+1) + (k-3+m)\Delta^2\varphi(-m+2) + \dots \\
& + 3\Delta^2\varphi(k-4) + 2\Delta^2\varphi(k-3) + \Delta^2\varphi(k-2)] \\
& = (1-k-m)\varphi(-m) + (k+m)\varphi(-m+1) \\
& + (k-1+m)[\varphi(-m+2) - 2\varphi(-m+1) + \varphi(-m)] \\
& + (k-2+m)[\varphi(-m+3) - 2\varphi(-m+2) + \varphi(-m+1)] \\
& + (k-3+m)[\varphi(-m+4) - 2\varphi(-m+3) + \varphi(-m+2)] + \dots \\
& + 3[\varphi(k-2) - 2\varphi(k-3) + \varphi(k-4)] + 2[\varphi(k-1) - 2\varphi(k-2) + \varphi(k-3)] \\
& + [\varphi(k) - 2\varphi(k-1) + \varphi(k-2)] \\
& = \varphi(-m)[1-k-m+k-1+m] \\
& + \varphi(-m+1)[k+m-2(k-1+m)+k-2+m] \\
& + \varphi(-m+2)[k-1+m-2(k-2+m)+k-3+m] \\
& + \varphi(-m+3)[k-2+m-2(k-3+m)+k-4+m] + \dots \\
& + \varphi(k-3)[4-2\cdot 3+2] + \varphi(k-2)[3-2\cdot 2+1] + \varphi(k-1)[2-2] + \varphi(k) \\
& = \varphi(-m)\cdot 0 + \varphi(-m+1)\cdot 0 + \dots + \varphi(k-1)\cdot 0 + \varphi(k) = \varphi(k).
\end{aligned}$$

So we shown that  $x(k) = \varphi(k)$  for any  $k \in \mathbb{Z}_{-m}^1$ .

□

### 3 REPRESENTATION OF SOLUTIONS OF NONHOMOGENOUS SYSTEM

Consider a nonhomogenous equation (1)

$$\Delta^2 x(k) + B^2 x(k-m) = f(k), \quad k \in \mathbb{Z}_0^\infty,$$

with the zero initial conditions

$$x(k) = \theta, \quad k \in \mathbb{Z}_{-m}^1, \quad (13)$$

where  $f(k) = (f_1(k), \dots, f_n(k))^T$ .

**Theorem 3.** *If  $B$  is a regular  $n \times n$  matrix, solution  $x_p(k)$  of nonhomogenous equation (1) with the zero initial condition (13) has form*

$$x_p(k) = B^{-1} \cdot \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j), \quad (14)$$

where  $k \in \mathbb{Z}_2^\infty$ .

*Proof.* We show that the expression (14) satisfies the nonhomogenous equation (1), i.e.

$$\Delta^2 x_p(k) + B^2 \cdot x_p(k - m) = f(k), \quad k \in \mathbb{Z}_2^\infty. \quad (15)$$

The left-hand side of (15) equals:

$$\begin{aligned} L &:= \Delta^2 \left( B^{-1} \cdot \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j) \right) \\ &\quad + B^2 \cdot B^{-1} \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\ &= B^{-1} \cdot \Delta^2 \left( \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j) \right) \\ &\quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\ &= B^{-1} \cdot \Delta \left( \sum_{j=0}^{k-1} \text{Sin}_m B(k-m-j) \cdot f(j) - \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j) \right) \\ &\quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\ &= B^{-1} \cdot \Delta \left( \text{Sin}_m B(-m+1) \cdot f(k-1) + \sum_{j=0}^{k-2} \text{Sin}_m B(k-m-j) \cdot f(j) \right. \\ &\quad \left. - \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j) \right) + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j). \end{aligned}$$

By Definition 2,  $\text{Sin}_m B(-m+1) = B$ . So we get:

$$\begin{aligned} L &= B^{-1} \cdot \Delta \left[ B \cdot f(k-1) + \sum_{j=0}^{k-2} \left( \text{Sin}_m B(k-m-j) \cdot f(j) \right. \right. \\ &\quad \left. \left. - \text{Sin}_m B(k-1-m-j) \cdot f(j) \right) \right] + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\ &= \Delta f(k-1) + B^{-1} \cdot \Delta \sum_{j=0}^{k-2} \Delta \text{Sin}_m B(k-1-m-j) \cdot f(j) \\ &\quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j). \end{aligned}$$

By Theorem 1,

$$L = \Delta f(k-1) + B^{-1} \cdot \Delta \sum_{j=0}^{k-2} b \cdot \text{Cos}_m B(k-1-m-j) \cdot f(j)$$

$$\begin{aligned}
& + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = \Delta f(k-1) + \Delta \sum_{j=0}^{k-2} \text{Cos}_m B(k-1-m-j) \cdot f(j) \\
& \quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = \Delta f(k-1) + \sum_{j=0}^{k-1} \text{Cos}_m B(k-m-j) \cdot f(j) - \sum_{j=0}^{k-2} \text{Cos}_m B(k-1-m-j) \cdot f(j) \\
& \quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = \Delta f(k-1) + \text{Cos}_m B(-m+1) \cdot f(k-1) + \sum_{j=0}^{k-2} \text{Cos}_m B(k-m-j) \cdot f(j) \\
& \quad - \sum_{j=0}^{k-2} \text{Cos}_m B(k-1-m-j) \cdot f(j) + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j).
\end{aligned}$$

By Definition 1,  $\text{Cos}_m B(-m+1) = I$ . Therefore

$$\begin{aligned}
L & = \Delta f(k-1) + f(k-1) \\
& \quad + \sum_{j=0}^{k-2} \left( \text{Cos}_m B(k-m-j) \cdot f(j) - \text{Cos}_m B(k-1-m-j) \cdot f(j) \right) \\
& \quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = f(k) - f(k-1) + f(k-1) + \sum_{j=0}^{k-2} \Delta \text{Cos}_m B(k-1-m-j) \cdot f(j) \\
& \quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B((k-1-2m-j)) \cdot f(j).
\end{aligned}$$

We use Theorem 1:

$$\begin{aligned}
L & = f(k) + \sum_{j=0}^{k-2} \left( -B \cdot \text{Sin}_m B(k-1-2m-j) \cdot f(j) \right) \\
& \quad + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = f(k) - B \cdot \left[ \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \right]
\end{aligned}$$

$$\begin{aligned}
& + \sum_{j=k-m-1}^{k-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \Big] + B \cdot \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = f(k) + B \cdot \left[ \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \right. \\
& \quad \left. - \sum_{j=0}^{k-m-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \right] - B \cdot \sum_{j=k-m-1}^{k-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = f(k) - B \cdot \sum_{j=k-m-1}^{k-2} \text{Sin}_m B(k-1-2m-j) \cdot f(j) \\
& = f(k) - B \cdot \left[ \text{Sin}_m B(-m) \cdot f(k-m-1) + \cdots + \text{Sin}_m B(-2m+1) \cdot f(k-2) \right].
\end{aligned}$$

Since, by Definition 2,

$$\text{Sin}_m B(-m) = \cdots = \text{Sin}_m B(-2m+1) = \Theta,$$

we get

$$L = f(k) = R,$$

where  $R$  is the right-hand side of (15). □

**Theorem 4.** *Solution  $x(k)$  of the problem (1), (11) can be represented in the form*

$$\begin{aligned}
x(k) = & (\text{Cos}_m Bk) \varphi(-m) + B^{-1} \cdot \left[ (\text{Sin}_m Bk) \Delta \varphi(-m) \right. \\
& + \sum_{j=-m+1}^0 \text{Sin}_m B(k-m-j) \cdot \Delta^2 \varphi(j-1) \Big] \\
& + B^{-1} \cdot \sum_{j=0}^{k-2} \text{Sin}_m B(k-1-m-j) \cdot f(j),
\end{aligned} \tag{16}$$

where  $k \in \mathbb{Z}_{-m}^\infty$ .

## 4 EXAMPLE

**Example 1.** It is given a two-dimensional ( $n = 2$ ) nonhomogenous system (1):

$$\Delta^2 x(k) + Bx(k-3) = f(k) \tag{17}$$

with initial function

$$\varphi(-3) = \varphi(-2) = \varphi(-1) = \varphi(0) = (0, 0)^T, \varphi(1) = (0.001, 0.001)^T, \tag{18}$$

where

$$B = \begin{pmatrix} 0.001 & -0.001 \\ 0.001 & 0 \end{pmatrix} \quad \text{and} \quad f(k) = \begin{cases} (1, 1)^T & \text{if } k = 0, \\ (0, 0)^T & \text{otherwise.} \end{cases}$$

*Solution.* We used program Maple 13 for computation 250 points of the given discrete equation. At the first we get calculate values of delayed cosine  $\text{Cos}_m Bk$  and delayed sine  $\text{Sin}_m Bk$  (in Maple 13, they are called simply  $\text{Cos}[k]$  and  $\text{Sin}[k]$ ) for  $k \in \mathbb{Z}_{-3}^{250}$  by Definition 1 and 2. Then we compute values of  $x(k)$  (called  $x[k]$ ) by Theorem 4.

```
> s:=ceil((k-1)/(m+2)):

> for k from -m to 250 do Cos[k]:=evalf[100](sum
  ((-1)^j*B^(2*j)*binomial(k-j*m+m, 2*j), j=0..s)) od:

> for k from -m to 250 do Sin[k]:=evalf[100](sum
  ((-1)^j*B^(2*j+1)*binomial(k-j*m+m, 2*j+1), j=0..s))
od:

> for k from 2 to 250 do x[k]:=evalf[40](Cos[k].x[-m]
  + B^(-1).(Sin[k].(x[-m+1]-x[-m])
  + sum(Sin[k-m-j].(x[j+1]-2*x[j]+x[j-1]), j=-m+1..0))
  + B^(-1).sum(Sin[k-1-m-j].f[j], j=0..k-2)) od:
```

In Figure 1 there is given the graph of the solution of the problem (17), (18), where  $x(k) = (x_1(k), x_2(k))^T$ . Points of the solution are represented by black color, their plan view by blue and side view by red color.

## CONCLUSION

In the paper, linear systems of discrete equations (1) of higher-order are considered. New formulas (14), (16) were derived for solutions of initial-value problems for homogenous and nonhomogenous systems (1) with the aid of special discrete matrix functions called the delayed discrete matrix cosine and sine. The formulas can be used (unlike known numerical algorithms) to qualitative analysis of solutions (such as impact estimation of initial data on properties of solutions or large-time behaviour ( $k \rightarrow \infty$ ) of solutions). An example illustrating obtained results is worked out (by Maple software) and graphically demonstrated as well.

## References

- [1] Diblík, J., Khusainov, D. Ya.: *Representation of solutions of linear discrete systems with constant coefficients and pure delay*, Advances in Difference Equations, Volume 2006, Article ID 80825, Pages 1–13.
- [2] Diblík, J., Mencáková, K. Formula for explicit solutions of a class of linear discrete equations with delay. In: *Mathematics, Information Technologies and Applied Sciences 2016*,

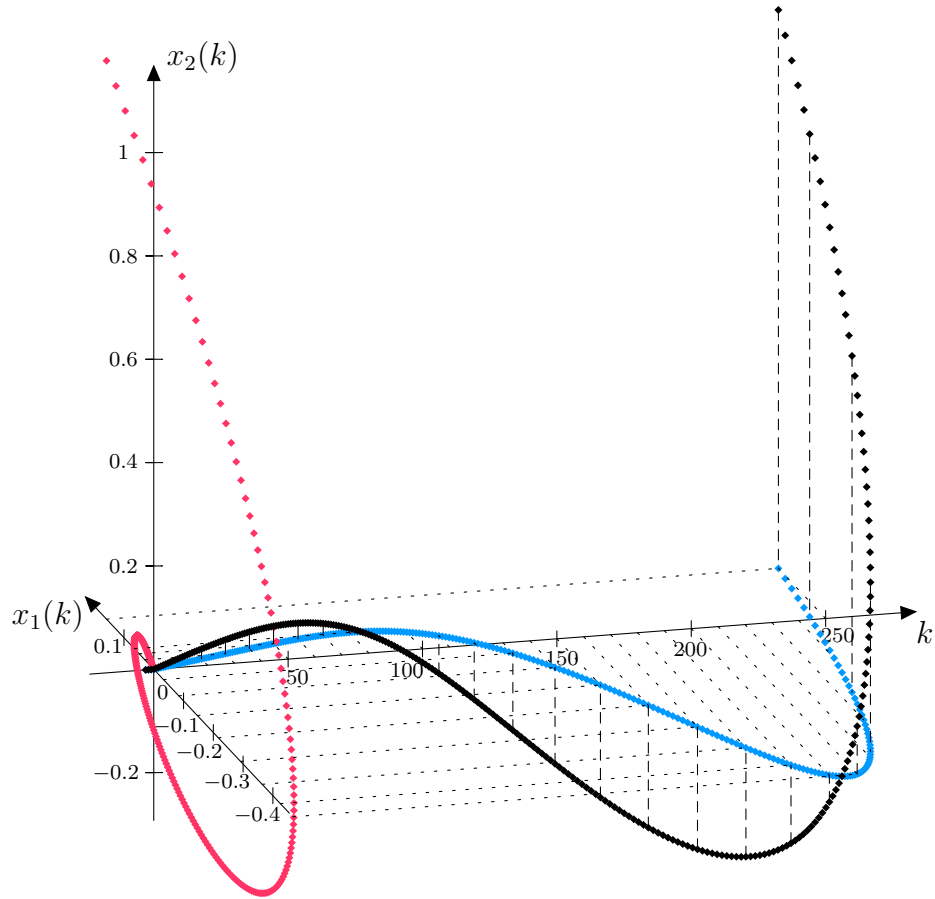


Figure 1: The graph of the solution of Example 1.

*post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 42–55. [Online]. [Cit.2017-07-26]. Available at: <http://mitav.unob.cz/data/MITAV%202016%20Proceedings.pdf>. ISBN 978-80-7231-400-3.

- [3] Khusainov, D. Ya., Diblík, J., Růžicková, M., Lukáčová, J.: *Representation of a solution of the Cauchy problem for an oscillating system with pure delay*, Nonlinear Oscillations, Volume 11, No. 2, 2008, Pages 276–285.
- [4] Diblík, J., Fečkan, M., Pospíšil, M.: *Representation of a solution of the Cauchy problem for an oscillating system with multiple delays and pairwise permutable matrices*, Hindawi Publishing Corporation, Abstract and Applied Analysis, Volume 2013, Article ID 931493, 10 pages, <http://dx.doi.org/10.1155/2013/931493>.

## Acknowledgement

The authors were supported by Grant FEKT-S-17-4225 of Faculty of Electrical Engineering and Communication, BUT.

# ON A QUASILINEAR PDE MODEL OF POPULATION DYNAMICS WITH RANDOM PARAMETERS

**Irada Dzhalladova**

Kyiv National Economic University named after Vadym Hetman  
Department of Computer Mathematics and Information Security  
54/1 Peremohy Ave.  
Kyiv, UA-03068, Ukraine  
Email: idzhalladova@gmail.com

**Michael Pokojovy**

The University of Texas at El Paso  
Department of Mathematical Sciences  
500 West University Ave.  
El Paso, TX 79968, USA  
Email: mpokojovy@utep.edu

**Abstract:** *A second-order quasilinear parabolic PDE system with random parameters is proposed to model the spatial-temporal evolution of multiple species dwelling on a common territory. For the associated stochastic Cauchy problem, the global well-posedness and long-time behavior are studied in a probabilistic weak functional-analytic framework under appropriate conditions on the data.*

**Keywords:** second-order parabolic PDE, random parameters, mild solutions, well-posedness, long-time behavior

## INTRODUCTION

Modeling and investigating the dynamics of populations is commonly viewed as one of central topics of modern mathematical demography, population biology and ecology (cf. [12]). Having its origin in the works of Malthus dating back to 1798 and historically preceded by Fibonacci's elementary considerations from 1202, the mathematical theory of population dynamics underwent a rapid growth during the 19<sup>th</sup> and 20<sup>th</sup> centuries. Among others, one should mention the works of Sharpe (1911), Lotka (1911 and 1924), Volterra (1926), McKendrick (1926), Kositzin (late 1930s), Fisher (1937), Kolmogorov (1937), Leslie (1945), Skellam (1950-s and 1970-s), Keyfitz (1950-s through 1980-s), Fredrickson & Hoppensteadt (1971 and 1975), Gurtin (1973), Gurtin & MacCamy (1981), etc. An age- and sex-structured model has recently been proposed by Pokojovy & Skvarkovsyi in [12]. For a detailed historical overview, we refer the reader to the monographs by Ianelli *et al.* [8] and Okubo & Levin [11] and references therein.

While a vast number of deterministic models are available in the literature, stochastic models are still rather scarce and mainly represented by Kolmogorov-type deterministic equations for the probability density of underlying Markovian diffusions (cf. [10]). Genuine (finite-dimensional) stochastic models are also available [9, 13].



Next, we present our new stochastic spatial-temporal population dynamics model. Consider a macroscopic description of the temporal evolution of  $m \in \mathbb{N}$  biological species dwelling on a common territory parametrized by a bounded domain  $G \subset \mathbb{R}^d$ . Whereas we require our model to account for the spatial distribution of the species including the diffusion and drifting phenomena, for the sake of simplicity, the age structure and (possible) intra-species morphological differences are neglected, etc. In addition to nonlinear local interactions between the species, stochastic parameters are incorporated into the model to better describe the environmental impact on the species. A detailed overview on related (stochastic and deterministic) models is given in [11].

Let  $(\xi_t)_{t \geq 0}$  and  $(\eta_t)_{t \geq 0}$  be random processes taking their values in some spaces of  $x$ -dependent functions such that  $\xi_t(\cdot, x)$  and  $\eta_t(\cdot, x)$  describe the environmental, climatic or any other conditions at time  $t \geq 0$  and place  $x \in \bar{G}$ . Let  $u_i(t, x)$  denote the population density of the  $i$ -th species at time  $t \geq 0$  at point  $x \in \bar{G}$ . Consider the vector function  $u := (u_1, \dots, u_m)^T$  and its Jacobian  $\nabla u = (\partial_{x_j} u_i)_{i=1, \dots, m}^{j=1, \dots, d}$ . In the following, we employ the Einstein's summation convention. For the indices  $i, j, k, l$ , we have  $i, k = 1, \dots, m$  and  $j, l = 1, \dots, d$ . Imposing a continuity equation and a Fick-type relation (reminiscent of the Fourier law of heat conduction) between the flux and the concentration gradient, we arrive at the equation

$$\begin{aligned} \partial_t u_i(t, x) = & \partial_{x_j} \left( a_{ijkl}(t, x, u(t, x), \nabla u(t, x), \xi_t) \partial_{x_l} u_k(t, x) \right) \\ & + b_i(t, x, \eta_t, u(t, x), \nabla u(t, x)) \text{ for } (t, x) \in (0, \infty) \times G, \end{aligned} \quad (1)$$

where  $b_i(t, x)$  stands for the local animal (net) ‘‘creation’’ intensity typically given as a polynomial in  $u_i$ 's. Since additive noise terms are less realistic for macroscopic population dynamics phenomena, we let the diffusion and the drift depend on stochastic ‘parameter’ processes.

Let  $\Gamma_0, \Gamma_1$  be relatively open, disjoint subsets of  $\partial G$  and let  $\nu(x)$  denote the outer unit normal vector to  $G$  at point  $x \in \partial G$ . With  $\bar{u}_i$  standing for the size of the  $i$ -th species at part  $\Gamma_0$  of the boundary and  $\bar{q}_i$  denoting the flow of  $i$ -th population in the direction of the outer normal at  $\Gamma_1$ , the boundary conditions read as

$$\begin{aligned} u_i(t, x) = & \bar{u}_i(t, x) \\ & \text{for } (t, x) \in (0, \infty) \times \Gamma_0, \\ \nu_j(x) \left( a_{ijkl}(t, x, u(t, x), \nabla u(t, x), \xi_t) \partial_{x_l} u_k(t, x) \right) = & \bar{q}_i(t, x) \\ & \text{for } (t, x) \in (0, \infty) \times \Gamma_1. \end{aligned} \quad (2)$$

Usually,  $\bar{u}_i \equiv \text{const}$  and  $\bar{q}_i \equiv 0$ . Finally, the initial conditions are given as

$$u_i(0, x) = u_i^0(x) \text{ for } x \in G, \quad (3)$$

where  $u_i^0$  is the size of the  $i$ -th species at the initial point of time.

The goal is to analyze Equations (1)–(3), discuss their well-posedness and study the long-time behavior in an appropriate probabilistic functional-analytic framework proposed below. See [7] for details. In contrast to the vast majority of stochastic partial differential equation (SPDE) models with additive – white or colored – noise studied in the recent literature, Equation (2) rather depends on (possibly) quite irregular stochastic data and/or parameter processes.

## WELL-POSEDNESS

Let  $H, V$  be separable Hilbert spaces such that the embedding  $V \hookrightarrow H$  is dense and continuous and let  $(\Omega, \mathcal{F}, \mathbf{P})$  be a probability space. Further, let  $\Theta_1, \Theta_2$  be separable Hilbert spaces (or closed subsets thereof). Let

$$\{\mathcal{A}(t, w; \theta_1) \mid t \geq 0, w \in V, \theta_1 \in \Theta_1\} \subset L(V, V')$$

be a family of bounded, linear, self-adjoint, positive operators with

$$(t, w, \theta_1) \mapsto \mathcal{A}(t, w; \theta_1) \in L_{\text{loc}}^\infty(0, \infty; \text{Lip}(V \times \Theta_1, L(V, V')))$$

such that for any  $T > 0$  there exists a number  $\kappa = \kappa(T) > 0$  with

$$\langle \mathcal{A}(t, w; \theta_1)v, v \rangle_{V'; V} \geq \kappa \|v\|_V^2 \text{ for } t \in [0, T], v, w \in V, \theta_1 \in \Theta_1.$$

Further, let

$$(t, w, \theta_2) \mapsto f(t, w; \theta_2) \in L_{\text{loc}}^2(0, \infty; \text{Lip}(V \times \Theta_1, V')).$$

For given  $L^2(\Omega, L_{\text{loc}}^2(0, \infty; \Theta_1); \mathbf{P}, \mathcal{F})$ - and  $L^2(\Omega, L_{\text{loc}}^2(0, \infty; \Theta_2); \mathbf{P}, \mathcal{F})$ -stochastic processes  $(\xi_t^1)_{t \geq 0}$  and  $(\xi_t^2)_{t \geq 0}$ , consider a quasilinear stochastic Cauchy problem

$$\partial_t u(t) + \mathcal{A}(t, u(t); \xi_t^1)u(t) = f(t, u(t); \xi_t^2) \text{ in } V' \text{ for a.e. } t \geq 0 \text{ } \mathbf{P}\text{-a.s. in } \Omega, \quad (4)$$

$$u(0) = u^0 \text{ in } H \text{ } \mathbf{P}\text{-a.s. in } \Omega. \quad (5)$$

**Theorem 1.** *For any initial data  $u^0 \in L^2(\Omega, H; \mathbf{P}, \mathcal{F})$ , there exists a unique weak solution*

$$u \in L^2\left(\Omega, H_{\text{loc}}^1(0, \infty; V') \cap L_{\text{loc}}^2(0, \infty; V); \mathbf{P}, \mathcal{F}\right)$$

*to Equations (4)–(5). Moreover,  $u$  continuously depends on  $u^0$  in respective topologies.*

*Sketch of the proof.* The proof is based on a Kato-type linearization and application of the classical linear variational parabolic theory (see, e.g., [2, Chapter XVIII]). For a given realization of  $(\xi_t^1, \xi_t^2)$ , using standard techniques, the resulting quasilinear deterministic problem can be solved using Banach's fixed-point theorem. Next, using the Lipschitz-continuity of nonlinearities, the solution is shown to be a Lipschitzian function of  $(\xi_t^1, \xi_t^2)$ . Since the solution process can uniquely be represented as a composition of the solution operator with the data processes  $(\xi_t^1, \xi_t^2)$ , the unique existence and measurability of the solution process follow, while the integrability is a direct consequence of the solution operator Lipschitzianity.  $\square$

**Remark 2.** *Under additional regularity assumptions on the “data” process  $\xi_t^1, \xi_t^2$  (cf. [5]), the weak solution in Theorem 1 can be shown to possess the regularity of a strong solution, i.e.,*

$$u \in L^2\left(\Omega, H_{\text{loc}}^1(0, \infty; V); \mathbf{P}, \mathcal{F}\right) \text{ with } \mathcal{A}(\cdot, u; \xi^1)u \in L^2\left(\Omega, L_{\text{loc}}^2(0, \infty; H); \mathbf{P}, \mathcal{F}\right).$$

To put the original problem from Equations (1)–(3) into the framework above, we let

$$H := \begin{cases} (L^2(G))^m, & \Gamma_0 \neq \emptyset \\ (L^2(G)/\{1\})^m, & \Gamma_0 = \emptyset \end{cases} \text{ and } V := (H_{\Gamma_0}^1(G))^m \cap H$$

and consider for  $t \geq 0$ ,  $w \in V$  and  $\theta_1 \in \Theta_1$

$$\mathcal{A}(t, w, \theta_1): V \rightarrow V', \quad u \mapsto -\operatorname{div}(a(t, \cdot, w, \nabla w, \theta_1) \nabla u)$$

and

$$f(t, w, \theta_2) := b(t, \cdot, w, \nabla w, \theta_2) \text{ for } t \geq 0, w \in V, \theta_2 \in \Theta_2.$$

Now, the conditions of Theorem 1 can easily be interpreted in terms of appropriate assumptions on the functions/operators  $a = (a_{ijkl})$  and  $b = (b_i)$ .

**Remark 3.** For Equations (1)–(3) to possess a weak solution, both processes  $\xi_t^1, \xi_t^2$  need to be time-square-integrable. While the later is true, e.g., for respective Hilbert-space-valued Wiener processes, the “white noise” would violate this property. If the solution process is additionally required to be ( $\mathbf{P}$ -a.s.) a strong solution, an extra regularity assumption on  $\xi_t^1$  such as the boundedness of the total variation (cf. [5]) or the Hölder-continuity of degree  $\alpha > \frac{1}{2}$  becomes important. This rules out the possibility of  $\xi_t^1$  being a Wiener process. At the same time, an integrated Wiener process or a vast class of semi-Markovian processes would comply with this requirement.

## LONG-TIME BEHAVIOR

Stability of stochastic systems has attracted considerable attention in the recent literature [1, 3, 4, 6], etc. The more prominent solution approaches include Lyapunov energy methods, spectral techniques, moment equations, stochastic observability instruments, etc. In the present work, we adopt the classical Lyapunov’s method.

**Theorem 4.** Suppose there exists a number  $\kappa > 0$  such that

$$\langle \mathcal{A}(t, v; \theta_1) v, v \rangle_{V'; V} \geq \kappa \|v\|_V^2 \text{ for all } t \geq 0, v \in V, \theta_1 \in \Theta_1.$$

Further, let

$$\langle f(t, v; \theta_2), v \rangle_H \leq 0 \text{ for } t \geq 0, v \in V \text{ and } \theta_2 \in \Theta_2. \quad (6)$$

Then, under conditions of Theorem 1, the unique weak solution  $u$  to Equations (4)–(5) is exponentially stable on  $H$  in the 2-mean, i.e., there exists a number  $\alpha > 0$  such that

$$\mathbf{E}[\|u(t)\|_H^2] \leq \exp(-2\alpha t) \mathbf{E}[\|u^0\|_H^2] \text{ for } t \geq 0. \quad (7)$$

*Sketch of the proof.* Assuming for the moment, Equations (4)–(5) possess a strong solution, consider the Lyapunov functional

$$E(t) := \frac{1}{2} \|u(t)\|^2,$$

where the Einstein’s summation convention is employed. Multiplying Equation (4) in  $H$  with  $u(t)$ , taking the expectation with respect to the probability measure  $\mathbf{P}$ , ‘integrating by parts’ and using the uniform coercivity of  $\mathcal{A}$  along with Equation (6), we arrive at the estimate

$$\frac{d}{dt} \mathbf{E}[E(t)] \leq -2\alpha \mathbf{E}[E(t)] \text{ for a.e. } t \geq 0$$

for an appropriate constant  $\alpha > 0$  that neither depends on  $u^0$  nor on  $(\xi_t^1, \xi_t^2)$ . Using Gronwall's inequality, Equation (7) follows.

Turning to the general case, i.e.,  $u$  is a weak solution, select sequences to approximate the initial data  $u^0$  and the data process  $(\xi_t^1, \xi_t^2)$  such that every element of the approximating sequence admits a strong solution. The respective solutions satisfy Equation (7). Recalling the solution map is Lipschitzian in  $(\xi_t^1, \xi_t^2)$  and continuous in  $u^0$ , we pass to the limit and observe that the limiting (weak) solution satisfies Equation (7) as well.  $\square$

## CONCLUSION

We presented a new quasilinear stochastic PDE model to describe the spatial-temporal evolution of multiple animal species and proposed an approach to studying the well-posedness for the underlying stochastic Cauchy problem. Further, under additional conditions on the diffusion and source terms, the exponential stability in the 2-mean was discussed. Future research directions will include deduction of moment equations for the case of Markovian and semi-Markovian data processes  $(\xi_t^1, \xi_t^2)$  and their application to stabilization and optimal control problems, etc.

## References

- [1] Bařtinec, J., Dzhalladova, I. Sufficient conditions for stability of solutions of systems of nonlinear differential equations with right-hand side depending on Markov's process. In: 7. konference o matematice a fyzice na vysokých školách technických s mezinárodní účastí. 2011. pp. 23–29. ISBN: 978-80-7231-815-5.
- [2] Dautray, R., Lions, J.-L. *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol - 5, Springer-Verlag, Berlin, 1992, pp. I-XIV, 1–739.
- [3] Diblík, J., Dzhalladova, I., Růžicková, M. The stability of nonlinear differential systems with random parameters. *Abstract and Applied Analysis*, 2012, vol. 2012, pp. 1–27.
- [4] Diblík, J., Dzhalladova, I., Růžicková, M. Stabilization of company's income modeled by a system of discrete stochastic equations. *Advances in Difference Equations*, 2014, vol. 2014, no. 2014, pp. 1–8.
- [5] Dier, D., Non-autonomous maximal regularity for forms of bounded variation, *Journal of Mathematical Analysis and Applications*, 2015, vol. 425, pp. 33–54.
- [6] Dzhalladova, I., Bařtinec, J., Diblík, J., Khusainov, D. Estimates of exponential stability for solutions of stochastic control systems with delay. *Abstract and Applied Analysis*, 2011, vol. 2011, no. 1, pp. 1–14.
- [7] Dzhalladova, I., Jobe, J. M., Pokojovy, M. On a spatially distributed stochastic population dynamics model (in preparation).
- [8] Ianelli, M., Martcheva, M., Milner, F. A. Gender-structured population modeling: Mathematical Methods, Numerics, and Simulations, in *Frontiers in Applied Mathematics*, SIAM, Philadelphia, 2005.
- [9] Lande, R., Engen, S., Saether, B.-E., *Stochastic Population Dynamics in Ecology and Conservation*, Oxford University Press, Oxford – New York, 2003.
- [10] Kunisch, K., Schappacher, W., Webb, G. F. Nonlinear age-dependent population dynamics with random diffusion, *Comput. Math. Appl.*, 11, (1985), pp. 155–173.

- [11] Okubo, A., Levin, S. A.: *Diffusion and Ecological Problems. Modern perspectives*, Springer Verlag, New York, Berlin, Heidelberg, 2001, pp. 1–467.
- [12] Pokojovy, M., Skvarkovskyi, E.: Analysis and numerics for an age- and sex-structured population model, *Numerical Methods for Partial Differential Equations*, 2015, vol. 32, issue 2, pp. 706–736.
- [13] Vainstein, M. H., Rubí, J. M., Vilar, J. M. G. Stochastic population dynamics in turbulent fields, *Eur. Phys. J. Special Topics*, 2007, vol. 146, pp. 177–187.

### **Acknowledgement**

This work has been partially funded by a research grant from the Young Scholar Fund supported by the Deutsche Forschungsgemeinschaft (ZUK 52/2) at the University of Konstanz, Konstanz, Germany.

# SOME PROPERTIES OF COMPOSITIONS OF CONFORMAL AND GEODESIC MAPPINGS

Irena Hinterleitner

Faculty of Civil Engineering, Brno University of Technology

Veveri 95, Brno, Czech Republic

`hinterleitner.i@fce.vutbr.cz`

**Abstract:** *In the paper we studied fundamental properties of conformal and geodesic mappings of (pseudo-) Riemannian spaces. We study in detail compositions of conformal and geodesic mappings. In the case that an assembling of conformal and geodesic mappings is commuting, then this composition is either only conformal or only geodesic. We discuss also the exceptional case of dimension 2.*

**Keywords:** conformal mapping, geodesic mapping, composition of mappings, (pseudo-) Riemannian manifold.

## INTRODUCTION

In differential geometry conformal and geodesic mappings play a very important role in the theory of surfaces, Riemannian and pseudo-Riemannian manifolds.

J. Lagrange [11] began to study problems in cartography in 1779. He presented stereographic and gnomonic projections of a sphere onto a plane. These projections are examples of conformal and geodesic mappings. The above mentioned mappings of surfaces, Riemannian and pseudo-Riemannian spaces are used in many applications, for example in theoretical mechanics, physics and especially in the general theory of relativity [4, 5, 6, 7, 8, 9, 15, 16].

The general theory of conformal and geodesic mappings of (pseudo-) Riemannian manifolds was studied in [7, 8, 9, 10, 12, 13, 14, 15, 17, 18]. Further conformal and geodesic mappings of special spaces were studied for example in [4, 8, 16]. This problem is connected with the solution of differential equations, for example [2].

Since from the time of T. Levi-Civita [10] it is known that geodesic mappings which in the same time are conformal and geodesic are homothetic, i.e. the metrics of these spaces are proportional.

As known, conformal as well as geodesic mappings give rise to classes of conformally and geodesically equivalent metrics, i.e. they are reflexive, symmetric and transitive.

We prove that “commutativity” of the composition of conformal and geodesic mappings leads to triviality, i.e. this composition is either conformal or geodesic.

# 1 CONFORMAL AND GEODESIC MAPPINGS

## 1.1 Conformal mappings

*Conformal mappings* are mappings which preserve angles. These mappings are characterized by the condition that their metrics are proportional, i.e. the following equation

$$\bar{g}_{ij}(x) = e^{2\sigma(x)} g_{ij}(x) \quad (1)$$

holds, where  $x = (x^1, \dots, x^n)$  are common coordinates respective to the conformal mapping  $V_n \rightarrow \bar{V}_n$ ,  $g_{ij}(x)$  and  $\bar{g}_{ij}(x)$  are the metric tensors of the (pseudo-) Riemannian manifolds  $V_n$  and  $\bar{V}_n$ , respectively.

In the coordinate free form we can rewrite formula (1) in the following form

$$\bar{g} = e^{2\sigma} g.$$

If  $\sigma = \text{const}$ , then the conformal mapping is called a *homothetic*, and if  $\sigma = 0$  then this mapping is *isometric*.

From equation (1) follows that the Levi-Civita connections of  $V_n$  and  $\bar{V}_n$  are in the relation:

$$\bar{\Gamma}_{ij}^h(x) = \Gamma_{ij}^h(x) + \delta_i^h \sigma_j + \delta_j^h \sigma_i - \sigma^h g_{ij}, \quad (2)$$

where  $\sigma^h = g^{h\alpha} \sigma_\alpha$ ,  $\sigma_i = \nabla_i \sigma$ ,  $\delta_i^h$  is the Kronecker symbol, and  $\Gamma_{ij}^h$  and  $\bar{\Gamma}_{ij}^h$  are the Christoffel symbols of  $V_n$  and  $\bar{V}_n$ .

In equivalent form under the conformal mappings the following relation for any vector fields  $X, Y$  holds

$$\bar{\nabla}(X, Y) = \nabla(X, Y) + \sigma(X)Y + \sigma(Y)X - g(X, Y)\sigma,$$

where  $\sigma$  is a gradient one-form  $\sigma(X) = \nabla_X \sigma$ ,  $\sigma$  is a vector field for which  $\sigma(X) = g(X, \sigma)$ ,  $\nabla$  and  $\bar{\nabla}$  are the Levi-Civita connection on  $V_n$  and  $\bar{V}_n$ , respectively.

## 1.2 Geodesic mappings

A diffeomorphism  $f: V_n \rightarrow \bar{V}_n$  is called a *geodesic mapping*, if any geodesic on  $V_n$  is mapped onto a geodesic on  $\bar{V}_n$ .

A diffeomorphism  $f: V_n \rightarrow \bar{V}_n$  is geodesic if and only if the Levi-Civita equation holds

$$\bar{\Gamma}_{ij}^h(x) = \Gamma_{ij}^h(x) + \delta_i^h \psi_j(x) + \delta_j^h \psi_i(x), \quad (3)$$

where  $\Gamma_{ij}^h$  and  $\bar{\Gamma}_{ij}^h$  are Christoffel symbols of  $V_n$  and  $\bar{V}_n$ , and  $\psi_i(x)$  are components of a linear form  $\psi$ . Geodesic mapping for which  $\psi_i \equiv 0$  is called *trivial* or *affine*. Evidently, a homothetic mappings is a special affine mappings.

In the coordinate free form we can rewrite formula (3) as follows

$$\bar{\nabla}_X Y = \nabla_X Y + \psi(X)Y + \psi(Y)X$$

for any vector fields  $X, Y$ .

If  $V_n$  and  $\bar{V}_n$  are (pseudo-) Riemannian spaces, then  $\psi$  is a gradient like form

$$\psi_i = \frac{1}{n+1} \partial_i \ln \sqrt{\left| \frac{\bar{g}}{g} \right|},$$

where  $g = \det(g_{ij})$  and  $\bar{g} = \det(\bar{g}_{ij})$ .

### 1.3 Geodesic mappings which are conformal

We prove the following lemma, see [18, p. 75].

**Lemma 1** *A diffeomorphism  $f: V_n \rightarrow \bar{V}_n$  ( $n \geq 2$ ) which is at the same time conformal and geodesic is homothetic.*

*Proof.* It follows that a geodesic and conformal mapping must at the same time satisfy conditions (2) and (3):

$$\bar{\Gamma}_{ij}^h(x) - \Gamma_{ij}^h(x) = \delta_i^h \sigma_j(x) + \delta_j^h \sigma_i(x) - \sigma^h g_{ij} = \delta_i^h \psi_j(x) + \delta_j^h \psi_i(x).$$

From that follows

$$\delta_i^h w_j + \delta_j^h w_i - \sigma^h g_{ij} = 0,$$

where  $w_i = \sigma_i - \psi_i$ .

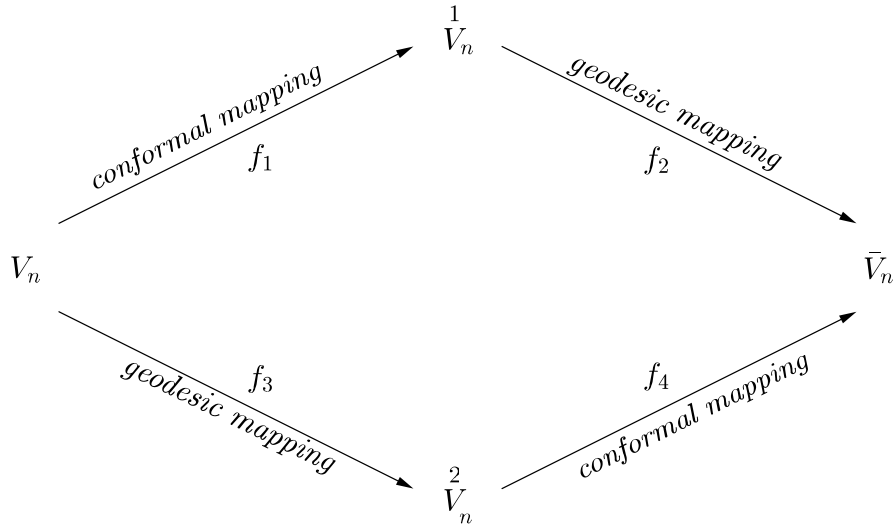
We can see that if  $n \geq 2$ , then from the last formula follows  $\psi_i = \sigma_i = 0$ , and  $\sigma = \text{const}$ .

## 2 COMPOSITION OF CONFORMAL AND GEODESIC MAPPINGS

### 2.1 General properties

As we have said earlier conformal and geodesic mappings are very important. We will be interested what happens, if we make a composition of these mappings. One of the important property is that their composition is commutative. Hereafter we prove that in the case when the mappings commute, the result is either conformal or geodesic.

Now we will study the compositions of conformal and geodesic mappings, and their “commutativity”. We demonstrate the composition at the following diagram



**Fig. 1.** The composition of conformal and geodesic mappings



Here

$$\begin{aligned}
f_1: V_n &\rightarrow \overset{1}{V}_n \text{ is a conformal,} \\
f_2: \overset{1}{V}_n &\rightarrow \bar{V}_n \text{ is a geodesic,} \\
f_3: V_n &\rightarrow \overset{2}{V}_n \text{ is a geodesic, and} \\
f_4: \overset{2}{V}_n &\rightarrow \bar{V}_n \text{ is a conformal mapping,}
\end{aligned}$$

where  $V_n$ ,  $\overset{1}{V}_n$ ,  $\overset{2}{V}_n$  and  $\bar{V}_n$  are (pseudo-) Riemannian spaces with metric tensors  $g$ ,  $\overset{1}{g}$ ,  $\overset{2}{g}$  and  $\bar{g}$ , respectively, and Christoffel symbols  $\Gamma$ ,  $\overset{1}{\Gamma}$ ,  $\overset{2}{\Gamma}$  and  $\bar{\Gamma}$ , respectively.

## 2.2 Main theorem of commutativity of composition of conformal and geodesic mappings

We prove the following theorem

**Theorem 1** *If the dimension  $n > 3$  and mapping*

$$f = f_1 \circ f_2 = f_3 \circ f_4 : V_n \rightarrow \bar{V}_n, \quad (4)$$

*then  $f$  is conformal or geodesic.*

*Moreover,  $f_2$  and  $f_3$ , or  $f_1$  and  $f_4$ , are homothetic mappings.*

**Note.** From condition (4) follows that the conformal and the geodesics mapping commute.

*Proof.* These mappings be can seen at the above diagram. From the condition of the theorem follows that the Christoffel's symbols of the spaces  $V_n, \overset{1}{V}_n, \overset{2}{V}_n, \bar{V}_n$  satisfy the following conditions

$$\begin{aligned}
\overset{1}{\Gamma}_{ij}^h(x) &= \Gamma_{ij}^h(x) + \delta_i^h \overset{1}{\sigma}_j + \delta_j^h \overset{1}{\sigma}_i - \overset{1}{\sigma}^h g_{ij} \\
\bar{\Gamma}_{ij}^h(x) &= \overset{1}{\Gamma}_{ij}^h(x) + \delta_i^h \overset{2}{\psi}_j + \delta_j^h \overset{2}{\psi}_i \\
\overset{2}{\Gamma}_{ij}^h(x) &= \Gamma_{ij}^h(x) + \delta_i^h \overset{3}{\psi}_j + \delta_j^h \overset{3}{\psi}_i \\
\bar{\Gamma}_{ij}^h(x) &= \overset{2}{\Gamma}_{ij}^h(x) + \delta_i^h \overset{4}{\sigma}_j + \delta_j^h \overset{4}{\sigma}_i - \overset{4}{\sigma}^h \overset{2}{g}_{ij},
\end{aligned} \quad (5)$$

where  $\overset{1}{\sigma}_i, \overset{2}{\psi}_i, \overset{3}{\psi}_i, \overset{4}{\sigma}_i$  are gradient covectors,  $g_{ij}$  and  $\overset{2}{g}_{ij}$  are metric tensors on  $V_n$  and  $\bar{V}_n$ .

We add the first two equations in (5) and subtract the third and the fourth.

After the calculation we get

$$\overset{4}{\sigma}^h \overset{2}{g}_{ij} - \overset{4}{\sigma}^h g_{ij} + \delta_i^h w_j + \delta_j^h w_i = 0, \quad (6)$$

where  $w_i = \overset{1}{\sigma}_i + \overset{2}{\psi}_i - \overset{3}{\psi}_i - \overset{4}{\sigma}_i$ .

Now we analyze equation (6). If  $w_i \neq 0$  then there exists a vector  $a^i$  such that  $w_i a^i = 1$ . We contract (6) with  $a^j$  and obtain

$$\delta_i^h = -a^h w_i - \sigma^h \bar{g}_{i\alpha} a^\alpha + \sigma^h g_{i\alpha} a^\alpha.$$

From that follows in the case  $n > 3$  a contradiction. This means that if  $w_i = 0$ , from (6) follows that:

$$\sigma^h \bar{g}_{ij} = \sigma^h g_{ij}. \quad (7)$$

Because  $f_4$  is conformal and  $\bar{g}_{ij} = e^{2\sigma} \cdot \bar{g}_{ij}$ , from equation (7) follow two possibilities:

- a)  $\bar{g}_{ij}$  is proportional to  $g_{ij}$ , i.e.  $V_n$  and  $\bar{V}_n$  are conformally equivalent, or
- b)  $\sigma^h \equiv \sigma^h \equiv 0$ .

In the case b) the mappings  $f_1$  and  $f_4$  are homothetic and  $f_2$  and  $f_3$  are geodesic (in fact identical) mappings. From the condition  $w_i = 0$  in that case follows  $\psi_i \equiv \psi_i$ .

In the case a) we have that the mapping

$$f = f_1 \circ f_2 = f_3 \circ f_4 \quad (8)$$

is conformal.

Because the mappings  $f_1$  and  $f_4$  are also conformal, from (8) follows

$$f_2 = f_1^{-1} \circ f \quad \text{and} \quad f_3 = f \circ f_4^{-1}$$

are conformal.

On the other hand  $f_2$  and  $f_3$  are a priori geodesic mappings. From Lemma 1 follows that these mappings  $f_2$  and  $f_3$  are homothetic. Therefore  $\psi_i = \psi_i = 0$ .

From the above analysis, it can be observed that in the considered “commutative” combinations of geodesic and conformal mappings either the geodesic or the conformal mappings is homothetic. The theorem is proved.

## 2.3 Notes about compositions of conformal and geodesic mappings

By analysis of equations (1) of conformal mappings we can convince ourselves that the composition of conformal mappings is commutative. More generally it is known that (pseudo-) Riemannian spaces form closed equivalence classes with respect to conformal mappings [14], p. 238.

From the Levi-Civita equation (3) follow analogical properties for geodesic mappings of (pseudo-) Riemannian spaces, so we can speak about geodesic classes [14], p. 262.

In general Theorem 1 is not valid for  $n = 2$ . This follows from the fact that all two-dimensional Riemannian spaces are locally conformal equivalent. Analogically this holds for two-dimensional pseudo-Riemannian spaces. We can easily see that under geodesic mappings the signatura of the metric need not be conserved.

# CONCLUSION

In the article we introduced conformal and geodesic mappings and some relations between them. Afterwards we studied the composition of geodesic and conformal mappings and we proved that if a composition of a conformal and a geodesic mapping commutes then this mapping is only conformal or geodesic. We showed that this is not true for the dimension equal to two.

## References

- [1] Chudá, H., Shiha, M.: Conformal holomorphically projective mappings satisfying a certain initial condition. *Miskolc Math. Notes*, Vol. 14, No. 2, 2013, 569-574. ISSN: 1787-2405, 1787-2413/e.
- [2] Diblík, J., Mencáková, K.: Formula for explicit solutions of a class of linear discrete equations with delay. In: *Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 42-45. [Online]. [Cit. 2017-07-26]. Available at: <http://mitav.unob.cz/data/MITAV2016Proceedings.pdf>. ISBN 978-80-7231-400-3.
- [3] Eisenhart, L.P.: *Riemannian geometry*. Princeton: Univ. Press, 1926, pp. 306.
- [4] Evtushik, L.E., Hinterleitner, I., Guseva, N.I., Mikeš, J.: Conformal mappings onto Einstein spaces. *Russ. Math.*, Vol. 60, No. 10, 2016, 5-9. ISSN 1066-369X, 1934-810X/e.
- [5] Hinterleitner, I.: Conformally-projective harmonic diffeomorphisms of equidistant manifolds. In: *Proc. of the XV Int. workshop on Geometry and Physics*, Puerto de la Cruz, Spain, 2006. *Publ. de la RSME*, Vol. 11, 2007, p. 298-303. ISBN 978-84-935196-1-2/pbk.
- [6] Hinterleitner, I., Mikeš, J.: On the equations of conformally-projective harmonic mappings. In: *XXVI Int. workshop on Geometrical Methods in Physics*, Biaowiea, Poland, *AIP Conf. Proc.*, Vol. 956, 2007, p. 141-148. ISBN 978-0-7354-0470-0/hbk.
- [7] Hinterleitner, I., Mikeš, J.: Projective equivalence and spaces with equiaffine connection. *J. Math. Sci. (N.Y.)*, Vol. 177, 2011, p. 546-550. ISSN: 1072-3374.
- [8] Hinterleitner, I., Mikeš, J.: *Geodesic mappings and Einstein spaces*. In *Geometric methods in physics, Trends Math.*, Basel: Birkhäuser/Springer, 2013, p. 331-335. ISBN 978-3-0348-0447-9/hbk; 978-3-0348-0448-6/ebook.
- [9] Hinterleitner, I., Mikeš, J.: Geodesic mappings and differentiability of metrics, affine and projective connections. *Filomat*, Vol. 29, 2015, p. 1245-1249. ISSN: 0354-5180; 2406-0933/e.
- [10] Levi-Civita, T.: Sulle trasformazioni delle equazioni dinamiche. *Ann. di Mat. (2)*, Vol. 24, 1896, p. 255-300. ISSN: 0373-3114, 1618-1891/e.
- [11] Lagrange, J.L.: Sur la construction des cartes géographiques. *Nouveaux mémoires de l'Académie royale des sciences et belles-lettres de Berlin*, Vol. 4, 1779, p. 637-692.
- [12] Mikeš J. Geodesic mappings of affine-connected and Riemannian spaces. *J. Math. Sci. (New York)*, Vol. 78, No. 3, 1996, p. 311-333. ISSN: 1072-3374.
- [13] Mikeš, J., Vanžurová, A., Hinterleitner, I.: *Geodesic mappings and some generalizations*. Olomouc: Palacky University, 2009, pp. 304. ISBN 978-80-244-2524-5/pbk.
- [14] Mikeš, J. et al: *Differential geometry of special mappings*. Olomouc: Palacky University, 2015, pp. 566. ISBN 978-80-244-4671-4/pbk.
- [15] Mikeš, J., Berezovski, V., Stepanova, E., Chudá, H.: Geodesic mappings and their generalizations. *J. Math. Sci. (N.Y.)*, Vol. 217, No. 5, 2016, p. 607-623. ISSN: 1072-3374.

- [16] Najdanović, M., Zlatanović, M., Hinterleitner, I.: Conformal and geodesic mappings of generalized equidistant spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, Vol. 98(112), 2015, p. 71-84. ISSN: 0350-1302.
- [17] Petrov, A.Z.: *New methods in the general theory of relativity*. Moscow: Nauka, 1966, pp. 496.
- [18] Sinyukov N.S.: *Geodesic mappings of Riemannian spaces*. Moscow: Nauka, 1979. pp. 256.

## Acknowledgement

The paper was supported by the project No. LO1408 “AdMaS UP - Advanced Materials, Structures and Technologies”, supported by the Ministry of Education, Youth and Sports under the “National Sustainability Programme I” of the Brno University of Technology.

# FINDING THE SPECTRAL SENSITIVITY OF A PHOTODIODE WITH HELP OF ORTHOGONAL PROJECTION

Irena Hlavičková, Martin Motyčka, Jan Škoda

Faculty of Electrical Engineering and Communication, Brno University of Technology,  
Technická 3058/10, Brno, Czech Republic

hlavicka@feec.vutbr.cz, motyckam@feec.vutbr.cz, skoda@feec.vutbr.cz

**Abstract:** The paper deals with the mathematical description of the problem of finding the spectral sensitivity  $S$  of a photodiode with a linear response. Knowing the responses of the photodiode on testing lights, we try to find the function  $S$ . The lights are represented as functions depending on the wave length. It is shown that one of the possible solutions is to use an orthogonal projection to the space of certain functions.

**Keywords:** orthogonal projection, spectral sensitivity, quantum efficiency

## INTRODUCTION

We are searching for the so called spectral sensitivity – the relative quantum efficiency of light detection of a silicon photodiode or luxmeter. Luxmeter is a measuring device of illuminance and it is basically a photodiode with correction filter to human eye sensitivity  $V(\lambda)$  (see figure 1, for more about this theme see, e.g. [1]). We will denote this sensitivity function as  $S(\lambda)$  where  $\lambda$  means the wavelength of the light. We are trying to find  $S$  on the basis of knowing the responses of the photodiode or luxmeter on testing lights. Each of the lights is described by its spectrum. For the  $i$ -th light, it is the function

$$\phi_i(\lambda), \quad i = 1, \dots, n.$$

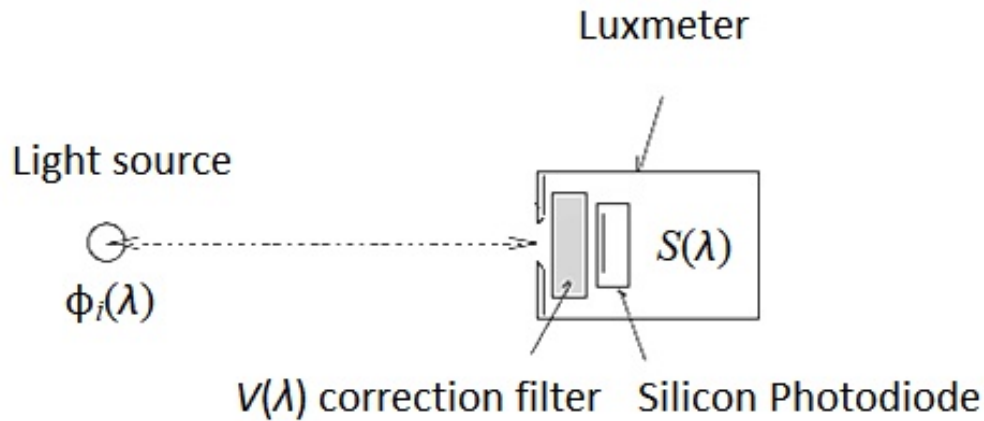


Figure 1: The measuring system scheme

The responses of the photodiode are

$$R_i = \int_a^b S(\lambda) \phi_i(\lambda) d\lambda, \quad i = 1, \dots, n. \quad (1)$$

The values  $a$  and  $b$  represent the end points of the range of the wavelengths, practically they can be set to approximately

$$a = 380[\text{nm}], \quad b = 780[\text{nm}].$$

So the task is: Find the function  $S$  if we know the values of integrals of the products of  $S$  with  $n$  known functions  $\phi_i$ . One of the possibilities how to do it is to use the orthogonal projection.

## 1 ORTHOGONAL PROJECTION ONTO A SUBSPACE

Let  $V$  be a vector space with the inner product  $\langle \cdot, \cdot \rangle$  and let  $L$  be its subspace with the basis  $h_1, h_2, \dots, h_n$ . To find the projection of the vector  $u \in V$  into  $L$ , we have to find the vectors  $v, w \in V$  such that

- (i)  $u = v + w$ ,
- (ii)  $v \in L$ ,
- (iii)  $\langle w, x \rangle = 0$  for every  $x \in L$ .

At the same time, the orthogonal projection  $v$  is the best approximation of the vector  $u$  in  $L$ .

It is well known that the vector  $v$  can be found as

$$v = \alpha_1 h_1 + \dots + \alpha_n h_n$$

where the coefficients  $\alpha_i$  are computed as the solution of the system of linear equations

$$\langle h_1, h_i \rangle \alpha_1 + \langle h_2, h_i \rangle \alpha_2 + \dots + \langle h_n, h_i \rangle \alpha_n = \langle u, h_i \rangle, \quad i = 1, \dots, n. \quad (2)$$

It is also well known that on the space of functions that are continuous on the interval  $\langle a, b \rangle$ , the inner product can be introduced as

$$\langle f, g \rangle = \int_a^b f(x)g(x) \, dx.$$

## 2 APPLICATION TO THE SOLVED PROBLEM

Looking at the system (2) and at (1), we can realize that the responses  $R_i$  can be seen as the right-hand sides of (2). The left-hand side, i.e. the matrix of the system, can be constructed by computing the integrals

$$\int_a^b \phi_i(\lambda) \phi_j(\lambda) d\lambda. \quad (3)$$

Finally, the approximation of the function  $S$  can be found as

$$\tilde{S}(\lambda) = \alpha_1 \phi_1(\lambda) + \alpha_2 \phi_2(\lambda) + \dots + \alpha_n \phi_n(\lambda).$$

Theoretically, all seems to be nice and clear. But practically, several problems arise.

## 2.1 Problems with practical implementation

The main problem is that all the values  $R_i$  and the functions  $\phi_i$  are obtained by measurement and thus they contain errors (noise). In fact, the functions  $\phi_i$  are given as tables of (measured) function values. Hence, the integrals (3) cannot be computed analytically. We have to compute them by some numerical method.

Another problem is the quality of the testing light functions  $\phi_i$ . They can be gained from various sources. “Nice” lights come from a programmable light source device which is able to produce any visible monochromatic light, but the system operation is very expensive. A cheaper variant is to use an incandescent light source (bulb) or xenon lamp with a colour filter, but the lights obtained this way can have unpleasant properties. For comparison, see figures 2, 3. It is very important to design the experiment properly, see, e.g. [2].

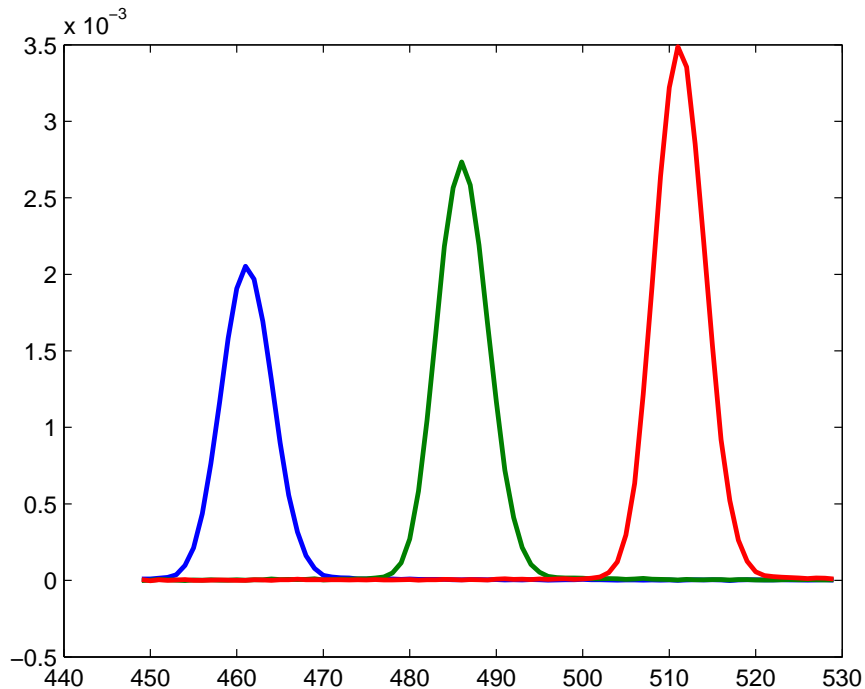


Figure 2: Three monochromatic lights

In figure 4 we can see the result of a numeric experiment. The computation was performed in Matlab, using 62 monochromatic testing lights  $\phi_i$ ,  $i = 1, \dots, 62$ . The integrals in system (2) were computed numerically with help of the Simpson method. The values of the resulting approximation of the function  $S(\lambda)$  are depicted as blue points. The red curve was obtained by smoothing these values by the moving average method.

## CONCLUSION

We have to admit that the numerical experiments still do not give results with the desired precision. The main cause lies in measured noise within the data and the fact that the used “monochromatic”

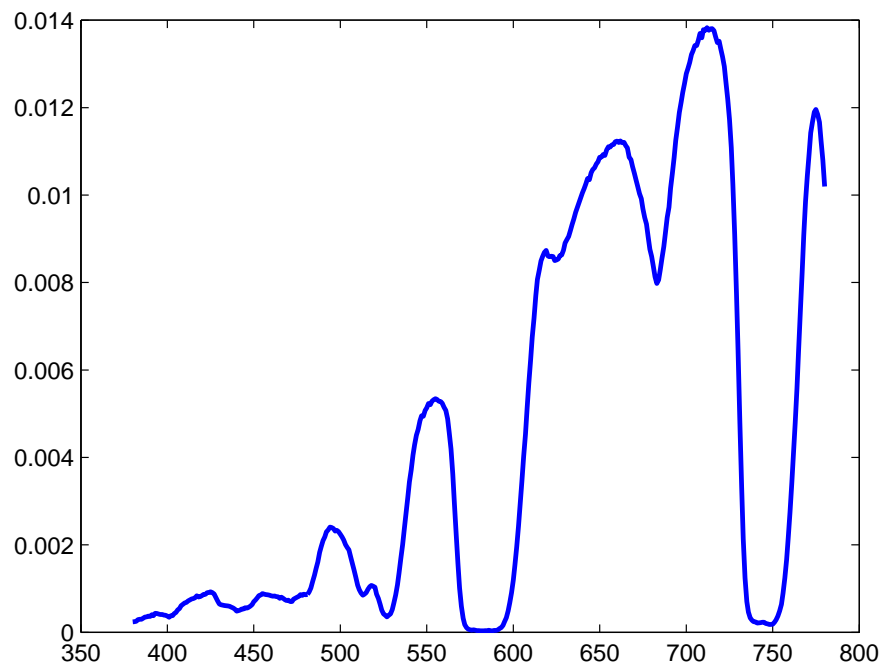


Figure 3: An example of an incandescent bulb light with filter

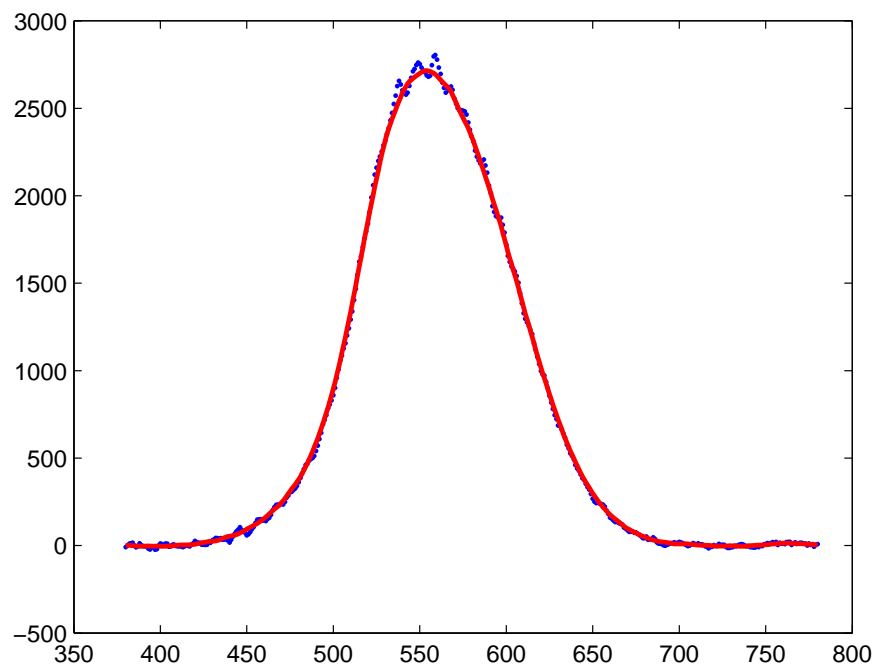


Figure 4: The result of a numeric experiment



test lights are not truly monochromatic. However, calculated sensitivity function fits well with expectations. The research goes on and some other possibilities how to find the spectral sensitivity  $S$  are in consideration, too. But, if nothing else, such an application of basic concepts of linear algebra can be shown to the students which permanently doubt about the usefulness of mathematics.

## References

- [1] Škoda J., Motyčka M.: Porovnávací měření fotometrických parametrů svítidla. In: *Sborník odborného semináře Kurz osvětlovací techniky XXXII. první*. Ostrava: VŠB - TECHNICKÁ UNIVERSITA OSTRAVA, 2016. pp. 279-288. ISBN: 978-80-248-3969- 1.
- [2] Vagaská A., Gombár M.: The Application of Design of Experiments to Analyze Operating Conditions of Technological Process. In: *Post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016. pp. 105-111. ISBN: 978-80-7231-400-3.

# SENSITIVITY ASSESSMENT AND COMPARISON OF MAXIMA METHODS IN THE ESTIMATION OF EXTREMAL INDEX

Jan Holešovský

Faculty of Civil Engineering, Brno University of Technology  
Veveří 95, 60200 Brno, Czech Republic  
holesovsky.j@fce.vutbr.cz

**Abstract:** *Extremal index is the primary measure of local dependence of extreme values and plays important role in extreme value estimation for stationary processes. The maxima estimators are often preferred in practical situations. These estimators, based on properties of the block maxima, are asymptotically characterized by the Generalized extreme value distribution. In contrast to other methods, the maxima estimators gain advantage in stability to the choice of auxiliary parameters. Still the main part of the maxima methods is selection of a proper approximation to the marginal distribution of the underlying process. Although the suitability of the approximation may significantly affect the estimation quality, to the effect of available approaches has not been paid a great interest in the literature. The aim of this contribution is the comparison of available sampling schemes and the assessment of sensitivity of existing maxima estimates of the extremal index.*

**Keywords:** extreme value, extremal index, stationary series, block maxima, resampling.

## INTRODUCTION

Characterization of rare events in natural processes is the objective in many application areas. Mostly, the practitioners restrict the inference to observations of independent and identically distributed (IID) random variables. In such cases the estimation of extreme values can be obtained through classical results of extreme value theory (see [11]). Recently the peak-over-threshold model is often preferred - a high threshold is selected and the independent threshold exceedances are modelled by the Generalized Pareto distribution. However, provided a time series is available, the requirement of independence enforces application of auxiliary techniques first in order to draw out an approximately IID series. This usually requires the use of a suitable sampling scheme leading to excessive data reduction, while the assumption of independence may still be harmed.

More efficient seems to deal with the raw time series. In order to be able to reach some specific inference, the attention is usually limited only to a stationary series satisfying the  $D(u_n)$  condition of Leadbetter et al. [11]. The  $D(u_n)$  condition restricts the long-range dependence at extremal levels, so that the distant observations can be considered approximately independent. The extremal behaviour of such series is managed by its marginal distribution and by its dependence structure capturing the tendency of extreme values to cluster. Extremal index  $\theta$ ,  $0 < \theta \leq 1$ , is thereby the primary measure of the short-range extremal dependence. Review of this area is given, for example, in [5]. The case  $\theta = 0$  is also possible, but in some sense degenerative (see [2] for details), and will not be further considered. There has been provided many interpretations to extremal index. One of the most descriptive is that  $\theta^{-1}$  represents the expected value of the cluster size distribution

[11]. Thus as  $\theta \rightarrow 0$ , the extremes tend to cluster. Clearly, for an IID series one obtains  $\theta = 1$ . The reverse implication however does not hold.

There have been proposed various approaches to the extremal index estimation. In summary the most methods deal with suitable identification of clusters. Lately, the interest is lied in estimation under the framework of the peaks-over-threshold model (see e.g. [2, 6, 8]). However it is typical that such methods are extremely sensitive to the choice of auxiliary parameters such as the threshold value or separation period. Still the most stable estimates are usually obtained by one of the maxima methods within consideration of the block model. For the maxima method, as it was introduced by Gomes [7], there is only one parameter to choose - the block size. Its sensitivity to the estimation quality has been already the object of study in [9], for example. Although it was not considered in [7], it turns out that proper estimation of the marginal distribution is also crucial. The main objective of the paper is to compare available methodologies for the estimation of the marginal distribution of the underlying stationary series. We aim to the sensitivity assessment of the maxima methods under various conditions, and study particularly the bias of extremal index estimates.

Let  $X_1, \dots, X_n$  be a stationary series satisfying the  $D(u_n)$  condition with marginal cumulative distribution function (CDF)  $F(x)$ . In the following denote  $X_1^*, \dots, X_n^*$  an IID series associated to the underlying sequence  $X_1, \dots, X_n$ , i.e. an IID series drawn from the same distribution  $F(x)$ . Let  $M_n = \max\{X_1, \dots, X_n\}$  and  $M_n^* = \max\{X_1^*, \dots, X_n^*\}$  be the sample maxima, and denote  $F_{M_n^*}(x)$  the CDF of  $M_n^*$ . From the extreme value theory follows that, if there exist normalizing constants  $a_n > 0, b_n$  such that

$$F_{M_n^*}(a_n x + b_n) = F^n(a_n x + b_n) \rightarrow G(x) \quad (1)$$

for some non-degenerate CDF  $G(x)$ , then  $G(x)$  is CDF of the Generalized extreme value (GEV) distribution. Hence, the function  $G(x)$  is of the form

$$G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \quad (2)$$

where  $a_+ := \max(a, 0)$ , and  $\mu, \sigma > 0, \xi$  are the parameters of location, scale, and shape, respectively. Occasionally we write  $\text{GEV}(\mu, \sigma, \xi)$  to emphasize the parameters of the distribution. The corresponding result for the distribution of  $M_n$  gives, under the conditions lied on (1),  $P(M_n \leq a_n x + b_n) \rightarrow G_\theta(x)$ , where  $G(x)$  and  $G_\theta(x)$  are related by the equation

$$G_\theta(x) = [G(x)]^\theta. \quad (3)$$

Thus the limiting distributions of  $M_n$  and  $M_n^*$  are both GEV with respective parameters  $(\mu_\theta, \sigma_\theta, \xi_\theta)$  and  $(\mu, \sigma, \xi)$ . From (3) it can be easily derived that the parameters, assuming  $\xi \neq 0$ , are further related by the following equalities

$$\mu_\theta = \mu - \frac{\sigma}{\xi}(1 - \theta^\xi), \quad \sigma_\theta = \sigma \theta^\xi, \quad \xi_\theta = \xi. \quad (4)$$

Note, the particular form of  $G(x)$  for  $\xi = 0$  (i.e. the Gumbel distribution) can be obtained by taking limit of (2) with  $\xi \rightarrow 0$ . In that case can be the parameters  $(\mu_\theta, \sigma_\theta)$  again rewritten in terms of  $(\mu, \sigma)$

similar to (4). Nevertheless, the case  $\xi = 0$  will not be explicitly emphasized in the paper. Hence the corresponding relations can be found in [3] for example.

## 1 MAXIMA METHODS

The main idea of the maxima estimates of  $\theta$  consists in comparison of the two triples of parameters  $(\mu, \sigma, \xi)$  and  $(\mu_\theta, \sigma_\theta, \xi_\theta)$ . Given sequences  $X_1, \dots, X_{mn}$  and  $X_1^*, \dots, X_{mn}^*$ , denote  $M_{n,i}, M_{n,i}^*$ ,  $i = 1, \dots, m$ , the corresponding block maxima of size  $n$ , i.e.  $M_{n,i} = \max\{X_{(i-1)n+1}, \dots, X_{in}\}$  and  $M_{n,i}^* = \max\{X_{(i-1)n+1}^*, \dots, X_{in}^*\}$ . For block size  $n$  large enough, the distribution of both entities can be approximated by a limiting GEV distribution, as it follows from equation (1). The  $D(u_n)$  condition limits the dependence at extreme levels, and hence for large  $n$  the block maxima  $M_{n,i}$ s are approximately independent, too. Such assumptions about block maxima of a time series are usually taken into account in practical situations (see e.g. [10, 1, 12]). Hereby standard techniques can be applied to fit a GEV distribution to  $M_{n,i}$ s. Typically the maximum likelihood (ML) method is applied.

In order to estimate the extremal index  $\theta$ , Gomes [7] in her first paper to the maxima methods proposed to fit a  $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$  distribution to the series of  $M_{n,i}$ s and a  $\text{GEV}(\mu, \sigma, \xi)$  distribution to  $M_{n,i}^*$ s. Hereby are obtained the ML estimates  $(\hat{\mu}_\theta, \hat{\sigma}_\theta, \hat{\xi}_\theta)$  and  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ . The particular parameter estimates are then combined at the basis of relations (4) to get the extremal index estimator

$$\hat{\theta}_G = \left( \frac{\hat{\sigma}}{\hat{\sigma}_\theta} \right)^{-1/\bar{\xi}}, \quad (5)$$

where  $\bar{\xi} = (\hat{\sigma} - \hat{\sigma}_\theta)/(\hat{\mu} - \hat{\mu}_\theta)$ .

Ancona-Navarrete and Tawn [2] combine the two GEV fits into one by maximizing a joint likelihood function of a sample  $(M_{n,1}, \dots, M_{n,m}, M_{n,1}^*, \dots, M_{n,m}^*)$ . According to (4) can be the parameters  $(\mu_\theta, \sigma_\theta, \xi_\theta)$  rewritten in terms of  $(\mu, \sigma, \xi, \theta)$ . The joint maximum likelihood function is hence maximized with respect to the parameters  $(\mu, \sigma, \xi, \theta)$ , yielding the extremal index estimator  $\hat{\theta}_{AT}$ . Actually, the block maxima  $M_{n,i}, M_{n,j}^*$  for  $i, j = 1, \dots, m$  are not independent. This is because of the construction of  $M_{n,i}^*$ s as described below. However, the authors of [2] argue the dependence is asymptotic insignificant.

In contrast to  $\hat{\theta}_G$ , the estimate  $\hat{\theta}_{AT}$  is obtained directly. Nevertheless, there are still present several nuisance parameters  $\mu, \sigma$ , and  $\xi$ , which are estimated with no specific purpose to be embedded into an estimator of  $\theta$ . To avoid this, Northrop [13] proposed a semiparametric maxima estimator  $\hat{\theta}_N$ . The estimator is based on factorization of a GEV likelihood function to be able to make independent inferences about  $\theta$  and  $(\mu_\theta, \sigma_\theta, \xi_\theta)$ . The maxima  $M_{n,i}$  are rewritten in terms of order statistics within  $X_1, \dots, X_{mn}$ . Northrop [13] further suggests an approximation to the part of the likelihood function associated to the rank. This is based on properties of the variable  $V_i = -n \ln F(M_{n,i})$ , whose distribution can be according to (1) subasymptotically approximated by exponential distribution; see the paper [13] for details.

Northrop [13] also suggests an extension to the framework of sliding block maxima, i.e. except the (disjoint) block maxima  $M_{n,i}$  are considered the sliding block maxima  $M_{n,j}^s = \max\{X_j, \dots, X_{j+n-1}\}$ . Similarly to the above, the inference about  $\theta$  based on the  $(mn - n + 1)$  sliding blocks is done by factorization of the GEV likelihood. Note, the disjoint block maxima form a subsequence of the sliding blocks, i.e.  $M_{n,i} = M_{(i-1)n+1}^s$  for  $i = 1, \dots, m$ . The  $M_{n,j}^s$ s should contain more information about  $\theta$  than  $M_{n,i}$ s. Hence the sliding-block estimator of  $\theta$ , say  $\hat{\theta}_N^s$ , is expected to be more efficient, particularly in terms of variability. However, besides of the dependence structure of the underlying series  $X_1, \dots, X_{mn}$ , the maxima  $M_{n,j}^s$  from nearby blocks are strongly positively associated. Hence the suitability of the approximative likelihood estimation is weakened.

The foregoing considerations relied on the assumption that an associated IID series of  $X_i^*$ s is available, or equivalently the marginal distribution  $F(x)$  of the series  $X_1, \dots, X_{mn}$  is known. Gomes [7] suggested to obtain approximation to  $X_1^*, \dots, X_{mn}^*$  by randomizing the index of the original series. This way the series of  $X_i^*$ s should follow the same marginal distribution as  $X_i$ s, nevertheless due to the independence there is no need to preserve the order of the variables. The same approach was applied in [2]. However, to the suitability of this *random resampling* has not been paid a great interest in the literature. Obviously the block size may play here an important role. For a reasonable large block size  $n$  (with respect to the series length  $mn$ ) one could consider the underlying blocks should be dispersed rather uniformly after the resampling. A pragmatic choice  $n = m$  has been early adapted by Gomes [7] and later used also in [2]. The maxima estimators of  $\theta$  show here a good balance between bias and variability (both depending on  $n, m$ ; see [9] for details). On the other hand, as it is pointed out in [14], under the random resampling one can expect strong intra-block dependence. Even for the choice  $n = m$  the dependence within block remains significant. Possibility to overcome this issue may lie in repetitive permutation, say  $K$  times, of the underlying series and taking an estimate  $\hat{\theta}$  as mean or median of individual  $\hat{\theta}_k, k = 1, \dots, K$ .

Other approach was discussed in [14]. Under *regular resampling* the IID series is obtained as  $X_{(s-1)m+r}^* := X_i$ , where  $s = (i \bmod n)$  and  $r = \lfloor i/n \rfloor$ . For  $s = 0$  we set  $X_{(n-1)m+r}^* := X_i$ . Thus, the values of the original series are placed exactly  $m$  points apart, so that the observations within an underlying block are spread uniformly in the IID series. As discussed in [14], the intra-block dependence should be minimized. However in comparison to random resampling, one would expect significant inter-block dependence between the block maxima  $M_{n,i}^*$ .

A different approach of  $F(x)$  estimation for the purpose of the semiparametric estimator  $\hat{\theta}_N$  was discussed by in [13]. Typically, there is a need for determination of the value  $F(M_{n,i})$ . To avoid the intra-block dependence, Northrop [13] suggests to construct an estimator  $\hat{F}_{-i}$  of  $F$  which is determined as empirical CDF of the  $(mn - n)$  values  $X_{j,s}$  *not present* in block corresponding to  $M_{n,i}$ . If the rank of  $M_{n,i}$  is  $R_i$  then the value  $F(M_{n,i})$  is estimated at the basis of out-of-block distribution by  $\hat{F}_{-i}(M_{n,i}) = (mn - n + 1 - R_i)/(mn - n + 1)$ . To ensure positivity of  $\hat{F}_{-i}$ , for  $x$  below the out-of-block observations  $X_{j,s}$  is the value  $\hat{F}_{-i}(x)$  set to  $1/(mn - n + m + 1)$ . Nevertheless this case is unlikely to occur unless the block size  $n$  is small. Similarly it is proceeded for the sliding block maxima  $M_{n,i}^s$ .

Some of the properties of the above discussed estimators are already known. The study [9], for example, shows that  $\hat{\theta}_G$  rather overperforms  $\hat{\theta}_{AT}$  in terms of bias with exception  $\theta$  being close to its bounds  $\theta \approx 1$  or  $\theta \approx 0$ . On the other hand,  $\hat{\theta}_{AT}$  has uniformly slightly smaller variability than  $\hat{\theta}_G$ . Nevertheless, all the maxima methods rely on proper estimation of the marginal CDF  $F(x)$  of the underlying process  $X_1, \dots, X_{mn}$  to construct an IID series  $X_1^*, \dots, X_{mn}^*$ .

## 2 SIMULATION STUDY

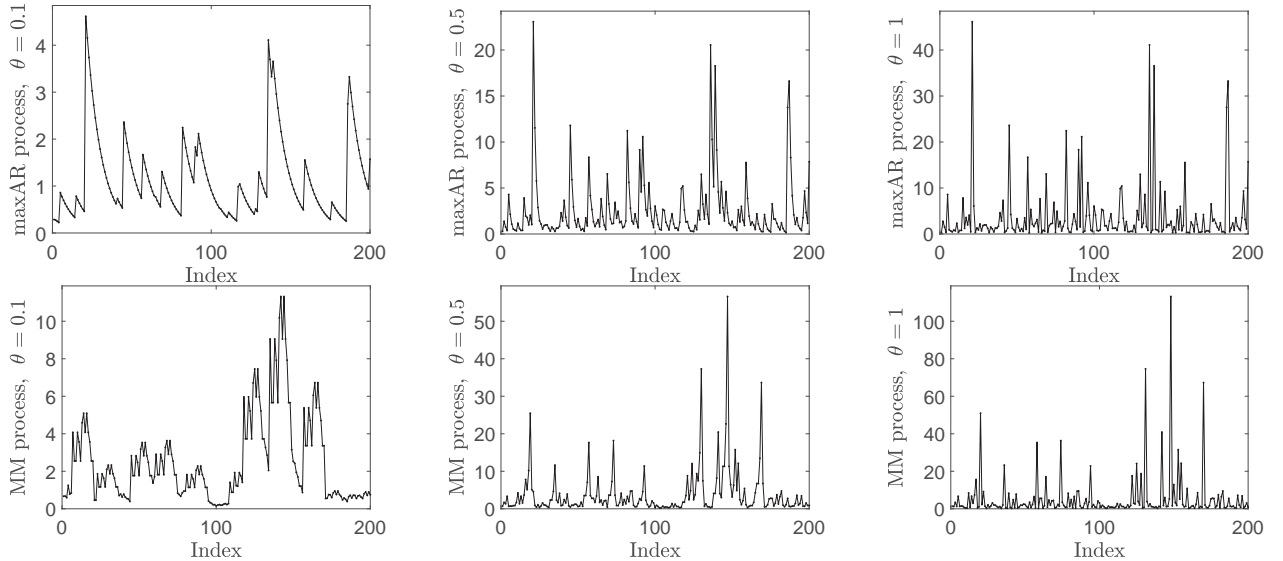
For simulation study we consider two stationary processes satisfying the  $D(u_n)$  condition to meet the extremal short-range dependence. Let  $Z_1, Z_2, \dots$  be an IID random sequence drawn from the standard Fréchet distribution, i.e. with CDF  $F_Z(z) = \exp(-1/z)$  for  $z > 0$ . First, we construct the max-autoregressive (maxAR) process  $X_1, X_2, \dots$  which is defined by

$$X_i = \max\{\beta X_{i-1}, (1 - \beta)Z_i\}, \quad i = 1, 2, \dots, \quad (6)$$

where  $0 \leq \beta < 1$ , and  $X_1 = Z_1$ . The extremal index of the maxAR process is equal to  $\theta = 1 - \beta$  (see [3]). Further, we also consider the moving maxima (MM) process

$$X_i = \max_{j=0, \dots, p} \{\alpha_j Z_{i+j}\}, \quad i = 1, 2, \dots, \quad (7)$$

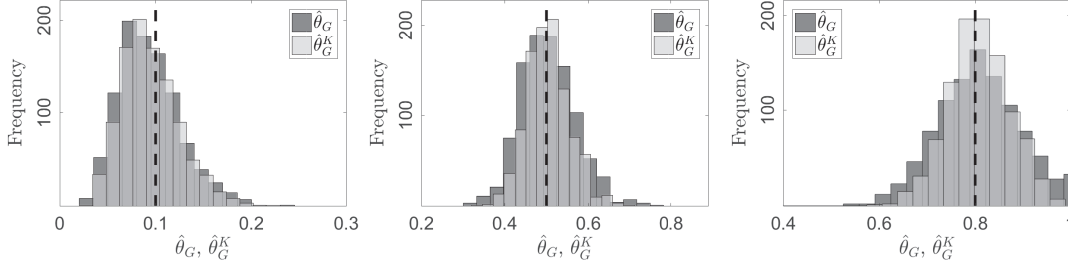
where  $\alpha_0, \alpha_1, \dots, \alpha_p$  are constants such that  $\alpha_0 > 0$ ,  $\alpha_p > 0$ , and  $\alpha_j \geq 0$  for  $j = 1, \dots, p - 1$ . Moreover, it need to be fulfilled  $\sum_{j=0}^p \alpha_j = 1$ . MM process is generalization of maxAR, and it can be shown that the extremal index is  $\theta = \max_{j=0, \dots, p} \{\alpha_j\}$  [3]. Several realizations of maxAR and MM processes for various  $\theta$  are visualized in Fig. 1.



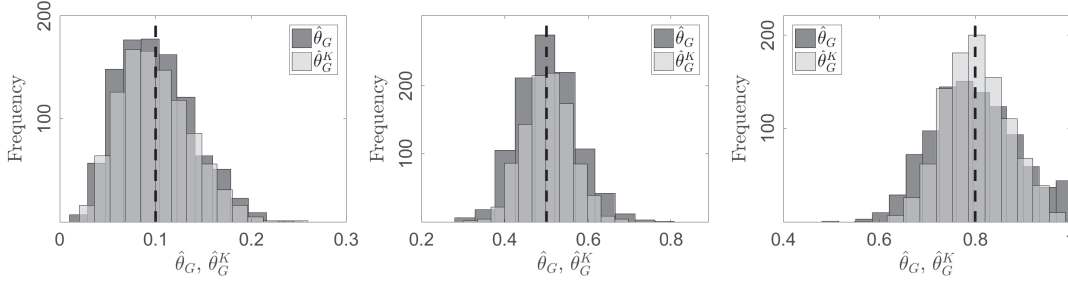
**Fig. 1.** Realization of maxAR and MM processes with extremal index  $\theta = 0.1, 0.5$  and 1 (the IID series). Small  $\theta$  leads to clustering of extreme values.

## 2.1 The effect of repetitive random sampling

For sensitivity comparison of simple and repetitive random resampling were drawn 1000 realizations of the processes above. Extremal index was estimated by the estimators  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$ , always taking either single ( $K = 1$ ) or  $K = 300$  permutations of the realization to obtain an IID series approximation. Extremal index was estimated for each permutation, and the corresponding estimate, say  $K$ -mean estimator  $\hat{\theta}_G^K$ , is observed by mean of such  $K$  estimates. The results for maxAR and MM processes with  $n = m = 100$  and the estimator  $\hat{\theta}_G$  are shown in Fig. 2 and Fig. 3, respectively.



**Fig. 2.** (maxAR process) Extremal index estimated by simple random resampling ( $\hat{\theta}_G$ ;  $K = 1$ ) and mean of  $K = 300$  repetitive random resamplings ( $\hat{\theta}_G^K$ ). True value  $\theta$  indicated by dashed line.

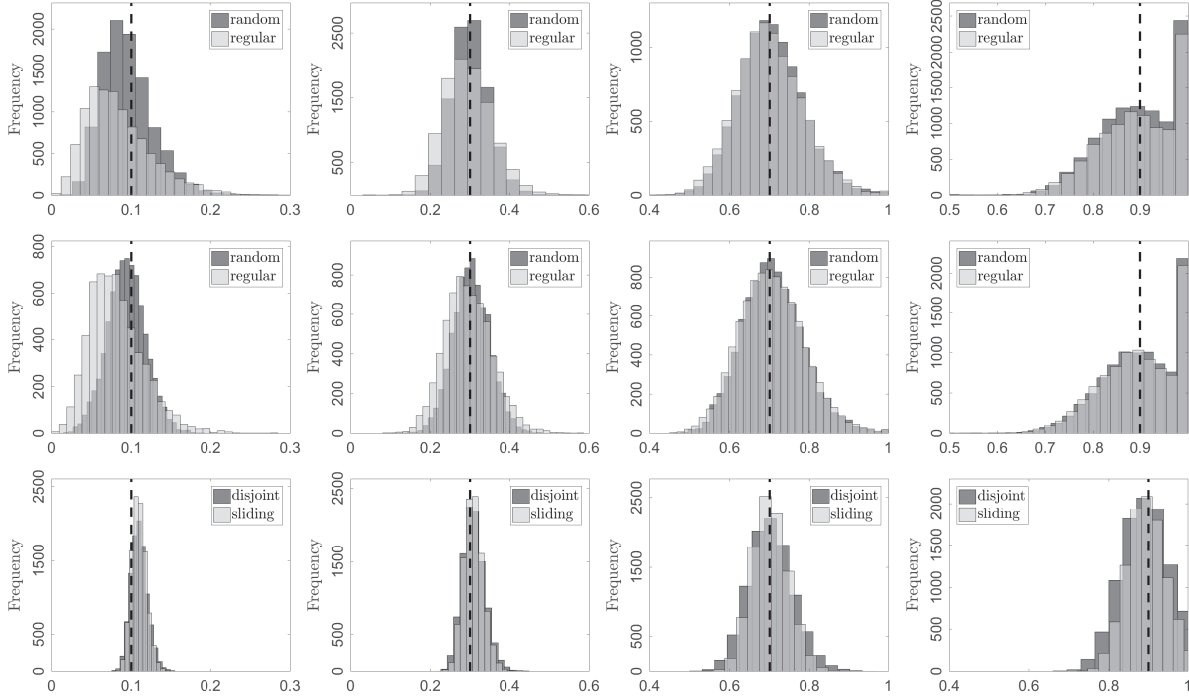


**Fig. 3.** (MM process) Extremal index estimated by simple random resampling ( $\hat{\theta}_G$ ;  $K = 1$ ) and mean of  $K = 300$  repetitive random resamplings ( $\hat{\theta}_G^K$ ). True value  $\theta$  indicated by dashed line.

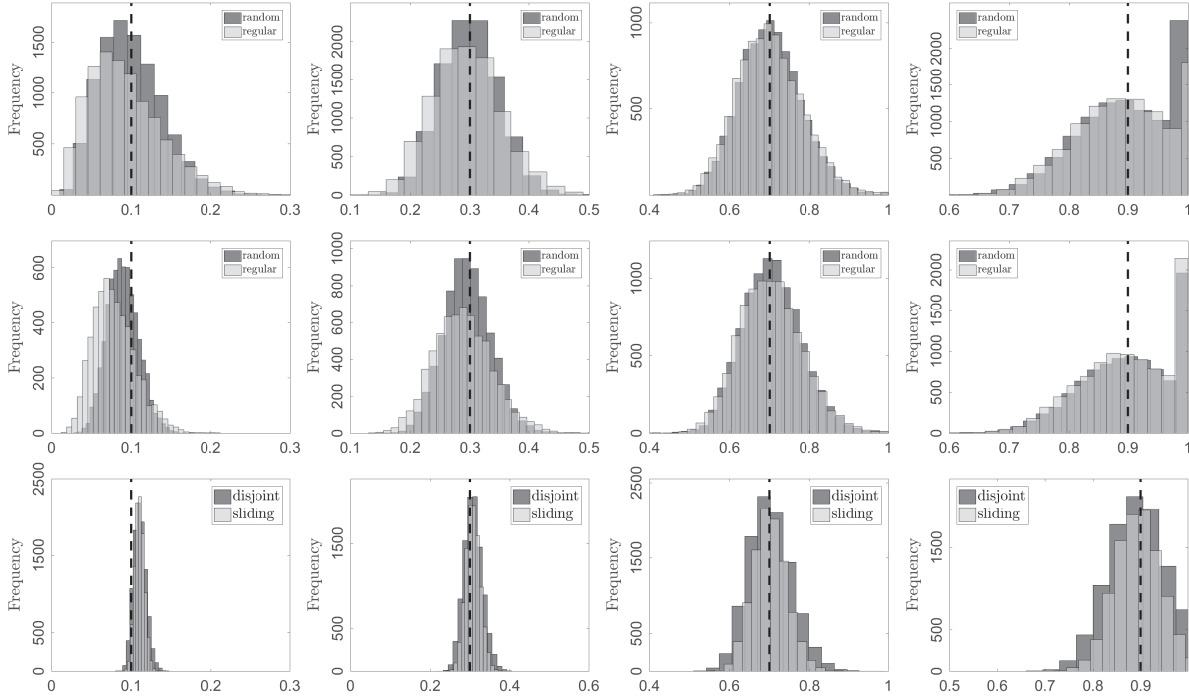
In the above plots there is evident small improvement in estimation stability of  $\theta$  under the repetitive random resampling. The estimator  $\hat{\theta}_G^K$  exhibits slightly smaller variability as the histograms are clustered closer to the true value of  $\theta$ . This is because of the nature of its construction as the mean value. However the little gain in precision is compensated by significantly higher computational demands. For this reason we will further omit the repetitive case and consider only the simple random resampling ( $K = 1$ ), which shows very comparable properties - particularly in terms of bias. Note that similar results were obtained also for the estimator  $\hat{\theta}_{AT}$ .

## 2.2 Extremal index estimation under diverse estimation of the marginal distribution

Next we assess the sensitivity of random and regular resampling applied to the estimators  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$ . We draw 10,000 independent realizations of maxAR and MM processes with again  $n = m = 100$ . Hereby we follow the selection of optimal block size determined in [9]. The specific results are visualized in Fig. 4 and Fig. 5 for various values of  $\theta$ .



**Fig. 4.** (maxAR process) Estimates of  $\theta$  obtained from 10,000 simulations of maxAR process with  $n = m = 100$ . Rows: estimators  $\hat{\theta}_G$ ,  $\hat{\theta}_{AT}$  with random and regular resampling, and  $\hat{\theta}_N$  along with  $\hat{\theta}_N^s$ . Columns:  $\theta = 0.1, 0.3, 0.7, 0.9$ . True value of  $\theta$  indicated by dashed line.

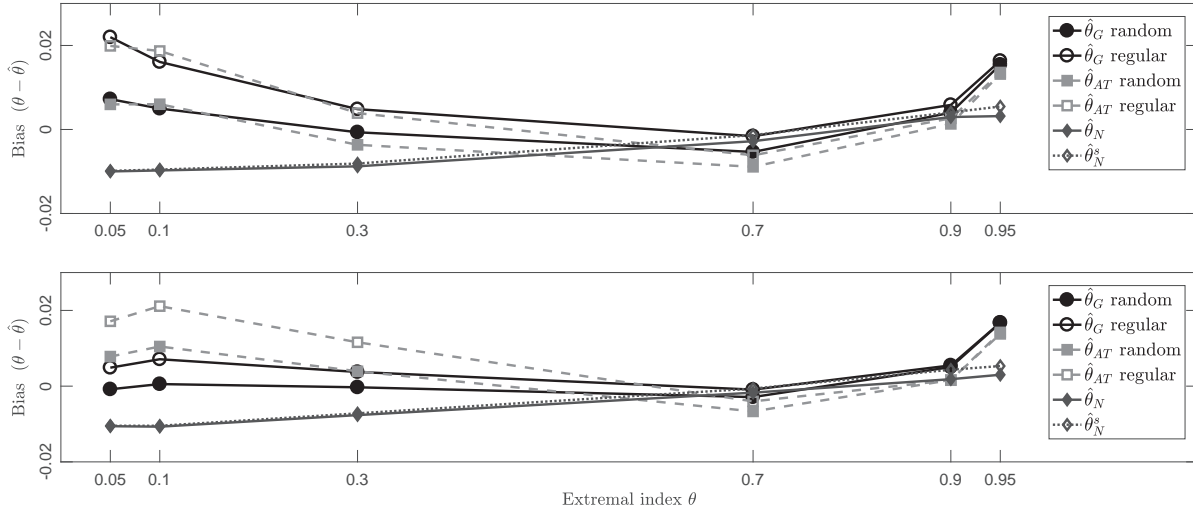


**Fig. 5.** (MM process) Estimates of  $\theta$  obtained from 10,000 simulations of MM process with  $n = m = 100$ . Rows: estimators  $\hat{\theta}_G$ ,  $\hat{\theta}_{AT}$  with random and regular resampling, and  $\hat{\theta}_N$  along with  $\hat{\theta}_N^s$ . Columns:  $\theta = 0.1, 0.3, 0.7, 0.9$ . True value of  $\theta$  indicated by dashed line.



Mostly, the estimators  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$  (the first and the second row of the plots) exhibit similar behaviour if random or regular resampling is employed. For large  $\theta$  both resampling methodologies result in very comparable estimates. On the other hand for  $\theta$  small, typically  $\theta \leq 0.3$ , the regular resampling leads to underestimation of the extremal index (see Fig. 6). Remind, the value  $\theta^{-1}$  is related to expected cluster size [11]. Hence the stationary series exhibits for small  $\theta$  extensive clusters of extremes which are followed by strong inter-block dependence after regular resampling. The lower rows in Fig. 4 and 5 show the estimates obtained by  $\hat{\theta}_N$  and  $\hat{\theta}_N^s$  with marginal CDF estimation based on out-of-block observations. Especially for small  $\theta$  this estimators exhibits much smaller variance. As it is visible in Fig. 6 for  $\theta$  small, in terms of bias are both  $\hat{\theta}_N$ ,  $\hat{\theta}_N^s$  comparable with the other estimators under the framework of regular resampling, i.e. they show rather poor performance. On the other hand,  $\hat{\theta}_N$  and  $\hat{\theta}_N^s$  exhibit only small bias if  $\theta$  is taken close to 1. This is also in agreement with [13], where the approximation of the likelihood function was constructed under the assumption of block maxima independence.

Note that by the nature of construction, both  $\hat{\theta}_G$ ,  $\hat{\theta}_{AT}$  are not constrained to be less than or equal to 1. So in practice we use  $\min\{\hat{\theta}_G, 1\}$  instead; for  $\hat{\theta}_{AT}$  is the boundary enforced by additional constraint in the ML maximization procedure. Hence, the restriction  $\theta \leq 1$  is significantly reflected in high concentration of the estimates near the upper boundary (Fig. 4 and 5 right).



**Fig. 6.** Mean bias obtained from 10,000 simulations of maxAR (upper fig.) and MM process (lower fig.) with various  $\theta$ . Extremal index estimated by  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$  with random and regular resampling, and by  $\hat{\theta}_N$  and  $\hat{\theta}_N^s$ .

### 3 CONFIDENCE INTERVALS AND THEIR COVERAGE PROBABILITIES

The properties of the estimators in practical situations usually approximated by its limiting behaviour. Especially, under the consideration of large number of blocks  $m$ , one deals with the limiting normal distribution of ML estimates. The variance of  $\hat{\theta}_{AT}$  is directly estimated from the inverse of the observed Fisher information matrix (FIM), i.e. matrix of negative second partial derivatives of the joint log-likelihood function evaluated at the obtained ML estimates. For  $\hat{\theta}_G$  the variance estimation can be obtained by delta method [4] from the observed FIMs related respectively to the

GEV distributions with parameters  $(\mu, \sigma, \xi)$  and  $(\mu_\theta, \sigma_\theta, \xi_\theta)$ . Hereby the pairs  $(\tau, \tau_\theta), \tau \in \{\mu, \sigma, \xi\}$ , are usually assumed to be independent.

Variance of  $\hat{\theta}_N$  is estimated by its “naive estimator”  $m^2 \hat{\theta}_N^2 (m-2)^{-1} (m-1)^{-2}$  emerging from the exponential approximation to the likelihood function (see [13] for wider discussion). Similarly, such naive estimator can be applied also for  $\hat{\theta}_N^s$ . Nevertheless, as already discussed in [13], this variance estimator shows rather poor performance in the latter case. The naive estimator is constructed under the assumption of block maxima independence. Thus, for  $\hat{\theta}_N^s$  that is based on the sliding blocks becomes such assumption completely unrealistic. Moreover, the dependence between the block maxima is related to  $\theta$ . Hence the use of the naive variance estimator for  $\hat{\theta}_N^s$  is totally inappropriate for practical purposes. For this reason we omit  $\hat{\theta}_N^s$  from our further considerations. Reader interested in this topic can find more information in [13] where are discussed another possibilities for the variance estimation of  $\hat{\theta}_N^s$ , e.g. block bootstrap method or the sandwich estimator.

The confidence interval of any considered estimate  $\hat{\theta}$  is determined at the basis of asymptotic normality, i.e. of the form

$$\left\langle \hat{\theta} - u_{1-\alpha/2} \cdot \widehat{\text{var}}(\hat{\theta}), \hat{\theta} + u_{1-\alpha/2} \cdot \widehat{\text{var}}(\hat{\theta}) \right\rangle,$$

where  $u_{1-\alpha/2}$  is the  $(1 - \frac{\alpha}{2})$  quantile of standardized normal distribution, and  $\widehat{\text{var}}(\hat{\theta})$  is the variance estimator of  $\hat{\theta}$ . In Table 1 are summarized coverage probabilities of 95% confidence intervals of particular estimators of  $\theta$ . Clearly, although it depends on the specific estimator, for small  $\theta$  the asymptotic confidence intervals exhibit overall poor performance. Nevertheless, random resampling overperforms the regular resampling in coverage in the majority of cases. In Fig. 4 and 5 for small  $\theta$ , there was revealed significant non-symmetry in the distribution of both  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$  in the regular case. The ML estimates show here quite slow convergence to the asymptotic normal distribution. There are thus serious doubts about the suitability of approximative normality. The estimator  $\hat{\theta}_N$  shows overall good coverage except very small values of  $\theta$ . On the other hand, large coverage proportions for large or even intermediate  $\theta$  indicate relatively high variances of the estimators that should be refined. Here lie possibilities for further research.

## CONCLUSION

Recently, there is significant interest in development of proper methods for extreme value estimation for stationary series. The extremal dependence in such sequences is hereby characterized by the extremal index  $\theta$ . Since the first paper of Gomes [7] belong the maxima methods to the most common techniques for estimation of extremal index. Besides other tuning parameters, the estimation of the marginal distribution  $F(x)$  of the series significantly affects the properties of maxima estimators. At the same time, there is lack of discussion about the suitability of various estimates to  $F(x)$ .

In this contribution were compared two resampling schemes, the regular and the random resampling, and a semiparameric methodology meant for replacement of  $F(x)$ . Despite some logical

**Table 1.** Extremal index coverage probabilities by 95 % asymptotic confidence interval for  $\hat{\theta}_G$ ,  $\hat{\theta}_{AT}$  and  $\hat{\theta}_N$ . 10,000 simulations drawn from maxAR and MM process and  $\theta$  estimated with various resampling methods.

	Estimator	Resampling	Extremal index $\theta$						
			0.05	0.1	0.3	0.5	0.7	0.9	0.95
maxAR process	$\hat{\theta}_G$	random	0.57	0.76	0.97	0.98	0.99	0.99	1.00
		regular	0.29	0.53	0.89	0.97	0.99	0.99	1.00
	$\hat{\theta}_{AT}$	random	0.60	0.62	0.71	0.83	0.92	0.96	0.96
		regular	0.19	0.31	0.60	0.79	0.91	0.96	0.97
	$\hat{\theta}_N$	disjoint	0.69	0.92	0.97	0.98	0.99	0.99	0.99
MM process	$\hat{\theta}_G$	random	0.50	0.69	0.96	0.98	0.99	0.99	1.00
		regular	0.39	0.55	0.90	0.96	0.99	0.99	1.00
	$\hat{\theta}_{AT}$	random	0.45	0.51	0.68	0.81	0.91	0.96	0.97
		regular	0.21	0.29	0.59	0.77	0.91	0.96	0.97
	$\hat{\theta}_N$	disjoint	0.67	0.96	0.99	0.99	0.99	0.99	0.99

arguments, the regular resampling for  $\hat{\theta}_G$  and  $\hat{\theta}_{AT}$  shows either worse or similar behaviour if compared to its random counterpart. Especially for small values of  $\theta$ , the regular resampling leads to introduction of extra bias. Moreover, the regular resampling results in very poor coverage probabilities for  $\theta \leq 0.3 - 0.5$  (dependent on the specific estimator). However, it must be kept in mind that the behaviour of particular estimators under various resampling approaches is strongly related to the block size selection. The bias-variance trade-off is typical for this issue. The above results were observed under the suggestions derived in [9], where the authors dealt purely with random resampling.

All over the best properties were observed in semiparametric estimation by  $\hat{\theta}_N$  and  $\hat{\theta}_N^s$  proposed by Northrop [13]. Particularly, for small  $\theta$  the use of any of those estimators leads to significant reduction of the estimation variance. Both estimator exhibit also good properties in terms of bias for large extremal index. This emerges from the nature of their construction. Nevertheless, for small  $\theta$  the bias of both  $\hat{\theta}_N$  and  $\hat{\theta}_N^s$  increases, and the estimators perform as poor as  $\hat{\theta}_G$  or  $\hat{\theta}_{AT}$  under the regular resampling. For such cases could be recommended one of the estimators above under the scheme of random resampling. On the other hand,  $\hat{\theta}_N$  shows its strength in suitable coverage of the asymptotic normal confidence intervals. This holds even for  $\theta$  relatively small (about  $\theta \geq 0.1$ ). Inappropriateness of the naive variance estimator for  $\hat{\theta}_N^s$  makes its use more difficult, and requires advanced - mostly computational demanding - techniques.

## References

- [1] Adamowski, K.: Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology*, Vol. 229, 2000, pp. 219-231.

- [2] Ancona-Navarrete, M. A., Tawn, J. A.: A comparison of methods for estimating the extremal index. *Extremes*, Vol. 3, 2000, pp. 5-38.
- [3] Beirlant, J., Geoghebeur, Y., Segers, J., Teugels, J., de Waal, D., Ferro, C.: *Statistics of Extremes: Theory and Application*. Hoboken: Wiley, 2004.
- [4] Casella, G., Berger, R. L.: *Statistical Inference*. Pacific Grove: Thomson Learning, 2002.
- [5] Chavez-Demoulin, V., Davison, A. C.: Modelling time series extremes. *REVSTAT*, Vol. 10, 2012, pp. 109-133.
- [6] Ferro, C. A. T., Segers, J.: Inference for clusters of extreme values. *Journal of Royal Statistical Society: Series B*, Vol. 65, 2003, pp. 545-556.
- [7] Gomes, M. I.: On the estimation of parameters of rare events in environmental time series. *Statistics for the Environment 2: Water Related Issues*. Chichester: Wiley, 1993, pp. 225-241.
- [8] Gomes, M. I., Hall, A., Miranda, M. C.: Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. *Computational Statistics & Data Analysis*, Vol. 52, 2008, pp. 2022-2041.
- [9] Holešovský, J., Fusek, M., Michálek, J.: Extreme value estimation for correlated observations. In: *20th International Conference on Soft Computing, MENDEL 2014*, Brno, BUT, 2014, pp. 359-364.
- [10] Khaliq, M. N., Ouarda, T. B. M. J., Ondo J.-C., Gachon, P., Bobée, B.: Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, Vol. 329, 2006, pp. 534-552.
- [11] Leadbetter, M., Lindgren, G., Rootzén, H.: *Extremes and related properties of random sequences and series*. New York: Springer, 1983.
- [12] Madsen, H., Pearson, C. P., Rosbjerg, D.: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 2. Regional modeling. *Water Resources Research*, Vol. 33, 1997, pp. 795-769.
- [13] Northrop, P.: An efficient semiparametric maxima estimator of the extremal index. *Extremes*, Vol. 18, 2015, pp. 585-603.
- [14] Northrop, P.: Semiparametric estimation of the extremal index using block maxima. *University College London*, Tech. Rep. 259, 2005.

## Acknowledgement

This paper was supported by the specific research project No. FAST-S-16-3385 at Brno University of Technology.

# RISK ASSESSMENT OF EMERGENCY OCCURRENCE AT RAILWAY CARGO TRANSPORT DUE TO HAZARDOUS SUBSTANCE LEAKAGE

**Šárka Hošková-Mayerová**

Department of Mathematics, Faculty of Military Technology, University of Defence

Kounicova 65, 662 10 Brno, Czech Republic

sarka.mayerova@unob.cz

**Abstract:** *The paper is dealing with risk assessment of emergency cases occurring at cargo transport of hazardous substances by rail. 2008-2016 data provided by the company ČD Cargo, a.s. cover incidents related with leakage: data were sorted out, analysed, processed statistically and discussed in terms of possible risk segments and sources of threats resulting from handling hazardous material. The paper finally presents trends and offer possible measures to prevent risks and reduce threats to the population and environment.*

**Keywords:** risk assessment, incidents, transport by rail, hazardous substances

## INTRODUCTION

This paper is based on the results of the PhD thesis [1] and follows up the conference paper [6] from MITAV conference 2017 focused on the risk assessment of emergency cases occurrence in freight rail transport, in particular, cases arising due to the hazardous substance leakage. The paper also shows the risk of emergency case origin and current trend when hazardous material is transported by rail. For calculations, data from 2008-2016 period were used. All graphs and figures were processed by the authors [1,7] and data from [8] were used for all calculations. Program MAPLE and EXCEL was used to calculate the risk.

Transport by rail is the most efficient type of land transport compared to other modes of transport. Its characteristics consists in ability to transport economically people, goods and bulk material over long distances. [1,10] One of the railway transport advantages is the relief of high-congested highways and roads. Thus, the transport by rail improves the traffic fluency and safety, affects the safety of people and goods against damage or loss. However, accidents, incidents and emergency cases have become undesirable and inseparable part of the transport process in railway transport. [9] Causes of their occurrence result from a number of various interrelated and combined factors.

## 1. RISK ASSESSMENT AT TRANSPORT OF HAZARDOUS SUBSTANCES

The term risk is linked to the probability or possibility of damage. Actually, it is the result of triggering a particular hazard, resulting in a certain negative result or damage. Risk is therefore a function of the probability that the frequency, intensity and duration of the activation will be sufficient to transform potential state of danger into a negative consequence (damage to health, environment or property). This term expresses the likelihood of a negative phenomenon as well as consequences of this phenomenon. The risk has always two dimensions:

- likelihood of a dangerous situation occurrence (threat),
- the severity of the possible consequence. [4,5,6,7]

The risk analysis is a process of detailed identification and analysis of risks, determining their sources and size, examining mutual interrelationship and predicting the range of negative effect on the system in case the security incident occurs and associated security situation. The analysis is a risk assessment and management; it also provides a rational basis for decision-making considering the fact that the assessment is strongly subjective where there are emphasized likelihood, number and even explicit quantification of uncertainty. The objective of the analysis is to provide sufficient ability to respond to upcoming adverse situations and restrict the impact of security incidents. [3,9,10] Risk assessment is a systematic reviewing of all aspects of the system. Its principle is to assign a numerical value or verbal evaluation to each risk identified.

For risk assessment purposes, the following groups of methods are used:

- quantitative methods using numerical risk assessment,
- qualitative methods using verbal evaluation,
- semi-quantitative methods using qualitative scale descriptions with assigned numerical values.

The risk assessment process is the first step of the health and safety management approach; if this process is not carried out properly or not at all, identifying and adopting preventive measures is unlikely. Risk assessment is a dynamic process that enables an enterprise to adopt a proactive risk management policy at the workplace. It is very important for any type and size of the enterprise to make regular evaluations. Assessment management comprises, among others, the assurance that all the relevant risk had been considered (and not only immediate or obvious), checking the effectiveness of the security measures taken, recording the evaluation results and regular reviewing accomplished. [1,4,5,10]

Railway accidents with hazardous substances presence are characterized by a variety of factors affecting the emergency case occurrence. In order to identify the risks for a particular emergency, at first, it is necessary to review input data available, possible methodologies and analysis objectives.

When characterizing a hazardous substance, it is necessary to realize that these substances become hazardous only after an emergency occurs. Some substances are considered automatically hazardous, such as chemicals, radioactive or petroleum products; others become hazardous depending on how, where and under what conditions they are transported and stored.

Hazardous substances are susceptible to explosion, fire, gas leakage or other threats, which at particular conditions or after disturbance can seriously affect the safety of people and cause material damage and damage to the environment. It is the characteristics determined by physical and chemical properties of a substance, which are inseparably related to the substance itself. In terms of the hazard (dangerous consequences), hazardous chemicals can be divided into:

- energy class, which includes explosive and flammables substances (substance turbulent reacting with water, oxidants, liquids with explosive vapours, etc.) ,

- toxic class, which are further divided into substances toxic to humans (pose a health risk) and eco-toxic substances, i.e., toxic to the environment (pose ecological risk) .

This classification of hazardous substances proves that the most significant hazardous characteristics of leaked substances present at incidents are as follows:

- explosiveness,
- flammability,
- toxicity,
- solubility,
- reactivity. [1,2,10,11]

**In terms of emergency case prediction, the most critical are phenomena as follows:**

1. Insufficient data storage.
2. Different forms of data storage, therefore further data compatibility problems.
3. Frequent changes in data categorization and registration.
4. Complications caused by secondary effects associated with the main cause of an emergency.

Table 1 shows the number of railway carriages with hazardous substances, which had been transported within the Czech Republic territory by the transport company ČD Cargo, a.s. The following 10 companies had been selected after detailed examination of the database available: Česká rafinérská, a.s., Terminal oil a.s., Metrans, a.s., DEZA, a.s., BorsodChem MCHZ, s.r.o., České dráhy, a.s., Czech Airlines Handling, a.s., ArcelorMittal Ostrava a.s., Synthesia, a.s., Lovochemie, a.s. ČD Cargo, a.s. provided a database, therefore the data could be analysed for a 3-year period (2014-2016). The paper considers only the data related to the transport within the Czech Republic territory and the assessment of transported hazardous substances covers the total of 141,229 railway carriages.

Year	2014	2015	2016	Total
Nr. of railway carriages	2049,461	2053,381	2044,684	102,526
Česká rafinérská, a.s.	12,656	13,490	3,842	29,988

Table 1 Number of transported railway carriages with hazardous substances by ten most significant companies and number of the most important manufacturers of chemicals in the Czech Republic (Česká rafinérská, a.s.)

## 2. EMERGENCY CASES DUE TO HAZARDOUS SUBSTANCE LEAKAGE

Leakages of hazardous substances represent a significant share of all emergency cases at rail transport. Despite the seemingly decreasing number of these case, every single incident has to be investigated and analysed. After examining every case in question, it is found out that the leaked substance is not classified hazardous because it is either plain water or frequently leaking operating fluids. Nevertheless, every leakage has to be investigated thoroughly.

The following tables present leakages of hazardous substances at railways, which had occurred at the operation of the company ČD Cargo, a.s. v ČR. SŽDC (Management of

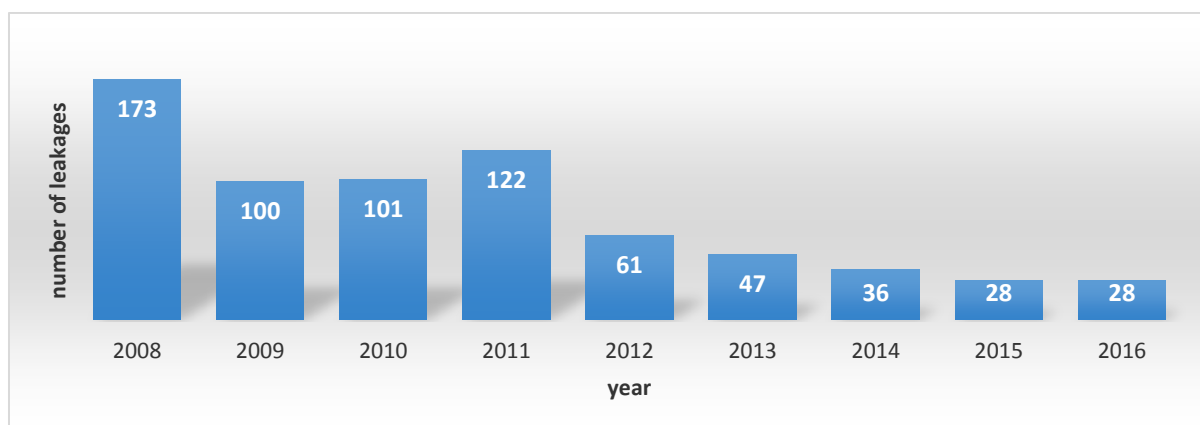
Railway Network Company) provided us the access to its database, which was examined in detail, and the following data could be processed afterwards.

In a 9-year period, 2008-2016, 597 leakages occurred on the Czech Republic railways.[8] There are recorded all emergency cases available where various quantities of hazardous substances occurred. In accordance with the Regulations for international rail transport RID, dangerous substances are classified into categories depending on their hazard class. The most hazardous substance leakage according to this categorization was in the hazard class 3 – flammable liquids, class 8 – corrosive substances, and class 2 – gasses. The table 20 presents number of leakages in a particular hazard class, which had occurred within the Czech Republic territory.

Year	Number of leakages in a particular hazard class								Total
	2	3	4.1	4.2	5.1	6.1	8	9	
2008	25	107	1	1	1	2	31	5	173
2009	9	82	0	0	5	0	3	1	100
2010	15	73	1	0	0	2	9	1	101
2011	8	94	0	0	2	1	16	1	122
2012	2	45	1	0	4	0	9	0	61
2013	4	33	0	0	0	0	9	1	47
2014	2	18	2	0	5	0	6	3	36
2015	11	12	0	0	1	0	4	0	28
2016	4	14	0	0	5	1	6	0	28
Total	80	478	5	1	23	4	93	12	696

Table 2 Number of leakages at hazardous substances transport in 2008-2016

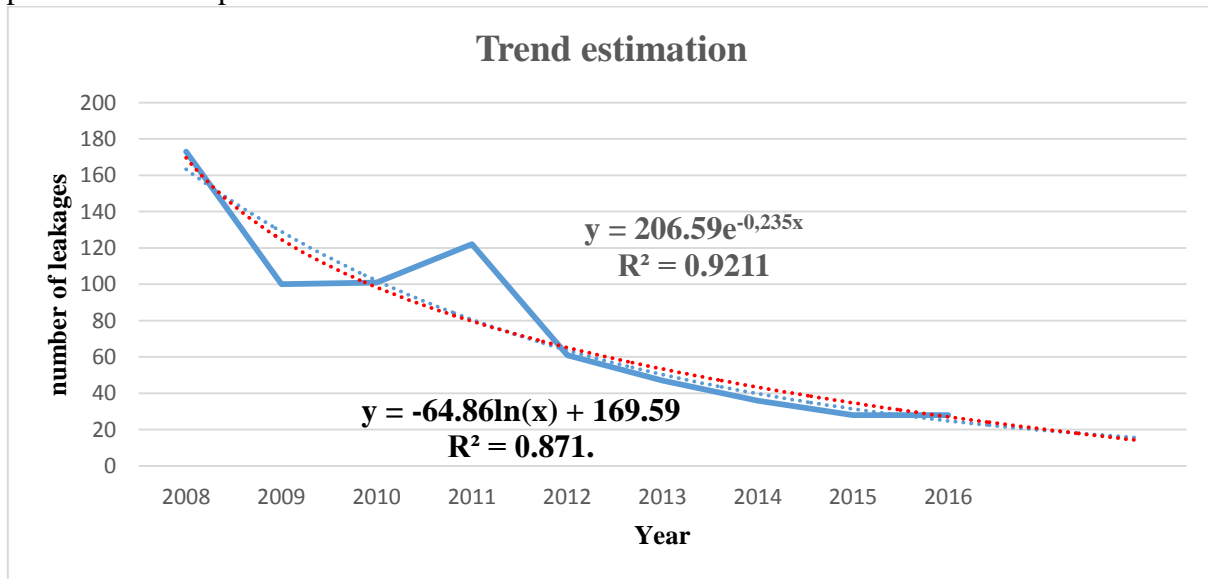
From the above presented table becomes evident that the highest number of leakages in the Czech Republic belong to class 3 (according to RID), i.e. flammable liquids. The transport was arranged by the company ČD Cargo, a.s. Graph 1 illustrates the development trend in terms of emergency cases occurrence due to hazardous substance leakage on the Czech Republic railways at ČD Cargo, a.s. operation. Leakages cover a 9-year monitored period from 2008 to 2016. [8]



Graph 1 Emergency cases due to hazardous substances leakage on the Czech Republic railways at ČD Cargo, a.s. operation



There is a significant number of leakages in 2008, however, in the following years there is a significant downward trend of such cases. The estimation of the trend for the next two years is presented in Graph 2.



Graph 2 Emergency cases due to hazardous substances leakage on the Czech Republic railways at ČD Cargo, a.s. operation

The estimation of the trend was determined by the exponential equation, which in this case has the form

$$y = 206.59e^{-0.235x}$$

The graph of this function is the red one in Graph 2. The determination coefficient in this case is high  $R^2 = 0.921$ .

The logarithmic trend has the equation

$$y = -64.86\ln(x) + 169.59$$

and is in the Graph 2 marked in blue dotted line. Its determination coefficient is lower  $R^2 = 0.8711$ . For more details see [1,7].

### 3. PREDICTION OF ACCIDENT DISTRIBUTION TREND USING MAPLE

In order to create an accident distribution trend, the least squares approximation can also be used by the first, second and third algebraic level polynomials; the LeastSquares command from the package CurveFitting. For illustration, we provide the source code. [1,7]

The source code for the above presented approximation:

```
restart;
with(CurveFitting):
with(plots):
data:=[[2008,173],[2009,100],[2010,101],[2011,122],[2012,62],[2013,47],[2014,36],
[2015,28],[2016,30]];
p:=LeastSquares(data,x);
p2:=LeastSquares(data,x,curve=a*x^2+b*x+c);
p3:=LeastSquares(data,x,curve=a*x^3+b*x^2+c*x+d);
plot([p,p2,p3,data],x=2007..2017,style=[line,line,line,point],color=[red,green,blue,
brown],symbol=solidcircle,symbolsize=15,thickness=2);
with(Student[LinearAlgebra]):
```

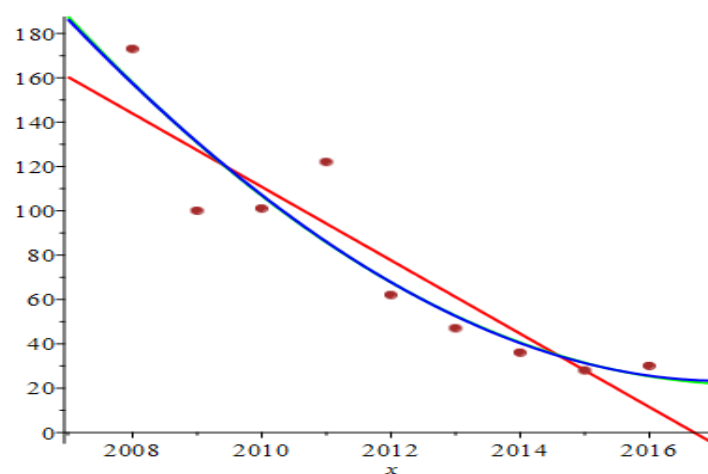
Resulting approximation by the first level polynomial:  $p = \frac{500644}{15} - \frac{331}{20}x$ .

Resulting approximation by the second level polynomial:  $p_2 = \frac{996577849}{165} - \frac{9220327}{1540}x + \frac{457}{308}x^2$ .

Resulting approximation by the third level polynomial:

$$p_3 = \frac{-14334263599}{99} + \frac{1820450945}{8316}x - \frac{305735}{2772}x^2 + \frac{1}{54}x^3$$

The graphical illustration of the least squares polynomials is in Graph 3. Polynomial  $p_1$  is red, polynomial  $p_2$  is blue, and polynomial  $p_3$  is green. The difference between blue and the green graphs is minimal, which means that the approximation by the second or third level polynomial is similar. To find the best fitting approximation the least squares errors and maximum errors of each approximation were calculated.



Graph 3 The least squares approximation

Comparing the errors that are made by approximating the function  $f$  by the least squares method by algebraic polynomial of the first, second and third level was done, again using MAPLE.

The source code for the above presented approximation:

```
infolevel[Student[LinearAlgebra]] := 1:
data1:=[[8,173],[9,100],[10,101],[11,122],[12,62],[13,47],[14,36],[15,28],[16,30]];
LeastSquaresPlot(data1,[x,y],curve=a*x^2+b*x+c,axes=boxed);
LeastSquaresPlot(data1,[x,y],curve=a*x^3+b*x^2+c*x+d,axes=boxed);
plot([p,data],x=2008..2016,style=[line,point],color=[red,green],symbol=solidcircle,symbolsize=15,thickness=2);
```

Final comparison – MAPLE program output:

Fitting curve:  $480.0 - 52.16x + 1.484x^2$

Least squares error: 51.44

Maximum error: 36.19

Fitting curve:  $450.7 - 44.38x + .8171x^2 + .1852e-1x^3$

Least squares error: 51.43

Maximum error: 35.99

When comparing approximations by polynomials of the second and third levels, from the figure and resulting values of minimal and maximum errors becomes evident that the approximation by the polynomial of the third level is more accurate in this case. Both types of error resulted smaller using the approximation by the third level polynomial comparing to the approximation by the second level polynomial. More or less, they differ slightly. See [1,7].

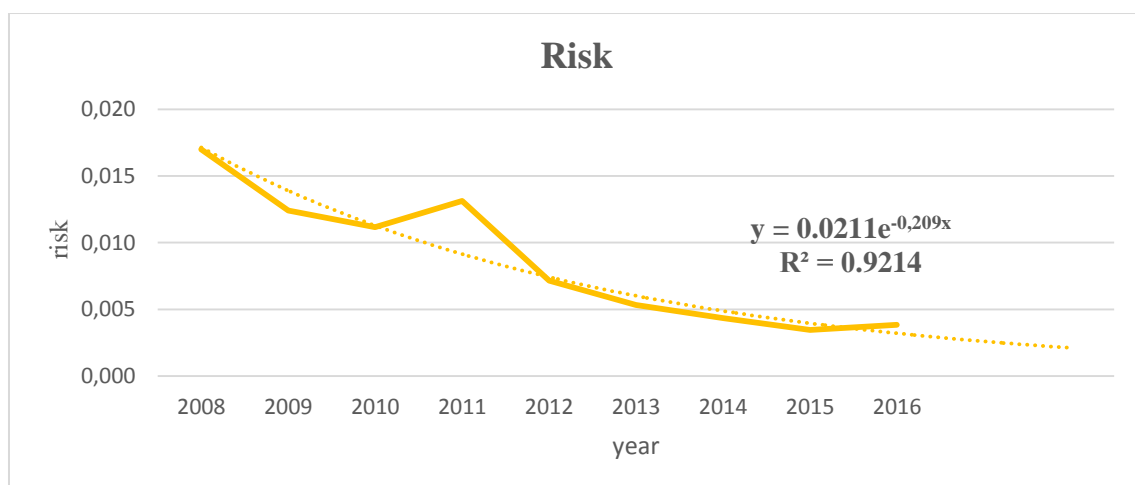
#### 4. RISK OF EMERGENCY CASE OCCURRENCE AT TRANSPORT OF HAZARDOUS MATERIAL

The following overview displayed in Table 3 presents the risk of an emergency case occurrence at transport of hazardous material. The degree of risk is determined by the relationship R, which is in fact the quotient of number of affected units and total number of units.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016
Nr. of accidents	<b>173</b>	<b>100</b>	<b>101</b>	<b>122</b>	<b>62</b>	<b>47</b>	<b>36</b>	<b>28</b>	<b>30</b>
Nr. of trains/day	848	672	754	774	721	736	689	675	650
Nr. of trains/year	10,176	8,064	9,048	9,288	8,652	8,832	8,268	8,100	7,800
Risk	0.017	0.012	0.011	0.013	0.007	0.005	0.004	0.003	0.004

Table 3 Risk of emergency case occurrence at transport of hazardous material

The risk trend was calculated by the Excel. It is determined by the exponential equation in the form:  $y = 0.0211e^{-0.209x}$  and the degree of determination is high; it is  $R^2 = 0.9214$ . Graph of the exponential function is in Graph 4 mark by the yellow dotted line. The trend line also displays graphically the prediction for the two following years. It is not appropriate to use this trend prediction for a higher number of periods because it is evident that although the trend is decreasing, the rate of decline expressed by this curve is high, and therefore unrealistic in the next time horizon. The risk is marked in yellow line. The commas in the graph of risk values means in this case the decimal dots.



Graph 4 Graphical illustration of risk and exponential risk of emergency case occurrence

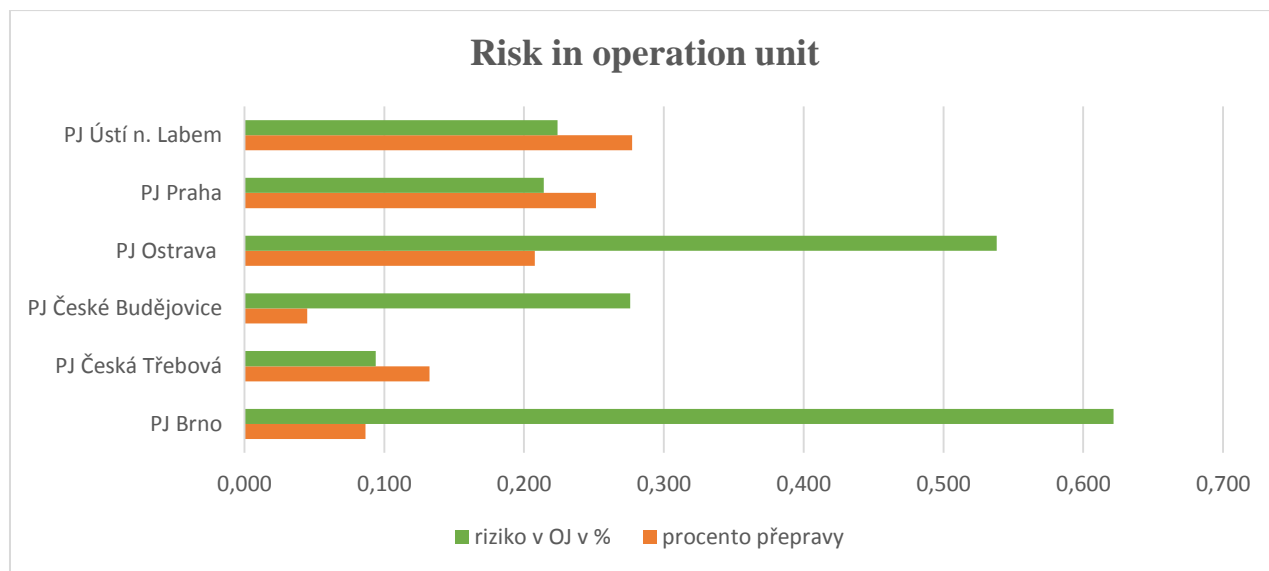
The accident distribution analysis in critical operating units in 2014-2016 presents following Table 4.

Operating unit	Number of loadings	Transport percentage	Number of trains	Number of accidents	Risk in operating unit	Risk in operating unit in %
PJ Brno	1,048	0.087	2,091	13	0.0062	0.622
PJ Česká Třebová	1,602	0.132	3,196	3	0.0009	0.094
PJ České Budějovice	545	0.045	1,087	3	0.0028	0.276
PJ Ostrava	2,515	0.208	5,018	27	0.0054	0.538
PJ Praha	3,045	0.251	6,075	13	0.0021	0.214
PJ Ústí n. Labem	3,359	0.277	6,701	15	0.0022	0.224

Table 4 Accident distribution analysis in operating units in 2014-2016

After analysing risks of emergency case occurrence becomes evident that the risk of emergency case occurrence is distributed among the operating units very unequally. The risk in Ostrava and Brno is significantly higher comparing to other operating units.

Another interesting finding based on this analysis is the fact that the operating unit Ústí nad Labem is no longer at high risk in terms of the number of emergency cases at hazardous material transport; however, it was on the top in the last years. On the contrary, it is the operating unit Brno, which did not belong among the risky ones in the previous years. It would also be very helpful to identify the cause of this change. However, there are many factors affecting this situation, and the detailed analysis would require a number of other, sometimes difficult available data. The risk in particular operation unit is illustrated at Graph 5.



Graf 5 Risk illustration by operating units

## CONCLUSION

When handling hazardous substances, whether stored or transported, there is a greater risk of an emergency and its consequences because they pose the risk of leakage to the environment and subsequent negative effects resulting from hazardous material characteristics.

The paper is based on the PhD thesis [1]; during the time of work was found out that the transport companies have poor quality of emergency cases records. Frequently, there is a lack of precise data on particular emergency cases occurred, what material was transported and what route the cargo train passed. This problem is being solved step by step, the records are edited, the emergency cases are documented with higher responsibility and the system is improving.

Within an 8-year period 2009-2016, 2,384 emergency cases occurred on the Czech Republic railways at ČD Cargo, a.s. operation. Due to different categorization in the monitored years, the most frequent cause of an emergency case occurrence could be determined only for the 2010-2014 period. The analysis showed that derailment was the most frequent accident over the reference period. In 2009-2012, the highest number of emergency cases in railway cargo occurred in the operating units Ústí nad Labem and Praha. In 2013-2016, it was again the operating unit in Praha and Ostrava; we can assume that the reason consisted in higher number of shipments.

In order to review the company professional specialization, there was made a list of companies cooperating with ČD Cargo, a.s. The company ČESKÁ RAFINÉRSKA, a.s., uses the company ČD Cargo, a.s. most frequently: in a 3-year period 2014-2016, it transported 29,988 cargo carriages. Further results of the analysis can be found in [1].

In order to increase the traffic safety, the emergency cases issues have to be always solved. The risk analysis is focused on identifying risks involved in transporting hazardous substances by rail, and further assessment so that critical locations on the railways could be specified. It is also essential to pay the attention to prevention, early warning and rapid intervention. The irreplaceable issue is also a link between information systems within the company, high-quality and trained staff, i.e., a crucial and critical entity in the entire transport process. Having followed these necessary requirements, the safety of the population can significantly be affected.

## References

- [1] Becherová, O., *Predikcia mimoriadnych udalosti na železnici*, disertačná práca, Univerzita obrany, 2017, pp.140.
- [2] Becherová, O., Hošková-Mayerová, Š. Rail infrastructure as a part of critical infrastructure. In: *Safety and Reliability - Theory and Applications - Epin & Briš (Eds) © 2017*. London: Taylor & Francis Group, 2017, pp. 1615-1619. ISBN 978-1-138-62937-0.
- [3] Bekesiene, S., Hošková-Mayerová, Š., Becherová, O., Accidents and Emergency Events in Railway Transport while Transporting Hazardous Items. In: *Proceedings of 20th International Scientific Conference. Transport Means*. Kaunas: Kaunas University of Technology, 2016, pp. 936-941. ISSN 1822-296X
- [4] Čapoun, T., *Chemické havárie*. Praha: MV – generální ředitelství Hasičského záchranného sboru ČR, 2009, s. 149, ISBN 978-80-86640-64-8.
- [5] Hasilová, K.; Vališ, D. Non-parametric estimates of the first hitting time of Li-ion battery. *Measurement*, 2018, 113, no. January 2018, p. 82-91. ISSN 0263-2241.
- [6] Hošková-Mayerová, Š., Becherová, O. Risk of probable incidents during railways transport, Uniwersytet Szczeciński, *Problemy Transportu i Logistyki*, 2017, 33, no. 1/2016, pp. 15-23. ISSN 1644-275X. (Zeszyty Naukowe)
- [7] Hošková-Mayerová, Š., Becherová, O. Risk assessment of emergency occurrence at railway cargo transport, In: *Mathematics, Information Technologies and Applied Sciences* Brno: University of Defence, 2017 ISBN 978-80-7231-400-3. MITAV 2017
- [8] Interné materiály SŽDC, (Internal material )

- [9] Rosická, Z., Beneš, L. 2007. Transport Engineering as an Important Part of the Economy. *Improvement of Quality Regarding Process and Materials*. Wydawnictwo Menedżerskie PTM, Warszawa, 2007, pp. 81-84.
- [10] Sakal, P. at al. *Envirometally oriented crisis management in strategic busines units*, Trnava, 2005, pp. 158, ISBN 80 -227-2286-3.
- [11] Vališ, D.; Hasilová, K.; Leuchter, J.. Assessment and estimation of energy power sources availability. In: *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, 2017, pp. 2054-2060. ISBN 978-1-138-02997-2.

### **Acknowledgement**

The work presented in this paper was supported by MŠMT ČR, research project no. *SV17-FVL\_K106-BEN*: Identification and security of places with high population movement.

# PROPOSAL MATHEMATICAL MODEL FOR CALCULATION OF MODAL AND SPECTRAL PROPERTIES

**Petr Hrubý, Tomáš Náhlík, Dana Smetanová**

Faculty of Technology,

Okružní 10, 370 01 České Budějovice, Czech Republic

Emails: hruby@mail.vstecb.cz, nahlik@mail.vstecb.cz,  
smetanova@mail.vstecb.cz

**Abstract:** *In the paper there is presented mathematical model of combined bending-gyratory vibration. Especially the paper is devoted a finite element for 1-dimensional linear continuum in the state of combined bending-gyratory vibration. An application of the finite element method is designed and tuned a method for calculating eigenvalues and vectors of a stepped shaft in the state of combined bending-gyratory vibration. The comparison of analytical and numerical methods is discussed.*

**Keywords:** torque and lateral vibrations, one-dimensional linear continuum, finite element method, eigenvalues and eigenvectors.

## INTRODUCTION

The shafts have a lot of interesting technical properties and they are studied intensively from different point of view. In Ben Arab et al. [1] there is study of the vibratory behaviour of rotating composite shafts and the effects of stacking sequences and shear-normal coupling on natural frequencies and critical speeds by using Equivalent Single Layer Theory.

The work [4] by Lanzutti et al. presents a failure analysis of transmission gearbox (and its components) used in motor of a food centrifugal dryer tested with a life test procedure developed by Electrolux Professional.

Sinitsin and Shestakov [6] present comprehensive analysis of the angular and linear accelerations of moving elements (shafts, gears) by wireless acceleration sensor of moving elements.

The coupling problems between shafting torsional vibration and speed control system of diesel engine is studied by Yibin et al. [8]. The torque is transmitted to relatively long distances by shafts in engines.

Leidich et al. [5] present current research results for polygonal connections with hypocycloidal profiles (H-profiles). A comparison with conventional shaft-hub connections reveals the benefits of new polygonal connections.

The shafts are constructed long and slim. They are stressed by torque and lateral vibration (= bending vibrations). It is necessary that the construction of the shafts must include the solving of torque and vibration problem.

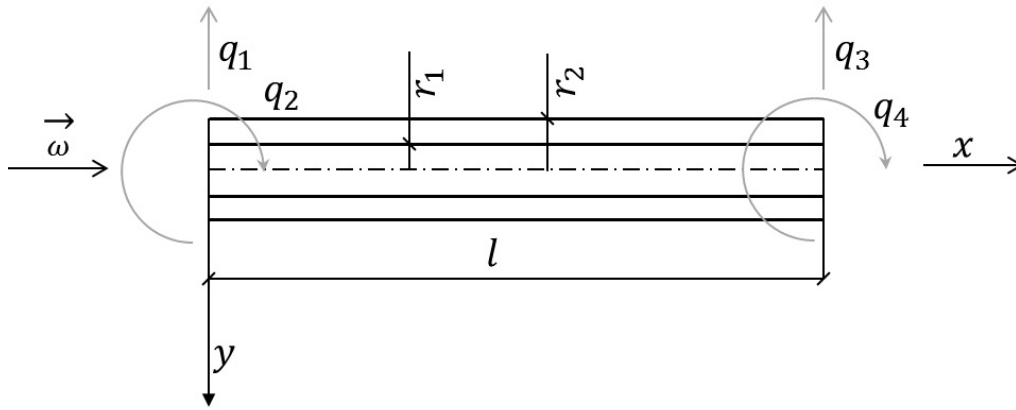
The aim of the paper is to study above properties of propeller shafts by construction on dynamical models. The first one is model of the shaft element as one dimensional linear continuum. The second model describes graduated shaft portions determined by “n” parts.

The natural frequencies are found as eigenvalues of the mathematical model of bending-gyratory vibrations of graduated shaft. They depend on angular speed rotation and the shape of graph of the dependence is circle. The different analytical and numerical methods are discussed at the end of the chapter Natural frequencies.

The presented mathematical model is useful for all mechanisms with shafts (e.g. shafts in cars, gear pumps, ...). The individual parts of the model serve to construct the entire mechanism.

## 1 MATHEMATICAL MODEL OF SHAFT ELEMENT

Consider an element of the propshaft in the shape of a prismatic section with the circular cross-section (Fig. 1).



**Fig. 1.** The element of shaft in state of combined bending-gyratory vibration  
Source: own

We denote the generalized coordinates by  $q_i$ , where  $i = 1, 2, \dots, 4$ . For the coordinates we choose immediately displacements and rotations at the edges of the cross section. The deflection  $y$  of the range  $0 \leq x \leq l$  (see Fig. 1) is expressed as

$$y(x, t) = \sum_{i=1}^4 q_i(t) \Phi_i(x) \quad (1)$$

where  $\Phi_i(x)$  are 3rd order polynomials

$$\Phi_i(x) = a_{3i}x^3 + a_{2i}x^2 + a_{1i}x + a_{0i}$$

with coefficient  $a_{3i}, a_{2i}, a_{1i}, a_{0i}$ .

The above coefficients we find from calculation of following boundary conditions:

$$\begin{aligned} \Phi_1(0) &= 1, \Phi_1(l) = 0, \Phi_1'(0) = 0, \Phi_1'(l) = 0, \Phi_2(0) = 0, \Phi_2(l) = 0, \\ \Phi_2'(0) &= 1, \Phi_2'(l) = 0, \Phi_3(0) = 0, \Phi_3(l) = 1, \Phi_3'(0) = 0, \Phi_3'(l) = 0, \\ \Phi_4(0) &= 0, \Phi_4(l) = 0, \Phi_4'(0) = 0, \Phi_4'(l) = 1. \end{aligned}$$



Hence, the polynomials  $\Phi_i(x)$  have the following forms

$$\begin{aligned}\Phi_1(x) &= 2 \left(\frac{x}{l}\right)^3 - 3 \left(\frac{x}{l}\right)^2 + 1, \quad \Phi_2(x) = \frac{x^3}{l^2} - 2 \frac{x^2}{l} + x, \\ \Phi_3(x) &= -2 \left(\frac{x}{l}\right)^3 + 3 \left(\frac{x}{l}\right)^2, \quad \Phi_4(x) = \left(\frac{x}{l}\right)^3 - \frac{x^2}{l}.\end{aligned}$$

Rewritting of the equation (1) to the matrix form we get:  $y(x, t) = [\Phi(x)] [q]$  where  $[\Phi(x)] = [\Phi_1(x), \Phi_2(x), \Phi_3(x), \Phi_4(x)]$ ,  $[q] = [q_1, q_2, q_3, q_4]^T$ .

The potential energy of an element is equal to the strain energy:

$$E_p = \frac{1}{2} E J \int_0^1 \left( \frac{\partial^2 y}{(\partial x)^2} \right)^2 dx \quad (2)$$

subtituing (1) to (2) we get

$$E_p = \frac{1}{2} E J \int_0^1 ([\Phi''(x)] [q])^2 dx, \quad (3)$$

where  $[\Phi''(x)] = [\Phi_1''(x), \Phi_2''(x), \Phi_3''(x), \Phi_4''(x)]$  and  $[q] = [q_1, q_2, q_3, q_4]^T$ .

The kinetic energy of above element can be expressed by formula

$$E_k = \frac{1}{2} \mu \int_0^1 \left( \left( \frac{\partial y}{\partial t} \right)^2 + (y\omega)^2 \right) dx + \frac{1}{2} \bar{\mu} \int_0^1 \left( \frac{\partial^2 y}{\partial t \partial x} \right)^2 dx, \quad (4)$$

where  $\mu = \rho \pi (r_2^2 - r_1^2)$ ,  $\bar{\mu} = \frac{\rho \pi}{4} (r_2^4 - r_1^4)$  and  $J = \frac{\pi}{4} (r_2^4 - r_1^4)$  (see Fig. 1).

Rewritting (4) to the matrix form with respect (1) we obtain

$$\begin{aligned}E_k &= \frac{1}{2} \mu \omega^2 \int_0^1 ([\Phi(x)] [q])^2 dx + \frac{1}{2} \mu \int_0^1 ([\Phi(x)] [\dot{q}])^2 dx \\ &+ \frac{1}{2} \bar{\mu} \int_0^1 ([\Phi'(x)] [\dot{q}])^2 dx,\end{aligned} \quad (5)$$

where  $\dot{q}$  denotes time derivation and  $\Phi'$  denotes derivation with respect to  $x$ .

Mathematical model of the element is represented by the “evolution equations”. The equation one can easily obtain from calculus of variation. They are well known as the Euler–Lagrange equations. The Lagrange function  $L$  is expressed by formula

$$L = E_k - E_p, \quad (6)$$

where  $E_k$ , resp.  $E_p$  are the forms (5), resp. (3).

The expresions (3), (5) and (6) we substitute into the Euler-Lagrange equations

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = 0, \quad (7)$$

where  $i = 1, 2, \dots, 4$ .

Hence, above Euler–Lagrange equations (7) take the form

$$\{[M_1] + [M_2]\}[\ddot{q}] - \{[K_1] - [K_2]\}[q] = 0, \quad (8)$$

where

$$[M_1] = \mu \int_0^1 [\Phi]^T [\Phi] dx = \frac{\mu l}{420} \begin{bmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{bmatrix}, \quad (9)$$

$$[M_2] = \bar{\mu} \int_0^1 [\Phi']^T [\Phi'] dx = \frac{\bar{\mu}}{30l} \begin{bmatrix} 36 & 3l & -36 & 3l \\ 3l & 4 & -3l & -l^2 \\ -36 & -3l & 36 & -3l \\ 3l & -l^2 & -3l & 4l^2 \end{bmatrix}, \quad (10)$$

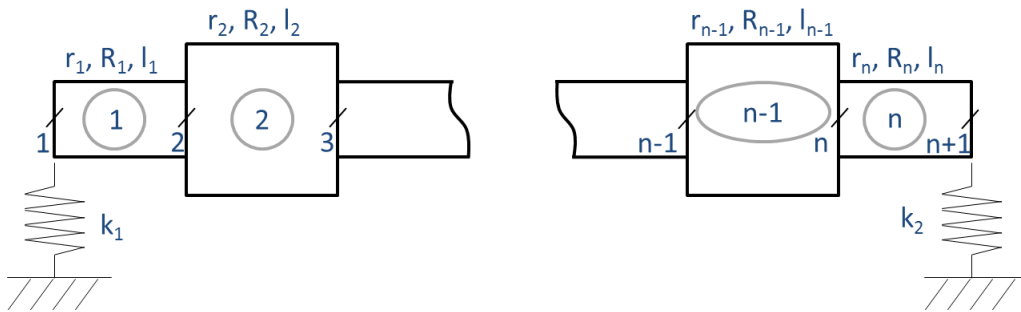
$$[K_1] = EJ \int_0^1 [\Phi'']^T [\Phi''] dx = \frac{EJ}{l^3} \begin{bmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{bmatrix}, \quad (11)$$

$$[K_1] = \mu \omega^2 \int_0^1 [\Phi]^T [\Phi] dx = \mu \omega^2 \begin{bmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{bmatrix}. \quad (12)$$

The equation (8) represents the mathematical model of the shaft element according to Fig. 1 in a state of bending-gyratory vibration.

## 2 BENDING-GYRATORY VIBRATIONS OF THE GRADUATED SHAFT

In this section we generalize previous situation to the finite element method. A dynamic model of the graduated shaft portions is determined by “n” sections of the annular cross-section of the “n” parts (outer radii  $R_i$ , inner radii  $r_i$ , lengths  $l_i$ ,  $1 \leq i \leq n$ , stiffnesses  $k_1, k_2$ ), mounted on bearings transversely deformable rigidity (left and right) and rotating angular velocity (see Fig. 2).



**Fig. 2.** Dynamic model of the graduated shaft

Source: own

If we choose generalized coordinate  $q_{2i-1}$  for lateral displacement of the  $i$ -th node and generalized coordinate  $q_{2i}$  ( $i = 1, \dots, n+1$ ) for angle of the  $i$ -th section then the kinetic and potential energy of the system have following forms

$$E_k = \frac{1}{2} \sum_{i=1}^n \dot{q}_i^T M_i \dot{q}_i, \quad E_p = \frac{1}{2} \sum_{i=1}^n q_i^T K_i q_i + \frac{1}{2} k_1 q_1^2 + \frac{1}{2} k_2 q_{2n+1}^2, \quad (13)$$

where  $M_i$  resp.  $K_i$  are mass resp. stiffness matrices of the  $i$ -th element (c.f. (9)-(12)) and we use  $r_i, R_i, l_i$  instead of  $r, R, l$ . Subvector  $q_i$  of the generalized coordinates vector  $q$  has the following form  $q_i = [q_{2i-1}, q_{2i}, q_{2i+1}, q_{2i+2}]^T$ .

We can rewrite energies (13) to the matrix expression by following way

$$E_k = \frac{1}{2} \sum_{i=1}^n \dot{q}^T M \dot{q}, \quad E_p = \frac{1}{2} \sum_{i=1}^n q^T K q, \quad (14)$$

where  $M$  and  $K$  are mass and stiffness matrices of the entire system. The total mass and stiffness matrices have blocks - “tridiagonal form” (compare with (13) and (14)). They have the following forms

$$M = \begin{bmatrix} M_{11} & M_{12} & 0 & \dots & 0 & 0 & 0 \\ M_{12}^T & M_{22} & M_{23} & 0 & \dots & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & \dots & 0 & M_{n-1,n}^T & M_{n,n} & M_{n,n+1} \\ 0 & 0 & 0 & \dots & 0 & M_{n,n+1}^T & M_{n+1,n+1} \end{bmatrix}, \quad (15)$$

$$K = \begin{bmatrix} K_{11} & K_{12} & 0 & \dots & 0 & 0 & 0 \\ K_{12}^T & K_{22} & K_{23} & 0 & \dots & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & \dots & 0 & K_{n-1,n}^T & K_{n,n} & K_{n,n+1} \\ 0 & 0 & 0 & \dots & 0 & K_{n,n+1}^T & K_{n+1,n+1} \end{bmatrix}. \quad (16)$$

The submatrices of above matrices  $M$  and  $K$  are square matrices of order 2 and they have the following form

$$K_{11} = \begin{bmatrix} k_1 + k_{11}^{(1)} & k_{12}^{(1)} \\ k_{12}^{(1)} & k_{22}^{(1)} \end{bmatrix}, \quad M_{11} = \begin{bmatrix} m_{11}^{(1)} & m_{12}^{(1)} \\ m_{12}^{(1)} & m_{22}^{(1)} \end{bmatrix},$$

$$K_{jj} = \begin{bmatrix} k_{33}^{(j-1)} + k_{11}^{(j)} & k_{34}^{(j-1)} + k_{12}^{(j)} \\ k_{34}^{(j-1)} + k_{12}^{(j)} & k_{44}^{(j-1)} + k_{22}^{(j)} \end{bmatrix}, \quad M_{jj} = \begin{bmatrix} m_{33}^{(j-1)} + m_{11}^{(j)} & m_{34}^{(j-1)} + m_{12}^{(j)} \\ m_{34}^{(j-1)} + m_{11}^{(j)} & m_{44}^{(j-1)} + m_{22}^{(j)} \end{bmatrix},$$

for  $j = 2, \dots, n$ ,

$$K_{j,j+1} = \begin{bmatrix} k_{13}^{(j)} & k_{14}^{(j)} \\ k_{23}^{(j)} & k_{24}^{(j)} \end{bmatrix}, \quad M_{j,j+1} = \begin{bmatrix} m_{13}^{(j)} & m_{14}^{(j)} \\ m_{23}^{(j)} & m_{24}^{(j)} \end{bmatrix},$$

for  $j = 1, \dots, n$ ,

$$K_{n+1,n+1} = \begin{bmatrix} k_{33}^{(n)} + k_2 & k_{34}^{(n)} \\ k_{34}^{(n)} & k_{44}^{(n)} \end{bmatrix}, \quad M_{n+1,n+1} = \begin{bmatrix} m_{33}^{(n)} & m_{34}^{(n)} \\ m_{34}^{(n)} & m_{44}^{(n)} \end{bmatrix}.$$

The mass and the stiffness matrices of the single elements ((9)-(12)) are expressed by  $M_p = [m_{ij}^{(p)}]_{i,j=1}^4$ ,  $K_p = [k_{ij}^{(p)}]_{i,j=1}^4$ , where  $p = 1, \dots, n$ .

Mathematical model of bending-gyratory vibrations of the graduated shaft has the following form

$$M\ddot{q} + Kq = \mathbf{0}. \quad (17)$$

### 3 NATURAL FREQUENCIES

Natural frequencies  $\Omega_i$  of the above system (17) satisfy the equation of frequencies

$$\det(-M\Omega_i^2 + K) = 0. \quad (18)$$

The eigenvectors (belonging to the  $i$ -th natural frequency) satisfy the relation

$$(-M\Omega_i^2 + K) \mathbf{v}_i = \mathbf{0}. \quad (19)$$

Because this relation is indefinite, for the uniqueness we normalize above vectors using the so-called M-norm, i.e. by the relation

$$\mathbf{v}_i^T M \mathbf{v}_j = \delta_{ij}, \quad (20)$$

where  $\delta_{ij}$  are the Kronecker symbols. The natural frequencies and eigenvectors are found in the base of the coordinates  $y$  which are related with the original coordinates  $q$  through transformation

$$y = B^T q, \quad (21)$$

where  $B$  is lower triangular matrix which satisfies  $M = BB^T$ . Such matrix exists due to the regularity and the positive definiteness of the mass matrix  $M$ . In the coordinates  $y$  the mathematical model (17) has the form

$$\ddot{y} + B^{-1}K(B^T)^{-1}y = \mathbf{0}. \quad (22)$$

Because a symmetric matrix  $A = B^{-1}K(B^T)^{-1}$  is similar matrix to the matrix  $M^{-1}K = (B^T)^{-1}B^{-1}K$  (via the matrix  $(B^T)^{-1}$ ). The eigenvalues of the original model (18) are the same as eigenvalues of model (22). Hence, the generalized eigenvalue problem (18) is transferred to the eigenvalue problem of the matrix  $A$

$$\det(-E\Omega_i^2 + A) = 0, \quad (23)$$

where  $E$  denotes the identity matrix.

The eigenvectors  $\mathbf{v}_i$  of (18) and the eigenvector  $\mathbf{u}_i$  of (23) satisfy relation  $\mathbf{u}_i = B^T \mathbf{v}_i$  (resp.  $\mathbf{v}_i = (B^T)^{-1} \mathbf{u}_i$ ) analogous to (21). Note that the  $\mathbf{u}_i$  are solution of equations  $(-E\Omega_i^2 + A) \mathbf{u}_i = \mathbf{0}$  (see (19)) and we normalize  $\mathbf{u}_i$  by Euklidean norm  $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$ . The utilization of Euklidean norm follows from substitution  $M = BB^T$  to (20).

The eigenvalue problem of matrix  $A$  is solved by standard procedure by utilization of linear algebra tools (see [7]). At solutions of the problems we calculate the eigenvectors  $\mathbf{u}_i$  and the eigenvalues  $\Omega_i$  (i.e., natural frequencies).

The natural frequency depends on the angular velocity of shaft rotation. The frequency decreasing with increasing speed. In the case of  $k_1 = k_2 = 0$  the system is isolated system and the first natural frequency vanishes. If the first natural frequency goes to zero and the system becomes unstable, an evaluation of the situation makes sense.

### Discrete physical model. [3]

Using physical discretization methods we obtain that natural frequency satisfies  $\ddot{y} + \Omega^2 y = 0$  (equation of relative vibrations in rotating plane) and

$$\Omega = \sqrt{\frac{k}{m} - \omega^2}, \quad (24)$$

where  $\omega$  is angular speed,  $k$  stiffness and  $m$  mass. For more details we recommend (see [3]).

If we rewrite (24) to the form

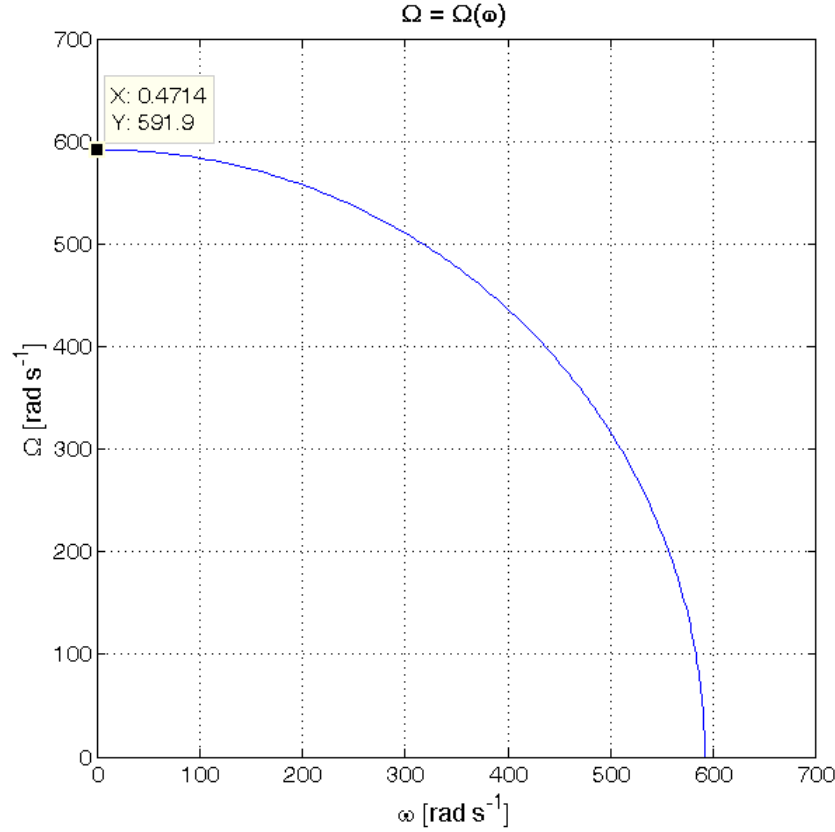
$$\Omega^2 + \omega^2 = \frac{k}{m} \quad (25)$$

we obtain equation of a circle with centre in origin of coordinate system  $O(\Omega, \omega)$  with radius  $\sqrt{\frac{k}{m}}$ .

### Example. Propeller shaft of prototype Škoda 781

Parameters of the car are  $r = 0.0105 [m]$ ,  $l = 0.65 [m]$ ,  $E = 2.1 \cdot 10^{11} [Pa]$ ,  $\rho = 7.8 \cdot 10^3 [kg \cdot m^{-3}]$  (c.f. [2]).

In program Matlab all physical quantities were calculated and Fig. 3 was created. Fig. 3 describes how the natural frequency of relative vibrations depends on the angular speed of rotation in discrete physical model.



**Fig. 3.** The dependence of natural frequency upon angular speed- discrete physical mode  
Source: own

#### Analytic model. [3]

The equation of motion of the vibrating 1-dimensional linear continuum in the rotating plane is given by formula

$$\frac{\partial^4 y}{\partial x^4} - \frac{\rho S r^4}{4 E J} \cdot \frac{\partial^4 y}{\partial x^2 \partial t^2} - \frac{\rho S r^4 \omega^2}{4 E J} \cdot \frac{\partial^2 y}{\partial x^2} + \frac{\rho S}{E J} \cdot \frac{\partial^2 y}{\partial t^2} - \frac{\rho S \omega^2}{E J} \cdot y = 0.$$

In procedure of solving above equation we obtain natural frequency formula

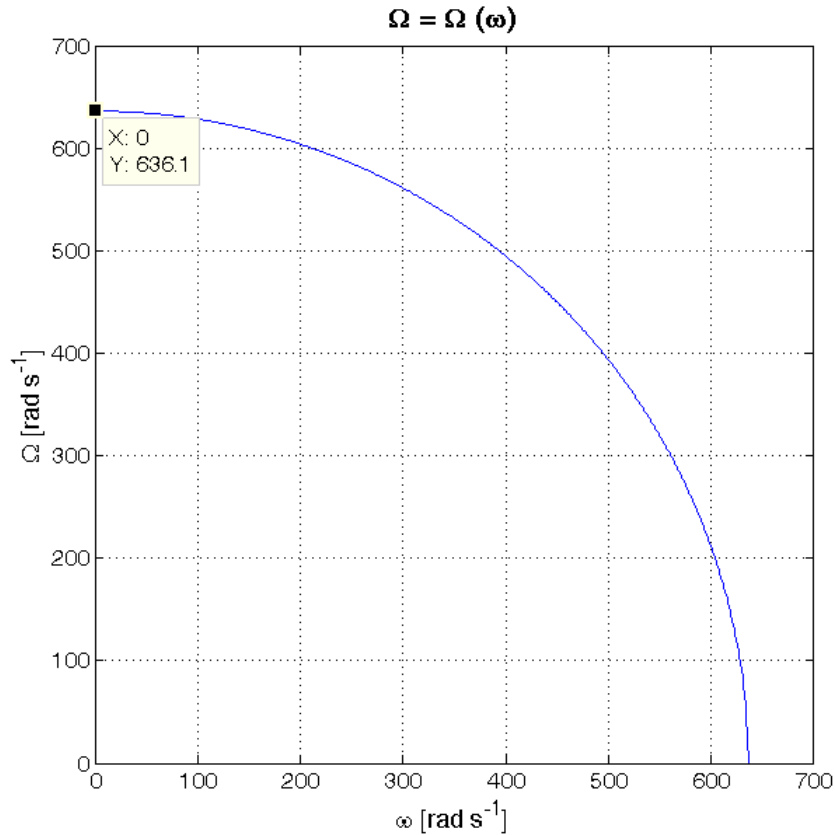
$$\Omega_n = \left( \frac{E J}{\rho S} \right)^{\frac{1}{2}} \left[ \frac{\left( \frac{\pi}{l} \right)^4 - \frac{\rho S \omega^2}{E J} \left[ 1 - \left( \frac{\pi n r}{2 l} \right)^2 \right]}{1 + \left( \frac{\pi n r}{2 l} \right)^2} \right]^{\frac{1}{2}}. \quad (26)$$

We can easily see for first natural frequency and  $l \gg r$  (long and slim shafts) we have

$$\Omega = \left( \frac{E J}{\rho S} \right)^{\frac{1}{2}} \left[ \left( \frac{\pi}{l} \right)^4 - \frac{\rho S \omega^2}{E J} \right]^{\frac{1}{2}}. \quad (27)$$

In [3] there is obtained that dependence frequency (27) and angular speed has circle shape.

We apply same parameters (see example - propeller shaft of prototype Škoda 781) for calculation (27) to Matlab programme (Fig. 4).



noindent

**Fig. 4.** The dependence of natural frequency upon angular speed - analytic model

Source: own

**Remark.** Also we obtain (26) from solution by the method of the transfer matrices.

It is very interesting that a significant shift of the natural frequency (the lowest rate) occurs a transition from  $n = 1$  to  $n = 2$ . Further increases in the number of elements doesn't bring changes. Graph for each of the elements with considerable precision approaching a circle centered at the origin of  $O(\Omega, \omega)$  with a radius  $\Omega_1$ .

## CONCLUSION

The propeller shaft represents a dynamic evolutive system. The natural frequency of oscillations depends on the angular speed of the rotation. For calculations and creation graphs the utilization of programs (e.g. Matlab) is very useful. Standardly the one type of vibrations (torque or lateral) is studied. The paper is devoted the model with combined types of vibrations.

## References

- [1] Ben Arab, S., Dias Rodrigues, J., Bouaziz, S., Haddar, M. A finite element based on Equivalent Single Layer Theory for rotating composite shafts dynamic analysis. *Composite structures*, No.178, 2017, p. 135-144.

- [2] Hrubý, P., Hlaváč, Z. *Aplikace MKP při řešení spektrálních a modálních vlastností* Výzkumná zpráva č. 102 07 91 (Research Report), ZČU Plzeň, 1991, p. 36
- [3] Hrubý, P., Hlaváč, Z., Žídková, P. Physical and mathematical models of shafts in drives with Hook's joints. In Michael McGreevy, Robert Rita. *Proceedings of the 5th biannual CER Comparative European Research Conference: International scientific conference for Ph.D. students of EU countries*. London, Sciemcee Publishing, 2016. p. 136-140 ISBN 978-0-9928772-9-3
- [4] Lanzutti, A., Gagliardi, A., Raffaelli, A., Simonato, M., Furlanetto, R., Mgnan, M., Andreatta, F., Fedrizzi, L. Failure analysis of gear, shafts and keys of centrifugal washers failed during life test *Engineering Failure Analysis*, No. 79, 2017, p. 634-641.
- [5] Leidich, E., Reiß, F., Schreiter, R. Investigations of hypocycloidal shaft and hub connections. *Materialwissenschaft und Werkstofftechnik*, Vol. 48, No. 8, 2017, p. 760-766.
- [6] Sinitsin, V. V., Shestakov, A. L. Wireless acceleration sensor of moving elements for condition monitoring of mechanism. *Measurement Science and Technology*, No. 28, 2017, p. 1-8.
- [7] Wilkinson, J.H., Reinsch, C.. *Handbook for Automatic Computation: Linear Algebra* New York, Springer Verlag, 2014 (reprint)
- [8] Yibin, G., Wanyou, L., Shuwen, Y., Xiao, H., Yunbo, Y., Zhipeng, W., Xiuzhen, M. Diesel engine torsional vibration control coupling with speed control system. *Mechanical Systems and Signal Processing*, No. 94, 2017, p. 1-13.

## Acknowledgement

The work presented in this paper was supported by project TA 04010579 of Technology Agency of the Czech Republic.



# The intransitive Lie group actions with variable structure constants

Veronika Chrastinová

Brno University of Technology, Faculty of Civil Engineering,  
Institute of Mathematics and Descriptive Geometry,  
602 00 Brno, Veveří 331/95, Czech Republic  
mailto:chrastinova.v@fce.vutbr.cz

**Abstract:** *In traditional Lie theory of transformation groups, the infinitesimal transformations constitute a classical Lie algebra with certain structure constants. However in the case of quite general intransitive transformation groups, with the presence of invariants, the structure of transformations may in fact depend on the invariants and need not be constant since the Lie groups may turn into the pseudogroups. The article starts with simple examples and finish with corresponding adaptation of the Lie fundamental theorems for the pseudogroups which is new.*

**Keywords:** Lie transformation group, infinitesimal transformation, Lie bracket, structure constants.

## INTRODUCTION

A somewhat provocative title is intended as an invitation for the nonspecialists. While there are excellent textbooks on the Lie group theory, the general pseudogroups are unknown. We start with the simplest possible nonabelian Lie group. A slight change of the notation paradoxically gives "variable" structure constants. The contradiction is subsequently clarified: the Lie groups turn into the pseudogroups. This provides the occasion to discuss the Lie fundamental theorems and to raise the main open problem: what is the true interrelation between the Sophus Lie and the Élie Cartan theories of continuous groups.

Let  $G$  be a (local) Lie group. In terms of coordinates we denote

$$a \sim (a^1, \dots, a^n), \quad b \sim (b^1, \dots, b^n), \quad \dots \in G \quad (a^i, b^i \in \mathbb{R}) \quad (1)$$

and there is multiplication

$$a \circ b = c \sim (c^1, \dots, c^n), \quad c^i = g^i(a, b); \quad i = 1, \dots, n \quad (2)$$

satisfying the well-known axioms. We recall the unit element

$$e \sim (e^1, \dots, e^n) \in G, \quad a \circ e = e \circ a = a. \quad (3)$$

In practical applications, every Lie group  $G$  is moreover represented by transformations on certain space  $M$ . Then a point  $[x] \in M$  is transformed into  $[x] \longrightarrow [y] = a \bullet [x] \in M$  where the rules

$$(a \circ b) \bullet [x] = (a \bullet (b \bullet [x])), \quad e \bullet [x] = [x]$$

hold true.

## 1 EXAMPLES

**Example 1** Let  $\mathbf{M} = \mathbb{R}$  and  $\mathbf{G}$  be the Lie group of all invertible linear transformations

$$[x] \sim x \in \mathbb{R} \longrightarrow [y] = a \bullet [x] \sim y = a^1 x + a^2 \in \mathbb{R} \quad (a \sim (a^1, a^2) \in \mathbf{G}).$$

One can find the composition rule, the unit

$$a \circ b = c \sim (c^1, c^2) = (a^1 b^1, a^1 b^2 + a^2), \quad e \sim (1, 0)$$

and the infinitesimal transformations

$$Z_1 = x \frac{\partial}{\partial x}, \quad Z_2 = \frac{\partial}{\partial x}, \quad [Z_1, Z_2] = -Z_2 \quad (4)$$

appearing by differentiation with respect to  $a^1$  and  $a^2$ , with the structure constant  $c_2^{12} = -1$ . See also the general formula (9) below.

**Example 2** Let  $\mathbf{M}$  and  $\mathbf{G}$  be as above but we change the coordinates on  $\mathbf{G}$  as follows

$$\bar{a}^1 = \frac{a^1}{k}, \quad \bar{a}^2 = a^2 \quad (k \neq 0)$$

with a certain constant  $k$ . The same transformation as above reads

$$x \longrightarrow \bar{a}^1 k x + \bar{a}^2, \quad \bar{c}^1 = \frac{c^1}{k} = \frac{a^1 b^1}{k} = k \bar{a}^1 \bar{b}^1, \quad \bar{c}^2 = c^2 = a^1 b^1 + b^2 = k \bar{a}^1 \bar{b}^1 + \bar{b}^2$$

in terms of new coordinates and we have infinitesimal transformations

$$\bar{Z}_1 = k x \frac{\partial}{\partial x}, \quad \bar{Z}_2 = \frac{\partial}{\partial x}, \quad [\bar{Z}_1, \bar{Z}_2] = -k \bar{Z}_2 \quad (5)$$

with the structure constant  $c_2^{12} = -k$ .

**Example 3** Let  $\mathbf{M} = \mathbb{R}^2$  and  $\mathbf{G}$  be as above. We introduce the transformations

$$[x] \sim (x^1, x^2) \in \mathbb{R}^2 \longrightarrow [y] = a \bullet [x] \sim (a^1 x^2 x^1 + a^2, x^2) \in \mathbb{R}^2. \quad (6)$$

This is the intransitive action, the coordinate  $x^2$  is invariant. With  $x^2 = k$  kept fixed, there are the same formulae as in the Example 2 (the bars are formally omitted here). It follows that

$$Z_1 = x^2 x^1 \frac{\partial}{\partial x^1} + 0 \cdot \frac{\partial}{\partial x^2}, \quad Z_2 = \frac{\partial}{\partial x^1} + 0 \cdot \frac{\partial}{\partial x^2} \quad (7)$$

by using (5) whence

$$[Z_1, Z_2] = -x^2 Z_2$$

with *variable* “structure constant”  $c_2^{12} = -x^2$  which is a *function* on the space  $\mathbf{M} = \mathbb{R}^2$ . This is in seeming contradiction with the Second Fundamental Theorem of the Lie theory stated below.

The true substance of this fact is of deep nature and can only be informally explained here: though the transformations (6) belong to the group  $\mathbf{G}$  separately on every leaf  $x^2 = \text{const.}$ , this is not the case on the total space  $\mathbb{R}^2$ . Indeed, the composition

$$a \circ b \bullet [x] \sim (x^2 a^1 b^1 + b^2, x^2 a^1 b^1 + b^2) \in \mathbb{R}^2$$

is *not* of the form  $c \bullet [x]$  for any  $c \in \mathbf{G}$ . In order to preserve the composition property  $a \circ b = c$ , the primary group  $\mathbf{G}$  should be included into the large pseudogroup  $\mathcal{G}$  of all transformations

$$(x^1, x^2) \in \mathbb{R}^2 \longrightarrow (f(x^2)x^1 + g(x^2), x^2) \in \mathbb{R}^2 \quad (f(x^2) \neq 0)$$

when regarded on the total space  $\mathbb{R}^2$ , see below.

## 2 FUNDAMENTAL THEOREMS

At this place, some elements of the Lie theory are worth mentioning in order to clarify our observations.

We recall the group  $\mathbf{G}$  with the notation (1), (2) and (3). Let moreover  $\mathbf{M}$  be a manifold of points

$$[x] \sim (x^1, \dots, x^m), \quad [y] \sim (y^1, \dots, y^m), \quad \dots \in \mathbf{M} \quad (x^j, y^j \in \mathbb{R})$$

in terms of coordinates. In classical theory, the (local) action of  $\mathbf{G}$  on  $\mathbf{M}$  is described by certain formulae

$$[x] \longrightarrow [x] = a \bullet [x] = [y] \sim (y^1, \dots, y^m), \quad y^j = f^j(a, [x]). \quad (8)$$

We also recall the *infinitesimal transformations*

$$Z_i = \sum z_i^j \frac{\partial}{\partial x^j}, \quad z_i^j = \frac{\partial f^j}{\partial a^i}(e, [x]); \quad j = 1, \dots, m \quad (9)$$

of the group action. These vector fields are of the special kind completely described in the famous

**Theorem 1 (Second Fundamental Theorem)** *Linearly independent over  $\mathbb{R}$  vector fields*

$$Z_i = \sum z_i^j [x] \frac{\partial}{\partial x^j} \quad (i = 1, \dots, n) \quad (10)$$

*are infinitesimal transformations (9) of action of a certain Lie group  $\mathbf{G}$  if and only if their Lie brackets  $[Z_i, Z_{i'}] = Z_i \cdot Z_{i'} - Z_{i'} \cdot Z_i$  satisfy certain identities*

$$[Z_i, Z_{i'}] = \sum c_{i''}^{ii'} Z_{i''} \quad (i, i', i'' = 1, \dots, n)$$

where  $c_{i''}^{ii'} \in \mathbb{R}$  are constants.

On this occasion, let us moreover mention

**Theorem 2 (First Fundamental Theorem)** *Functions  $f^j$  in the above transformation formula (8) satisfy the Lie system*

$$\frac{\partial f^j(a, [x])}{\partial a^i} = \sum A_i^{i'}(a) z_{i'}^j(f^1(a, [x]), \dots, f^m(a, [x])) \quad (11)$$

with appropriate (fixed) functions  $A_i^{i'}$ .

The actual literature on the Lie transformation groups systematically rests on the mechanisms of the infinitesimal transformations and the above stated Fundamental Theorems, cf. [2], [3] and large literature therein. On the contrary the alternative É. Cartan's approach [4], [5] is expressed in terms of invariant differential forms and invariant functions and involves the pseudogroups as well, however, then certain results and concepts look somewhat intricate if compared with the more elementary Lie theory [1]. The monumental task therefore appears.

**Open problem.** *To include the appropriately generalized mechanisms of infinitesimal transformations into the É. Cartan's pseudogroup theory.*

The problem was also briefly raised in lecture [6]. In this article, we will discuss only a very particular subcase related to the above Examples.

### 3 THE INTRANSITIVE ACTION

While the action (8) of a Lie group  $\mathbf{G}$  on the manifold  $\mathbf{M}$  depends on a finite number of parameters, namely the coordinates  $a^1, \dots, a^n$ , we shall introduce the action of a pseudogroup  $\mathcal{G}$  depending on arbitrary functions  $a^1(t), \dots, a^n(t)$  of one independent variable  $t$ . This is a very special pseudogroup which may be informally regarded for "a group  $\mathbf{G}$  depending on parameter  $t$ ." Roughly saying, we return to Example 3 which will be discussed in full generality.

Let us complete, in a way, the collection of our concepts. We introduce the manifold  $\mathbf{N}$  of points

$$[x, k] \sim (x^1, \dots, x^m, k), [y, k] \sim (y^1, \dots, y^m, k), \dots \in \mathbf{N} \quad (x^j, y^j, k \in \mathbb{R})$$

and moreover the space  $\mathcal{G}$  of  $n$ -tuples of smooth functions

$$a(t) \sim (a^1(t), \dots, a^n(t)), b(t) \sim (b^1(t), \dots, b^n(t)), \dots \in \mathcal{G} \quad (-\delta < t < \delta)$$

where  $\delta > 0$ . For every  $t$  fixed and near enough to  $t = 0$ , the multiplication (2) may be applied, hence

$$a(t) \circ b(t) = c(t), \quad c^i(t) = g^i(a(t), b(t), t) \quad (12)$$

and there is a unit element  $e(t) \sim (e^1(t), \dots, e^n(t)) \in \mathcal{G}$ . With these assumptions, we speak of a *pseudogroup  $\mathcal{G}$  modelled on Lie groups*. (This is a slight change, the coordinates  $a^j$  in  $\mathbf{G}$  are "supplied" with parameter  $t$  and we may even denote  $\mathcal{G} = \mathbf{G}(t)$ .)

Let us suppose that  $\mathcal{G}$  naturally acts on  $\mathbf{N}$  with the invariant  $k$ . In more detail,

$$[x, k] \longrightarrow a(k) \bullet [x, k] = [y, k], \quad y^j = f^j(a(k), [x, k], k) \quad (13)$$

where the common composition rules hold true. It follows that the leaves  $\mathbf{N}(k) \subset \mathbf{N}$  defined by  $k = \text{const.}$  are preserved in the action: on every such leaf, the formulae (8) with  $a^j = a^j(k)$  substituted hold true. (The functions  $f^j$  in (13) moreover depend on the last parameter  $k$ .)

While the multiplications (2) on  $\mathbf{G}$  and (12) on  $\mathcal{G}$  do not essentially differ one from another very much, the action formulae (8) and (13) are of other nature. This is demonstrated by comparing the formulae (9) with the *infinitesimal transformations* of the pseudogroup  $\mathcal{G}$  denoted  $Z_i$ .

Inserting arbitrary function

$$a^i(k, \epsilon) \quad (a^i(k, 0) = e^i(k), \quad -\delta < \epsilon < \delta)$$

for  $a^i(k)$  into (13), it follows that

$$Z_i = \sum Z_i^j \frac{\partial}{\partial x^j} (+ 0 \cdot \frac{\partial}{\partial k}), \quad Z_i^j = b^i(k) \frac{\partial f^j}{\partial a^i}(e(k), [x, k], k) \quad (14)$$

where

$$b^i(k) = \frac{\partial a^i}{\partial \epsilon}(t, 0) \quad (i = 1, \dots, n).$$

Altogether we have the general infinitesimal transformation

$$Z = \sum Z_i = \sum b^i(k) Z_i^j \frac{\partial}{\partial x^j} \quad (15)$$

of the pseudogroup  $\mathcal{G}$ . One can, e.g., choose  $a^i(k, \epsilon) = e^i(k) + \epsilon b^i(k)$  and it follows that  $b^1(k), \dots, b^n(k)$  ( $-\delta < k < \delta$ ) may be quite arbitrary functions.

**Theorem 3** *Linearly independent over  $\mathbb{R}$  vector fields  $Z_i$  ( $i = 1, \dots, n$ ) are infinitesimal transformations of a pseudogroup  $\mathcal{G}$  acting on the space  $\mathbf{N}$  if and only if*

$$[Z_i, Z_{i'}] = \sum c_{ii'}^{i''}(k) Z_{i''} \quad (i, i', i'' = 1, \dots, n) \quad (16)$$

*with the structure constant depending on the invariant  $k$ .*

**Theorem 4** *Functions  $f^j$  in formula (13) satisfy the generalised Lie system*

$$\frac{\partial f^j}{\partial a^i}(a(k), [x, k], k) = \sum A_i^{i'}(a(k), k) \frac{\partial f^j}{\partial a^{i'}}(e(k), [x, k], k) \quad (17)$$

*with appropriate (fixed) functions  $A_i^{i'}$ .*

The proofs are routine but somewhat lengthy if presented with details. They rest on the observation that the action of the pseudogroup  $\mathcal{G}$  on the space  $\mathbf{N}$  preserves every leaf  $\mathbf{N}(k) \subset \mathbf{N}$ . On every such leaf, the classical Theorem 1 and Theorem 2 can be applied. The procedure described in [2] leads to certain Lie group  $\mathbf{G}(k)$  which altogether determine the pseudogroup  $\mathcal{G}$  on  $\mathbf{N}$  by using the action formula (13). The choice of functions  $b^i(k)$  in formula (14) on a fixed leaf is clearly irrelevant and one can suppose  $b^i(k) = 1$  which provides the simple final formula (17).

We conclude with the remark that quite analogous results can be obtained if there are more invariants, however, this is still very far from the solution of the general *Open problem*.

## CONCLUSION

The pseudogroups modelled on Lie groups are introduced in order to describe the interrelation between the well-known theory of Lie groups and actually the rather vague and involved theory of pseudogroups. The First and the Second Fundamental Theorems with *variable* structure constants then appear.

## Acknowledgements

The paper was supported by the project of the specific university research FAST-S-16-3385 at the Brno University of Technology.

## References

- [1] Lie S.: *Theorie der Transformationsgruppen*. Leipzig (1888, 1890, 1893).
- [2] Eisenhart L. P.: *Continuous groups of Transformations*. Princeton Univ. Press (1933); Dover Publications, Inc., New York 1961 ix+301 pp.
- [3] Mikeš J., Stepanova E., Vanžurová A. et al.: *Differential geometry of special mappings*. Palacký University Olomouc, Faculty of Science, Olomouc, 2015, 568 pp., ISBN: 978-80-244-4671-4.
- [4] Cartan É.: *Sur la structure des groupes infinis de transformations*. Ann. Ec. Norm. XXI (1904), 153–206 and XXII (1905), 219–308.

- [5] Cartan É.: *Seminaire de Mathématiques*. 4-e année (1936–1937).
- [6] Chrastinová V.: *Lie algebra structure constants need not be constant*. MITAV Brno (2017).

# An application of stochastic partial differential equations to transmission line modelling

Edita Kolářová, Lubomír Brancík

Fac. of Electrical Engineering and Communication, Brno University of Technology

Technická 8, 616 00 Brno, Czech Republic

kolara@feec.vutbr.cz, brancik@feec.vutbr.cz

**Abstract:** *In this paper we deal with stochastic partial differential equations (SPDEs). We shortly introduction the variational approach to SPDEs. Finally we apply the theory to the model of transmission line with stochastic source.*

**Keywords:** stochastic partial differential equation, Wiener process, cylindrical Wiener process, transmission line.

## INTRODUCTION

A stochastic partial differential equation (SPDE) is a partial differential equation containing a random term. The theory of SPDEs brings together techniques from probability theory, functional analysis, and the theory of partial differential equations.

Most of dynamics with stochastic influence in the nature or man-made complex systems can be modelled by stochastic partial differential equations (SPDEs). The state spaces of their solutions are infinite dimensional spaces of functions, mostly Hilbert spaces or separable Banach spaces. The representation of the white noise in SPDEs is the cylindrical Wiener process  $W(t, x)$ , which has some spatial correlation. First we introduce the standard Wiener process or Brownian motion.

**Definition 1**  $\beta(t) = \{\beta(t, \omega), t \geq 0, \omega \in \Omega\}$ , a real-valued, continuous stochastic process on probability space  $(\Omega, \mathcal{A}, P)$  is called the **Wiener process** if  $\beta(0) = 0$ ,  $\beta(t) - \beta(s)$  is  $N(0, t - s)$  for all  $t \geq s \geq 0$  and the random variables  $\beta(t_1), \beta(t_2) - \beta(t_1), \dots, \beta(t_n) - \beta(t_{n-1})$  for all  $0 < t_1 < t_2 < \dots < t_n$ , are independent.

Note:  $\beta(t) = \beta(t) - \beta(0) \sim N(0, t)$ ,  $E[\beta(t)] = 0$  and  $E[\beta^2(t)] = t$  for  $t \geq 0$ .

For SPDEs we have to introduce space dependence into Wiener process. Let  $U$  be a separable Hilbert space with norm  $\|\cdot\|_U$  and inner product  $\langle \cdot, \cdot \rangle_U$ . We define the cylindrical Wiener process  $W(t) = W(t, x)$  as an  $U$ -valued process:

**Definition 2** Let  $U$  be a separable Hilbert space. **The cylindrical Wiener process** (also called space-time white noise) is the process

$$W(t) = \sum_{j=1}^{\infty} \chi_j \beta_j(t),$$

where  $\{\chi_j\}_{j=1}^{\infty}$  is any orthonormal basis of  $U$  and  $\beta_j(t)$  are Wiener processes.

If  $U \subset U_1$  for a second Hilbert space  $U_1$ , the series converges in  $L^2(\Omega, U_1)$  if the inclusion  $\iota : U \rightarrow U_1$  is Hilbert-Schmidt.

We want to study stochastic differential equations on a real, separable, infinite dimensional Hilbert space  $H$  with a cylindrical Wiener process  $W(t)$  on another separable Hilbert space  $U$ .

$$dX(t) = A(t, X(t)) dt + B(t, X(t)) dW(t), \quad (1)$$

where  $A : [0, T] \times H \rightarrow H$  and  $B : [0, T] \times H \rightarrow L_2(U, H)$ . Here  $L_2(U, H)$  denotes the space of all Hilbert-Schmidt operators from  $U$  to  $H$ . The solution  $X(t)$  is a  $H$  valued stochastic process, that satisfies (1) in integral form (see [1], p. 73).

**Remark.** In the case, when  $B$  is independent on  $X(t)$  we call (1) an equation with additive noise, otherwise it is an equation with multiplicative noise.

**Example 1** Let  $W(t)$ ,  $t \in [0, T]$ ,  $T > 0$  be the  $m$  dimensional standard Wiener process on the probability space  $(\Omega, \mathcal{A}, P)$ . In terminology of the introduction  $U := \mathbb{R}^m$  and  $H := \mathbb{R}^n$ ,  $m, n \in \mathbb{N}$ . We denote  $M(n \times m)$  the set of all real  $n \times m$  matrices and define the maps

$$A : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad B : [0, T] \times \mathbb{R}^n \rightarrow M(n \times m),$$

that are continuous in  $x \in \mathbb{R}^n$  for fixed  $t \in [0, T]$ . Let's the initial condition  $X(0)$  is a given vector in  $\mathbb{R}^n$ . The equation (1) with these functions is an ordinary stochastic differential equation. Applications of ordinary stochastic differential equations to electrical network, including analytic and numeric solutions, can be found in [2].

## 1 Stochastic partial differential equation

### 1.1 Stochastic partial differential equation with additive noise

Let  $H$  be a Hilbert space, we denote its elements as  $u_t(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ ,  $n \in \mathbb{N}$  and  $W_t$  the cylindrical Wiener process as in definition 1 for every  $t \in [0, T]$ ,  $T > 0$ . Then a stochastic partial differential equation with additive noise has the form

$$du_t(\mathbf{x}) = L(t, u_t(\mathbf{x}), D_{\mathbf{x}}u_t(\mathbf{x}), D_{\mathbf{x}}^2u_t(\mathbf{x})) dt + B_t(\mathbf{x}) dW_t, \quad (2)$$

where  $D_{\mathbf{x}}$  denotes the first,  $D_{\mathbf{x}}^2$  the second total derivative of  $u_t$  with respect to  $\mathbf{x}$ .

**Example 2** Let  $\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$  be the Laplace operator. The stochastic version of the heat equation, see [1], can be written as

$$du_t = \Delta u_t dt + \sigma_t dW_t.$$

### 1.2 Linear first order SPDE with additive noise

If we have the operator  $L(t, u_t(\mathbf{x}), D_{\mathbf{x}}u_t(\mathbf{x}), D_{\mathbf{x}}^2u_t(\mathbf{x})) = L(t, u_t(\mathbf{x}), D_{\mathbf{x}}u_t(\mathbf{x}))$  in equation (2) and it is linear, we get a linear first order SPDE with additive noise.

**Example 3** Let  $u_t : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\sigma_t : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$ , are functions,  $\mathbf{a}$  is an  $n$ -dimensional vector and  $b$  a real number. Then the following equation is a linear SPDE with additive noise:

$$du_t(\mathbf{x}) = (\mathbf{a} \cdot \nabla u_t(\mathbf{x})^T + b \cdot u_t(\mathbf{x})) dt + \sigma_t(\mathbf{x}) dW_t.$$



## 2 Transmission line model with stochastic source

### 2.1 Deterministic transmission line model

The uniform transmission line (TL) of length  $l$  is described with per-unit-length primary parameters  $R, L, G$  and  $C$  as telegraphic partial differential equations for current and voltage as

$$\begin{aligned} -\frac{\partial i(t, x)}{\partial x} &= G v(t, x) + C \frac{\partial v(t, x)}{\partial t}, \\ -\frac{\partial v(t, x)}{\partial x} &= R i(t, x) + L \frac{\partial i(t, x)}{\partial t}, \end{aligned}$$

where  $x$  is the length from the TL's beginning. This equation has the following matrix form

$$-\frac{\partial}{\partial x} \begin{pmatrix} i \\ v \end{pmatrix} = \begin{pmatrix} 0 & G \\ R & 0 \end{pmatrix} \begin{pmatrix} i \\ v \end{pmatrix} + \begin{pmatrix} 0 & C \\ L & 0 \end{pmatrix} \frac{\partial}{\partial t} \begin{pmatrix} i \\ v \end{pmatrix}. \quad (3)$$

We can solve this equation with given boundary conditions by analytical and by numerical methods as well, see [3].

### 2.2 Stochastic transmission line model

For the stochastic model first we rewrite the equation (3) as

$$\frac{\partial}{\partial t} u_t(x) = P \frac{\partial}{\partial x} u_t(x) + S u_t(x), \quad (4)$$

where

$$u_t(x) = \begin{pmatrix} i(t, x) \\ v(t, x) \end{pmatrix}, \quad P = - \begin{pmatrix} 0 & C \\ L & 0 \end{pmatrix}^{-1} \text{ and } S = - \begin{pmatrix} 0 & C \\ L & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 & G \\ R & 0 \end{pmatrix}.$$

Let us to allow the source be influenced by some randomness. We will consider

$$v^*(t, 0) = v(t, 0) + \text{"noise"}.$$

Substituting this into the deterministic model we get the stochastic model of transmission line with random source as

$$du_t(x) = L \left( u_t(x), \frac{\partial}{\partial x} u_t(x) \right) dt + \sigma_t(x) dW_t, \quad (5)$$

where  $L \left( u_t(x), \frac{\partial}{\partial x} u_t(x) \right) = P \frac{\partial}{\partial x} u_t(x) + S u_t(x)$ ,  $x \in (0, \infty)$ ,  $t \in [0, T]$ ,  $T > 0$  as in equation (4) and  $\sigma_t(x)$  is a  $2 \times 2$  matrix function on  $\mathbb{R}$ .

## 3 Conclusion

The theory of SPDEs is an interdisciplinary subject in mathematics and there is a very rich literature in all three main "approaches" to this theory, as the martingale approach, the semigroup approach and the variational approach. In this paper we gave a short introduction to the variational approach. Our aim is to create and solve transmission line models effected by randomness in source. So

far we solved the problem by modeling the transmission line as a cascade connection of lumped-parameter circuits, the RLGC cells, which led to a system of ordinary differential equations after the state-variable method was applied. If we consider this system having noisy source, we get a system of stochastic ordinary differential equations (see [4] and [5]). We deal and solve such systems in [6]. But this was only approximate solution as the mathematical model of the transmission line leads to a linear partial differential equation. If we allow some randomness in source, it leads to SPDE described in this paper. Our next goal is to solve such SPDEs by numerical methods.

## References

- [1] Prévôt C., Röckner M.: *A Concise Course on Stochastic Partial Differential Equations*. Lecture Notes in Mathematics, Springer, 2007.
- [2] Kolářová E.: Applications of second order stochastic integral equations to electrical networks. *Tatra Mountains Mathematical Publications*, 2015, vol. 63, p. 163-173.
- [3] Granzow K. D.: *Digital Transmission Lines: Computer Modelling and Analysis*. New York: Oxford University Press, 1998.
- [4] Øksendal B.: *Stochastic Differential Equations, An Introduction with Applications*, New York: Springer-Verlag, 2000.
- [5] Baštinec, J., Klimešová, M. Stability of the Zero Solution of Stochastic Differential Systems with Four-Dimensional Brownian Motion. In: *Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 7-30. [Online]. [Cit. 2017-07-26]. Available at: <http://mitav.unob.cz/data/MITAV2016Proceedings.pdf>. ISBN 978-80-7231-400-3.
- [6] Brančík, L. and Kolářová, E.: Simulation of multiconductor transmission lines with random parameters via stochastic differential equations approach, *SIMULATION-Transactions of the Society for Modeling and Simulation International*, 2016, vol. 92, no. 6, p. 521-533.
- [7] Right, A. *Distance Learning*. Prague: Charles University, 2008, 350 pp. ISBN 978-80-7231-615-1.
- [8] Kolmanovskii, V.B. Delay equations and mathematical modelling. *Soros Educational Journal*, No. 4, 1996, p. 122-127. ISSN 1511-1100.

## Acknowledgement

This work was supported by Czech Science Foundation under grant 15-18288S.

# PRIESTLEY-CHAO ESTIMATOR OF CONDITIONAL DENSITY

Kateřina Konečná<sup>1, 2</sup>

<sup>1</sup> Faculty of Civil Engineering, Brno University of Technology

Žižkova 17, Brno, Czech Republic

konecna.k@fce.vutbr.cz

<sup>2</sup> Faculty of Science, Masaryk University,

Kotlářská 2, Brno, Czech Republic

**Abstract:** *This contribution is focused on a non-parametric estimation of conditional density. Several types of kernel estimators of conditional density are known, the Nadaraya-Watson and the local linear estimators are the widest used ones. We focus on a new estimator - the Priestley-Chao estimator of conditional density. As conditional density can be regarded as a generalization of regression, the Priestley-Chao estimator, proposed initially for kernel regression, is extended for kernel estimation of conditional density. The conditional characteristics and the statistical properties of the suggested estimator are derived. The estimator depends on the smoothing parameters called bandwidths which influence the final quality of the estimate significantly. The cross-validation method is suggested for their estimation and the expression for the cross-validation function is derived. The theoretical approach is supplemented by a simulation study.*

**Keywords:** kernel smoothing, conditional density, Priestley-Chao estimator, statistical properties, bandwidth selection, cross-validation method.

## INTRODUCTION

A conditional density estimation provides a very comprehensive information about the data set. The conditional density expresses the probability  $f(y|x)$  of a random variable  $Y|(X = x)$ , it can be regarded as a generalization of regression. While regression models the conditional expectation, conditional density models the distribution in a fixed point  $x$ , including conditional expectation and uncertainty.

The conditional density estimator generally depends on the smoothing parameters, called bandwidths. The widths of the smoothing parameters influence the final estimation significantly. This is the reason why so much attention is paid to their detection. While the optimal values of the smoothing parameters depend on the unknown conditional (and marginal) density, a data-driven method is needed for their practical estimation. Such one method, the cross-validation method, is suggested. The performance of the Priestley-Chao estimator and the cross-validation method is included via a simulation study.

## 1 THE PRIESTLEY-CHAO ESTIMATOR OF CONDITIONAL DENSITY

Conditional density  $f(y|x)$  models the probability of a random variable  $Y$  given a random variable  $X$ , represented by a fixed observation  $X = x$ . The conditional density estimations provide a

detailed information about the data distribution. Besides modelling the distribution in fixed observations, conditional density produces also the conditional expectation and its uncertainty.

In kernel smoothing generally, the main building block is a kernel function, which plays a role of a weighting function.

**Definition 1.1** [17] *Let  $K$  be a real valued function satisfying:*

1.  $K \in \text{Lip}[-1, 1]$ , i.e.  $|K(x) - K(y)| \leq L|x - y|$ ,  $\forall x, y \in [-1, 1]$ ,  $L > 0$ ,
2.  $\text{supp}(K) = [-1, 1]$ ,
3. *moment conditions:*

$$\int_{-1}^1 K(x) dx = 1, \int_{-1}^1 xK(x) dx = 0, \int_{-1}^1 x^2 K(x) dx = \beta_2(K) \neq 0.$$

*Such a function  $K$  is called a kernel of order 2.*

The Epanechnikov, quartic, uniform, triangular kernel etc. are the examples of the kernel functions. In practice as well as in our simulation study, the Gaussian kernel is used due to computational aspects, although the second condition is not satisfied because of the unconstrained support of the kernel.

The smoothing parameters, called bandwidths, play a very important role in kernel smoothing. The smoothing parameters in the  $x$  and  $y$  direction are denoted as  $h_x$  and  $h_y$ , and they control the smoothness of the estimate. It is very important to work with the "appropriate" values of the smoothing parameters, otherwise the final estimate could tend to be undersmoothed or over-smoothed.

This is the reason, why so much attention is paid to bandwidth selection. There are many publications dealing with this topic, a lot of them proceeds from the methods basically developed for kernel regression and/or kernel density estimation. We can mention a reference rule method [2], based on the assumption of uniform or normal marginal density and normal conditional density with linear mean and linear variance. An iterative method [12] is the extension of the iterative method suggested for kernel density estimations and kernel regression (see [8], [7] and [11]). A method of penalizing functions [2] and a bootstrap method [2], [4] can also be mentioned.

The beginnings of kernel conditional density estimations date back to 1969 when the classical conditional density estimator was proposed by Rosenblatt ([16]). Despite this, kernel smoothing is still used in both, theoretical and practical cases. For example in [13], conditional density estimator was suggested for a left-truncated and right-censored mode, whereas [14] discuss a class of estimators in the cases when the conditioning variable is either circular or linear. The theoretical as well as the practical application of kernel smoothing can be found in [10], the authors are focused on a new estimator of  $f(y|\mathbf{x})$  that adapts to sparse structure in  $\mathbf{x}$ . They also show applications of ZIP Code data, Galaxy spectra, and photometric redshift estimation. Another example of the application can be seen in [1], in which authors use kernel conditional density estimation with the incorporation of

a decay parameter to forecast electricity smart meter data. Thus, their results can help consumers analyse and minimize their excess electricity usage, and the estimates can be used to devise innovative pricing strategies for suppliers.

Let  $(X, Y)$  be a random vector and  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  its observations. Our aim is to construct a new estimator of conditional density, there are several types of the already known estimators. Generally, the kernel conditional density estimator takes the form

$$\hat{f}(y|x) = \sum_{i=1}^n w_i(x) K_{h_y}(y - Y_i),$$

where  $w_i(x)$  is a weight function in the point  $x$ . The type of the estimator depends on the choice of the weighting function. The commonly used estimator is the Nadaraya-Watson estimator ([16]) with the weighting function in the form

$$w_i^{NW}(x) = \frac{K_{h_x}(x - X_i)}{\sum_{i=1}^n K_{h_x}(x - X_i)}.$$

The name of the estimator comes from the popular Nadaraya-Watson estimator of regression. In [9], the Nadaraya-Watson estimator was improved by a two-step estimator, characterized by lower bias. The other (but not so widely used as the Nadaraya-Watson estimator) estimator is the local linear estimator ([3]) with the weighting function of the form

$$w_i^{LL}(x) = \frac{K_{h_x}(x - X_i) (\hat{s}_2(x) - (x - X_i) \hat{s}_1(x))}{\hat{s}_0(x) \hat{s}_2(x) - \hat{s}_1^2(x)}$$

and the auxiliary function  $\hat{s}_j(x) = \frac{1}{n} \sum_{i=1}^n (x - X_i)^j K_{h_x}(x - X_i)$ . Better statistical properties are the reason for using the local linear estimator.

The Nadaraya-Watson estimator is mostly used for non-uniformly distributed design variable  $X$ . The random design with the non-uniformly distributed variable  $X$  is convenient especially when deriving the asymptotic properties. For equally spaced designs, we introduce a new estimator of conditional density. The estimator is an extension of the Priestley-Chao estimator of the regression function, suggested by Priestley and Chao in [15].

The fixed design is supposed in the Priestley-Chao estimator construction. Although the design of  $n$  observations is supposed in the form  $x_i = \frac{i}{n}, i = 1 \dots, n$ , the design points can not be restricted only on the interval  $[0, 1]$  but generally on  $[a, b], a < b$ . In [15], Priestley and Chao suggested even the kernel regression estimator removing the restriction of equally spaced design.

The Priestley-Chao estimator of conditional density is defined as

$$\hat{f}_{PC}(y|x) = \delta \sum_{i=1}^n K_{h_x}(x - x_i) K_{h_y}(y - Y_i). \quad (1)$$

As the conditional density estimation is a generalization of regression, the regression function is represented by the conditional mean

$$\hat{m}_{PC}(x) = \delta \sum_i K_{h_x}(x - x_i) Y_i. \quad (2)$$

The estimator (2) is the Priestley-Chao estimator of regression function introduced by Priestley and Chao in [15].

## 2 STATISTICAL PROPERTIES OF THE PRIESTLEY-CHAO ESTIMATOR

In this section, the statistical properties of the Priestley-Chao estimator are focused on. The expressions of the statistical properties are necessary for appraisal of a suitability of the estimator in both, the local as well as the global view of the quality measure of the estimator. At first, bias and variance of the estimator are given.

**Theorem 1** *Let  $x$  be a fixed design,  $Y$  random variable with conditional density  $f(y|x)$  being at least twice continuously differentiable, and  $K$  be a kernel function satisfying Definition 1.1. For  $h_x \rightarrow 0$ ,  $h_y \rightarrow 0$  and  $nh_x h_y \rightarrow \infty$  as  $n \rightarrow \infty$ , asymptotic bias (AB) and asymptotic variance (AV) of the Priestley-Chao estimator are given by the expressions*

$$\text{AB} \left\{ \hat{f}_{PC}(y|x) \right\} = \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2}, \quad (3)$$

$$\text{AV} \left\{ \hat{f}_{PC}(y|x) \right\} = \frac{\delta}{h_x h_y} R^2(K) f(y|x), \quad (4)$$

where  $R(K) = \int K^2(u) du$ .

**Proof.** The proof is given in the Appendix.

The Asymptotic Mean Squared Error (AMSE) is the local measure of the quality of the estimator at the point  $[x, y]$ . AMSE is defined as a summation of the Asymptotic Squared Bias (ASB, the main term of squared bias) and Asymptotic Variance (AV, the main term of variance) by the expression

$$\begin{aligned} \text{AMSE} \left\{ \hat{f}_{PC}(y|x) \right\} &= \text{AV} \left\{ \hat{f}_{PC}(y|x) \right\} + \text{ASB} \left\{ \hat{f}_{PC}(y|x) \right\} \\ &= \frac{\delta}{h_x h_y} R^2(K) f(y|x) + \left( \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2} \right)^2. \end{aligned}$$

The global measure of the quality of the estimator is given by the Asymptotic Mean Integrated Squared Error (AMISE) by the expression

$$\text{AMISE} \left\{ \hat{f}_{PC}(\cdot|\cdot) \right\} = \iint \text{AMSE} \left\{ \hat{f}_{PC}(y|x) \right\} dx dy = \frac{\delta}{h_x h_y} c_1 + c_2 h_x^4 + c_3 h_y^4 + c_4 h_x^2 h_y^2 \quad (5)$$

with the constants  $c_1, c_2, c_3, c_4$  in the forms

$$\begin{aligned} c_1 &= \int R^2(K) dx, \\ c_2 &= \frac{1}{4}\beta_2^2(K) \iint \left( \frac{\partial^2 f(y|x)}{\partial x^2} \right)^2 dx dy, \\ c_3 &= \frac{1}{4}\beta_2^2(K) \iint \left( \frac{\partial^2 f(y|x)}{\partial y^2} \right)^2 dx dy, \\ c_4 &= \frac{1}{2}\beta_2^2(K) \iint \frac{\partial^2 f(y|x)}{\partial x^2} \frac{\partial^2 f(y|x)}{\partial y^2} dx dy. \end{aligned}$$

### 3 METHODS FOR ESTIMATING THE BANDWIDTHS

The values of the smoothing parameters have the essential significance on the final estimate of conditional density. While choosing too small bandwidths, the final estimate will tend to undersmooth and will contain an abundance of information. On the other hand, the oversmoothed estimate with the lack of information will be obtained in the case of choosing too large bandwidths.

At first, the optimal bandwidths are derived as the values which minimize the global measure of the quality (5). The optimal bandwidths depend on the unknown conditional density, thus the data-driven method is needed for their estimations. The classical approach, the cross-validation method, is introduced.

In literature, there are plenty of methods for bandwidth selection. The cross-validation method is a widely used method, it was suggested by Fan and Yim [4], Hansen [6], and Hall, Racine and Li [5]. We use the original idea of the method and we derive the cross-validation function for the Priestley-Chao estimator.

#### 3.1 Optimal values of the smoothing parameters

The optimal values of the smoothing parameters are the values minimizing the global measure  $\text{AMISE}\{\hat{f}_{PC}(\cdot|\cdot)\}$  of the quality of the estimate. The optimal bandwidths can be derived by differentiating (5) with respect to  $h_x$  and  $h_y$  and setting the derivatives to 0. Thus, we get the following system of two non-linear equations

$$-\frac{\delta}{h_x^2 h_y} c_1 + 4c_2 h_x^3 + 2c_4 h_x h_y^2 = 0 \quad (6)$$

$$-\frac{\delta}{h_x h_y^2} c_1 + 4c_3 h_y^3 + 2c_4 h_x^2 h_y = 0. \quad (7)$$

Further, making several algebraic simplifications and then adding the equations (6) and (7) together, we get

$$4c_2 h_x^5 h_y - 4c_3 h_x h_y^5 = 0. \quad (8)$$

Solving the equation (8) with respect to  $h_y$  and substituting this expression to (6), the optimal values of the smoothing parameters are given by

$$h_x^* = \delta^{1/6} c_1^{1/6} \left( 4 \left( \frac{c_2^5}{c_3} \right)^{1/4} + 2c_4 \left( \frac{c_2}{c_3} \right)^{3/4} \right)^{-1/6}$$

$$h_y^* = \left( \frac{c_2}{c_3} \right)^{1/4} h_x^*.$$

### 3.2 Cross-validation method

In kernel smoothing generally, the cross-validation method is a standard procedure widely used for bandwidth detection. This method is based on the minimization of the cross-validation function, which is represented by the global quality measure ISE (Integrated Squared Error). Its derivation follows the method proposed by Fan and Yim ([4]), the error measure is given by

$$\begin{aligned} \text{ISE} \left\{ \hat{f}_{PC}(\cdot|\cdot) \right\} &= \iint \left( \hat{f}_{PC}(y|x) - f(y|x) \right)^2 dx dy \\ &= \iint \hat{f}_{PC}^2(y|x) dx dy - 2 \iint \hat{f}_{PC}(y|x) f(y|x) dx dy + \iint f^2(y|x) dx dy \\ &=: I_1 - 2I_2 + I_3. \end{aligned}$$

As the term  $I_3$  does not depend on the unknown parameters  $h_x$  and  $h_y$ , the function being minimized is formed by terms  $I_1$  and  $I_2$  only. This function as called the cross-validation function and it can be defined as

$$CV(h_x, h_y) = I_1 - 2I_2.$$

The Gaussian kernel, the symmetry of the kernel function and the symmetry of the kernel convolution (denoted by  $*$ ) are used in the computations. The term  $I_1$  can be derived as follows

$$\begin{aligned} I_1 &= \iint \hat{f}_{PC}^2(y|x) dx dy \\ &= \iint \sum_i \sum_j \delta^2 K_{h_x}(x - x_i) K_{h_x}(x - x_j) K_{h_y}(y - Y_i) K_{h_y}(y - Y_j) dx dy \\ &= \delta^2 \sum_i \sum_j \iint K(t) K\left(t - \frac{x_j - x_i}{h_x}\right) K(v) K\left(v - \frac{Y_j - Y_i}{h_y}\right) dt dv \\ &= \delta^2 \sum_i \sum_j \int K(t) K\left(\frac{x_j - x_i}{h_x} - t\right) dt \cdot \int K(v) K\left(\frac{Y_j - Y_i}{h_y} - v\right) dv \\ &= \delta^2 \sum_i \sum_j (K * K)\left(\frac{x_i - x_j}{h_x}\right) (K * K)\left(\frac{Y_i - Y_j}{h_y}\right) \\ &= \delta^2 \sum_i \sum_j h_x h_y K_{h_x \sqrt{2}}(x_i - x_j) K_{h_y \sqrt{2}}(Y_i - Y_j). \end{aligned}$$



The term  $I_2$  is given by

$$\begin{aligned}
I_2 &= \iint \hat{f}_{PC}(y|x) f(y|x) dx dy \\
&= \iint \delta \sum_i K_{h_x}(x - x_i) K_{h_y}(y - Y_i) f(y|x) dx dy \\
&= \iint \delta \sum_i K(t)K(v) f(Y_i + h_y v | x_i + h_x t) dt dv \\
&= \delta \sum_i \iint K(t)K(v) \hat{f}(Y_i | x_i) dt dv \\
&= \delta \sum_i \hat{f}_{PC}(Y_i | x_i).
\end{aligned}$$

In our computations, we use the leave-one out cross-validation method. It means, that we use the estimation in the pair of points  $(x_i, Y_i)$  using points  $\{(x_j, Y_j), i \neq j\}$ . Finally, the cross-validation function is of the form

$$CV(h_x, h_y) = I_1 - 2I_2 = \delta^2 \sum_i \sum_{j \neq i} h_x h_y K_{h_x \sqrt{2}}(x_i - x_j) K_{h_y \sqrt{2}}(Y_i - Y_j) - 2\delta \sum_i \hat{f}_{PC}(Y_i | x_i).$$

The optimal values of the bandwidths are given by minimizing of  $CV(h_x^*, h_y^*)$

$$(h_{x,PC}^{CV}, h_{y,PC}^{CV}) = \arg \min_{(h_x, h_y)} CV(h_x, h_y).$$

## 4 SIMULATION STUDY

In this section, we conduct a simulation study introducing the cross-validation method. The simulation study involves two models defined as

$$\begin{aligned}
M_1 : Y_i &= e^{x_i} + \varepsilon_i, \quad x_i = \frac{i}{n}, \quad i = 1 \dots, 100, \quad \varepsilon_i \sim N(0, 0.5^2) \\
M_2 : Y_i &= \sin(3\pi x_i^2) + \varepsilon_i, \quad x_i = \frac{i}{n}, \quad i = 1 \dots, 100, \quad \varepsilon_i \sim N(1, 1)
\end{aligned}$$

At first, one hundred observations are generated from each model to apply the Priestley-Chao estimator for detection of conditional density. For both simulation studies, an exactly given grid of 100 times 100 points is considered to construct an estimation and measure of the error term. The  $x$  grid is formed by the observations  $x_i$ , the  $y$  grid is formed by the exact equidistant points at the range of  $Y$  values.

We perform the cross-validation method from several points of view, we assess the accuracy of the estimates of the smoothing parameters to the optimal bandwidths as well as the measure of quality is focused on. The measure of quality of the estimate is given by integrated squared error

$$ISE \left\{ \hat{f}_{PC}(y|x) \right\} = \iint \left\{ \hat{f}_{PC}(y|x) - f(y|x) \right\}^2 dx dy.$$

Due to computational aspect, we use its estimation

$$\widehat{\text{ISE}} \left\{ \hat{f}_{PC}(y|x) \right\} = \frac{\Delta}{n} \sum_{j=1}^N \sum_{i=1}^n \left( \hat{f}_{PC}(y_j|x_i) - f(y_j|x_i) \right)^2,$$

where  $\mathbf{y} = (y_1, \dots, y_N)$  is a vector of equally spaced values over the sample space of  $Y$  and  $\Delta$  is the distance between two consecutive values of  $\mathbf{y}$ . The expression  $\widehat{\text{ISE}}_{\text{opt}}$  stands for the integrated squared error of the estimator with the optimal values of the smoothing parameters.

The procedure of simulating the observations and computing the bandwidths, constructing the conditional density estimation and measuring the quality was repeated two hundred times to get the required characteristics.

The numerical results for the model  $M_1$  are given in Tab. 1.

	$\hat{h}_x$	$\hat{h}_y$	$\widehat{\text{ISE}}$	$\widehat{\text{ISE}}_{\text{opt}}$
mean	0.0757	0.0540	0.5501	0.0488
median	0.0746	0.0247	0.4346	0.0487
sd	0.0042	0.0542	0.4432	0.0112
IQR	0.0058	0.0809	0.8727	0.0149

**Tab. 1.** The statistical characteristics of the results for the model  $M_1$  with the optimal bandwidths  $h_x^* = 0.1157$  and  $h_y^* = 0.2236$ .

Source: own

It can be seen, that the cross-validation method undervalues the smoothing parameters in both, the  $x$  and  $y$  direction. Undersmoothing in the  $y$  direction is more significant than in the  $x$  direction. The reason for estimating such small value of the parameter  $h_y$  is in the shape of the cross-validation function, which minimum lies near the lower bound of the range of  $h_y$ . This is caused by the very smooth conditional expectation in the definition of the model  $M_1$ .

The numerical results for the model  $M_2$  are given in Tab. 2.

	$\hat{h}_x$	$\hat{h}_y$	$\widehat{\text{ISE}}$	$\widehat{\text{ISE}}_{\text{opt}}$
mean	0.0669	0.1544	0.0864	0.0272
median	0.0658	0.1217	0.0821	0.0266
sd	0.0027	0.0618	0.0358	0.0060
IQR	0.0022	0.1130	0.0600	0.0080

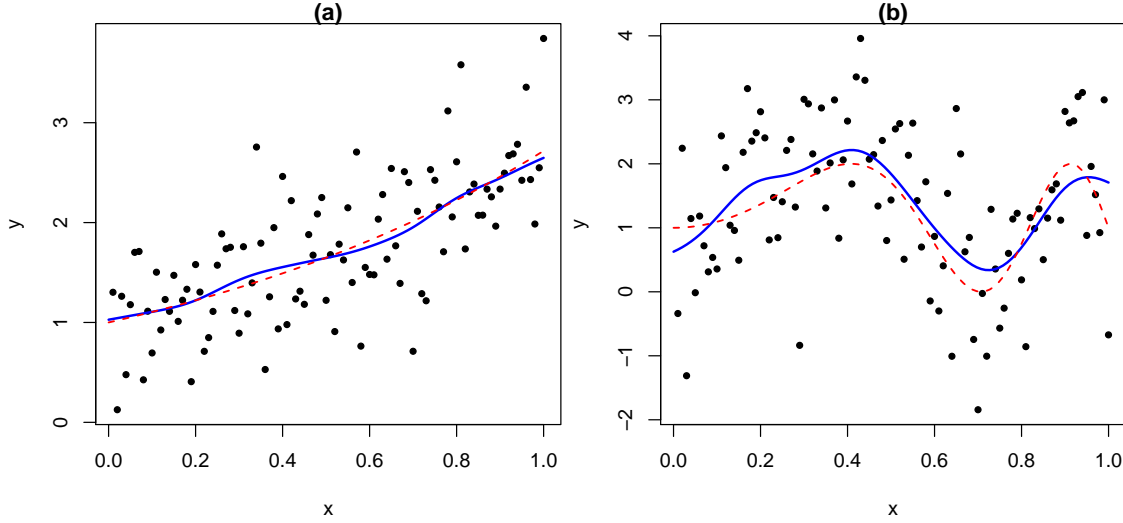
**Tab. 2.** The statistical characteristics of the results for the model  $M_2$  with the optimal bandwidths  $h_x^* = 0.0482$  and  $h_y^* = 0.5604$ .

Source: own

The simulation  $M_2$  is represented by the true conditional expectation changing in shape instead of very smooth conditional expectation as in the case of the model  $M_1$ . The cross-validation function has quite distinctive minimum in the  $x$  direction whereas finding the minimum is not so clear in the

$y$  direction due to its flat distribution. The results show slightly overvalued values of the smoothing parameter  $h_x$  and undervalued estimations of the smoothing parameter  $h_y$ . This simulation study also gives much better values for the error estimations.

As conditional density can be regarded as a generalization of regression, the estimation of the regression function is displayed for both models in Fig. 1.



**Fig. 1.** Data (black dots) simulated from the model (a)  $M_1$  and (b)  $M_2$ , the true conditional expectation (red dashed line) and the estimation of the regression function (blue solid line).

Source: own

## CONCLUSION

In this contribution, a new estimator - the Priestley-Chao estimator of conditional density - was focused on. The motivation for introducing the estimator was the Priestley-Chao estimator of the regression function, its simple construction and implementation. The statistical properties of the estimator like bias, variance, local and global measures of the quality of the estimates were derived. As the smoothing parameters play a pivotal role in kernel smoothing, the expressions for optimal bandwidths were computed. For bandwidth detection, the cross-validation method was suggested.

The appropriateness of the cross-validation method was explored via a simulation study. The simulations showed that the cross-validation approach can be the satisfactory method for bandwidth selection, especially in the cases with variable true regression function.

The future research should improve the bandwidth estimation for data with smooth regression function. The improvement should consist in a modified cross-validation function penalizing small values of the smoothing parameters. The future research should also be focused on developing other methods for bandwidth estimation. A method of reference rule given by [2] or iterative method by [12], both suggested for the Nadaraya-Watson estimator, can be extended for the new estimator. The future work could also include an improved estimator for data without restriction on equally-spaced design as proposed for the kernel regression in [15].

## APPENDIX

Here, you can find the detailed proof of Theorem 1.

**Proof.** At first, we prove the expression (3). We start with the derivation of the expectation of the estimator.

$$\begin{aligned} \mathbb{E} \left\{ \hat{f}_{PC}(y|x) \right\} &= n \delta \mathbb{E} \left\{ K_{h_x}(x - x_i) K_{h_y}(y - Y_i) \right\} \\ &= \iint K_{h_x}(x - u) K_{h_y}(y - v) f(v|u) \, du \, dv \\ &= f(y|x) + \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2} + \mathcal{O}(h_x^4) + \mathcal{O}(h_y^4) + \mathcal{O}(h_x^2 h_y^2). \end{aligned}$$

Then, bias is given as

$$\begin{aligned} \text{bias} \left\{ \hat{f}_{PC}(y|x) \right\} &= \mathbb{E} \left\{ \hat{f}_{PC}(y|x) \right\} - \hat{f}_{PC}(y|x) \\ &= \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2} + \mathcal{O}(h_x^4) + \mathcal{O}(h_y^4) + \mathcal{O}(h_x^2 h_y^2). \end{aligned}$$

As the asymptotic bias includes only the main term of  $\text{bias}\{\hat{f}_{PC}(y|x)\}$ , the expression is proved. For variance derivation of the estimator (1), the following expression is needed. Let  $X$  and  $Y$  be a random variables, variance of  $Y$  is stated by a well known law of total variance

$$\text{var} \{Y\} = \mathbb{E} \left\{ \text{var}_{Y|X} \{Y|X\} \right\} + \text{var} \left\{ \mathbb{E}_{Y|X} \{Y|X\} \right\}. \quad (9)$$

At first, the expression (9) is used for deriving the expression of variance for the  $i$ -th term of the estimator (1), followed by using the expression for variance of the summation of all the terms of the estimator.

Conditional expectation of the  $i$ -th term of the estimator  $\hat{f}_{PC}(y|x)$  is given by

$$\begin{aligned} \mathbb{E}_{f(y|x_i)} \left\{ \delta K_{h_x}(x - x_i) K_{h_y}(y - Y_i) | x_i \right\} &= \int \delta K_{h_x}(x - x_i) K_{h_y}(y - v) f(v|x_i) \, dv \\ &= \delta K_{h_x}(x - x_i) \int K(w) f(y - h_y w | x_i) \, dw \\ &= \delta K_{h_x} \left( f(y|x_i) + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x_i)}{\partial y^2} + \mathcal{O}(h_y^4) \right). \end{aligned} \quad (10)$$

The conditional expectation of the squared  $i$ -th term of (1) can be expressed by

$$\begin{aligned} \mathbb{E}_{f(y|x_i)} \left\{ \delta^2 K_{h_x}^2(x - x_i) K_{h_y}^2(y - Y_i) | x_i \right\} &= \delta^2 K_{h_x}^2(x - x_i) \frac{1}{h_y} \int K^2(v) f(y - h_y v | x_i) \, dv \\ &= \delta^2 K_{h_x}^2(x - x_i) \frac{1}{h_y} \left( R(K) f(y|x_i) + \frac{1}{2} h_y^2 G(K) \frac{\partial^2 f(y|x_i)}{\partial y^2} + \mathcal{O}(h_y^4) \right), \end{aligned} \quad (11)$$

where  $G(K) = \int u^2 K^2(u) \, du$ . The conditional variance is equal to the subtraction of the expressions (11) and the second power of (10)

$$\begin{aligned} \text{var}_{f(y|x_i)} \left\{ \delta K_{h_x}(x - x_i) K_{h_y}(y - Y_i) | x_i \right\} &= \delta^2 K_{h_x}^2(x - x_i) \left( \frac{1}{h_y} R(K) f(y|x_i) - f^2(y|x_i) + \frac{1}{2} h_y G(K) \frac{\partial^2 f(y|x_i)}{\partial y^2} + \mathcal{O}(h_y^2) \right). \end{aligned} \quad (12)$$

By an application of the expected value to the expression (12), we obtain

$$\begin{aligned}
& \mathbb{E} \left\{ \text{var}_{f(y|x_i)} \right\} \left\{ \delta K_{h_x} (x - x_i) K_{h_y} (y - Y_i) \right\} \\
&= \int \delta^2 K_{h_x}^2 (x - u) \left( \frac{1}{h_y} R(K) f(y|u) - f^2(y|u) + \frac{1}{2} h_y G(K) \frac{\partial^2 f(y|u)}{\partial y^2} \right) du \\
&= \frac{\delta^2 R^2(K) f(y|x)}{h_x h_y} - \frac{\delta^2}{h_x} R(K) f^2(y|x) + \frac{\delta^2 h_y}{2 h_x} R(K) G(K) \frac{\partial^2 f(y|x)}{\partial y^2} + \mathcal{O}(h_y^2). \quad (13)
\end{aligned}$$

This is a derivation of the first term of the expression (9). Now, we focus on the derivation of the second term of the expression (9). At first, we express the expectation of (10).

$$\begin{aligned}
& \mathbb{E} \left\{ \delta K_{h_x} (x - x_i) \left( f(y|x_i) + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x_i)}{\partial y^2} + \mathcal{O}(h_y^4) \right) \right\} \\
&= \int \delta K(t) \left( f(y|x - h_x t) + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x - h_x t)}{\partial y^2} + \mathcal{O}(h_y^4) \right) dt \\
&= \delta \left( f(y|x) + \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2} + \mathcal{O}(h_y^4) \right). \quad (14)
\end{aligned}$$

Further, the expected value for the second power of the expression (10) is equal to

$$\begin{aligned}
& \mathbb{E} \left\{ \delta^2 K_{h_x}^2 (x - x_i) \left( f(y|x_i) + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x_i)}{\partial y^2} + \mathcal{O}(h_y^4) \right)^2 \right\} \\
&= \int \delta^2 \frac{1}{h_x} K^2(t) \left( f^2(y|x - h_x t) + h_y^2 \beta_2(K) f(y|x - h_x t) \frac{\partial^2 f(y|x - h_x t)}{\partial y^2} \right. \\
&\quad \left. + \frac{1}{4} h_y^4 \beta_2^2(K) \left( \frac{\partial^2 f(y|x - h_x t)}{\partial y^2} \right)^2 + \mathcal{O}(h_y^4) \right) dt \\
&= \frac{\delta^2}{h_x} R(K) f^2(y|x) + \mathcal{O}(\delta^2 h_x) + \mathcal{O}(\delta^2 h_y^2). \quad (15)
\end{aligned}$$

The variance of the  $\mathbb{E}_{f(y|x_i)} \left\{ \delta K_{h_x} (x - x_i) K_{h_y} (y - Y_i) \right\}$  expression is derived by subtraction of (15) and (14) squared

$$\begin{aligned}
& \text{var} \left\{ \mathbb{E}_{f(y|x_i)} \left\{ \delta K_{h_x} (x - x_i) K_{h_y} (y - Y_i) \right\} \right\} \\
&= \frac{\delta^2}{h_x} R(K) f^2(y|x) - \delta^2 f^2(y|x) + \mathcal{O}(\delta^2 h_x) + \mathcal{O}(\delta^2 h_y^2). \quad (16)
\end{aligned}$$

The expression (16) is the desired second term of the expression (9). Thus, the variance of the  $i$ -th term of the Priestley-Chao estimator is given by a summation of (13) and (16)

$$\begin{aligned}
& \text{var} \left\{ \delta K_{h_x} (x - x_i) K_{h_y} (y - Y_i) \right\} \\
&= \frac{\delta^2}{h_x h_y} R^2(K) f(y|x) - \delta^2 f^2(y|x) + \frac{1}{2} \delta^2 \frac{h_y}{h_x} R(K) G(K) \frac{\partial^2 f(y|x)}{\partial y^2} + \mathcal{O}(\delta^2) + \mathcal{O}\left(\frac{\delta^2}{h_x}\right) + \mathcal{O}\left(\frac{\delta^2}{h_y}\right).
\end{aligned}$$

As  $\delta K_{h_x}(x - x_i) K_{h_y}(y - Y_1)$  and  $\delta K_{h_x}(x - x_2) K_{h_y}(y - Y_2)$  are stochastically independent, their covariance can be expressed as

$$\text{cov} \{ \delta K_{h_x}(x - x_i) K_{h_y}(y - Y_1), \delta K_{h_x}(x - x_2) K_{h_y}(y - Y_2) \} = 0.$$

Finally, the variance of the Priestley-Chao estimator is equal to

$$\begin{aligned} \text{var} \left\{ \delta \sum_i K_{h_x}(x - x_i) K_{h_y}(y - Y_i) \right\} &= \sum_{i=1}^n \text{var} \{ \delta K_{h_x}(x - x_i) K_{h_y}(y - Y_i) \} \\ &- 2 \sum_{i=1}^n \sum_{j>i} \text{cov} \{ \delta K_{h_x}(x - x_i) K_{h_y}(y - Y_i), \delta K_{h_x}(x - x_j) K_{h_y}(y - Y_j) \} \\ &= \frac{\delta}{h_x h_y} R^2(K) f(y|x) + \mathcal{O}(\delta) + \mathcal{O}\left(\frac{\delta}{h_x}\right) + \mathcal{O}\left(\frac{\delta}{h_y}\right). \end{aligned}$$

□

## References

- [1] Arora, S., Taylor, J. W. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, Vol. 59, Part A, 2016, p. 47 – 59. ISSN 0305-0483
- [2] Bashtannyk, D. M., Hyndman, R. J. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, Vol. 36, No. 3, 2001, p. 279 – 298. ISSN 0167-9473
- [3] Fan, J., Yao, Q., Tong, H. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, Vol. 83, No. 1, 1996, p. 189–206. ISSN 0006-3444
- [4] Fan, J., Yim, T. H. A crossvalidation method for estimating conditional densities. *Biometrika*, Vol. 91, No. 4, 2004, p. 819–834. ISSN 0006-3444
- [5] Hall, P. Cross-validation in density estimation. *Biometrika*, Vol. 69, No. 2, 1982, p. 383–390.
- [6] Hansen, B. E. Nonparametric conditional density estimation. Unpublished manuscript, 2004.
- [7] Horová, I., Koláček, J., Vopatová, K.. Full bandwidth matrix selectors for gradient kernel density estimate. *Computational Statistics & Data Analysis*, Vol. 57, No. 1, 2013, p. 364 – 376. ISSN 0167-9473
- [8] Horová, I., Zelinka, J. Contribution to the bandwidth choice for kernel density estimates. *Computational Statistics*, Vol. 22, No. 1, 2007, p. 31–47. ISSN 1613-9658
- [9] Hyndman, R. J., Bashtannyk, D. M., Grunwald, G. K. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, Vol. 5, No. 4, 1996, p. 315–336.
- [10] Izbicki, R., Lee, A. B. Nonparametric conditional density estimation in a high-dimensional regression setting. *Journal of Computational and Graphical Statistics*, Vol. 25, No. 4, 2016, p. 1297–1316. ISSN 1061-860
- [11] Koláček, J., Horová, I. Bandwidth matrix selectors for kernel regression. *Computational Statistics*, Vol. 32, No. 3, 2017, p. 1027–1046. ISSN 1613-9658
- [12] Konečná, K., Horová, I. Conditional Density Estimations. In: *Theoretical and Applied Issues in Statistics and Demography*, Athens: International Society for the Advancement of Science and Technology (ISAST), 2014, p. 15–31. ISBN 978-618-81257-7-3

- [13] Liang, H.-Y., Liu, A.-A. Kernel estimation of conditional density with truncated, censored and dependent data. *Journal of Multivariate Analysis*, Vol. 120, 2013, p. 40 – 58. ISSN 0047-259X
- [14] Marzio, M. D., Fensore, S., Panzera, A., Taylor, C. A note on nonparametric estimation of circular conditional densities. *Journal of Statistical Computation and Simulation*, Vol. 86, No. 13, 2016, p. 2573–2582. ISSN 0094-9655.
- [15] Priestley, M. B., Chao, M. T. Non-parametric function fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 3, 1972, p. 385–392. ISSN 0035-9246
- [16] Rosenblatt, M. Conditional probability density and regression estimators. *Multivariate analysis II*, New York: Academic Press, 1969, p. 25–31. ISBN 0124266525
- [17] Wand, M. P., Jones, M. C. *Kernel smoothing*. Crc Press, 1994, 224 pp. ISBN 978-041-255-2700

### **Acknowledgement**

This research was supported by the Project FAST-S-16-3385 (Brno University of Technology) and by the Czech Science Foundation no. GA15-06991S (Masaryk University).

## 3D PRINTING – LEARNING AND MASTERING

**Martin Kopecek, Petr Voda, Pravoslav Stransky, Josef Hanus**

Department of Medical Biophysics, Medical Faculty in Hradec Kralove, Charles University  
Simkova 870, 500 03 Hradec Kralove, Czech Republic  
kopecema@lfhk.cuni.cz, vodap@lfhk.cuni.cz, str@lfhk.cuni.cz,  
hanus@lfhk.cuni.cz

**Abstract:** *The use of the latest 3D technologies is increasingly gaining ground in biomedicine and clinical practice. The experimental biophysical laboratory of 3D printing was created for a better understanding of the practical impact of these innovations by students in dentistry and general medicine with the support of E-learning. The aim of the Lab is to provide students with the theoretical basis and practical skills, especially in the development of dental crowns and implants. The laboratory has been innovated thanks to closer cooperation with the Stomatology clinic at the University Hospital in Hradec Kralove. This allows students to solve real-life situations from practice using the latest treatment approaches. Innovated Lab expands the elective subject – Medical Biophysics Seminar. The impact of enhancement of the seminar by the topic of 3D printing has been analysed.*

**Keywords:** biophysics, 3D printing, 3D technologies, education, E-learning, prosthetics

### INTRODUCTION

Students during the studies at the Faculty of Medicine obtain through the education specialization of the Dentistry and General Medicine basic knowledge of physical and biophysical principles of the physiological processes in a human body. Biophysics leads the students to logical reasoning in finding solutions to the tasks built on a basis of the physics. This subject is not one of the most popular and, to students, it is therefore necessary to lend a hand. Complex processes should be explained to students by attractive and creative forms of education. Possibilities of physics in medicine are far-reaching and students should learn the correct orientation in the maze of physical concepts, methods, equipment and processes. The latest devices and the latest techniques are necessary to include into specific lecture for a preparation of the future physicians. The preparation should directly correspond with the possibilities of their future workplace, and current trends in a medicine.

The cornerstone for the attractiveness on the field of the physics is the use of the modern information and communication technologies and the practical demonstrations that have specific practical outcomes. Department of Medical Biophysics has been involved in research and development of E-learning systems. These were found more effective in education of medical systems [1, 2] and - especially for laboratory experiments [3]. For example, in the practicum in areas such as nitinol stents [4], materials for dentistry [5] and other several topics, E-learning courses were created to deepen the knowledge and combine an interesting education environment with the practical examples. Also, the knowledge of basic branches of the statistic according to Kordek [6, 7, 8] is an important part of the practical training of undergraduate students for their future research. In addition to the core topics of biophysics, students can expand their knowledge in the elective subject – Seminar of



medical biophysics. There is an effort to attract teaching about the latest trends in medicine and enable students to express their creativity.

There are a lot of medical branches where it is possible to enforce applications of the 3D printing. It is known in implantology, dentistry, printing of parts of artificial skeleton, the newest development in 3D Bio-printing (e.g., artificial blood-vessels, heart, livers) [9], etc.

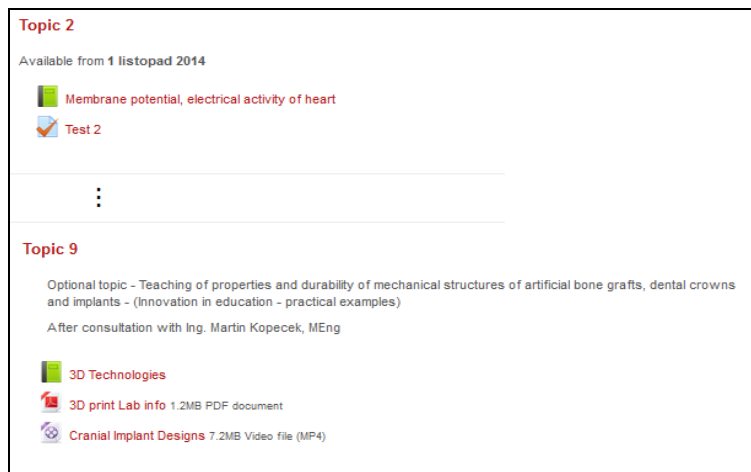
The specialized laboratory for practical experience with 3D technology was innovated within the elective course with the use of the E-learning. Practicum is based on design, creating, and testing of the artificial 3D printing of the cranial implants and it is also extended for dental crowns and implants. The aim of this work is to introduce the possibility of the 3D printing technology in combination with the education platform. Students can identify the advantages and disadvantages of 3D printing, generally applied in medicine, especially in implantology and dentistry. The laboratory is designed to show how to prepare final real dental crown element by the stereolithography (DLP) 3D printer with special software. The virtual preparation with the data pre-processing is used. The real output from the lab is tested on the printed sample.

There are many scientific articles, where the techniques and methods of the 3D printing are described (a simple search for “3D printing” in the Web of Science database produces 7. 965 results). This work is over these quite different and it is unique in accessing these technologies for educational purposes for the students of medicine.

## **1. DENTISTRY IN THE MODERN CONCEPTION OF LEARNING AND MASTERING**

### **1.1 Elective seminar – extended topics**

The seminar was split into the 8 branches – topics. The 3D technologies were added as the ninth one. This topic is optional and is intended only to candidates who wish to extend their knowledge in this field. Topics include several parts which can or must students fulfill. For a better idea, below in the Figure 1 are the examples of the classification each topic in multiple step learning (MSL) concepts. The MSL is a way of E-learning course creation in Moodle sw which reflects the previous knowledge of enrolled students. MSL allow to the readers choose the optimal level of the text difficulty. There are possible three text setups from elementary to extended knowledge of the problematic. Student can click on one button and immediately speedy the learning skills and the amount of displayed information.



**Fig. 1.** Categorization of the topics in the MSL system Moodle  
Source: own

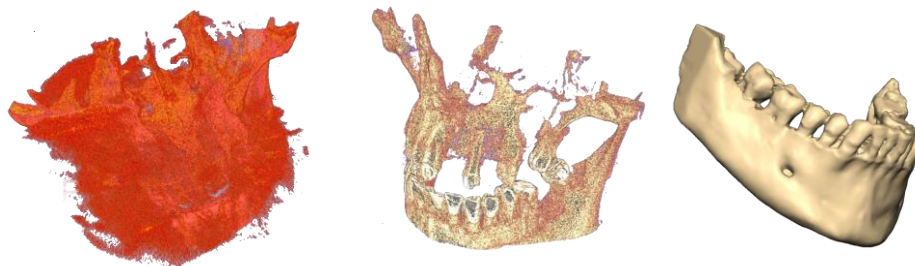
### 1.1.1 3D printing in dentistry

As it was written in the previous text, the course in the laboratory is optional. According to the number of the students what are interested in the 3D printing, the individual time windows for 2-3 students are created.

The structure of the course is possible to split into several parts: *Theoretical preparation* (E-learning form - the study materials, manuals and presentation are prepared in the system LMS Moodle); *Design of the implant, dental crown* (the real and virtual skull with the cranial defect is pre-prepared for the students, design of the implant or crown); *Preparation of the print area* (sw CreationWorkshop and individual settings of the printer); *Print itself* (the time of printing depends on the position of the printed element in the 3D printer); *Finalising the implant* (cleaning by isopropyl alcohol, water, UV lamp hardening); *Testing and measuring the implant* (use of the push-pull device); *Protocol* (working with the data).

### 1.1.2 Pre-processing

Anonymised data (CT pictures) of real patient are pre-processed because it is very time consuming. DICOM CT data model with some anatomical defect was used for those interested in the studies of General medicine and the real CT data were also used for design of the dental crown and implant for Dentistry. The pre-processing of the dentistry model is shown in Figure 2. For 3D computer model of the skeleton (reading and editing of data) and the final .stl data format has been used 3D Slicer sw, Autodesk Inventor sw and Meshmixer sw.



**Fig. 2.** Example of the dentistry task with pre-processed model; *General DICOM data* (on the left side), *Filtering and editing* (at the center); *Final model* (on the right side).

Source: own

### **1.1.3 Model creating instruction**

Theoretical preparation (LMS Moodle) can take place distantly. The form of theory, module, is called “Book” and the advantage is also an advanced way of dividing books for chapters and subchapters. Practical, laboratory part of the course is with the teacher in the specialized laboratory. There are more workflows how to accomplish the task. There may be many solutions or can be used prepared datasheet (manual) and E-learning video course. The video course shows all necessary steps for correct design. The mathematical functions, calculations and measurements are described and explained. It is not therefore necessary have the experiences and knowledge in the field of 3D modelling and design.

Virtual solution is different for all groups of the students. Creativity has no limits, but it is necessary to stick to the task - therefore suggest workable solutions.

Practical solution is realized by 3D printer. The settings of the printer are predefined. The system of the printing is individual and depends on the size of the printed element.

### **1.1.4 Materials for printing**

It is important for work with the 3D printer to know the basic materials what we can use. Generally in these days exists materials as (PLA, ABS, photo resins, many types of metals etc.) Between the processing techniques these materials e.g. melting of plastics (FDM), curing the photo resins (DLP, SLA, PolyJet, CLIP), sintering of the various type of the metals (SLS). The DLP printer 3DWARF uses the photo-resin. The model is printed in reverse, pulled out of the beaker filled with the coloured polymer.

## **1.2 Evaluation of measurement**

The printed implant is tested on the push-pull device “Intron”. The material static testing is used in a compression mode within a single frame on the Instron 1 kN head device. For example, students with teacher sets the device for simulation the cranial defect caused by fall to blunt edge. The design of the methods for measuring the mechanical properties of the elements was taken into account Navrátil [10] in the stress - strain characteristic at a certain temperature. Students have to find the critical points and calculate rigidity of the implant or other element by the linearization.

## **1.3 Statistical evaluation**

The results were compared, processed, and statistically analyzed using MS Excel 2007 (Microsoft Corp, Redmond WA, USA). The basic descriptive statistics was used.

The cumulative level of active participant, related to the topic “Teaching of properties and durability of mechanical structures of artificial bone grafts and implants” (*3D Technologies*) after application of the modified MSL e-learnig course Medical Biophysics seminar - elective subject was compared with participant activity in other eight topics of this subject.

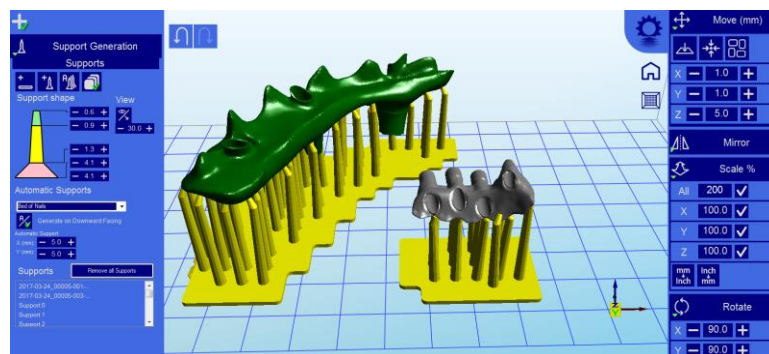
## 2. RESULTS

The biggest benefit allow to the student compare the virtual solution of their design with the real prototype of the dental crowns or jaw implants as is shown at the Figure 3. Everything can be tested and further described in the Protocol to the topic, also with all design flaws which students find.



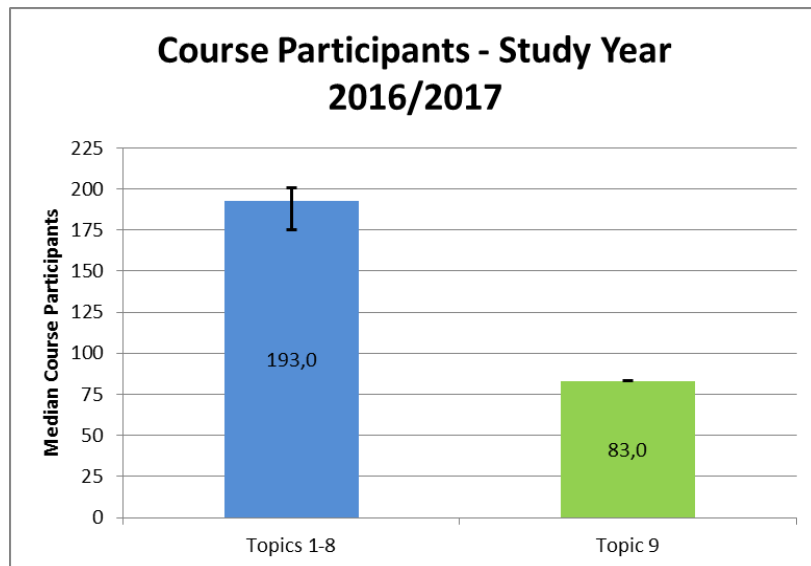
**Fig. 3.** Final printed example – jaw (red) and the manual fitting test with the dental crown.  
Source: own

The time structure of the course may be changed, because the duration of the measuring lab class is between 50-150 min and the teacher need to know the schedule in advance. The printing of the element most affects the time course index (Fig.4). There could be set the same printing position method. This however leads to standardization of the task, because the same system of the position for printing of the crowns has similar material properties. Individual solutions so slightly lost.



**Fig. 4.** Optional printed position of crowns set in the CreationWorkshop sw of the 3D printer.  
Printing is performed by layers.  
Source: own

To compare the number of participants for the topic 9 – 3D technologies and other topics were selected only accesses into theoretical parts – books in all topics, because there is no test in optional innovative topic. Median approach in the course of all topics was 193 total, 3D printing had 83 participants which is 43 % of students (Fig. 5). Even though the course was optional, 83 students signed up, which can be regarded as high number of participants and it clearly shows that students want this type of education. Rating has not even closed, as the course is open to students into late September 2017. In any case, we can say that the course has been completed successfully (credit) for more than 90 % of students and the number of users will not significantly change.



**Fig. 5.** Comparison of cumulative activities (participants access) vs the access into the innovative topic 3D Technologies (Topic 9).

Source: own

All results as design of the implants and crowns, measurement of the rigidity are archived and may be later analysed. Because the topic of the 3D printing technology was implemented into the course last year there are still not enough data for more detailed statistical analyses.

## CONCLUSION

The innovative 3D Technology course combines the E-learning and distance method of the theoretical preparation with the practical use of the advanced modern design, production and measurement methods of the implants, artificial bone structures and dental crowns. The course is innovated especially for the student in dentistry and teaches students to use their theoretical knowledge to the design of real solutions that can be “touched”. Teamwork and discussion of the final prototype simulate the environment for their future careers. Students significantly improve their knowledge of the biophysics by entertaining and modern form of E-learning MSL teaching.

## References

- [1] Hanuš, J., Nosek, T., Záhora, J., *et al.* On-line integration of computer controlled diagnostic devices and medical information systems in undergraduate medical physics education for physicians. *Physica Medica-European Journal of Medical Physics*, vol. 29, no. 1, 2013, p. 83–90. ISSN 1120-1797.
- [2] Hanuš, J., Záhora, J., Mašín, V., *et al.* On-Line Incorporation of Study and Medical Information System in Undergraduate Medical Education. In: *6th International Conference of Education, Research and Innovation (iceri 2013). Proceedings*, Seville, Spain, 2013, p. 1500–1507. ISBN 978-84-616-3847-5
- [3] Záhora, J., Hanuš, J., Jezbera, D., *et al.* Remotely Controlled Laboratory and Virtual Experiments in Teaching Medical Biophysics. In: *6th International Conference of Education, Research and Innovation (iceri 2013). Proceedings*, Seville, Spain, 2013, p. 900–906. ISBN 978-84-616-3847-5

- [4] Záhora, J., Bezrouk, A., Hanuš, J. Models of stents - Comparison and applications. *Physiological Research*, vol. 56, 2007, p. 115–121. ISSN 0862-8408.
- [5] Bezrouk, A., Balský, L., Smutný, M., et al. Thermomechanical properties of nickel-titanium closed-coil springs and their implications for clinical practice. *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 146, no. 3, 2014, p. 319–327. ISSN 0889-5406.
- [6] Kordek, D. Statistical Analysis of Subconscious Human Behaviour. In: *APLIMAT 2009: 8TH INTERNATIONAL CONFERENCE. Proceedings*, Bratislava, Slovakia, 2009, p. 783–789.
- [7] Jezbera, D., Kordek, D., Kříž, J., et al. Walkers on the circle. *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2010, no. 01, 2010. [Online]. [Cit. 2017-02-21]. Available at: doi:10.1088/1742-5468/2010/01/L01001 ISSN 1742-5468
- [8] Kordek, D. The definition of optical systems aberrations to secondary school students regarding their knowledge of mathematics. In: *AIP Conference Proceedings*, vol. 1804, 2017, p.030004-1 - 030004-6. Available at: <http://doi.org/10.1063/1.4974375>. ISBN: 9780735414723
- [9] CHua, Ch. K., Yeong, W. Y. Bioprinting: principles and applications. Singapore: *World Scientific Publishing*, 2015. ISBN 981-4612103.
- [10] Navrátil, V. Yield point phenomena in metals and alloys. In: *Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers*. Brno: University of Defence, 2016, p. 62-70. [Online]. [Cit. 2017-07-26]. Available at: <[http://mitav.unob.cz/data/MITAV 2016 Proceedings.pdf](http://mitav.unob.cz/data/MITAV_2016_Proceedings.pdf)>. ISBN 978-80-7231-400-3.

## Acknowledgement

The work presented in this paper has been supported by the PROGRES Q40-09 and SVV-2016-260287.

# THE LMS MOODLE AND THE MOODLE MOBILE APPLICATION IN EDUCATIONAL PROCESS OF BIOPHYSICS

David Kordek<sup>1</sup>, Martin Kopeček<sup>1</sup>, Kristýna Čáňová<sup>2</sup>, Klára Habartová<sup>3</sup>, Monika Pospíšilová<sup>3</sup>

<sup>1</sup>Department of Medical Biophysics, <sup>2</sup>Department of Medical Biology and Genetics,

<sup>3</sup>Department of Medical Biochemistry, Faculty of Medicine in Hradec Králové, Charles University

Šimkova 870, 500 03 Hradec Králové, Czech Republic

kordekd@lfhk.cuni.cz

**Abstract:** *The aim of the paper is to acquaint the readers with the process of teaching medical biophysics at the Faculty of Medicine in Hradec Králové. In its introduction, the paper describes the biophysics teaching process at our faculty, including the Moodle access statistics. The reader is also acquainted with the structure of the Moodle “LFHK” (Faculty of Medicine in Hradec Králové) web portal at moodle.lfhk.cuni.cz. The main part of the paper subsequently addresses the Moodle Mobile app and the possibilities of its use in education. There is also a detailed description of the interactive accessory of the application, which was created by our IT team in order to get feedback from the users (students). At the end, the advantages and disadvantages of the mobile application are evaluated and it is compared with the LMS Moodle version intended for PCs.*

**Keywords:** e-learning, Moodle “LFHK”, Moodle Mobile app, e-learning courses, students.

## INTRODUCTION

At first it's necessary to mention the growing influence of mobile devices (e.g. tablets, mobile phones, e-readers,...) in teaching. In general this increase is most evident in lower years of schools. The above mentioned increase is also related to the increased interest of the society in mobile technology and related mobile applications. With regards to education, e-learning is becoming more and more popular. E-learning is defined e.g. in [1]. A number of software tools is used to create e-learning courses as, e.g. WebCT, Blackboard, Adobe Connect, etc. [2]. There are more forms of e-learning and it's not the aim of this article to divide them and characterize them. In this contribution we will only concentrate on the LMS (Learning management system) Moodle and its mobile application Moodle Mobile. Currently, the system Moodle constitutes more than a half of all the installations of LMS (Learning Management System) systems in the world [3]. It's an application, that contains some online tools for lessons organization and communication with the students (e.g. chat, forum, news,...) and it also includes some components, that enable to get feedback about students' knowledge and attitudes (e.g. questionnaire, test, survey, ...). It also makes the studying materials available for students (e.g. book, lecture, ...). The students can hand in their homework as well. The basic unit is a Moodle course, that can include the above mentioned components. If a quality course is created (not only from the point of view of the content), the participation of the pedagogical staff is not necessary when filling in. Our faculty, especially the Department of Medical Biophysics has many years of experience with e-learning as it's obvious from the papers [4], [5], [6]. At the faculty e-learning in Moodle is carried out at the address:

moodle.lfhk.cuni.cz. The students have access to all e-learning courses at this address. In these courses there are e.g. interactive manuals for practical exercises, in which the scientific and didactic attitudes to the given problem are combined. As an example e.g. laboratory assignment of measuring rigidity of a nitinol stent, where the theory of this assignment is based on [7], [8], [9], [10].

## **1 THE MOODLE “LFHK” PORTAL**

The Moodle “LFHK” portal is divided into several basic categories: Czech courses, English courses, Preparatory course, the Dean's Advisory board, Study Division, etc. The Czech and English courses are important for medical students. In both categories there are subcategories called according to individual workplaces in alphabetical order. The workplaces are in charge of managing the categories and its courses. However the Moodle “LFHK” is not optimized for mobile devices and thus is not really appropriate e.g. for the use on a mobile phone, especially because of a long list of workplaces, that can be further divided into subcategories. The student can see all categories and courses including those, that they don't need yet. If the students want to use the courses in a comfortable and effective way in a mobile device, it's better to use the Moodle Mobile app.

## **2 TEACHING OF MEDICAL BIOPHYSICS**

The aim of the report is to present to the reader the learning process of biophysics at the Faculty of Medicine in Hradec Kralove, particularly to acquaint the reader with the Moodle Mobile app. The “Biophysics and Biostatistics” course at the Department of Medical Biophysics is implemented as a combination of traditional forms of teaching and e-learning. As indicated above, the e-learning system is conducted using LMS Moodle. This system allows students to both acquire information and complete tasks. All of the information regarding the subject “Biophysics and Biostatistics” is therefore available to the students on the Moodle “LFHK” in two courses: Biofyzika a biostatistika – všeobecné lékařství 2016/2017 (Biophysics and Biostatistics – General Medicine 2016/2017), and Biofyzika a biostatistika – zubní lékařství 2016/2017 (Biophysics and Biostatistics – Dentistry 2016/2017). These courses also involve handing\_in the assignments and taking exams. There are also instructions for practical exercises, which can be used as a walkthrough (without the presence of the teacher) prior to the actual workshop. In addition to this obligatory course, which each student has to sign up for, there are also various optional supplementary courses. These courses often involve a lecture activity, offering various structured branching and multi-choice questions. The combination of these elements allows students to take a course entirely on their own, simulating an actual class to the greatest possible extent. The aforementioned obligatory course is among the 3 courses with the highest student activity in Moodle “LFHK” during the past term, as shown in Table 1.

The highest number of views is shown for the obligatory activities of the course, such as credit exams, online exams for practical exercises, presentation submissions, etc. Instructions for practical exercises, also prepared in the form of e-learning courses, represented approximately 2/3 as many views in comparison to the obligatory parts of the course, as shown in Table 2.



Course full name	Number of activity
<i>Seminář z lékařské biofyziky 2016/2017</i> (Medical biophysics seminar-2016/2017)	58735
<i>Doporučené postupy pro VPL AH</i> (ESH-ESC Guidelines for AH)	56037
<i>Biofyzika a biostatistika – všeobecné lékařství 2016/2017</i> (Biophysics and Biostatistics – General Medicine 2016/2017)	47876

**Tab. 1.** Courses with the highest activity (winter term 2016/17)

Source: moodle.lfhk.cuni.cz

The highest number of views is shown for the obligatory activities of the course, such as credit exams, online exams for practical exercises, presentation submissions, etc. Instructions for practical exercises, also prepared in the form of e-learning courses, represented approximately 2/3 as many views in comparison to the obligatory parts of the course, as shown in Table 2.

Activity	Views / user
Test 1 – Statistics	1290 / 187
Test 2 – Biophysics	1449 / 180
CT instruction	545 / 139
Microscopy instruction	810 / 167
ECG + BP instruction	460 / 135
Ultrasound instruction	513 / 128

**Tab. 2.** View of selected activities in the course *Biofyzika a biostatistika všeobecné lékařství 2016/2017* (Biophysics and Biostatistics – General Medicine 2016/2017)

Source: moodle.lfhk.cuni.cz

From the described facts and the steady increase of views of Moodle “LFHK”, the affinity of students to use LMS Moodle to study is obvious.

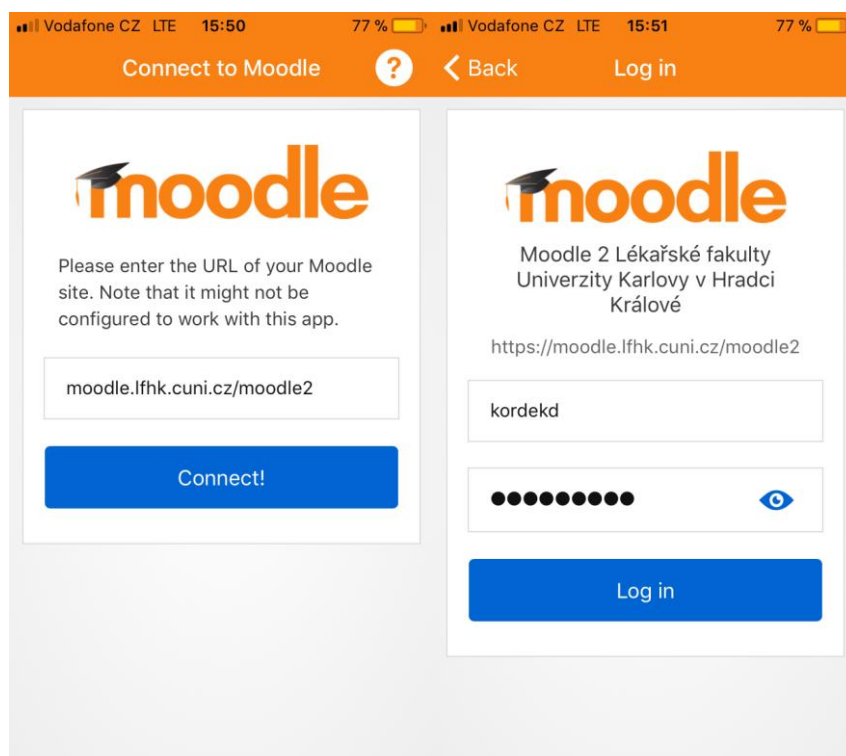
Year	Number of view of Moodle “LFHK”
2015	194 693
2016	201 788
2017	238 099

**Tab. 3.** Number of views of Moodle “LFHK” in the period from the 20<sup>th</sup> of February to the 20<sup>th</sup> of March

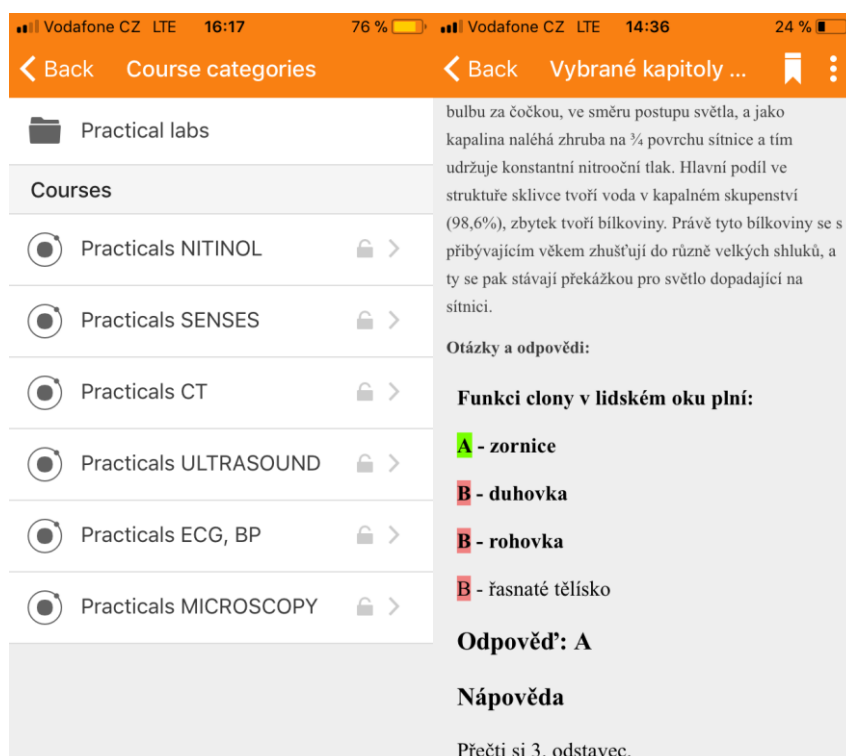
Source: moodle.lfhk.cuni.cz

### 3 THE MOODLE MOBILE APP

The Moodle Mobile app is an application, that is very well optimized for Android OS and iOS. The student can get the application for free through Google Play or App Store. Or it's possible to get access to the application at the link: <https://download.moodle.org/mobile/>, where is also some general information about the application. Installation of the Moodle Mobile app is a standard installation for both supported OS. After successfully installing the application into the mobile device the student must fill in the webpage address moodle.lfhk.cuni.cz/moodle2 on the initial screen of the application and then click on “connect” as can be seen in Figure 1a. As the next step the students must introduce their username, that they use to log in the web Moodle and their password, as in Figure 1b.



**Fig. 1.** a. Connecting to the Moodle “LFHK” web server, b. Introducing the login information in the Moodle Mobile app  
Source: own



**Fig. 2.** a. Virtual library, b. Interactive complement of the “book”  
Source: own

The student can see his/her grades in all activities of the given course and can also contact other participants of the course as can be seen in Figure 2a. If the study material “book” is part of the course, the student can see this component Moodle directly in the application.

Internet connection is necessary for initial loading of the book. However the advantage of the application is, that when reloading the book in the given course, the student doesn't need the Internet connection anymore. The study materials in the form of a book are available in the students' mobile device. As a part of the module "book" the authors of the courses offered a created complement, that enables to include multiple choice questions in the book, that the student can answer in an interactive way anywhere in the text of the book. An example of this complement is available in Figure 2b. This complement is since used in courses, that are created in the Czech language. The version Moodle 3.1 enables to fill in the tests in individual courses in the Moodle Mobile app. There was not this function in the previous versions and thus the student was redirected to the web Moodle "LFHK", where filling the test on a mobile device is not as optimized as in the application. And so this possibility can be considered another undeniable advantage of the Moodle Mobile app. It's not possible to attend a lecture in this application. The lecture is among the activities in Moodle, in which the presence of the teacher is not necessary, so the student can go through the whole topic including feedback. Unfortunately the activity "lecture" needs JavaScript, that is forbidden in the Moodle Mobile app. That's why it is not possible to attend the activity "lecture" directly in the application. This lack can be partially solved by using the mentioned material "book" with the interactive feedback, that we have created.

## CONCLUSION

In conclusion, it is appropriate to point out the advantages and disadvantages of the mentioned application, especially with regard to possible usage when studying. Generally speaking, Moodle Mobile is not for obvious reasons meant to be used to create courses, but rather to view already finished courses. Once a course for students is created in Moodle, each course can be viewed on mobile devices in the Moodle Mobile app. The courses may not be displayed properly as they are not optimized for the application. Problems may occur due to font size and data size of the images inserted into the studying materials – book, chapter titles, etc. As was already mentioned, a book is downloaded into the device for offline use, which means that images should be inserted in the text in a suitable resolution. Large files should be attached separately to the book or directly to the text body. The major disadvantage of Moodle Mobile is, that it doesn't enable the students to attend a "lecture" activity. In such case, the student needs to use the web version of Moodle, which is usually not optimised for mobile devices. As was mentioned above, the application allows users to complete exams. "Books" allow reading study material in offline mode as well. This means that the main advantage of the application is the ability to create custom virtual "libraries", even in the offline mode, providing students access to study materials in places where it was previously impossible. The own contribution of the team of authors to the issue of e-learning, that they deal with in the article is the proposal and implementation of the new concept of biophysics teaching as a combination of e-learning and contact teaching. The contribution presents a new point of view of this form of teaching, especially with regards to the Moodle Mobile Application (mobile solution of the Moodle application) and describes the unique innovation of this application created precisely by the team of authors for the needs of education at the Faculty of Medicine.

## References

- [1] Průcha, J., Walterová, E., Mareš, J. Pedagogický slovník, Prague: Portal, pp. 395, 2009, ISBN: 9788073676476.
- [2] Feberová, J., Dostálová, T., Hladíková, M. et al. Evaluation of 5-year Experience with E-learning Techniques at Charles University in Prague. Impact on Quality of Teaching and Students' Achievements. New Educ. Rev., vol. 21, no. 2, pp. 110-120, 2010.
- [3] Minovic, M., Stavljanin, V., Milovanovic, M. et al. Usability issues of e-learning systems: case-study for Moodle learning management system. On the Move to Meaningful Internet Systems: OTM 2008 Workshops, pp. 561-570, Nov. 2008. [http://dx.doi.org/10.1007/978-3-540-88875-8\\_79](http://dx.doi.org/10.1007/978-3-540-88875-8_79).
- [4] Hanuš, J., Nosek, T., Záhora, J. et al. On-line integration of computer controlled diagnostic devices and medical information systems in undergraduate medical physics education for physicians. Physica Medica-European Journal of Medical Physics, vol. 29, no. 1, 2013, p. 83–90. ISSN 1120-1797.
- [5] Hanuš, J., Záhora, J., Mašín, V. et al. On-Line Incorporation of Study and Medical Information System in Undergraduate Medical Education. In: 6th International Conference of Education, Research and Innovation (iceri 2013). Proceedings, Seville, Spain, 2013, p. 1500–1507. ISBN 978-84-616-3847-5.
- [6] Záhora, J., Hanuš, J., Jezbera, D. et al. Remotely Controlled Laboratory and Virtual Experiments in Teaching Medical Biophysics. In: 6th International Conference of Education, Research and Innovation (iceri 2013). Proceedings, Seville, Spain, 2013, p. 900–906. ISBN 978-84-616-3847-5.
- [7] Bezrouk, A., Balský, L., Smutný, M. et al. Thermomechanical properties of nickel-titanium closed-coil springs and their implications for clinical practice. American Journal of Orthodontics and Dentofacial Orthopedics, vol. 146, no. 3, 2014, p. 319–327. ISSN 0889-5406.
- [8] Záhora, J., Bezrouk, A., Hanuš, J. Models of stents - Comparison and applications. Physiological Research, vol. 56, 2007, p. 115–121. ISSN 0862-8408.
- [9] Bezrouk, A., Balský, L., Selke-Krulichová, I. et al., Nickel-titanium closed-coil springs: evaluation of the clinical plateau. Rev. Chim., vol. 68, no.5, pp. 1137–1142.
- [10] Navrátil, V. Yield point phenomena in metals and alloys. In: Mathematics, Information Technologies and Applied Sciences 2016, post-conference proceedings of extended versions of selected papers. Brno: University of Defence, 2016, p. 62-70. [Online]. [Cit. 2017-07-26]. Available at: <[http://mitav.unob.cz/data/MITAV\\_2016\\_Proceedings.pdf](http://mitav.unob.cz/data/MITAV_2016_Proceedings.pdf)>. ISBN 978-80-7231-400-3.

## Acknowledgement

The work presented in this paper has been supported by the project “Creating of multi-platform systems for Education support including tools for user friendly support”.

# On the theorem by Estrada and Kanwal

Ladislav Mišík

University of Ostrava,  
30. dubna 22, Ostrava

Email: ladislav.misik@osu.cz

**Abstract:** Let  $(x_n)$  be a sequence of positive real numbers such that the corresponding series  $\sum_{n=1}^{\infty} x_n$  diverges. Then, intuitively, its subseries along “small” sets of indices converge, while subseries along “large” sets of indices diverge. Extending the known results by Estrada and Kanwal, we will present that there are also some very small sets of indices along which the subseries diverges. On the other hand, we will show that these kind of results can not be strengthened in some natural direction.

**Keywords:** Divergent series, subseries, asymptotic density, lacunary sets

## Introduction

We will start with recalling some of the most frequently used characterizations of small subsets of positive integers. For a set  $A \subset \mathbb{N}$  and  $n \in \mathbb{N}$  denote by  $A(n)$  the number of elements of the set  $A \cap \{1, 2, \dots, n\}$  and define

$$\underline{d}(A) = \liminf_{n \rightarrow \infty} \frac{A(n)}{n}, \quad \bar{d}(A) = \limsup_{n \rightarrow \infty} \frac{A(n)}{n},$$

the *lower* and *upper asymptotic density* of  $A$ , respectively. If both these values equal, we denote the common value by  $d(A)$  and call it *asymptotic density* of  $A$ . Note that sets of asymptotic density 0 play in number theory a similar role as sets of measure 0 in analysis do.

Further, for  $A \subset \mathbb{N}$  and positive integers  $n$  and  $k$ , denote by  $A(n, k)$  the number of elements of  $A$  in the interval  $(n, n + k]$  and define

$$\underline{b}(A) = \lim_{k \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{A(n, k)}{n}, \quad \bar{b}(A) = \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{A(n, k)}{n},$$

the *lower* and *upper Banach density* of  $A$ , respectively. If both these values equal, we denote the common value by  $b(A)$  and call it *Banach density* of  $A$ . Note that

$b(A) = 0$  readily implies  $d(A) = 0$ , thus every set of Banach density 0 is also a set of asymptotic density 0, but the reverse implication does not hold. A set  $A = \{a_1 < a_2 < \dots\} \subset \mathbb{N}$  is *lacunary* if  $\lim_{n \rightarrow \infty} (a_{n+1} - a_n) = \infty$ . Denoting by  $\mathcal{Z}_1, \mathcal{B}_0, \mathcal{L}$  families of all sets of asymptotic density 0, all sets of Banach density 0 and all lacunary sets, respectively, we have the following inclusions

$$\mathcal{L} \subset \mathcal{B}_0 \subset \mathcal{Z}_1. \quad (1)$$

For every  $\alpha > 0$  put

$$\mathcal{Z}_\alpha = \left\{ A \subset \mathbb{N}; \lim_{n \rightarrow \infty} \frac{(A(n))^\alpha}{n} = 0 \right\}.$$

Note that for  $0 < \alpha < \beta$  the inclusion  $\mathcal{Z}_\beta \subset \mathcal{Z}_\alpha$  holds and  $\mathcal{Z}_1$  is the family of all sets of null asymptotic density.

In [1] the following beautiful theorem is proved.

**Theorem 1** (Estrada - Kanwal) *Let  $\sum_{n \in \mathbb{N}} x_n$  be a series with positive terms. Then the series converges if and only if for every set  $A \in \mathcal{Z}_1$  the subseries  $\sum_{n \in A} x_n$  converges.*

Inspired by the above theorem, let us say that a family  $\mathcal{F}$  is *potent* if the following condition holds for every series with positive terms.

$$\sum_{n \in \mathbb{N}} x_n < \infty \quad \text{if and only if} \quad \sum_{n \in F} x_n < \infty \quad \forall F \in \mathcal{F}. \quad (2)$$

Later in [2] authors simplified the proof of the Estrada - Kanwal theorem and added its negative counterpart.

**Theorem 2** *Let  $\alpha > 0$ . Then the family  $\mathcal{Z}_\alpha$  is potent if and only if  $\alpha \leq 1$ .*

A function  $m: \mathcal{P}(\mathbb{N}) \rightarrow [0, \infty)$  is said to be a *submeasure* if for all pairs  $A, B \subset \mathbb{N}$  the relations

$$A \subset B \quad \Rightarrow \quad m(A) \leq m(B) \quad (i)$$

and

$$m(A \cup B) \leq m(A) + m(B) \quad (ii)$$

hold. Let us denote by  $\mathcal{Z}(m)$  the set of all sets  $A \subset \mathbb{N}$  such that  $m(A) = 0$ . A submeasure  $m$  is *compact* if it possesses also the following two conditions.

$$m(\{a\}) = 0 \quad \text{for every } a \in \mathbb{N} \quad (\text{iii})$$

and for every  $\varepsilon > 0$  there exists a finite decomposition  $A_1 \cup A_2 \cup \dots \cup A_k$  of  $\mathbb{N}$  such that for all  $i = 1, 2, \dots, k$

$$m(A_i) < \varepsilon. \quad (\text{iv})$$

The following very nice and strong generalization of the Estrada - Kanwal theorem was proved in [3].

**Theorem 3** *Let  $m$  be a compact submeasure. Then  $\mathcal{Z}(m)$  is a potent family.*

Finally, let us mention two generalizations of the Estrada - Kanwal theorem in [4] and [5], see also [6] and [7].

**Theorem 4** *Let  $\mathcal{Z}$  be the set of all  $A \subset \mathbb{N}$  such that  $b(A) = 0$ . Then  $\mathcal{Z}$  is a potent family.*

**Theorem 5** *Let a positive sequence of weights  $(c_n)$  fulfill*

$$\sum_{n \in \mathbb{N}} c_n = \infty \quad (\text{d})$$

and

$$\sum_{n \in \mathbb{N}} |c_{n+1} - c_n| < \infty. \quad (\text{v})$$

*Then the set of all sets of null density with respect to  $(c_n)$  is a potent family.*

## 1 Some new results

### 1.1 Positive results

Now we are going to present a new proof of the Estrada - Kanwal theorem. This proof is simpler and more straightforward than those published in [1] and [2]. Moreover, its slight modification yields a more general result.

**Proof** (Estrada - Kanwal theorem) Let  $\sum_{n \in \mathbb{N}} x_n$  be a series with positive terms. First we define by induction an increasing sequence of positive integers  $(n_k)$

as follows. Put  $n_0 = 0$  and let  $n_1$  be the smallest positive integer such that  $\sum_{n=1}^{n_1} x_n > 1$ . Suppose that also  $n_2, \dots, n_{k-1}$  have already been defined. Let  $n_k$  be the smallest positive integer greater than  $n_{k-1}$  such that  $n_k - n_{k-1}$  is divisible by  $k$  and  $\sum_{n=n_{k-1}+1}^{n_k} x_n > 1$ . Now we will construct a set  $J \subset \mathbb{N}$  by induction determining  $J \cap (n_{k-1}, n_k]$  for arbitrary  $k \in \mathbb{N}$  as follows. Decompose each interval  $(n_{k-1}, n_k]$  into groups of exactly  $k$  consecutive numbers

$$\{n_{k-1}+1, \dots, n_{k-1}+k\}, \{n_{k-1}+k+1, \dots, n_{k-1}+2k\}, \dots, \{n_k-k+1, \dots, n_k\}$$

and pick up the only one element from each group into the set  $J$  by the following rule:

$$J \cap \{n_{k-1} + ik + 1, \dots, n_{k-1} + (i+1)k\} = \{j_i^k\}, \quad (3)$$

where  $x_{j_i^k} \geq x_s$  for all  $s \in \{n_{k-1} + ik + 1, \dots, n_{k-1} + (i+1)k\}$ . From this choice we have immediately  $\sum_{j \in J \cap (n_{k-1}, n_k]} x_j \geq \frac{1}{k}$  for every  $k \in \mathbb{N}$ . Consequently

$$\sum_{j \in J} x_j = \sum_{k \in \mathbb{N}} \sum_{j \in J \cap (n_{k-1}, n_k]} x_j \geq \sum_{k \in \mathbb{N}} \frac{1}{k} = \infty$$

Moreover, the construction yields directly  $J(n, k) \leq 2$  for all sufficiently large  $n$  and  $k$ , thus  $b(J) = 0$ , consequently  $d(J) = 0$  follows and the theorem is proved. Let us remark that from the above proof also the stronger result than Theorem 4 follows directly. Also note that by a slight modification of the proof we obtain the following stronger result.

**Theorem 6** *The family  $\mathcal{L}$  of all lacunary sets is potent.*

In fact, it is sufficient to modify the rule (3) in the above proof only at places where the distance  $j_{i+1}^k - j_i^k$  of elements  $j_i^k$  and  $j_{i+1}^k$  picked from two consecutive intervals  $\{n_{k-1} + ik + 1, \dots, n_{k-1} + (i+1)k\}$  and  $\{n_{k-1} + (i+1)k + 1, \dots, n_{k-1} + (i+2)k\}$  is less than  $k$ . In this case we remove from  $J$  the index with smaller value of  $x$ .

## 1.2 Negative results

First, let us mention the following negative result generalizing that of [2]. It says that the Estrada - Kanwal theorem is optimal in some sense.



**Theorem 7** Let  $f: \mathbb{N} \rightarrow \mathbb{N}$  be such that  $\lim_{n \rightarrow \infty} f(n) = \infty$  and  $\lim_{n \rightarrow \infty} \frac{f(n)}{n} = 0$  and

$$\mathcal{Z}_f = \{A \subset \mathbb{N}; \exists n_0 \forall n > n_0 \quad A(n) \leq f(n)\}.$$

Then  $\mathcal{Z}_f$  is not a potent family.

**Proof** Let  $f: \mathbb{N} \rightarrow \mathbb{R}^+$  be such that

$$\lim_{n \rightarrow \infty} f(n) = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{f(n)}{n} = 0. \quad (4)$$

We are going to construct a divergent series  $\sum_{n=1}^{\infty} x_n$  of positive terms such that  $\sum_{n \in A} x_n$  is convergent for every  $A \in \mathcal{Z}_f$ . First, let  $(k_n)$  be an increasing integer sequence such that  $k_1 = 1$  and the inequality

$$\frac{f(k_{n+1})}{k_{n+1} - k_n} < \frac{1}{(n+1)^2} \quad (5)$$

holds for every  $n = 1, 2, \dots$ . Put  $x_1 = 1$  and for every positive integer  $i \in (k_n, k_{n+1}]$  define  $x_i = \frac{1}{k_{n+1} - k_n}$  for each  $n = 1, 2, \dots$ . Then

$$\sum_{i \in \mathbb{N}} x_i = 1 + \sum_{n=1}^{\infty} \sum_{i \in (k_n, k_{n+1}] \cap \mathbb{N}} x_i = \sum_{n=1}^{\infty} 1 = \infty.$$

On the other hand, let  $J \in \mathcal{Z}_f$  be arbitrary. For simplicity we can assume that the condition  $J(n) \leq f(n)$  holds for all  $n \in \mathbb{N}$ , as the convergence or divergence of series does not depend on finite number of indices. Then

$$\begin{aligned} \sum_{j \in J} x_j &\leq x_1 + \sum_{n=1}^{\infty} \sum_{j \in (k_n, k_{n+1}] \cap J} x_j = 1 + \sum_{n=1}^{\infty} (J(k_{n+1}) - J(k_n)) \frac{1}{k_{n+1} - k_n} \leq \\ &\leq 1 + \sum_{n=1}^{\infty} f(k_{n+1}) \frac{1}{k_{n+1} - k_n} < \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \end{aligned}$$

what finishes the proof.

Note that the result of Theorem 2 for  $\alpha > 1$  easily follows from the above general theorem taking  $f(n) = n^{\frac{1}{\alpha}}$ . In addition, the following stronger corollary following from the previous theorem for  $f(n) = n(\ln n)^{-\alpha}$  was observed by G. Grekos.

**Corollary 1** Let  $\alpha > 1$  and

$$\mathcal{S} = \left\{ A \subset \mathbb{N}; \frac{A(n)}{n(\ln n)^{-\alpha}} \rightarrow 0 \right\}.$$

Then  $\mathcal{S}$  is not a potent family.

## 2 Conclusion

Estrada - Kanwal theorem says that divergence of a series with positive terms can be discovered by inspecting all its subseries along sets of asymptotic density zero. We have seen that there are some possible extensions of this theorem related to inspection of some subclasses of sets of density zero. On the other hand, whenever one limits the grow of the number of elements in these sets, the resulting class fails to be able to discover the divergence of all series with positive terms.

## Reference

- [1] Estrada R., Kanwal R.P.: *Series that converge on sets of null density*, Proc. Amer. Math. Soc. 97, No.4, 1986. p.682 - 686.
- [2] Greenberg P., Reséndis L., Rivaud J.J.: *Convergencia absoluta y densidad asintotica*, Aportacionnes Matemáticas Comunicaciones 5, 1988. p.25 - 30.
- [3] Paštéka M.: *Convergence of series and submeasures of the sets of positive integers*, Mathematica Slovaca, 40, No. 3, 1990. p.273 - 278.
- [4] Paštéka M., Šalát T., Visnyai T.: *Remarks on Buck's measure density and a generalization of asymptotic density*, Tatra Mountains Mathematical Publications, 31, 2005. p.87 - 101.
- [5] Šalát T., Visnyai T.: *Subadditive measures on  $\mathbb{N}$  and the convergence of series with positive terms*, Acta Mathematica 6, 2003. p.43 - 52.
- [6] Visnyai, T.: *Remarks on compact submeasures*, Mathematics, Information Technologies and Applied Sciences 2015, p. 156 – 161.
- [7] Visnyai, T.: *Convergence of series along the sets from ideals*, In Balko, Ľ. – Szarková, D. – Richtáriková, D. Aplimat 2016: Proceedings of the 15th Conference on Applied Mathematics 2016. Bratislava, 2.-4. 2. 2016. Bratislava, STU, 2016, s. 1105. ISBN 978-80-227-4531-4.

# EL-SEMIHYPERGROUPS IN WHICH THE QUASI-ORDERING IS NOT ANTISYMMETRIC

Michal Novák

Faculty of Electrical Engineering and Communication, Brno University of Technology,  
Technická 8, Brno, Czech Republic  
novakm@feec.vutbr.cz

**Abstract:** *EL-hyperstructures are a class of hyperstructures constructed from quasi-ordered semigroups. Their construction – in its original version – uses partial ordering. However, antisymmetry is often not needed to achieve desired results. In this paper we focus on such cases. We also discuss implications of the quasi-ordering being moreover symmetric, i.e. an equivalence.*

**Keywords:** *EL-hyperstructures, equivalence, hyperstructure theory, quasi-ordered semigroups, partially ordered semigroup.*

## INTRODUCTION

The concept of *relation* is prominent not only in classical algebra but also in the algebraic hyperstructure theory. Just as in classical algebra, the hyperstructure theory studies preordered / partially ordered sets, hyperstructure generalizations of (semi)lattices, extensions of BCK-algebras, preordered / partially ordered semi(hyper)groups, etc.

The concept of *EL-hyperstructures* is one of the concepts proposed by Chvalina [2]. Unlike *quasi-order hypergroups* introduced in [1] and included in [2], *EL-hyperstructures* were not discussed in [5], which is often considered (together with [4, 9]) to be a canonical book on the algebraic hyperstructure theory. Also, since [1] was written in English while [2] in Czech only, the concept of *EL-hyperstructures* did not spread as widely as the concept of quasi-order hypergroups used e.g. in [3, 5, 7, 13, 14]. However, since the idea is rather natural, its traces can be found in some earlier works as well. Already in Pickett [22] can we find an example based on hyperoperation (1). Also Phanthawimol and Kemprasit [21] in fact work in the *EL*-context (on top of that using equivalencies). Notice that their hyperoperation was originally defined by Corsini [4].

Notice that while quasi-order hypergroups are sets  $(H, *)$ , where “ $*$ ” is a hyperoperation, which is defined by means of a quasi-ordering (i.e. preordering) “ $\leq$ ” on a set  $(H, \leq)$ , to construct *EL-hyperstructures* we need a *quasi-ordered* (i.e. preordered) *semigroup*. Originally, *EL-hyperstructures* were constructed ad hoc for suitable semigroups endowed with a suitable compatible quasi-ordering. Later, in [16, 18, 19, 20] a theoretical background was provided for the idea. Shortly after this, Anvariye, Ghazavi and Mirvakili [10, 11, 12] extended the construction and studied some particular aspects of it.

The construction of *EL-semihypergroups* is based on two theorems (quoted below as lemmas) included in [2]. In their original version they assume that the relation “ $\leq$ ” is reflexive, transitive and antisymmetric, i.e. that it is a *partial ordering*. However, for many of the results obtained so far antisymmetry of the relation is not needed, i.e. they are valid also if “ $\leq$ ” is a *quasi-ordering* (i.e. a *preorder*). Notice that the abbreviation *EL* stands for “Ends lemma”, the name of the construction

used because of the fact that for an arbitrary  $a \in H$  the set  $[a]_{\leq} = \{x \in H \mid a \leq x\}$  is an “upper end” generated by  $a$ , i.e. we consider all elements of  $H$  which are, in relation “ $\leq$ ”, “above”  $a$ .

**Remark 1** Throughout the paper we prefer saying “quasi-ordering” to “preorder”. Also, we prefer saying “partial ordering” to “partial order” or simply “ordering”.

## 1 CONSTRUCTION, EXAMPLES AND SETTING THE GROUND

To be exact, we include the theorems from [2] and – in order to identify the exact place where antisymmetry (and reflexivity and transitivity) is needed – we also include their proofs.

**Lemma 1** Let  $(S, \cdot, \leq)$  be a partially ordered semigroup. Binary hyperoperation  $*$  :  $S \times S \rightarrow \mathcal{P}^*(S)$  defined by

$$a * b = [a \cdot b]_{\leq} = \{x \in S \mid a \cdot b \leq x\} \quad (1)$$

is associative. The semihypergroup  $(S, *)$  is commutative if and only if  $(S, \cdot)$  is commutative.

*Proof.* Suppose  $a, b, c \in S$  arbitrary. First of all, it is useful to show that the following equality holds:

$$\bigcup_{t \in [b \cdot c]_{\leq}} [a \cdot t]_{\leq} = \bigcup_{x \in [a \cdot b]_{\leq}} [x \cdot c]_{\leq}.$$

Suppose therefore an arbitrary  $s \in \bigcup_{t \in [b \cdot c]_{\leq}} [a \cdot t]_{\leq}$ . This means that  $s \geq a \cdot t_0$  for a suitable  $t_0 \in S$ ,  $t_0 \geq b \cdot c$ . Then  $a \cdot t_0 \geq a \cdot (b \cdot c) = (a \cdot b) \cdot c$  and if we set  $x_0 = a \cdot b$ , we get that  $x_0 \cdot c \leq s$ ,  $x_0 \in [a \cdot b]_{\leq}$ , i.e.  $s \in [x_0 \cdot c]_{\leq} \subseteq \bigcup_{x \in [a \cdot b]_{\leq}} [x \cdot c]_{\leq}$ . The other inclusion may be proved in the analogous way. Now we get that

$$a * (b * c) = \bigcup_{t \in b * c} a * t = \bigcup_{t \in [b \cdot c]_{\leq}} [a \cdot t]_{\leq} = \bigcup_{x \in [a \cdot b]_{\leq}} [x \cdot c]_{\leq} = \bigcup_{x \in a * b} x * c = (a * b) * c,$$

which completes the proof of associativity. Obviously, if  $(S, \cdot)$  is commutative, then also  $(S, *)$  is commutative. On the other hand, if  $(S, *)$  is commutative, then for an arbitrary pair of elements  $a, b \in S$  we have that  $a * b = b * a$ , i.e.  $[a \cdot b]_{\leq} = [b \cdot a]_{\leq}$ , which means that  $a \cdot b \leq b \cdot a$  and simultaneously  $b \cdot a \leq a \cdot b$ , i.e. – given the fact that “ $\leq$ ” is a partial ordering – means that  $a \cdot b = b \cdot a$ .

**Definition 1** A semihypergroup constructed using Lemma 1 is called *EL-semihypergroup*. We also say that  $(S, *)$  is the *EL-semihypergroup* of  $(S, \cdot, \leq)$ . If  $(S, *)$  is a hypergroup, we call it *EL-hypergroup*.

One can see that the only place in the proof of Lemma 1, where antisymmetry of the relation “ $\leq$ ” is used, is one of the implications on commutativity, in which we prove that the associativity of the hyperoperation “ $*$ ” implies the associativity of the single-valued operation “ $\cdot$ ”. On the other hand, reflexivity and transitivity of “ $\leq$ ” are essential. Notice that thanks to reflexivity of “ $\leq$ ” we have that  $[a]_{\leq} \neq \emptyset$  for all  $a \in S$ , i.e. that  $(S, *)$  is a hypergroupoid (not a partial hypergroupoid). Compatibility of the relation “ $\leq$ ” and the single-valued operation “ $\cdot$ ” are essential in the proof too. In the following examples we construct *EL-semihypergroups* using an equivalence relation, i.e. a quasi-ordering which is moreover symmetric. By a *proper semi(hyper)group* we mean a semi-(hyper)group which is not a (hyper)group.

**Example 1** Let  $(\mathbb{N}, \cdot, \equiv)$  be the multiplicative semigroup of natural numbers and “ $\equiv$ ” the relation of congruence modulo  $m$ . Obviously,  $(\mathbb{N}, \cdot, \equiv)$  is a quasi-ordered proper semigroup and “ $\equiv$ ” is not antisymmetric. If we, for a fixed  $m \in \mathbb{N}$ , define that for arbitrary  $a, b \in \mathbb{N}$  that  $a * b = \{x \in \mathbb{N} \mid a \cdot b \equiv x \pmod{m}\}$ , then  $(\mathbb{N}, *)$  is an  $EL$ -semihypergroup.

**Example 2** On the set  $\mathbb{C}$  of all complex numbers regard a binary operation “ $\cdot_{|z|}$ ” defined as multiplication of absolute values, i.e. for all  $z_1, z_2 \in \mathbb{C}$  define  $z_1 \cdot_{|z|} z_2 = |z_1| \cdot |z_2|$  and a relation “ $\leq_{|z|}$ ” defined as equality of absolute values, i.e. for all  $z_1, z_2 \in \mathbb{C}$  put  $z_1 \leq_{|z|} z_2$  whenever  $|z_1| = |z_2|$ . Obviously,  $(\mathbb{C}, \cdot_{|z|}, \leq_{|z|})$  is a quasi-ordered semigroup (and “ $\leq_{|z|}$ ” is not antisymmetric, yet it is symmetric). Thus if we define, for all  $z_1, z_2 \in \mathbb{C}$ ,  $z_1 * z_2 = \{x \in \mathbb{C} \mid |z_1| \cdot |z_2| = |x|\}$ , we get that  $(\mathbb{C}, *)$  is an  $EL$ -semihypergroup.

In the proof of Lemma 1 we used the fact that  $[a]_{\leq} = [b]_{\leq}$  implies  $a = b$ . However, this is true only on condition of antisymmetry.

**Example 3** Regard the  $EL$ -semihypergroup  $(\mathbb{N}, *)$  constructed in Example 1 from  $(\mathbb{N}, \cdot, \equiv)$ , in which we set  $m = 3$ . In this case

$$[4]_{\equiv} = \{x \in \mathbb{N} \mid x \equiv 4 \pmod{3}\} = [7]_{\equiv},$$

yet the fact that  $[4]_{\equiv} = [7]_{\equiv}$  does not imply that  $4 = 7$ . Also, in Example 2, the fact that two complex numbers have the same absolute value does not mean that they are the same.

The following lemma from [2] is a tool to find out whether a quasi-ordered (in its original version quoted from [2], partially ordered) semigroup generates a proper semihypergroup or a hypergroup.

**Lemma 2** Let  $(S, \cdot, \leq)$  be a partially ordered semigroup. The following conditions are equivalent:

- 1<sup>0</sup> For any pair  $a, b \in S$  there exists a pair  $c, c' \in S$  such that  $b \cdot c \leq a$  and  $c' \cdot b \leq a$ .
- 2<sup>0</sup> The semi-hypergroup  $(S, *)$  defined by 1 is a hypergroup.

*Proof.*

1<sup>0</sup>  $\Rightarrow$  2<sup>0</sup>: Suppose  $t \in S$  arbitrary. Since  $t * S \subseteq S$  and  $S * t \subseteq S$  obviously holds, we will prove the converse inclusions. Suppose  $s \in S$  arbitrary. We assume that for the pair  $s, t \in S$  there exists a pair  $c, c' \in S$  such that  $t \cdot c \leq s$ ,  $c' \cdot t \leq s$ , i.e.

$$\begin{aligned} s \in [t \cdot c]_{\leq} \cap [c' \cdot t]_{\leq} &= (t * c) \cap (c' * t) \subseteq \left( \bigcup_{x \in S} t * x \right) \cap \left( \bigcup_{x \in S} x * t \right) = \\ &= (t * S) \cap (S * t), \end{aligned}$$

which means that  $S \subseteq t * S$  and  $S \subseteq S * t$ .

2<sup>0</sup>  $\Rightarrow$  1<sup>0</sup>: Suppose that  $(S, *)$  is a hypergroup and  $a, b \in S$  are arbitrary. Since there is  $b * S = S * b = S$ , there is

$$a \in b * S = \bigcup_{t \in S} b * t = \bigcup_{t \in S} [b \cdot t]_{\leq},$$

which means that  $a \in [b \cdot c]_{\leq}$  for a suitable element  $c \in S$ , i.e.  $b \cdot c \leq a$ . In an analogous way,  $a \in S * b$ , i.e.  $c' \cdot b \leq a$  for a suitable element  $c' \in S$ , which is 2<sup>0</sup>.

One can see that antisymmetry of the relation “ $\leq$ ” is not needed anywhere in the proof. *Thus we can conclude that Lemma 2 is valid for quasi-ordered semigroups  $(S, \cdot, \leq)$  as well.*

Let us now include two important special cases which fulfill the condition of Lemma 2. The first of them can be found in [2] (for partially ordered groups) while the other one, included below as Theorem 1, has not been mentioned yet (even though it is truly trivial).

**Corollary 1** *If, in Lemma 2,  $(S, \cdot, \leq)$  is a quasi-ordered group, then  $(S, *)$ , constructed by means of (1), is a hypergroup.*

*Proof.* Obvious because it is sufficient to set  $c = b^{-1} \cdot a$ , which turns the condition  $b \cdot c \leq a$  of Lemma 2 into  $a \leq a$  which holds because “ $\leq$ ” is reflexive. In a similar way, we set  $c' = a \cdot b^{-1}$ . The corollary holds because we already know that Lemma 2 holds for quasi-ordered groups as well.

**Definition 2** *A hyperoperation “ $*$ ” on  $S$  is called extensive if, for all  $a, b \in S$ , there is  $\{a, b\} \subseteq a * b$ . A hypergroupoid  $(S, *)$  with an extensive hyperoperation is called an extensive hypergroupoid.*

**Theorem 1** *Every extensive  $EL$ -semihypergroup is a hypergroup.*

*Proof.* Obvious because extensivity in  $EL$ -semihypergroups means that  $a \cdot b \leq a$  for all  $a, b \in S$ . Thus it is sufficient to set  $c = c' = a$  and apply Lemma 2. Again, “ $\leq$ ” can be a quasi-ordering as well.

Alternatively, we could write that, if “ $\leq$ ” is extensive, then

$$a * b = \{a, b\} \cup [a \cdot b]_{\leq},$$

which means that the reproductive law  $a * S = S * a = S$  turns into

$$a * S = \bigcup_{b \in S} a * b = \bigcup_{b \in S} \{a, b\} \cup [a \cdot b]_{\leq} = S \cup [a \cdot b]_{\leq} = S$$

and likewise for  $S * a$ .

**Example 4**  *$EL$ -semihypergroups constructed from  $(S, \min, \leq)$ , where  $S \in \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\} \cup \langle a, b \rangle$ , where  $\langle a, b \rangle$  is an arbitrary interval of real numbers, and “ $\leq$ ” is the usual ordering of numbers by size (which is a partial ordering), are extensive.*

**Example 5** *The  $EL$ -semihypergroup constructed from quasi-ordered semigroup  $(\mathbb{N}, \gcd, |)$ , where “gcd” stands for the greatest common divisor of natural numbers and “ $|$ ” is the divisibility relation, is extensive.*

**Example 6** *If we in Example 4 change “min” to “max” or to “+”, then  $(S, *)$  are not extensive.*

Thus, one can expect that a great many results regarding  $EL$ -hyperstructures will be valid even in cases when the relation “ $\leq$ ” is not antisymmetric (yet maintains reflexivity, transitivity and compatibility with the single-valued operation “ $\cdot$ ”). As a result, we can consider relations “ $\leq$ ” which are equivalencies. Naturally, using the symbol “ $\leq$ ” for an equivalence is rather a misleading choice as “ $\leq$ ” suggests a partial ordering (just as “ $\preceq$ ” is reserved for proper quasi-ordering). In fact, we could use the general notation “ $aRb$ ” to suggest that  $a$  and  $b$  are in relation  $R$ . However,

we prefer either writing “ $a \leq b$ ” and specifying the relation (quasi-ordering, partial ordering, equivalence) or using standard symbols of a given type of relation such as “ $\equiv$ ” for congruence or “ $\subseteq$ ” for set inclusion. One could see that in the examples we have already given we also describe the relation in plain words wherever possible.

Naturally, the main obstacle with “downgrading” from partial ordering to quasi-ordering is the fact that we loose notions such as the *greatest* or the *lowest element* and encounter rather big difficulties when manipulating *maximal* or *minimal elements*. Notice that, in *EL*-hyperstructures, this is an important limitation because we work with sets  $[a]_{\leq} = \{x \in S \mid a \leq x\}$  and very often we need to prove that  $[a]_{\leq} = \{a\}$ , or rather that for no  $b \neq a$  there is  $b \in [a]_{\leq}$  or that if  $b \in [a]_{\leq}$ , then  $b = a$ . Even though the issue of maximal elements actually can be manipulated in quasi-ordered sets (for an example of use see [15]), in some contexts we can bypass the problem by using the following definition.

**Definition 3** Let  $(S, \leq)$  be a quasi-ordered set. An element  $a \in S$  such that  $[a]_{\leq} = \{x \in S \mid a \leq x\} = \{a\}$ , is called an *EL-maximal element*.

## 2 SOME RESULTS WHERE ANTISYMMETRY IS NOT NEEDED

In this section we gather a selection of few theorems, in which the antisymmetry of the binary relation “ $\leq$ ” is not needed. As a result, they are valid also in contexts in which “ $\leq$ ” is an equivalence relation. However, the implications of the symmetry of the relation must always be examined separately because symmetry of the relation might sometimes lead to trivialities.

First of all, we discuss the issue of *hyperstructure identities*, i.e. elements  $e \in S$  such that  $x \in x * e \cap e * x$  for all  $x \in S$ , where “ $*$ ” is a hyperoperation on  $S$ .

**Theorem 2** Let  $(S, *)$  be the *EL-semihypergroup* of a quasi-ordered monoid  $(S, \cdot, \leq)$  with the neutral element  $u$ . An element  $e \in S$  is an identity of  $(S, *)$  if and only if  $e \leq u$ .

*Proof.* “ $\Rightarrow$ ”: If  $e \in S$  is an identity of an *EL-semihypergroup*  $(S, *)$ , then there holds  $e \cdot a \leq a$  and  $a \cdot e \leq a$  for all  $a \in S$ . Specifically, this holds for  $a = u$ . In this case we get  $e \leq u$ .

“ $\Leftarrow$ ”: Suppose that  $e \leq u$ . Since  $(S, \cdot)$  is a quasi-ordered semigroup, this is equivalent to  $e \cdot a \leq a$  for any  $a \in S$ , which means that for any  $a \in S$  we have that  $a \in [e \cdot a]_{\leq} = e * a$ . In an analogous way we get that  $a \in a * e$ , i.e.  $e$  is an identity of  $(S, *)$ .

**Example 7** In Example 1,  $u = 1$ . Thus the set of hyperstructure identities of  $(N, *)$  from Example 1 is the set  $\{x \in \mathbb{N} \mid x \equiv 1 \pmod{m}\}$  for a given  $m \in \mathbb{N}$ .

It is also rather easy to describe all *hyperstructure inverses* in *EL*-hypergroups constructed from quasi-ordered groups. Recall that by a hyperstrucre inverse of  $a \in S$  we mean such an element  $a' \in S$  for which there exists a hyperstructure identity  $e \in S$  such that  $e \in a * a' \cap a' * a$ . Hypergroups in which to every element there exists a *unique* inverse are called *canonical*. In [20] one can find a proof that *EL-semihypergroups* cannot be canonical hypergroups. The following theorem is included in [20]; notice that  $i(a)$  is a notation reserved for the set of hyperstructure inverses of an element  $a \in S$ .

**Theorem 3** Let  $(S, *)$  be the  $EL$ -hypergroup of a quasi-ordered group  $(S, \cdot, \leq)$ . Then for an arbitrary  $a \in S$  there is

$$i(a) = \{a' \in S \mid a' \leq a^{-1}\} = (a^{-1})_{\leq},$$

where  $a, a^{-1}$  are inverses in  $(S, \cdot)$ .

*Proof.* In order to prove the theorem, we have to prove the following implications:

1. If  $a' \leq a^{-1}$ , then  $a'$  is an inverse of  $a$  in  $(S, *)$ .  
 Suppose that  $a' \leq a^{-1}$ . This means that  $a' \cdot a \leq a^{-1} \cdot a = u$ , where  $u$  is the neutral element of  $(S, \cdot)$ . It does not matter whether we multiply from the left or from the right. Since  $u$  is an identity of  $(S, *)$ ,  $a'$  is an inverse of  $a$  in  $(S, *)$ .
2. If  $a' \in S$  is an inverse of  $a$  in  $(S, *)$ , then  $a' \leq a^{-1}$ .  
 Since  $a, a' \in S$  are inverses in  $(S, *)$ , there exists an identity  $e \in S$  such that  $e \in a * a' \cap a' * a$ . This means that there simultaneously holds  $a \cdot a' \leq e$  and  $a' \cdot a \leq e$ . Denote  $u$  the neutral element of  $(S, \cdot)$ . Since from Theorem 2 there follows that  $e \leq u$ , and since “ $\leq$ ” is transitive, we altogether get that  $a \cdot a' \leq u$  and  $a' \cdot a \leq u$ , which implies  $a' \leq a^{-1}$ .

In other words, if  $(S, \cdot)$  is a group and “ $\leq$ ” is an equivalence relation, then  $i(a)$  is the set of elements equivalent to  $a^{-1}$ . Thus, one can see that if we define an operation on the set of equivalence classes of  $S$ , we get a group. This can be regarded as an example of how the hyperstructure theory can generalize some classical concepts. Notice that when Vougiouklis [23] introduced  $H_v$ -structures (i.e. *weak hyperstructures*), it was the equivalence relation on a hyperstructure such that “almost all but some problematic” elements had the desired property that was the background motivation. *Zero scalars* (often called *absorbing elements*) are elements  $e \in S$  such that, for all  $x \in S$ , there is  $x * e = \{e\} = e * x$ , where “ $*$ ” is a hyperoperation on  $S$ .

**Theorem 4** Let  $(S, *)$  be the  $EL$ -semihypergroup of a non-trivial quasi-ordered semigroup  $(S, \cdot, \leq)$ . Then  $(S, *)$  has zero scalars if and only if  $(S, \cdot, \leq)$  has an element which is simultaneously  $EL$ -maximal with respect to “ $\leq$ ” and absorbing with respect to “ $\cdot$ ”.

*Proof.* If we realize that “ $\leq$ ” must be reflexive, the proof becomes obvious.

**Corollary 2** Let  $(S, *)$  be the  $EL$ -semihypergroup of a non-trivial quasi-ordered semigroup  $(S, \cdot, \leq)$ . Then  $(S, *)$  has at most one zero scalar element. To be more precise, if  $(S, \cdot)$  is a monoid, then it is its neutral element that can be the only zero scalar of  $(S, *)$ . If  $(S, \cdot)$  is not a monoid, then  $(S, *)$  is without zero scalars.

*Proof.* An obvious rewording of Theorem 4.

**Example 8** In Example 1,  $u = 1$ . However,  $u = 1$  is neither absorbing with respect to “ $\leq$ ” nor  $EL$ -maximal. Therefore,  $(\mathbb{N}, *)$  is without zero scalars. We get the same conclusion in Example 2. However, the reasoning is different because  $(\mathbb{C}, \cdot_{|z|})$  is not a monoid.

By a *hyperstructure idempotent element* we mean such an element  $a \in S$  that  $a \in a * a$ , where “ $*$ ” is a hyperoperation on  $S$ . If  $(S, \cdot)$  is a group, then, in the  $EL$ -context, the notions of hyperstructure idempotent elements and hyperstructure identities coincide.



**Theorem 5** Let  $(S, *)$  be the  $EL$ -hypergroup of a quasi-ordered group  $(S, \cdot, \leq)$ . An element  $a \in S$  is idempotent in  $(S, *)$  if and only if it is an identity of  $(S, *)$ .

*Proof.* In the “Ends lemma” context,  $a \in a * a$  rewrites to  $a \cdot a \leq a$ . In a group, this means that  $a \leq u$ , where  $u$  is the neutral element of  $(S, \cdot)$ . According to Theorem 2 this is equivalent to the fact that  $a$  is an identity of  $(S, *)$ .

The proof of the following theorem, which discusses the case when  $S$  is a semigroup only, can be found in [18]. Notice that by  $a^n$  we mean the hyperproduct of  $n$  elements  $a$ , i.e.  $a^n = \underbrace{a * \dots * a}_n$ .

**Theorem 6** Let  $(S, *)$  be the  $EL$ -semihypergroup of a quasi-ordered semigroup  $(S, \cdot, \leq)$ . Then for an arbitrary idempotent element  $a$  in  $(S, \cdot)$  we have:

- (i)  $a$  is an idempotent of  $(S, *)$ ,
- (ii)  $a * a$  is a subsemihypergroup of  $(S, *)$ ,
- (iii)  $[a]_{\leq} = a^2 = a^3 = \dots = a^n$  for all  $n \in \mathbb{N}, n \geq 2$ .

**Remark 2** Obviously, if the hyperoperation “ $*$ ” is extensive, then every element of an  $EL$ -semihypergroup (or rather, thanks to Theorem 1,  $EL$ -hypergroup) is idempotent.

In both of the following examples the relation is antisymmetric.

**Example 9** Suppose the  $EL$ -semihypergroup  $(\langle 0, 1 \rangle, *)$  constructed from the quasi-ordered semigroup  $(\langle 0, 1 \rangle, \cdot, \leq)$ , where “ $\cdot$ ” and “ $\leq$ ” are the usual multiplication and ordering of real numbers. In this case, by Theorem 2 every element is an identity. Further,  $u = 1$  is a zero scalar, 0 and 1 are idempotent elements and  $0 * 0$  and  $1 * 1$  are (trivial) subsemihypergroups of  $(\langle 0, 1 \rangle, *)$ .

**Example 10** Suppose the  $EL$ -semihypergroup  $(\mathbb{N}, *)$  constructed from the quasi-ordered semigroup  $(\mathbb{N}, \gcd, |)$ , where “ $\gcd$ ” stands for the greatest common divisor and “ $|$ ” is the usual divisibility relation. In this case, every element of  $(\mathbb{N}, *)$  is idempotent. By Theorem 6, every set

$$a * a = \{x \in \mathbb{N} \mid \gcd\{a, a\} | x\} = \{x \in \mathbb{N} \mid a | x\}$$

is a subsemihypergroup of  $(\mathbb{N}, *)$ . Indeed, if e.g.  $a = 3$ , then obviously e.g.  $12 \in 3 * 3$ ,  $18 \in 3 * 3$ . Now,

$$12 * 18 = \{x \in \mathbb{N} \mid \gcd\{12, 18\} | x\} = \{x \in \mathbb{N} \mid 6 | x\} \subseteq 3 * 3 = \{x \in \mathbb{N} \mid 3 | x\}.$$

If the relation “ $\leq$ ” is an equivalence, then by the definition of idempotent elements  $a \cdot a$  and  $a$  must be in the same equivalence class.

**Example 11** In Example 2, the set of idempotent elements coincides with the set  $\{z \in \mathbb{C} \mid |z| = 1\}$ , i.e. with a unit circle of the Gaussian plane.

Finally, we discuss one hyperstructure generalization of a concept from the lattice theory. The notion of a *semilattice* is based on a relation which is partial ordering. However, in  $EL$ -hyperstructures, antisymmetry of this relation is not needed. First of all we include the definition of concepts introduced by Xiao and Zhao in [24] and studied by Dehghan Nezhad and Davvaz in [8].

**Definition 4** Let  $L$  be a nonempty set with a binary hyperoperation “ $*$ ” on  $L$  such that, for all  $a, b, c \in L$ , the following conditions hold:

1.  $a \in a * a$  (idempotency)
2.  $a * b = b * a$  (commutativity)
3.  $(a * b) * c \cap a * (b * c) \neq \emptyset$  (weak associativity)

Then  $(L, *)$  is called an  $H_v$ -semilattice. When in the condition 3 we have equality, then  $(L, *)$  is called a hypersemilattice.

The connection between  $EL$ -(semi)hypergroups and hypersemilattices /  $H_v$ -semilattices is rather straightforward.

**Theorem 7** Let  $(L, *)$  be the  $EL$ -semihypergroup of a quasi-ordered semigroup  $(L, \cdot, \leq)$ .

1. If “ $\cdot$ ” is commutative and  $(L, \cdot)$  is a proper semigroup, then the condition that for all  $a \in L$  there holds  $a \cdot a \leq a$  is equivalent to the fact that  $(L, *)$  is a hypersemilattice.
2. If “ $\cdot$ ” is not commutative and “ $\leq$ ” is antisymmetric, then  $(L, *)$  is neither a hypersemilattice nor an  $H_v$ -semilattice.
3. If  $(L, \cdot)$  is a non-trivial group and “ $\leq$ ” is antisymmetric, then  $(L, *)$  is neither a hypersemilattice nor an  $H_v$ -semilattice.

*Proof.* Condition 3 of Definition 4 (in its strong associative version) is secured by default. Therefore, the question of whether our construction gives rise to hypersemilattices, is for commutative “ $*$ ” equivalent to the question of validity of condition in statement 1. Moreover, in our context, the idempotency condition rewrites to “ $a \cdot a \leq a$  for all  $a \in L$ ”.

If  $(L, \cdot)$  is a proper semigroup, this has no special implications and we obtain statement 1.

However, if  $(L, \cdot)$  is a group, then this is equivalent to  $a \leq u$  for all  $a \in L$ , where  $u$  is the neutral element of  $(L, \cdot)$ . On condition of antisymmetry of “ $\leq$ ” this means that  $u$  is the greatest element of  $(L, \leq)$ . Yet  $a \leq u$  is in a partially ordered group equivalent to  $u \leq a^{-1}$  for all  $a \in L$ , which is possible only if  $u = a^{-1}$ . Yet since this should hold for all  $a \in L$ , there is  $L = \{u\}$  and we obtain statement 3.

Finally, if “ $\leq$ ” is antisymmetric, then  $(L, \cdot, \leq)$  is a partially ordered semigroup and commutativity of  $(L, *)$  is equivalent to commutativity of  $(L, \cdot)$  and we obtain statement 2.

**Example 12** If we denote by  $|\mathbb{C}|_0^1$  the set of all complex numbers such that their absolute value is smaller than or equal to one 1 (i.e. we regard a unit disc of the Gaussian plane) and regard “ $\cdot_{|z|}$ ” multiplication of absolute values and set that  $z_1 \leq_{|z|} z_2$  whenever  $|z_1| \leq |z_2|$ , then we get that  $(|\mathbb{C}|_0^1, \cdot_{|z|}, \leq_{|z|})$  is a proper quasi-ordered semigroup. Moreover, “ $\leq_{|z|}$ ” is not antisymmetric. We define a hyperoperation on  $|\mathbb{C}|_0^1$  by

$$z_1 * z_2 = [z_1 \cdot_{|z|} z_2]_{\leq_{|z|}} = \{x \in |\mathbb{C}|_0^1 \mid |z_1| \cdot |z_2| \leq |x|\}.$$

Since  $z \in z * z$  for all  $z \in |\mathbb{C}|_0^1$ , we have that  $(\mathbb{C}, *)$  is a hypersemilattice.

The case of commutative quasi-ordered groups, where “ $\leq$ ” is *not* antisymmetric is not discussed in Theorem 7. Therefore, we include the following example.

**Example 13** *Regard the additive group of complex numbers  $(\mathbb{C}, +)$  and define, for all  $z_1, z_2 \in \mathbb{C}$ , relation “ $\leq_{|z|^{-1}}$ ” by  $z_1 \leq_{|z|^{-1}} z_2$  whenever  $|z_1| \geq |z_2|$ , where  $|z|$  stands for the absolute value of  $z \in \mathbb{C}$ . It is easy to verify that  $(\mathbb{C}, +, \leq_{|z|^{-1}})$  is a commutative quasi-ordered group, where “ $\leq_{|z|^{-1}}$ ” is obviously not antisymmetric. If we define, for all  $z_1, z_2 \in \mathbb{C}$ , that  $z_1 * z_2 = \{x \in \mathbb{C} \mid |x| \leq |z_1 + z_2|\}$ , then, by Definition 4 and Lemma 1,  $(\mathbb{C}, *)$  is a hypersemilattice. Indeed,  $|z| \leq |z + z|$  for all  $z \in \mathbb{C}$ , i.e.  $z \in z * z$  (and the rest is obvious). However, if we regard “ $\leq_{|z|}$ ” such that  $z_1 \leq_{|z|} z_2$  whenever  $|z_1| \leq |z_2|$  instead of  $\leq_{|z|^{-1}}$ , then “ $*$ ” is no longer idempotent, i.e.  $(\mathbb{C}, *)$  is neither a hypersemilattice nor an  $H_v$ -semilattice.*

Obviously, idempotent quasi-ordered semigroups always create hypersemilattices. In the following example, the semigroup is not idempotent. (Mind the difference between idepotence in semigroups and in semihypergroups, i.e.  $a \cdot a = a$  vs  $a \in a * a$ !)

**Example 14** *Denote the closed interval of real numbers  $\langle 0, 1 \rangle$  by  $L$ . Obviously,  $(L, \cdot, \leq)$ , where “ $\cdot$ ” is the usual multiplication and “ $\leq$ ” the usual ordering of real numbers, is a proper quasi-ordered semigroup. Also obviously, the condition that, for all  $a \in L$ ,  $a \cdot a \leq a$  holds in  $L$ . Therefore, the  $EL$ -semihypergroup  $(L, *)$ , is a hypersemilattice. Moreover, since “ $*$ ” defined on  $(L, \cdot, \leq)$  is extensive,  $(L, *)$  is, by Theorem 1, a hypergroup.*

Moreover – obviously again – if we consider a proper commutative quasi-ordered semigroup  $(L, \cdot, \leq)$  such that the relation “ $\leq$ ” is not antisymmetric, than extensivity of the hyperoperation “ $*$ ” implies that its  $EL$ -hyperstructure  $(L, *)$  is a hypersemilattice.

### 3 CONCLUSION

In this short paper we attempted to assign a new meaning to result which have (mostly) already been used. We have shown that, in some cases, antisymmetry of the binary relation which is used to construct  $EL$ -hyperstructures, is in fact not needed. This allows us to broaden the applicability of the construction and to consider another special class of  $EL$ -hyperstructures – those where the relation is an equivalence. We have included a variety of examples (based on some usual sets) to demonstrate the many possible uses of the construction.

### References

- [1] Chvalina J., Commutative hypergroups in the sense of Marty and ordered sets. In *Gen. Alg. and Ordered Sets, Proc. Int. Conf. Olomouc*. 1994, p. 19–30.
- [2] Chvalina J., *Functional Graphs, Quasi-ordered Sets and Commutative Hypergroups*. Brno, Masaryk University, 1995 (in Czech).
- [3] Chvalina J., Hořková-Mayerová Š., Dehghan Nezhad A., General actions of hypergroups and some applications. *An. Șt. Univ. Ovidius Constanța*, 21(1) (2013), p. 59–82.
- [4] Corsini P., *Prolegomena of Hypergroup Theory*. Aviani Editore, Tricesimo, 1993.
- [5] Corsini P., Leoreanu V., *Applications of Hyperstructure Theory*. Kluwer Academic Publishers, Dordrecht – Boston – London, 2003.

- [6] Corsini P., Hyperstructures associated with ordered sets. *Bull. Greek Math. Soc.* 48 (2003), p. 7–18.
- [7] Cristea I., Ştefănescu M., Binary relations and reduced hypergroups. *Discret. Math.* 308(16) (2008), p. 3537–3544.
- [8] Dehghan Nezhad A., Davvaz B., An Introduction to the Theory of  $H_v$ -Semilattices. *Bull. Malays. Math. Sci. Soc.* (2)32(3) (2009), p. 375–390.
- [9] Davvaz B., Leoreanu Fotea V., *Applications of Hyperring Theory*. International Academic Press, Palm Harbor, 2007.
- [10] Ghazavi S. H., Anvariye S. M.,  $EL$ -hyperstructures associated to  $n$ -ary relations. *Soft Comput.* (2016) (No. not assigned yet), <http://dx.doi.org/10.1007/s00500-016-2165-3>.
- [11] Ghazavi S. H., Anvariye S. M., Mirvakili S.,  $EL^2$ -hyperstructures derived from (partially) quasi-ordered hyperstructures. *Iran. J. Math. Sci. Inform.* 10(2) (2015), p. 99–114.
- [12] Ghazavi S. H., Anvariye S. M., Mirvakili S., Ideals in  $EL$ -semihypergroups associated to ordered semigroups. *Journal of Algebraic Systems* 3(2) (2016), p. 109–125.
- [13] Heidari D., B. Davvaz B., On ordered hyperstructures. *U.P.B. Sci. Bull. Series A*, 73(2) (2011), p. 85–96.
- [14] Hošková Š., Chvalina J., Discrete transformation hypergroups and transformation hypergroups with phase tolerance space. *Discret. Math.* 308(18) (2008), p. 4133–4143.
- [15] Kovár M., Chernikava A., On the proof of the existence of undominated strategies in normal form games. *Amer. Math. Monthly* 121(4) (2014), p. 332–337.
- [16] Novák M.,  $n$ -ary hyperstructures constructed from binary quasi-ordered semigroups. *An. Şt. Univ. Ovidius Constanţa*, 22(3) (2014), p. 147–168.
- [17] Novák M., Important elements of  $EL$ -hyperstructures. In: *APLIMAT: 10th International Conference*. Bratislava, STU in Bratislava, 2011, p. 151–158.
- [18] Novák M., On  $EL$ -semihypergroups. *European J. Combin.* 44 Part B, (2015), p. 274–286.
- [19] Novák M., Potential of the “Ends lemma” to create ring-like hyperstructures from quasi-ordered (semi)groups. *South Bohemia Mathem. Letters* 17(1) (2009), p. 39–50.
- [20] Novák M., Some basic properties of  $EL$ -hyperstructures. *European J. Combin.*, 34 (2013), p. 446–459.
- [21] Phanthawimol W., Kemprasit Y., Homomorphisms and epimorphisms of some hypergroups, *Ital. J. Pure Appl. Math.* 27 (2010), 305–312.
- [22] Pickett H. E., Homomorphisms and subalgebras of multialgebras, *Pac. J. Math.* 21(2) (1967), p. 327–342.
- [23] Vougiouklis T., *Hyperstructures and their Representations*. Monographs in Mathematics, Hadronic Press, 1994.
- [24] Xiao Y., Zhao B., Hypersemilattices and their ideals. *J. Shaanxi Normal Univ. Nat. Sci. Ed.* 33(1) (2005), p. 7–10.

## Acknowledgement

The work presented in this paper has been supported by Brno University of Technology (research project FEKT-S-14-2200).

# COMPARISON OF TWO POLYNOMIAL CALIBRATION METHODS

**Petra Ráboňová**

Faculty of Science, Masaryk University

Kotlářská 2, Brno, Czech Republic

324037@mail.muni.cz

**Abstract:** *In the contribution we focused on comparison of two polynomial calibration methods. We compare a method based on maximum likelihood method, and a method using linearised model with errors in variables and Kenward Roger's type of approximation. First, we introduce models for both procedures. Then we estimate unknown parameters of transformation function with use of these models. We compare both methods in a small simulation study.*

**Keywords:** kalibration, transformation function, transformation curve, maximum likelihood method, Kenward Roger's type of approximation.

## INTRODUCTION

Calibration is an important part of metrology. It is a set of tasks which gives relationship between a reference and a calibrated device if some special conditions are fulfilled. In the contribution we assume that the relationship between the calibrated device and reference device is polynomial. This relationship can be described by a transformation function and represented by a transformation curve. We can divide calibration process into two parts: 1) creation of calibration model and 2) measurement with calibrated device. In the contribution we estimate the parameters of the transformation function by two methods and compare them. Similar investigations are realized in [7] without comparisons of results.

We assume that we have  $m$  different objects. Each of these objects is measured with two different measuring devices (device A and device B), and we repeat the measurement  $n$  times. We assume that device A is less precise than device B. It is assumed that measured values on both devices are realizations of random variables with normal distribution, and for each of  $m$  objects, values measured errorlessly by device A are  $\mu = (\mu_1, \dots, \mu_m)$ , and by device B are  $\nu = (\nu_1, \dots, \nu_m)'$  where  $\nu_i = a_0 + a_1\mu_i + \dots + a_k\mu_i^k$  and  $a_0, a_1, \dots, a_k \in \mathbb{R}$ ,  $i = 1, 2, \dots, m$ . The function  $\nu_i = a_0 + a_1\mu_i + \dots + a_k\mu_i^k$  is called the transformation function. The next assumption is that measurements realised by device A and device B are independent and  $X_{ij} \sim N(\mu_i, \sigma_x^2)$  ( $X_{ij}$  is normally distributed with mean  $\mu_i$  and dispersion  $\sigma_x^2$ ),  $Y_{ij} \sim N(a_0 + a_1\mu_i + \dots + a_k\mu_i^k, \sigma_y^2)$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . The aim is to estimate the parameters  $(a_0, a_1, \dots, a_k, \sigma_x^2, \sigma_y^2)'$  with using of the maximum likelihood method and a method using a linearised model with errors in variables and the Kenward Roger's type of approximation, and compare obtained estimates based on the simulation study.

## 1 REPLICATED MODEL WITH ERRORS IN VARIABLES

At first we assume that with the devices A and B we realize only one measurement of each object. In this case we have a vector of random variables - measurements with device A as  $\mathbf{X} = (X_1, \dots, X_m)'$ . For each  $i = 1, \dots, m$  is  $X_i \sim N(\mu_i, \sigma_x^2)$ .  $\mathbf{Y} = (Y_1, \dots, Y_m)'$  measurements are obtained with the device B, where  $Y_i \sim N(\nu_i, \sigma_y^2)$ . We assume that the measurements are not correlated. So we can write the measurement model as:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N \left[ \begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_y^2 \mathbf{I}_m \end{pmatrix} \right].$$

If we denote  $\nu$  and  $\mu^b$  as  $\nu = a_0 \mathbf{1}_m + a_1 \mu + \dots + a_k \mu^k$ ,  $\mu^b = (\mu_1^b, \dots, \mu_m^b)'$ ,  $b = 1, 2, \dots, k$ ,  $\mathbf{1}_m = (1, \dots, 1)' \in \mathbb{R}^m$ , we obtain a linear regression model with nonlinear constraints on parameters.

$$\nu = \mathbf{1}_m a_0 + \mu a_1 + \dots + \mu^k a_k$$

Using the Taylor expansion in values  $a_{10}, \dots, a_{k0}, \mu_0 = (\mu_{10}, \dots, \mu_{m0})$ ,  $\mu_0^b = (\mu_{10}^b, \dots, \mu_{m0}^b)$ ,  $\delta \mu_i = \mu_i - \mu_{i0}$  for  $i = 1, \dots, m$ ,  $\delta \mu = (\delta \mu_1, \dots, \delta \mu_m)'$  and neglecting the terms of the second and higher order we obtain the linear regression model with linear constraints:

$$\begin{pmatrix} \mathbf{X} - \mu_0 \\ \mathbf{Y} \end{pmatrix} \sim N \left[ \begin{pmatrix} \delta \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_y^2 \mathbf{I}_m \end{pmatrix} \right]$$

$$(\text{diag}(a_{10} \mathbf{1}_m + \dots + k a_{k0} \mu_0^{k-1}), -\mathbf{I}_m) \begin{pmatrix} \delta \mu \\ \nu \end{pmatrix} + (\mathbf{1}_m, \mu_0, \dots, \mu_0^k) \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix} = \mathbf{0},$$

(*diag* is the diagonal matrix with elements of the vector  $(a_{10} \mathbf{1}_m + \dots + k a_{k0} \mu_0^{k-1})$  on the diagonal). If we repeat the measurement with devices A and B  $n$  times, we obtain the replicated model:

$$\begin{pmatrix} \mathbf{X}^1 - \mu_0 \\ \mathbf{Y}^1 \\ \vdots \\ \mathbf{X}^n - \mu_0 \\ \mathbf{Y}^n \end{pmatrix} \sim N \left[ \mathbf{1}_n \otimes \begin{pmatrix} \delta \mu \\ \nu \end{pmatrix}, \mathbf{I}_n \otimes \begin{pmatrix} \sigma_x^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_y^2 \mathbf{I}_m \end{pmatrix} \right] \quad (1)$$

with linear constraints on parameters

$$(\text{diag}(a_{10} \mathbf{1}_m + \dots + k a_{k0} \mu_0^{k-1}), -\mathbf{I}_m) \begin{pmatrix} \delta \mu \\ \nu \end{pmatrix} + (\mathbf{1}_m, \mu_0, \dots, \mu_0^k) \begin{pmatrix} a_0 \\ \vdots \\ a_k \end{pmatrix} = \mathbf{0},$$

where  $\otimes$  is the Kronecker's product of matrices,  $\mathbf{X}^k = (X_{1k}, \dots, X_{mk})'$ ,  $\mathbf{Y}^k = (Y_{1k}, \dots, Y_{mk})'$ .

Denote  $\beta_1 = \begin{pmatrix} \delta\mu \\ \nu \end{pmatrix}$ ,  $\beta_2 = \begin{pmatrix} a_0 \\ \vdots \\ a_k \end{pmatrix}$ .

### 1.1 Estimators of vectors of paramters $\beta_1, \beta_2$

In this section we focuse on the estimators of transformation function paramaters  $\begin{pmatrix} a_0 \\ \vdots \\ a_k \end{pmatrix}$  and estimators of vector  $\begin{pmatrix} \delta\mu \\ \nu \end{pmatrix}$ . For this purpose we use the procedure described in [2].

According to [2, page 129], let  $\mathbf{Y}_D \sim (\mathbf{X}_D \beta_1, \Sigma)$  be a model with constraint  $\mathbf{b} + \mathbf{B}_1 \beta_1 + \mathbf{B}_2 \beta_2 = \mathbf{0}$ , where  $r(\mathbf{X}_D) = k_1 < n$ ,  $r(\mathbf{B}_1, \mathbf{B}_2) = q < k_1 + k_2$ ,  $r(\mathbf{B}_2) = k_2 < q$  and  $\Sigma$  is a positive definite matrix, then BLUE (best linear unbiased estimation) of vector  $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$  is:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = - \begin{pmatrix} \mathbf{C}^{-1} \mathbf{B}_1' \mathbf{Q}_{11} \\ \mathbf{Q}_{21} \end{pmatrix} \mathbf{b} + \begin{pmatrix} \mathbf{I} - \mathbf{C}^{-1} \mathbf{B}_1' \mathbf{Q}_{11} \mathbf{B}_1 \\ -\mathbf{Q}_{21} \mathbf{B}_1 \end{pmatrix} \hat{\beta}_1,$$

where  $\mathbf{C} = \mathbf{X}_D' \Sigma^{-1} \mathbf{X}_D$  and  $\hat{\beta}_1 = \mathbf{C}^{-1} \mathbf{X}_D' \Sigma^{-1} \mathbf{Y}_D$ .

$$\begin{aligned} \text{var}(\hat{\beta}_2) &= -\mathbf{Q}_{22}, \\ \begin{pmatrix} \mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1' & \mathbf{B}_2 \\ \mathbf{B}_2' & \mathbf{0} \end{pmatrix}^{-1} &= \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}. \end{aligned}$$

Now we apply the procedure on the model (1). We denote  $\mathbf{B}_2 = (\mathbf{1}_m; \mu_0; \dots; \mu_0^k)$ ,  $\mathbf{B}_1 = (\mathbf{S}; -\mathbf{I}_m)$ ,  $\mathbf{S} = \text{diag}(a_{10} \mathbf{1}_m + \dots + k a_{k0} \mu_0^{k-1})$ ,  $\mathbf{A}_1 = \mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1' = \frac{1}{n} (\sigma_x^2 \mathbf{S} \mathbf{S}' + \sigma_y^2 \mathbf{I})$ ,  $\bar{\mathbf{X}} = \frac{1}{n} (\sum_{i=1}^n X_{1i}, \dots, \sum_{i=1}^n X_{mi})'$ ,  $\bar{\mathbf{Y}} = \frac{1}{n} (\sum_{i=1}^n Y_{1i}, \dots, \sum_{i=1}^n Y_{mi})'$ .

Firstly, we compute matrices  $\mathbf{Q}_{11}$ ,  $\mathbf{Q}_{21}$ ,  $\mathbf{Q}_{22}$  for our model with use of [5, page 65] and  $\hat{\beta}_1$ .

$$\mathbf{C}^{-1} = \frac{1}{n} \begin{pmatrix} \sigma_x^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_y^2 \mathbf{I}_m \end{pmatrix},$$

$$\mathbf{Q}_{11} = (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} - (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \left( \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \right)^{-1} \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1},$$

$$\mathbf{Q}_{12} = (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \left( \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \right)^{-1},$$

$$\mathbf{Q}_{12} = \mathbf{Q}_{21}',$$

$$\mathbf{Q}_{22} = - \left( \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \right)^{-1},$$

$$\hat{\beta}_1 = \begin{pmatrix} \bar{\mathbf{X}} - \mu_0 \\ \bar{\mathbf{Y}} \end{pmatrix}.$$

Therefore:

$$\begin{pmatrix} \hat{\delta\mu} \\ \hat{\nu} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{X}} - \mu_0 \\ \bar{\mathbf{Y}} \end{pmatrix} - \begin{pmatrix} \frac{\sigma_x^2}{n} \mathbf{S} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_0) - \bar{\mathbf{Y}}) \\ -\frac{\sigma_y^2}{n} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_0) - \bar{\mathbf{Y}}) \end{pmatrix},$$

$$\hat{\mu} - \mu_0 = \bar{\mathbf{X}} - \mu_0 - \frac{\sigma_x^2}{n} \mathbf{S} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_0) - \bar{\mathbf{Y}}),$$

$$\hat{\mu} = \bar{\mathbf{X}} - \frac{\sigma_x^2}{n} \mathbf{S} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_0) - \bar{\mathbf{Y}}),$$

$$\hat{\nu} = \bar{\mathbf{Y}} + \frac{\sigma_y^2}{n} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_0) - \bar{\mathbf{Y}}),$$

$$\begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_k \end{pmatrix} = -\mathbf{Q}_{21} \mathbf{B}_1 \hat{\beta}_1,$$

$$\text{var} \begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_k \end{pmatrix} = \text{var}(\hat{\beta}_2) = -\mathbf{Q}_{22} = (\mathbf{B}_2' \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1}.$$

For more details, see [1], [8],[9].

## 1.2 MINQUE estimates of the variance matrix

In subsection 1.1 we assumed that the variance matrix is known. In our case the variance matrix is unknown therefore we have to estimate it. We use MINQUE (minimum norm quadratic unbiased estimator) procedure described in [2, str. 97-101] for the estimation.

Before we apply the MINQUE procedure, we have to transform the model with constraints on parameters to the model without constraints on parameters. We find an arbitrary solution of the equation  $\mathbf{b} + \mathbf{B}_1 \beta_1 + \mathbf{B}_2 \beta_2 = \mathbf{0}$ , denote it  $\beta_0 = \begin{pmatrix} \beta_{0,1} \\ \beta_{0,2} \end{pmatrix}$ . With use of this solution and a solution of a homogeneous system of equations  $\mathbf{B}_1 \beta_1 + \mathbf{B}_2 \beta_2 = \mathbf{0}$ , we can find the arbitrary solution of a system of equations  $\mathbf{b} + \mathbf{B}_1 \beta_1 + \mathbf{B}_2 \beta_2 = \mathbf{0}$ . All solutions of system of equations  $\mathbf{B}_1 \beta_1 + \mathbf{B}_2 \beta_2 = \mathbf{0}$  form space  $\ker(\mathbf{B}_1, \mathbf{B}_2)$ . We search for matrices  $\mathbf{K}_1, \mathbf{K}_2$ , where the dimension of  $\mathbf{K}_1$  is  $k_1 \times (k_1 + k_2 - q)$ , the dimension of  $\mathbf{K}_2$  is  $k_2 \times (k_1 + k_2 - q)$  and  $r \begin{pmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{pmatrix} = k_1 + k_2 - q$  so that  $\mathcal{M} \begin{pmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{pmatrix}_{(k_1+k_2) \times (k_1+k_2-q)} = \ker(\mathbf{B}_1, \mathbf{B}_2)$ . Put together  $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_{0,1} \\ \beta_{0,2} \end{pmatrix} + \begin{pmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{pmatrix} \gamma$ , where  $\gamma \in \mathbb{R}^{k_1+k_2-q}$ .



We can rewrite the original model to form:  $\mathbf{Z} - \mathbf{T}\beta_{0,1} \sim (\mathbf{TK}_1\gamma, \Sigma)$  without constraints on parameters. We obtain a model (in our model  $\mathbf{b} = \mathbf{0}$  therefore also  $\beta_0 = \mathbf{0}$  and  $\beta_{0,1} = \mathbf{0}$ ):

$$\begin{pmatrix} \mathbf{X}^1 - \mu_0 \\ \mathbf{Y}^1 \\ \vdots \\ \mathbf{X}^n - \mu_0 \\ \mathbf{Y}^n \end{pmatrix} \sim N \left[ (\mathbf{1}_n \otimes \mathbf{I}_{2m}) \mathbf{K}_1 \gamma, \mathbf{I}_n \otimes \begin{pmatrix} \sigma_x^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_y^2 \mathbf{I}_m \end{pmatrix} \right], \quad (2)$$

where  $\gamma \in \mathbb{R}^{m+k+1}$ ,  $\mathbf{K}_1$  is the matrix of dimension  $2m \times (m+k+1)$ ,  $\mathbf{K}_2$  is matrix of dimension  $(k+1) \times (m+k+1)$ .

Now we focus on estimating the components of the variance matrix of the model 2. As

$$\begin{aligned} \begin{pmatrix} \hat{\sigma}_x^2 \\ \hat{\sigma}_y^2 \end{pmatrix} &= \mathbf{S}_{(\mathbf{M}_L \Sigma_0 \mathbf{M}_L)^+}^{-1} \mathbf{F}, \\ \mathbf{F} &= \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}, \\ F_1 &= \frac{1}{\sigma_{x0}^4} \left[ \sum_{j=1}^n (\mathbf{X}^j - \bar{\mathbf{X}})' (\mathbf{X}^j - \bar{\mathbf{X}}) + n (\bar{\mathbf{X}} - \hat{\mu})' (\bar{\mathbf{X}} - \hat{\mu}) \right], \\ F_2 &= \frac{1}{\sigma_{y0}^4} \left[ \sum_{j=1}^n (\mathbf{Y}^j - \bar{\mathbf{Y}})' (\mathbf{Y}^j - \bar{\mathbf{Y}}) + n (\bar{\mathbf{Y}} - \hat{\nu})' (\bar{\mathbf{Y}} - \hat{\nu}) \right], \\ \mathbf{S}_{(\mathbf{M}_L \Sigma_0 \mathbf{M}_L)^+} &= \begin{pmatrix} \frac{(n-1)m}{\sigma_{x0}^4} + \frac{1}{n^2} Tr(\mathbf{S} \mathbf{Q}_{11} \mathbf{S} \mathbf{S} \mathbf{Q}_{11} \mathbf{S}) & \frac{1}{n^2} Tr(\mathbf{Q}_{11} \mathbf{S} \mathbf{S} \mathbf{Q}_{11}) \\ \frac{1}{n^2} Tr(\mathbf{Q}_{11} \mathbf{S} \mathbf{S} \mathbf{Q}_{11}) & \frac{(n-1)m}{\sigma_{y0}^4} + \frac{1}{n^2} Tr(\mathbf{Q}_{11} \mathbf{Q}_{11}) \end{pmatrix}, \end{aligned}$$

where  $\sigma_{x0}^2, \sigma_{y0}^2$  are sample variances,  $\sigma_{x0}^2 = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$ ,  $\sigma_{y0}^2 = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2$ , the

variance matrix of  $\begin{pmatrix} \hat{\sigma}_x^2 \\ \hat{\sigma}_y^2 \end{pmatrix}$  is:

$$\mathbf{V} = 2\mathbf{S}_{(\mathbf{M}_L \Sigma_0 \mathbf{M}_L)^+}^{-1}. \quad (3)$$

### 1.3 Iterative procedure for estimating $\beta_1, \beta_2, \sigma_x$ , and $\sigma_y$

We estimate unknown parameters  $(a_0, \dots, a_k, \mu, \nu, \sigma_x, \sigma_y)'$  with use of the procedure described in subsections 1.1 and 1.2.

1. We make an initial estimate of parameters  $(a_0, \dots, a_k, \mu, \sigma_x, \sigma_y)'$  (denote it  $(a_{00}, \dots, a_{k0}, \mu_0, \sigma_{x0}, \sigma_{y0})'$ ) in two steps. At first we estimate  $\sigma_x$ , and  $\sigma_y$ :

$$\sigma_{x0}^2 = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \mu_{pi})^2,$$

$$\sigma_{y0}^2 = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \nu_{pi})^2,$$

$$\text{where } \mu_{pi} = \frac{1}{n} \sum_{j=1}^n X_{ij}, \nu_{pi} = \frac{1}{n} \sum_{u=1}^n Y_{iu}.$$

Then we calculate an estimate of vectors  $\beta_1, \beta_2$  by the procedure described in subsection 1.1, (in notation of subsection 1.1):

$$\mu_0 = \bar{\mathbf{X}} - \frac{\sigma_{x0}^2}{n} \mathbf{S} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_{0p}) - \bar{\mathbf{Y}}),$$

$$\nu_0 = \bar{\mathbf{Y}} + \frac{\sigma_{y0}^2}{n} \mathbf{Q}_{11} (\mathbf{S} (\bar{\mathbf{X}} - \mu_{0p}) - \bar{\mathbf{Y}}),$$

$$\begin{pmatrix} a_{00} \\ \vdots \\ a_{k0} \end{pmatrix} = -\mathbf{Q}_{21} \mathbf{B}_1 \hat{\beta}_1,$$

where  $(a_{0p}, \dots, a_{kp})' = (M_u' \cdot M_u)^{-1} \cdot M_u' \cdot \nu_{0p}$ ,  $M_u = (\mathbf{1}_m, \mu_{0p}, \dots, \mu_{0p}^k)'$ ,

$$\mu_{0p} = \left( \frac{1}{n} \sum_{j=1}^n X_{1j}, \dots, \frac{1}{n} \sum_{j=1}^n X_{mj} \right)', \quad \nu_{0p} = \left( \frac{1}{n} \sum_{j=1}^n Y_{1j}, \dots, \frac{1}{n} \sum_{j=1}^n Y_{mj} \right)', \quad \mathbf{B}_1 = (\mathbf{S}; -\mathbf{I}_m),$$

$$\mathbf{S} = \text{diag}(a_{1p} \mathbf{1}_m + \dots + k a_{kp} \mu_{0p}^{k-1}), \quad \mathbf{A}_1 = \mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1', \quad \bar{\mathbf{X}} = \frac{1}{n} (\sum_{i=1}^n X_{1i}, \dots, \sum_{i=1}^n X_{mi})',$$

$$\bar{\mathbf{Y}} = \frac{1}{n} (\sum_{i=1}^n Y_{1i}, \dots, \sum_{i=1}^n Y_{mi})', \quad \mathbf{C}^{-1} = \frac{1}{n} \begin{pmatrix} \sigma_{x0}^2 \mathbf{I}_m & \mathbf{0}_m \\ \mathbf{0}_m & \sigma_{y0}^2 \mathbf{I}_m \end{pmatrix}, \quad \hat{\beta}_1 = \begin{pmatrix} \bar{\mathbf{X}} - \mu_{0p} \\ \bar{\mathbf{Y}} \end{pmatrix}.$$

$$\mathbf{Q}_{11} = (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} - (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \left( \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \right)^{-1} \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1},$$

$$\mathbf{Q}_{12} = (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \left( \mathbf{B}_2' (\mathbf{B}_1 \mathbf{C}^{-1} \mathbf{B}_1')^{-1} \mathbf{B}_2 \right)^{-1}, \quad \mathbf{Q}_{12} = \mathbf{Q}_{21}'.$$

2. We use the initial values  $(a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \sigma_{x0}, \sigma_{y0})'$  of vector  $(a_0, \dots, a_k, \mu, \nu, \sigma_x, \sigma_y)'$  obtained in step 1., and calculate estimators of  $\sigma_x, \sigma_y$  (denote them  $\hat{\sigma}_x, \hat{\sigma}_y$ ) by the procedure described in subsection 1.2 (with use of notation in 1.2):

$$(\hat{\sigma}_x^2, \hat{\sigma}_y^2)' = S_{(M_L \Sigma_0 M_L)^+}^{-1} \mathbf{F}, \text{ from this equation we can easily obtain } \hat{\sigma}_x, \hat{\sigma}_y.$$

3. We calculate the estimate of  $(a_0, \dots, a_k, \mu, \nu)'$  with use of procedure 1.1 (with use of notation of 1.1), denote obtained estimates  $(\hat{a}_0, \dots, \hat{a}_k, \hat{\mu}, \hat{\nu})'$  where:

$$(a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \sigma_{x0}, \sigma_{y0})' = (a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \hat{\sigma}_x, \hat{\sigma}_y)'.$$

4. We have the estimate of vector  $(a_0, \dots, a_k, \mu, \nu, \sigma_x, \sigma_y)'$  in form  $(\hat{a}_0, \dots, \hat{a}_k, \hat{\mu}, \hat{\nu}, \hat{\sigma}_x, \hat{\sigma}_y)'$ . So we can refine this estimate, if we put  $(a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \sigma_{x0}, \sigma_{y0})' = (\hat{a}_0, \dots, \hat{a}_k, \hat{\mu}, \hat{\nu}, \hat{\sigma}_x, \hat{\sigma}_y)'$  and repeat steps 2 and 3. We assume that the estimate is accurate enough if:

$$\frac{|(a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \sigma_{x0}, \sigma_{y0})' - (\hat{a}_0, \dots, \hat{a}_k, \hat{\mu}, \hat{\nu}, \hat{\sigma}_x, \hat{\sigma}_y)'|}{|(a_{00}, \dots, a_{k0}, \mu_0, \nu_0, \sigma_{x0}, \sigma_{y0})'|} < 0.1$$

For this iterative procedure there are created scripts in Matlab software (*progam.m*, *beta\_12.m*, *poc\_odhad.m*, *sx\_sy.m*) available on websites <http://www.math.muni.cz/~xsirucko/>.

#### 1.4 Confidence region for $\beta_2$

In this subsection we derive confidence region for  $\beta_2$  with use of the Kenward Rogers method. According to 1.1:

$$\widehat{\beta}_2 \approx N \left[ \beta_2, (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \right].$$

Denote  $\Sigma = (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1}$  and  $\sigma^* = \begin{pmatrix} \sigma_x^2 \\ \sigma_y^2 \end{pmatrix}$ . According to Kenward Roger (see [3, page 985]) we calculate an approximate variance matrix:

$$\Sigma_A = \Sigma + 2\Sigma \left( \sum_{i \in \{x,y\}} \sum_{j \in \{x,y\}} \mathbf{V}_{ij} (\mathbf{Q}_{ij} - \mathbf{P}_i \Sigma \mathbf{P}_j - \frac{1}{4} \mathbf{R}_{ij}) \right) \Sigma, \quad (4)$$

where  $\mathbf{P}_i = \frac{\partial \Sigma^{-1}}{\partial \sigma_i}$ ,  $i = 1, 2$ ,  $\sigma_1 = \sigma_x^2$ ,  $\sigma_2 = \sigma_y^2$ ,  $\mathbf{B}_2 = (\mathbf{1}_m, \mu_0, \dots, \mu_0^k)$ ,  $\mathbf{A}_1 = \frac{1}{n}(\sigma_x^2 \mathbf{S}\mathbf{S} + \sigma_y^2 \mathbf{I}_m)$ .

$$\mathbf{P}_1 = \frac{\partial \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2}{\partial \sigma_x^2} = \mathbf{B}'_2 \frac{\partial \mathbf{A}_1^{-1}}{\partial \sigma_x^2} \mathbf{B}_2 = -\mathbf{B}'_2 \mathbf{A}_1^{-1} \frac{\partial \mathbf{A}_1}{\partial \sigma_x^2} \mathbf{A}_1^{-1} \mathbf{B}_2 = -\frac{1}{n} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$\mathbf{P}_2 = \frac{\partial \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2}{\partial \sigma_y^2} = \mathbf{B}'_2 \frac{\partial \mathbf{A}_1^{-1}}{\partial \sigma_y^2} \mathbf{B}_2 = -\mathbf{B}'_2 \mathbf{A}_1^{-1} \frac{\partial \mathbf{A}_1}{\partial \sigma_y^2} \mathbf{A}_1^{-1} \mathbf{B}_2 = -\frac{1}{n} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2.$$

$$\mathbf{Q}_{ij} = \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j},$$

$$\mathbf{Q}_{11} = \frac{1}{n^2} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$\mathbf{Q}_{12} = \frac{1}{n^2} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$\mathbf{Q}_{21} = \frac{1}{n^2} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$\mathbf{Q}_{22} = \frac{1}{n^2} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$\mathbf{R}_{ij} = \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \sigma_i \partial \sigma_j} \Sigma^{-1}.$$

$$\frac{\partial \Sigma}{\partial \sigma_x^2} = \frac{1}{n} (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1},$$

$$\frac{\partial \Sigma}{\partial \sigma_y^2} = \frac{1}{n} (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1},$$

$$\mathbf{R}_{11} = \frac{2}{n^2} \left[ \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 - \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 \right],$$

$$\mathbf{R}_{12} = \mathbf{R}_{21} = \frac{1}{n^2} \left[ \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 - \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 - \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 + \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{S} \mathbf{S} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 \right],$$

$$\mathbf{R}_{22} = \frac{2}{n^2} \left[ \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 (\mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{B}_2)^{-1} \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 - \mathbf{B}'_2 \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{A}_1^{-1} \mathbf{B}_2 \right].$$

Finally we obtain the wanted matrix  $\hat{\Sigma}_A$  by substituting in (4)  $\hat{\sigma}_x^2$  and  $\hat{\sigma}_y^2$  instead of  $\sigma_x^2$  and  $\sigma_y^2$ .  $\mathbf{V}_{ij}$  are elements of matrix  $\mathbf{V}$  (3).

We derive confidence region for  $\beta_2$  with use of statistic:

$F = \frac{1}{k+1} \left( \widehat{\beta}_2 - \beta_2 \right)' \widehat{\Sigma}_A^{-1} \left( \widehat{\beta}_2 - \beta_2 \right)$ , with distribution  $\frac{1}{\lambda} F_{k+1, v}$ . According to Kenward and Roger [3] we determine parameters  $\lambda$  and  $v$ . According to [3] we denote:

$$\Gamma = \Sigma^{-1} = \mathbf{B}_2' \mathbf{A}_1^{-1} \mathbf{B}_2,$$

$$H_1 = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\Gamma \Sigma \mathbf{P}_i^* \Sigma\} Tr\{\Gamma \Sigma \mathbf{P}_j^* \Sigma\},$$

$$H_1 = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\Sigma^{-1} \Sigma \mathbf{P}_i^* \Sigma\} Tr\{\Sigma^{-1} \Sigma \mathbf{P}_j^* \Sigma\} = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\mathbf{P}_i^* \Sigma\} Tr\{\mathbf{P}_j^* \Sigma\},$$

$$H_2 = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\Gamma \Sigma \mathbf{P}_i^* \Sigma \Gamma \Sigma \mathbf{P}_j^* \Sigma\},$$

$$H_2 = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\Sigma^{-1} \Sigma \mathbf{P}_i^* \Sigma \Sigma^{-1} \Sigma \mathbf{P}_j^* \Sigma\} = \sum_{i=1}^2 \sum_{j=1}^2 V_{ij}^* Tr\{\mathbf{P}_i^* \Sigma \mathbf{P}_j^* \Sigma\},$$

$$g = \frac{(k+2)H_1 - (k+5)H_2}{(k+3)H_2},$$

$$B^* = \frac{1}{2k+2} (H_1 + 6H_2),$$

$$c_1 = \frac{g}{3k+5-2g},$$

$$c_2 = \frac{k+1-g}{3k+5-2g},$$

$$c_3 = \frac{k+3-g}{3k+5-2g},$$

$$c_1 = \frac{(k+2)H_1 - (k+5)H_2}{-(2k+4)H_1 + (3k^2 + 16k + 25)H_2},$$

$$c_2 = \frac{-(k+2)H_1 + (k^2 + 5k + 8)H_2}{-(2k+4)H_1 + (3k^2 + 16k + 25)H_2},$$

$$c_3 = \frac{-(k+2)H_1 + (k^2 + 7k + 14)H_2}{-(2k+4)H_1 + (3k^2 + 16k + 25)H_2},$$

$$E^* = \left(1 - \frac{H_2}{k+1}\right)^{-1},$$

$$V^* = \frac{2}{k+1} \left( \frac{1 + c_1 B^*}{(1 - c_2 B^*)^2 (1 - c_3 B^*)} \right),$$

$$\rho = \frac{V^*}{2E^{*2}},$$

$$v = 4 + \frac{k+3}{(k+1)\rho - 1},$$

$$\lambda = \frac{v}{E^*(v-2)}.$$

The  $(1 - \alpha)$  confidence region for vector  $\beta_2$  with use of the Kenward Roger's method is:

$$C_{(1-\alpha)} = \left\{ \beta_2 : (\hat{\hat{\beta}}_2 - \beta_2)' \hat{\hat{\Sigma}}_A^{-1} (\hat{\hat{\beta}}_2 - \beta_2) \leq \frac{(k+1) \cdot F_{k+1,v}(1-\alpha)}{\lambda} \right\}.$$

## 2 ESTIMATION OF PARAMETERS OF TRANSFORMATION FUNCTION BY MAXIMUM LIKELIHOOD METHOD

We would like to find out an estimate of a vector of parameters  $\Theta = (a_0, a_1, \dots, a_k, \sigma_x^2, \sigma_y^2, \mu_1, \dots, \mu_m)'$ . Distribution of random variables  $X_{ij}, Y_{ij}$  is known, for  $i = 1, \dots, m, j = 1, \dots, n$  and we assume, that random variables are independent. Therefore we know their distribution and we can use maximum likelihood procedure for calculation of the estimate of the unknown vector of parameters  $\Theta$ .

We denote a joint probability density function  $\varphi = \varphi_1 \cdots \varphi_n$ , where  $\varphi_i$  is density of  $i$ -th measurement, then:

$$\begin{aligned}
\varphi_1(x_{11}, \dots, y_{m1}; \Theta) &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x_{i1}-\mu_i)^2}{2\sigma_x^2}} \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y_{i1}-a_k\mu_i^k-\dots-a_2\mu_i^2-a_1\mu_i-a_0)^2}{2\sigma_y^2}} \\
&\vdots \\
\varphi_n(x_{1n}, \dots, y_{mn}; \Theta) &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x_{in}-\mu_i)^2}{2\sigma_x^2}} \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y_{in}-a_k\mu_i^k-\dots-a_2\mu_i^2-a_1\mu_i-a_0)^2}{2\sigma_y^2}}.
\end{aligned}$$

We denote  $p_i = \nu_i = a_k\mu_i^k + \dots + a_2\mu_i^2 + a_1\mu_i + a_0$  and derivative  $dp_i = \frac{\partial p_i}{\partial \mu_i} = k \cdot a_k\mu_i^{k-1} + \dots + 2a_2\mu_i + a_1$ .

If  $x_{11}, \dots, y_{mn}$  are measured values, we obtain logarithmic likelihood function:

$$\begin{aligned}
l(x_{11}, \dots, y_{mn}; \Theta) &= \ln L(x_{11}, \dots, y_{mn}; \Theta) = -mn \ln(2\pi) - \frac{mn}{2} \ln \sigma_x^2 - \frac{mn}{2} \ln \sigma_y^2 - \\
&- \frac{1}{2\sigma_x^2} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \mu_i)^2 - \frac{1}{2\sigma_y^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - p_i)^2.
\end{aligned}$$

We compute a maximum likelihood estimate  $\Theta^*$ . We search for  $\Theta^*$ , where  $l(X_{11}, \dots, Y_{mn}, \Theta) \leq l(X_{11}, \dots, Y_{mn}, \Theta^*)$ ,  $\forall \Theta$ .  $\Theta^*$  is estimator of parameter  $\Theta$  obtained by maximum likelihood method, where we assume  $\ln 0 = -\infty$ . We gain estimates of parameters of the transformation function using Matlab software. All functions used in this contribution are available on website <http://www.math.muni.cz/~xsirucko/>.

## 2.1 Confidence region for vector $(a_0, a_1, \dots, a_k)'$

We derive a confidence region for  $\Theta$ , especially for parameters  $a_0, a_1, \dots, a_k$ . According to [4, pg. 160],  $\Theta^*$  is asymptotically unbiased estimator of  $\Theta$  and for  $n$  large enough:

$$\Theta^* \sim N\left(\Theta, \frac{1}{n} \mathbf{J}(\Theta)^{-1}\right),$$

where  $\mathbf{J}(\Theta)$  is the Fisher's information matrix of  $i$ -th measurements:

$$\{\mathbf{J}(\Theta)\}_{ij} = E_{\Theta} \left( -\frac{\partial^2 \ln \varphi_1(X_{11}, \dots, Y_{m1}, \Theta)}{\partial \Theta_i \partial \Theta_j} \right).$$

We focus on the estimate of parameters  $a_0, a_1, \dots, a_k$ . We denote this estimator  $\widehat{\Theta}^k = \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \vdots \\ \widehat{a}_k \end{pmatrix}$  and

denote the exact value of parameters  $a_0, a_1, \dots, a_k$  as  $\Theta^k$ , then according to [4, pg. 160]

$$\sqrt{n} \left( \widehat{\Theta}^k - \Theta^k \right) \xrightarrow[n]{\mathcal{D}} N \left( \mathbf{0}, \mathbf{J} \left( \Theta^k \right)^{-1} \right)$$

(convergence in distribution). For large enough  $n$  we can write:

$$\widehat{\Theta}^k \approx N \left( \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix}, \frac{1}{n} \mathbf{J} \left( \Theta^k \right)^{-1} \right),$$

where  $\mathbf{J} \left( \Theta^k \right)$  is a submatrix of the Fisher information matrix:

$$\mathbf{J} \left( \Theta^k \right) = \begin{pmatrix} \sum_{i=1}^m \frac{1}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i^2}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \cdots & \sum_{i=1}^m \frac{\mu_i^k}{\sigma_y^2 + \sigma_x^2 dp_i^2} \\ \sum_{i=1}^m \frac{\mu_i}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i^2}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i^3}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \vdots & \vdots \\ \sum_{i=1}^m \frac{\mu_i^2}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i^3}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \sum_{i=1}^m \frac{\mu_i^4}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m \frac{\mu_i^k}{\sigma_y^2 + \sigma_x^2 dp_i^2} & \cdots & \cdots & \cdots & \sum_{i=1}^m \frac{\mu_i^{2k}}{\sigma_y^2 + \sigma_x^2 dp_i^2} \end{pmatrix}.$$

We denote:

$$\text{Var}_{\sigma_x^2, \sigma_y^2} = \frac{1}{n} \begin{pmatrix} \frac{m}{2\sigma_x^4} & 0 \\ 0 & \frac{m}{2\sigma_y^4} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2\sigma_x^4}{mn} & 0 \\ 0 & \frac{2\sigma_y^4}{mn} \end{pmatrix}$$

and a covariance matrix

$$\Sigma^k = \frac{1}{n} \mathbf{J} \left( \Theta^k \right)^{-1}.$$

We obtain an asymptotic  $(1 - \alpha)$  confidence region for the vector of parameters  $(a_0, a_1, \dots, a_k)'$

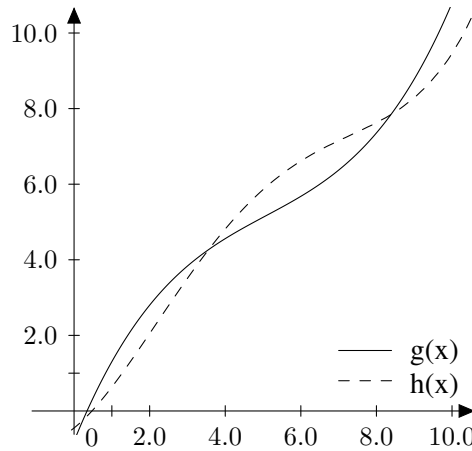
$$*C_{(1-\alpha)}^1 = \left\{ \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix} : \begin{pmatrix} \widehat{a}_0 - a_0 \\ \widehat{a}_1 - a_1 \\ \dots \\ \widehat{a}_k - a_k \end{pmatrix}' (\Sigma^k)^{-1} \begin{pmatrix} \widehat{a}_0 - a_0 \\ \widehat{a}_1 - a_1 \\ \dots \\ \widehat{a}_k - a_k \end{pmatrix} \leq \chi_{k+1}^2 (1 - \alpha) \right\},$$

where  $\chi_{k+1}^2(1 - \alpha)$  is quantile of chi square distribution with  $k + 1$  degrees of freedom. Matrix  $(\Sigma^k)^{-1} = n\mathbf{J}(\Theta^k)$  is unknown. We can replace parameters  $a_0, a_1, \dots, a_k$  by estimators  $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k$ , where  $(\widehat{\Sigma}^k)^{-1} \doteq n\mathbf{J}(\widehat{\Theta}^k)$ . This way we obtain the asymptotic  $(1 - \alpha)$  confidence region for the vector of parameters  $(a_0, \dots, a_n)'$ .

### 3 SIMULATION STUDY

In this part we compare both methods with use of the simulation study. We focused on empirical coverage of the confidence region derived for the replicated model with errors in variables (denoted KR) and empirical coverage of the confidence region based on the maximum likelihood method (denoted ML). We randomly generate matrices of measurement  $\mathbf{X}, \mathbf{Y}$  1000 times and find out if the confidence region covers the actual value of parameters  $a_0, \dots, a_k$  for each pair of matrices. We use the `simulace_mitav.m` program for computations.

Simulations are done for a third degree polynomial  $g(x) = -0,8 + 2,46x - 0,38x^2 + 0,025x^3$  and a polynomial of the fourth degree  $h(x) = -0,45 + 0,8x + 0,35x^2 - 0,07x^3 + 0,0037x^4$ . In the first part we select fixed  $\sigma_x = 0,25, \sigma_y = 0,125, \alpha = 0,05$  and observe the influence of the changing number of measured points. In the second part we have fixed number of the measuring points and we observe the influence of increasing dispersion of both measuring devices.



**Fig. 1.** Transformation curves

Source: Created for the contribution's purposes.



$\sigma_x = 0, 25, \sigma_y = 0, 125$	KR	ML
$\mu = (0; 10/3; 20/3; 10)'$		
n=2	0,858	0,612
n=5	0,918	0,826
n=10	0,946	0,913
n=50	0,951	0,945
$\mu = (0; 2; 5; 5; 7; 5; 10)'$		
n=2	0,878	0,671
n=5	0,941	0,877
n=10	0,948	0,929
n=50	0,948	0,946
$\mu = (0; 2; 4; 6; 8; 10)'$		
n=2	0,869	0,717
n=5	0,937	0,888
n=10	0,946	0,917
n=50	0,961	0,945
$\mu(0; 1; 2; 3; \dots; 10)'$		
n=2	0,916	0,810
n=5	0,941	0,915
n=10	0,950	0,929
n=50	0,956	0,939

$\mu = (1; 3; 5; 7; 9)'$	KR	ML
$\sigma_x = 0, 125, \sigma_y = 0, 0625$		
n=2	0,885	0,660
n=5	0,947	0,897
n=10	0,957	0,919
n=50	0,942	0,936
$\sigma_x = 0, 25, \sigma_y = 0, 125$		
n=2	0,874	0,647
n=5	0,936	0,854
n=10	0,945	0,910
n=50	0,949	0,944
$\sigma_x = 0, 5, \sigma_y = 0, 25$		
n=2	0,867	0,611
n=5	0,933	0,857
n=10	0,935	0,897
n=50	0,946	0,944
$\sigma_x = 1, \sigma_y = 0, 5$		
n=2	0,856	0,603
n=5	0,877	0,812
n=10	0,909	0,872
n=50	0,935	0,931

**Tab. 1.** Results of the simulation study for polynomial  $g(x)$

Source: Created for the contribution's purposes.

$\sigma_x = 0, 25, \sigma_y = 0, 125$	KR	ML
$\mu = (0; 2; 5; 5; 7; 5; 10)'$		
n=2	0,851	0,556
n=5	0,926	0,840
n=10	0,926	0,882
n=50	0,955	0,947
$\mu = (0; 2; 4; 6; 8; 10)'$		
n=2	0,858	0,595
n=5	0,927	0,842
n=10	0,932	0,889
n=50	0,945	0,933
$\mu = (0; 10/6; \dots; 10)'$		
n=2	0,889	0,642
n=5	0,930	0,856
n=10	0,951	0,905
n=50	0,962	0,949
$\mu(0; 10/11; \dots; 10)'$		
n=2	0,874	0,752
n=5	0,932	0,883
n=10	0,949	0,922
n=50	0,948	0,930

$\mu = (0; 2; 4; 6; 8; 10)$	KR	ML
$\sigma_x = 0, 125, \sigma_y = 0, 0625$		
n=2	0,855	0,573
n=5	0,920	0,834
n=10	0,958	0,910
n=50	0,956	0,930
$\sigma_x = 0, 25, \sigma_y = 0, 125$		
n=2	0,857	0,582
n=5	0,919	0,841
n=10	0,932	0,885
n=50	0,951	0,934
$\sigma_x = 0, 5, \sigma_y = 0, 25$		
n=2	0,863	0,607
n=5	0,919	0,845
n=10	0,923	0,872
n=50	0,943	0,922
$\sigma_x = 1, \sigma_y = 0, 5$		
n=2	0,901	0,655
n=5	0,910	0,839
n=10	0,912	0,878
n=50	0,918	0,913

**Tab. 2.** Results of the simulation study for polynomial  $h(x)$

Source: Created for the contribution's purposes.

## CONCLUSION

The aim of the contribution was to compare empirical coverage of the described method based on a simulation study. We can see that empirical coverage of the method based on replicated model with errors in variables is getting closer to the theoretical coverage faster than the empirical coverage of the maximum likelihood method. We obtained comparable result for both coverages for large number of repetitions of measurements. This is caused by asymptotic properties of the maximum likelihood method.

Results for polynomials  $g(x)$  and  $h(x)$  are very similar, therefore we can assume that the degree of a polynomial does not have a significant impact on the empirical coverage (only a minimal number of measuring points is changing due to the degree of the polynomial). Let's consider how number of measuring points can influence the empirical coverage (with a given dispersion). For the replicated model with errors in variables the empirical coverage is near to the theoretical coverage if we repeat the measurement 5 times, even for a minimal number of measuring points plus one. The maximum likelihood method gives a good result if we repeat the measurement 50 times.

Finally, let's assess the dispersion impact on the empirical coverage. The method based on the replicated model with errors in variables gives better empirical coverage than the maximum likelihood method. However, we can see that for some choices of dispersion the replicated model with errors in variables gives good results for  $n \geq 50$ . We can recommend picking up measuring points so that the distance between them is at least four standard deviations (otherwise it can cause computational problems or a large number of repeated measurements might be needed).

## References

- [1] RÁBOŇOVÁ, Petra. Polynomial calibration with use of linearised model with errors in variables and, Kenward Roger type of approximation. In *Dagmar Szarková, Peter Letavaj, Daniela Richtáriková, Monika Prašilová. 16th Conference on Applied Mathematics Aplimat 2017, Proceedings*. Bratislava: Spektrum STU, 2017. s. 1283-1293, 11 s. ISBN 978-80-227-4650-2.
- [2] KUBÁČEK, Lubomír a Ludmila KUBÁČKOVÁ. *Statistika a metrologie*. 1. vyd. Olomouc: Univerzita Palackého, 2000, 307 s. ISBN 8024400936.
- [3] KENWARD, Michael, G., ROGER, James H. *Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood*, Biometric, Volume 53, Issue 3 (Sep.,1997), 983-997.
- [4] ANDĚL, Jiří. *Základy matematické statistiky*. 1. vyd. Praha: Matfyzpress, 2005, 358 s. ISBN 8086732401.
- [5] ANDĚL, Jiří. *Matematická statistika*. 2. vyd. Praha: SNTL - nakladatelství technické literatury, Alfa, vydavatelstvo technickej a ekonomickej literatury, 1985, 346 s.
- [6] ŠIRŮČKOVÁ, Petra. Řešení problému polynommické kalibrace metodou maximalní věrohodnosti. *Forum Statisticum Slovacum*, Slovenská štatistická a demografická spoločnosť, 2014, ro. 6/2014, s. 148-158. ISSN 1336-7420.
- [7] WIMMER, Gejza, PALENČÁR, Rudolf, WITKOVSKÝ, Viktor, ĎURIŠ, Stanislav. *Vyhodnotenie kalibrácie meradiel: štatistické metódy pre analýzu neistôt v metrológii*. V Bratislavě:

Slovenská technická univerzita v Bratislavě, 2015. Edícia monografií, ISBN 978-80-227-4374-7.

- [8] WIMMER, G., WITKOVSKÝ, V., *Univariate linear calibration via replicated errors-in-variables model*, Journal of Statistical Computation and Simulation 77(3), (2007), 213-227
- [9] WIMMER, G., WITKOVSKÝ, V., *Linear comparative calibration with correlated measurements*, Kybernetika 43(4), (2007), 443-452

## **Acknowledgement**

The work presented in this paper has been supported by the MUNI/A/1194/2016.

# ROTARY MAPPINGS OF SURFACES OF REVOLUTION

**Lenka Rýparová, Josef Mikeš**

Department of Algebra and Geometry,  
Faculty of Science, Palacký University Olomouc,  
17. listopadu 1192/12, 774 16 Olomouc

lenka.ryparova01@upol.cz, josef.mikes@upol.cz

**Abstract:** *Presented paper concerns with rotary mappings of surfaces of revolution. It is proved that any surface of revolution with differentiable Gaussian curvature admits rotary mapping. Furthermore, same holds even for (pseudo-) Riemannian spaces.*

**Keywords:** isoperimetric extremal of rotation, rotary diffeomorphism, surface of revolution, (pseudo-) Riemannian space.

## INTRODUCTION

Special diffeomorphisms between (pseudo-) Riemannian manifold and manifold with affine or projective connection, for which any special curve maps onto a special curve, were studied in many works. For example, geodesic [1, 3, 5, 16, 22, 26], holomorphically-projective [4, 17, 18], and more special mappings [6, 14, 21, 23, 24]. These problems can be found in the more developed form in monographs [15, 20].

Leiko studied interesting questions about rotary mapping in [7, 8, 9, 10, 11, 12, 13]. Newly found results have their application in the theory about gravitational fields, for example [7, 10, 12]. He proved that certain surfaces of revolution which metric is in special form admit rotary mapping [8]. Leiko continued this research with Vinnik cooperation, see [25].

New results in theory of isoperimetric extremals of rotation and rotary diffeomorphisms were obtained by Mikeš, Stepanova and Sochor [19]. Generalization of these terms is in the work of Chudá, Mikeš and Sochor [2].

In this paper we deal with rotary mappings of surfaces of revolution and Riemannian manifolds which are isometric with these surfaces. We manage to prove that any surface of revolution with differentiable Gaussian curvature admits rotary diffeomorphism and same holds for (pseudo-) Riemannian spaces. These results have local validity.

## 1 ISOPERIMETRIC EXTREMALS OF ROTATION

Leiko [8] was the first one to introduce term of isoperimetric extremals of rotation on two-dimensional Riemannian spaces  $\mathbb{V}_2$  and surfaces  $\mathcal{S}_2$  with metric  $g$ .

**Definition 1** ([8]). A curve  $\ell: x = x(t)$  on surface or on two-dimensional (pseudo-) Riemannian space is called the *isoperimetric extremal of rotation* if  $\ell$  is extremal of functionals  $\theta[\ell]$  and  $s[\ell] = \text{const}$  with fixed ends.

Here

$$s[\ell] = \int_{t_0}^{t_1} |\lambda| \, dt \quad \text{and} \quad \theta[\ell] = \int_{t_0}^{t_1} k(t) \, dt,$$

where  $k(t)$  is the curvature and  $|\lambda|$  is the length of the tangent vector  $\lambda$  of  $\ell$ .

In [8, 11] Leiko proved that a curve  $\ell$  is an isoperimetric extremal of rotation if and only if its Frenet curvature  $k$  and Gaussian curvature  $K$  are proportional

$$k = c \cdot K,$$

where  $c$  is a constant. For  $c = 0$  we get a geodesic.

Mikeš, Stepanova and Sochor [19] found new simpler form of equations of isoperimetric extremal of rotation

$$\nabla_s \lambda = c \cdot K \cdot F \lambda,$$

where  $c$  is a constant,  $s$  is the arc length,  $F$  is a tensor  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  which satisfies the conditions

$$F^2 = -e \cdot Id, \quad g(X, FX) = 0, \quad \nabla F = 0. \quad (1)$$

For Riemannian manifold  $\mathbb{V}_2$  is  $e = +1$  and  $F$  is *complex structure* and for (pseudo-) Riemannian manifold is  $e = -1$  and  $F$  is a *product structure*. This tensor  $F$  is uniquely defined (with the respect to the sign) with using skew-symmetric and covariantly constant discriminant tensor  $\epsilon_{ij}$ , which is defined

$$F_j^h = g^{hi} \epsilon_{ij}, \quad \epsilon_{ij} = \sqrt{|g_{11}g_{22} - g_{12}^2|} \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

## 2 ROTARY DIFFEOMORPHISM

In [8] there was introduced the term of rotary diffeomorphism between two-dimensional Riemannian spaces  $\mathbb{V}_2$  and surfaces  $\mathcal{S}_2$  with metric  $g$ .

**Definition 2** ([8]). A diffeomorphism between two-dimensional (pseudo-) Riemannian manifolds  $\mathbb{V}_2$  and  $\bar{\mathbb{V}}_2$  is called *rotary* if any geodesic on  $\bar{\mathbb{V}}_2$  is mapped onto isoperimetric extremal of rotation on  $\mathbb{V}_2$ .

Chudá, Mikeš, Sochor later generalized the definition itself to the following form:

**Definition 3** ([2]). A diffeomorphism  $f: \mathbb{V}_2 \rightarrow \bar{\mathbb{A}}_2$  is called *rotary mapping* if any geodesic on manifold  $\bar{\mathbb{A}}_2$  with affine connection  $\bar{\nabla}$  is mapped onto isoperimetric extremal of rotation on two-dimensional (pseudo-) Riemannian manifold  $\mathbb{V}_2$ .

If the definition was formulated the other way around: A diffeomorphism between two-dimensional (pseudo-) Riemannian manifolds  $\mathbb{V}_2$  and  $\bar{\mathbb{V}}_2$  is called *rotary* if any isoperimetric extremal of rotation on  $\mathbb{V}_2$  is mapped onto geodesic on  $\bar{\mathbb{V}}_2$ . Then this mapping would be geodesic mapping.

Later, some new properties were proved, see [2]: When  $\mathbb{V}_2$  admits rotary mapping  $f$  onto  $\bar{\mathbb{A}}_2$  then if  $\mathbb{V}_2$  and  $\bar{\mathbb{A}}_2$  in common coordinate system belong differentiability class  $C^2$  and  $C^1$ , respectively, then Gaussian curvature  $K$  on  $\mathbb{V}_2$  is differentiable. As a result they formulated new theorem: Rotary diffeomorphism  $\mathbb{V}_2 \rightarrow \bar{\mathbb{A}}_2$  does not exist if Gaussian curvature  $K \notin C^1$ .

Chudá, Mikeš and Sochor [2] later proved that (pseudo-) Riemannian manifold  $\mathbb{V}_2$  admits rotary mapping onto  $\bar{\mathbb{A}}_2$  if and only if in  $\mathbb{V}_2$  holds equation

$$\theta^h_{,j} = \theta^h(\theta_j + \partial_j \ln |K|) + \nu \delta^h_j \quad (2)$$

where  $\theta_i = g_{i\alpha} \theta^\alpha$ ,  $\nu$  is a function on  $\mathbb{V}_2$  and vector field  $\theta^h$  is a special case of torse-forming field. Here and after comma denotes covariant derivative respective connection  $\nabla$ . and  $\partial_1 = \partial/\partial x^1$ .

### 3 ROTARY MAPPINGS OF SURFACES OF REVOLUTION

Leiko [8] has also studied rotary mappings of surfaces of revolution. Here, he used the metric of the surface of revolution  $\mathcal{S}_2$  in the following form

$$ds^2 = f(r) dr^2 + r^2 d\varphi^2. \quad (3)$$

Leiko analyzed the equations (2) and proved a theorem that vector fields (2) exist in Riemannian space  $\mathbb{V}_2$  if and only if  $\mathbb{V}_2$  is isometric with surface of revolution  $\mathcal{S}_2$  and the metrics of  $\mathbb{V}_2$  has one from the following forms:

$$(\tilde{g}_{ij}) = \frac{f(r)}{A^2(B + \sqrt{f(r)})^2} \text{diag}(f(r), r^2),$$

$$(\tilde{g}_{ij}) = B^2 f(r) \text{diag}(f(r), r^2), \quad \text{where } A \neq 0 \text{ and } B \text{ are const.}$$

Above mentioned was formulated in Theorem 2, see [8].

Let us remind that the metric of the Riemannian space  $\mathbb{V}_2$  (that is induced by the surface of revolution  $\mathcal{S}_2$ ) in certain coordinate system can be written in the form

$$ds^2 = (dx^1)^2 + f(x^1) (dx^2)^2, \quad (4)$$

where  $f (\neq 0)$  is a certain function of  $x^1$ .

Note that the metric (4) of the surface of revolution  $\mathcal{S}_2$  is more general than the metric in form (3), which was used by Leiko. The metric (4) also includes gorge circles, which are in (3) basically excluded.

We note that existence of coordinate system (4) is connected with existence of anisotropic concircular vector field  $\lambda$  which is characterized by the equations

$$\lambda_i^h = \rho \delta_i^h,$$

where  $\rho$  is a certain function on  $\mathbb{V}_n$ .

Concircular vector fields has been studied in 1940's by K. Yano and in 1950's by N.S. Sinyukov, who called spaces with those vector fields *equidistant*, see [15] pp. 140–155. Existence of these vector fields on the surface is a criteria of local isometry of surface of revolution.

In  $\mathbb{V}_2$  vector field  $\lambda$  generates Killing vector. This vector has the following form

$$\nu^h = \lambda^\alpha F_\alpha^h,$$

where  $F$  is from formula (1). Evidently, for  $\nu_i = \nu^\alpha g_{\alpha i}$  it holds  $\nu_{(i,j)} = 0$ . From the other side Killing vector  $\nu^h$  generates concircular vector field.

Locally,  $\mathbb{V}_2$  with metric (4) realises as surface of revolution in Euclidean space  $\mathbb{E}_3$  given by the equations

$$x = F(x^1) \cos x^2, \quad y = F(x^1) \sin x^2, \quad z = z(x^1),$$

here  $f = F^2$ .

In case the metric  $ds^2$  is indefinite the surface is given by the equations

$$x = F(x^1) \cosh x^2, \quad y = F(x^1) \sinh x^2, \quad z = z(x^1),$$

where  $(x, y, z)$  are coordinates in Minkowski space, which metric has the form

$$ds^2 = dx^2 - dy^2 + dz^2$$

therefore for our example

$$ds^2 = (F'^2 + z'^2) (dx^1)^2 - F^2 (dx^2)^2.$$

Further, we are going to prove that any surface of revolution admits rotary mapping. Moreover, any Riemannian space  $\mathbb{V}_2$  that is isometric with such surface of revolution  $\mathcal{S}_2$ , and also any pseudo-Riemannian space  $\mathbb{V}_2$  which metric has form (4) admits rotary mapping onto space  $\bar{\mathbb{A}}_2$ .

Note that all the results we obtained have local validity.

## 4 NEW RESULTS IN THEORY OF ROTARY MAPPINGS

In this section we are going to prove that vector fields (2) exist in any Riemannian space  $\mathbb{V}_2$  which is isometric with surface of revolution  $\mathcal{S}_2$ . Firstly, we formulate the following theorem.

**Theorem 1.** *Any surface of revolution  $\mathcal{S}_2$  with differentiable Gaussian curvature  $K$  admits rotary mapping onto  $\bar{\mathbb{A}}_2$ .*

*Proof.* In the proof of this theorem we use Theorem 5 from [2]. Therefore, if we prove that on any surface of revolution  $\mathcal{S}_2$  exist vector fields which satisfy condition

$$\theta_{,i}^h \equiv \partial_i \theta^h + \theta^\alpha \Gamma_{\alpha i}^h = \theta^h (\theta_i + \partial_i \ln |K|) + \nu \delta_i^h, \quad (5)$$

where  $\theta_h = g_{h\alpha} \theta^\alpha$  and  $\nu$  is a function on  $\mathcal{S}_2$ , then rotary mapping exists.

We choose the metric of the surface of revolution  $\mathcal{S}_2$  in the following form

$$ds^2 = (dx^1)^2 + f(x^1) (dx^2)^2. \quad (6)$$

Components of the metric tensor  $g$  and its inverse tensor have the following form

$$g_{11} = 1, \quad g_{12} = 0, \quad g_{22} = f(x^1) \quad \text{and} \quad g^{11} = 1, \quad g^{12} = 0, \quad g^{22} = 1/f(x^1)$$

Hence, we can calculate the Christoffel symbols of the first kind  $\Gamma_{ijk} = 1/2 (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})$  which are

$$\Gamma_{122} = \Gamma_{212} = \frac{1}{2} f'(x^1) \quad \text{and} \quad \Gamma_{221} = -\frac{1}{2} f'(x^1),$$

the others are vanishing. The Christoffel symbols of the second kind  $\Gamma_{ij}^h = g^{hk} \Gamma_{ijk}$  are

$$\Gamma_{12}^2 = \Gamma_{21}^2 = \frac{1}{2} \frac{f'(x^1)}{f(x^1)} \quad \text{and} \quad \Gamma_{22}^1 = -\frac{1}{2} f'(x^1).$$

Well known Gaussian curvature  $K$  satisfies formula  $R_{1212} = K \cdot (g_{11}g_{22} - g_{12}^2)$ , where

$$R_{hijk} = g_{h\alpha} R_{ijk}^\alpha$$

are components of Riemann tensor of first type and

$$R_{ijk}^h = \partial_j \Gamma_{ik}^h - \partial_k \Gamma_{ij}^h + \Gamma_{ik}^\alpha \Gamma_{\alpha j}^h - \Gamma_{ij}^\alpha \Gamma_{\alpha k}^h.$$

We calculate the Gaussian curvature  $K$  of the surface  $\mathcal{S}_2$

$$K = \frac{1}{4} \left( \frac{f'(x^1)}{f(x^1)} \right)^2 - \frac{1}{2} \frac{f''(x^1)}{f(x^1)}. \quad (7)$$

Let us suppose  $\theta^h = a(x^1) \delta_1^h$ , thus from (5) we obtain following equations

$$\begin{aligned} a'(x^1) &= a(x^1) \cdot \left( a(x^1) + \frac{K'}{K} \right) + \nu(x^1), \\ \frac{1}{2} a(x^1) \frac{f'(x^1)}{f(x^1)} &= \nu(x^1). \end{aligned}$$

Now we merge these equations and obtain the following relation

$$a' = a^2 + a \cdot \left( \frac{K'}{K} + \frac{1}{2} \frac{f'}{f} \right). \quad (8)$$



The equation (8) is an ordinary differential equation of Bernoulli type. We use the substitution  $u = \frac{1}{a}$  and get an inhomogeneous linear ordinary differential equation

$$u' = -1 - u \cdot \left( \frac{K'}{K} + \frac{1}{2} \frac{f'}{f} \right).$$

Now we use the method of variation of parameters to solve the equation and we get a particular solution  $u(x^1) = \frac{c(x^1)}{K\sqrt{f(x^1)}}$ . By substituting the particular solution into the inhomogeneous equation, we find  $c'(x^1) = -K\sqrt{f(x^1)}$ , thus

$$u(x^1) = \frac{1}{K\sqrt{f(x^1)}} \left( - \int K\sqrt{f(x^1)} \, dx^1 \right).$$

As Gaussian curvature  $K$  has a special form (7) then we obtain

$$u(x^1) = \frac{1}{K\sqrt{f(x^1)}} \left( C + \frac{f'(x^1)}{2\sqrt{f(x^1)}} \right),$$

where  $C$  is a constant of integration.

Consequently, the function  $a(x^1)$  has the following form

$$a(x^1) = \frac{2K \cdot f(x^1)}{f'(x^1) + 2C\sqrt{f(x^1)}}, \quad (9)$$

where  $C$  is the constant of integration.

Since the function  $a(x^1)$  in (8) has to be differentiable so does the Gaussian curvature  $K$  of the surface  $S_2$ . It is evident, that the function  $a(x^1)$  in (9) is the solution of differential equation (8). Therefore, vector fields exist for any surface of revolution and the theorem is proved.  $\square$

As was mentioned above, the metric (6) of the surface of revolution  $S_2$  is more general than the metric (3) used by Leiko in [8]. Unlike the metric (3), it includes gorge circles.

In the proof of Theorem 1 the metric is used in the form (6) therefore Theorem holds for any Riemannian space  $\mathbb{V}_2$  which is isometric with surface of revolution  $S_2$ . Moreover, this Theorem holds even for pseudo-Riemannian spaces which have indefinite metric for which  $f(x^1) < 0$ . In this case instead of  $\sqrt{f(x^1)}$  we write  $\sqrt{|f(x^1)|}$ .

General solution of (2) in system (6) is:

$$\theta^h = a \cdot \delta_1^h \quad \text{where} \quad a(x^1) = \frac{2K \cdot f}{f' + 2C\sqrt{f}}, \quad (10)$$

this solution depends on one parameter  $C$ .

In studied rotary mapping, Riemannian space  $\mathbb{V}_2$  maps onto manifold with affine connection  $\bar{\mathbb{A}}_2$ . For chosen constant  $C$  we denote corresponding manifold with affine connection  $\bar{\mathbb{A}}_2(C)$ . In the following part we are going to prove, that there does not exist geodesic mapping between any two manifolds  $\bar{\mathbb{A}}_2(C_i)$  and  $\bar{\mathbb{A}}_2(C_j)$  for  $i \neq j$ .

**Theorem 2.** *Manifolds  $\bar{\mathbb{A}}_2(C_1)$  and  $\bar{\mathbb{A}}_2(C_2)$  with affine connection for  $C_1 \neq C_2$  are not geodesically connected.*

*Proof.* Let us suppose that  $\bar{\mathbb{A}}_2(C_1)$  and  $\bar{\mathbb{A}}_2(C_2)$  are images of space  $\mathbb{V}_2$  in rotary mapping. Then the following equations hold

$$\mathcal{T}_{ij}^h(x) = \Gamma_{ij}^h(x) + \delta_{(i}^h \psi_{j)} + \theta^h \cdot g_{ij}$$

$$\bar{\mathcal{T}}_{ij}^h(x) = \Gamma_{ij}^h(x) + \delta_{(i}^h \bar{\psi}_{j)} + \bar{\theta}^h \cdot g_{ij},$$

where  $\psi_i, \theta^h$  are solutions of rotary mapping  $\mathbb{V}_2 \longrightarrow \bar{\mathbb{A}}_2(C_1)$  respective  $\bar{\psi}_i, \bar{\theta}^h$  are solutions of rotary mapping  $\mathbb{V}_2 \longrightarrow \bar{\mathbb{A}}_2(C_2)$ . Here  $\mathcal{T}_{ij}^h$  and  $\bar{\mathcal{T}}_{ij}^h$  are components of manifolds  $\bar{\mathbb{A}}_2(C_1)$  and  $\bar{\mathbb{A}}_2(C_2)$ .

We subtract these equations and get

$$\bar{\mathcal{T}}_{ij}^h(x) - \mathcal{T}_{ij}^h(x) = \delta_{(i}^h \bar{\psi}_{j)} - \delta_{(i}^h \psi_{j)} + (\bar{\theta}^h - \theta^h) \cdot g_{ij}$$

therefore

$$\delta_{(i}^h \omega_{j)} + (\bar{\theta}^h - \theta^h) \cdot g_{ij} = 0.$$

For indices  $h = 1$ , resp.  $h = 2$  we obtain

$$\delta_{(i}^1 \omega_{j)} + (\bar{\theta}^1 - \theta^1) \cdot g_{ij} = 0, \quad \text{resp.} \quad \delta_{(i}^2 \omega_{j)} = 0,$$

thus  $\omega_2 = 0, \omega_1 = 0$ . From  $\omega_i = 0$  it follows that  $\bar{\theta}^h = \theta^h$ . Using (10), we obtain

$$\frac{1}{K\sqrt{f(x^1)}} \left( C_2 + \frac{f'(x^1)}{2\sqrt{f(x^1)}} \right) = \frac{1}{K\sqrt{f(x^1)}} \left( C_1 + \frac{f'(x^1)}{2\sqrt{f(x^1)}} \right)$$

thus  $C_1 = C_2$  which is contradiction with assumption  $C_1 \neq C_2$  therefore the theorem is proved.  $\square$

## CONCLUSION

The paper is devoted to study of rotary mappings of two-dimensional Riemannian spaces, which are isometric to surfaces of revolution, and also two-dimensional pseudo-Riemannian spaces, which have analogical metric form. We manage to prove that any surface of revolution with differentiable Gaussian curvature admits rotary mapping. These results are more general than those presented by Leiko. Furthermore, there does not exists geodesic mapping between any two manifolds  $\bar{\mathbb{A}}_2(C_i)$  and  $\bar{\mathbb{A}}_2(C_j)$  (for  $i \neq j$ ) mentioned above. In addition, we also proved that any two-dimensional equidistant pseudo-Riemannian space  $\mathbb{V}_2$  with differentiable Gaussian curvature  $K$  admits rotary mapping onto  $\bar{\mathbb{A}}_2$ .

## References

- [1] Dini, U. On a problem in the general theory of the geographical representations of a surface on another. *Analisi di Mat.*, 3, 1869, p. 269–294.
- [2] Chudá, H., Mikeš, J., Sochor, M. Rotary diffeomorphism onto manifolds with affine connection. In: *Geometry, Integrability and Quantization 18, proc. of 18th Int. Conf.* Sofia: Bulgaria, 2017, p. 130–137. ISSN 1314-3247.
- [3] Hinterleitner, I. Geodesic mappings on compact Riemannian manifolds with conditions on sectional curvature. *Publ. Inst. Math.*, 94(108), 2013, p. 125–130.
- [4] Hinterleitner, I., Mikeš, J. Fundamental equations of geodesic mappings and their generalizations. *J. Math. Sci.*, 174, 2011, p. 537–554.
- [5] Hinterleitner, I., Mikeš, J. Geodesic mappings and differentiability of metrics, affine and projective connections. *Filomat*, 29, 2015, p. 1245–1249.
- [6] Kuzmina, I., Mikeš, J. On pseudoconformal models of fibrations determined by the algebra of antiquaternions and projectivization of them. *Ann. Math. Inform.*, 42, 2013, p. 57–64.
- [7] Leiko, S. Conservation laws for spin trajectories generated by isoperimetric extremals of rotation. *Gravitation and Theory of Relativity*, 26, 1988, p. 117–124.
- [8] Leiko, S. Rotary diffeomorphisms on Euclidean spaces. *Mat. Zametki*, 47(3), 1990, p. 52–57.
- [9] Leiko, S. Variational problems for rotation functionals, and spin-mappings of pseudo-Riemannian spaces. *Sov. Math.*, 34(10), 1990, p. 9–18.
- [10] Leiko, S. Extremals of rotation functionals of curves in a pseudo-Riemannian space, and trajectories of spinning particles in gravitational fields. *Russian Acad. Sci. Dokl. Math.*, 46, 1993, p. 84–87.
- [11] Leiko, S. Isoperimetric extremals of a turn on surfaces in Euclidean space  $\mathbb{E}^3$ . *Izv. Vyssh. Uchebn. Zaved. Mat.*, 6, 1996, p. 25–32.
- [12] Leiko, S. On the conformal, concircular, and spin mappings of gravitational fields. *J. Math. Sci.*, 90, 1998, p. 1941–1944.
- [13] Leiko, S. G. Isoperimetric problems for rotation functionals of the first and second orders in (pseudo) Riemannian manifolds. *Russ. Math.*, 49, 2005, p. 45–51.
- [14] Mencáková, K., Diblík, J. Formula for explicit solutions of a class of linear discrete equations with delay. In: *Mathematics, Information Technologies and Applied Sciences 2016*. Brno: Univerzita obrany, Brno, 2016, p. 42–55. ISBN: 978-80-7231-400-3.
- [15] Mikeš, J., et al. *Differential geometry of special mappings*. Olomouc: Palacky Univ. Press, 2015, 566 pp. ISBN 978-80-244-4671-4.
- [16] Mikeš, J. Geodesic mappings of affine-connected and Riemannian spaces. *J. Math. Sci.*, 78, 1996, p. 311–333.
- [17] Mikeš, J. Holomorphically Projective mappings and their generalizations. *J. Math. Sci.*, 89, 1998, p. 1334–1353.
- [18] Mikeš, J., Berezovski, V. E., Stepanova, E., Chudá, H. Geodesic mappings and their generalizations. *J. Math. Sci.*, 217(5), 2016, p. 607–623.
- [19] Mikeš, J., Sochor, M., Stepanova, E. On the existence of isoperimetric extremals of rotation and the fundamental equations of rotary diffeomorphism. *Filomat*, 29(3), 2015, p. 517–523.
- [20] Mikeš, J., Vanžurová, A., Hinterleitner, I. *Geodesic mappings and some generalizations*. Olomouc: Palacky Univ. Press, 2009, 304 pp. ISBN 978-8-244-2524-5.
- [21] Najdanović, M. S., Velimirović, L. S. On the Willmore energy of curves under second order infinitesimal bending. *Miskolc Mathematical Notes*, 17(2), 2016, p. 979–987.

- [22] Najdanović, M. S., Zlatanović, M., Hinterleitner, I. Conformal and geodesic mappings of generalized equidistant spaces. *Publ. Inst. Math*, 98(112), 2015, p. 71–84.
- [23] Petrov, A. Modeling of the paths of test particles in gravitation theory. *Gravit. and the Theory of Relativity*, 4(5), 1968, p. 7–21.
- [24] Stepanov, S., Shandra, I., Mikeš, J. Harmonic and projective diffeomorphisms. *J. Math. Sci.*, 207, 2015, p. 658–668.
- [25] Vinnik, A. V. Leiko, S. The property of reciprocity of rotary diffeomorphisms of two-dimensional Riemannian spaces. *Differ. Geom. Mnogoobr. Figur*, 29, 1998, p. 13–16.
- [26] Zlatanović, M., Velimirović, L., Stanković, M. Necessary and sufficient conditions for equitortion geodesic mapping. *J. Math. Anal. Appl.*, 435, 2016, p. 578–592.

### **Acknowledgement**

The paper was supported by the project IGA PrF 2017012 Palacký University Olomouc.

# AFFINE LAGRANGIANS IN SECOND ORDER FIELD THEORY

Dana Smetanová

Department of Informatics and Natural Sciences, Faculty of Technology,  
Okružní 10, 370 01 České Budějovice, Czech Republic  
email: smetanova@mail.vstecb.cz

**Abstract:** *In the present paper we consider an extension of the classical Hamilton–Cartan variational theory on fibred manifold. It is known that in field theory to a variational problem represented by a Lagrangian one can associate different Hamilton equations corresponding to different Lepagean equivalents of the Lagrangian. The case of Lagrangians affine in second derivatives is studied by tools of differential geometry. New regularity and strong regularity conditions and Legendre transformations are found.*

**Keywords:** Lagrangian, Legendre transformation, regularity, Hamilton equations.

## INTRODUCTION

The aim of this paper is to apply an extension of the classical Hamilton–Cartan variational theory on fibred manifolds to the case of class of second order Lagrangians affine in second derivatives. In the generalized Hamiltonian field theory, to a variational problem represented by a Lagrangian one can associate different Hamilton equations corresponding to different Lepagean equivalents of the Euler–Lagrange form. The arising Hamilton equations and regularity conditions depend not only on a Lagrangian, but also on some “free” functions, which correspond to the choice of a concrete Lepagean equivalent. Within this setting, a proper choice of a Lepagean equivalent can lead to a “regularization” of a Lagrangian.

A regularization (by different methods) of some interesting singular physical fields (the Dirac field, the Electromagnetic field and Scalar Curvature Lagrangians) has been studied in [2], [3], [5] and [7], some second order Lagrangians have been discussed also in [11], [12]. In [12] the regularization of non-affine Lagrangians (singular in standard Hamilton–De Donder sense) has been investigated. The multisymplectic approach has been proposed in [1], [9] and [15].

Note that an alternative approach to the study of “degenerated” Lagrangians (singular in standard sense) is the constraint theory in mechanics (see [13], [14]) and in the field (c.f. [8]).

In the paper [11] properties (e.g., regularity, Legendre transformation) of Hamilton  $p_2$ -equations for second order Lagrangian affine in second derivatives are studied. The Hamilton  $p_2$ -equations for second order Lagrangian are created from Lepagean equivalents whose order of contactness is maximal 2. This paper is generalization of the paper [11] to Hamilton equations whose arise from Lepagean equivalent more than 2-contact.

The paper is devoted to second order Lagrangians affine in second derivatives. All these Lagrangians are singular in the standard Hamilton–De Donder theory and its do not admit Legendre transformation. However, in the generalized setting, the question on existence of regular Hamilton equations has sense. For such Lagrangian the set of Lepagean equivalents (resp. family of Hamilton equations) regular in the generalized sense is found and a generalized Legendre transformation is proposed. Note that the generalized momenta  $p_\sigma^{ij}$  satisfy  $p_\sigma^{ij} \neq p_\sigma^{ji}$ .

The correspondence between solutions of Euler–Lagrange and Hamilton equations is studied. The regularity conditions are found (ensuring that the Hamilton extremals are holonomic up to the second order). These conditions depend on a choice of a Hamiltonian system (i.e., “free” functions).

## 1 PRELIMINARIES AND NOTATION

Throughout the paper all manifolds and mappings are smooth and summation convention is used. We consider a fibred manifold (i.e., surjective submersion)  $\pi : Y \rightarrow X$ ,  $\dim X = n$ ,  $\dim Y = n+m$ , its  $r$ -jet prolongation  $\pi_r : J^r Y \rightarrow X$ ,  $r \geq 1$  and canonical jet projections  $\pi_{r,k} : J^r Y \rightarrow J^k Y$ ,  $0 \leq k \leq r$  (with an obvious notations  $J^0 Y = Y$ ). A fibred char on  $Y$  (resp. associated fibred chart on  $J^r Y$ ) is denoted by  $(V, \psi)$ ,  $\psi = (x^i, y^\sigma)$  (resp.  $(V_r, \psi_r)$ ,  $\psi_r = (x^i, y^\sigma, y_i^\sigma, \dots, y_{i_1 \dots i_r}^\sigma)$ ). A vector field  $\xi$  on  $J^r Y$  is called  $\pi_r$ -vertical (resp.  $\pi_{r,k}$ -vertical) if it projects onto the zero vector field on  $X$  (resp. on  $J^k Y$ ).

Recall that every  $q$ -form  $\eta$  on  $J^r Y$  admits a unique (canonical) decomposition into a sum of  $q$ -forms on  $J^{r+1} Y$  as follows [4]:

$$\pi_{r+1,r}^* \eta = h\eta + \sum_{k=1}^q p_k \eta,$$

where  $h\eta$  is a horizontal form, called the *horizontal part of  $\eta$* , and  $p_k \eta$ ,  $1 \leq k \leq q$ , is a  *$k$ -contact part of  $\eta$* .

We use the following notations:

$$\omega_0 = dx^1 \wedge dx^2 \wedge \dots \wedge dx^n, \quad \omega_i = i_{\partial/\partial x^i} \omega_0, \quad \omega_{ij} = i_{\partial/\partial x^j} \omega_i,$$

and

$$\omega^\sigma = dy^\sigma - y_j^\sigma dx^j, \quad \dots, \quad \omega_{i_1 i_2 \dots i_k}^\sigma = dy_{i_1 i_2 \dots i_k}^\sigma - y_{i_1 i_2 \dots i_k j}^\sigma dx^j$$

For more details on fibred manifolds and the corresponding geometric structures we refer e.g. to [10].

We briefly recall basic concepts on Lepagean equivalents of Lagrangians, due to Krupka [4], and on Lepagean equivalents of Euler–Lagrange forms and generalized Hamiltonian field theory, due to Krupková [6].

By an  $r$ -th order Lagrangian we shall mean a horizontal  $n$ -form  $\lambda$  on  $J^r Y$ .

A  $n$ -form  $\rho$  is called a *Lepagean equivalent of a Lagrangian  $\lambda$*  if (up to a projection)  $h\rho = \lambda$ , and  $p_1 d\rho$  is a  $\pi_{r+1,0}$ -horizontal form.

For an  $r$ -th order Lagrangian we have all its Lepagean equivalents of order  $(2r - 1)$  characterized by the following formula

$$\rho = \Theta + \bar{\mu}, \tag{1}$$

where  $\Theta$  is a (global) Poincaré–Cartan form associated to  $\lambda$  and  $\bar{\mu}$  is an arbitrary  $n$ -form of order of contactness  $\geq 2$ , i.e., such that  $h\bar{\mu} = p_1 \bar{\mu} = 0$ . Recall that for a Lagrangian of order 1,  $\Theta = \theta_\lambda$

where  $\theta_\lambda$  is the classical Poincaré–Cartan form of  $\lambda$ . If  $r \geq 2$ ,  $\Theta$  is no more unique, however, there is an *non-invariant* decomposition

$$\Theta = \theta_\lambda + p_1 d\nu, \quad (2)$$

where

$$\theta_\lambda = L\omega_0 + \sum_{k=0}^{r-1} \left( \sum_{l=0}^{r-k-1} (-1)^l d_{p_1} d_{p_2} \dots d_{p_l} \frac{\partial L}{\partial y_{j_1 \dots j_k p_1 \dots p_l}^\sigma} \right) \omega_{j_1 \dots j_k}^\sigma \wedge \omega_i, \quad (3)$$

and  $\nu$  is an arbitrary at least 1-contact  $(n-1)$ -form.

A closed  $(n+1)$ -form  $\alpha$  is called a *Lepagean equivalent of an Euler–Lagrange form*  $E = E_\sigma \omega^\sigma \wedge \omega_0$  if  $p_1 \alpha = E$ .

Recall that the Euler–Lagrange form corresponding to an  $r$ -th order  $\lambda = L\omega_0$  is the following  $(n+1)$ -form of order  $\leq 2r$

$$E = \left( \frac{\partial L}{\partial y^\sigma} - \sum_{l=1}^r (-1)^l d_{p_1} d_{p_2} \dots d_{p_l} \frac{\partial L}{\partial y_{p_1 \dots p_l}^\sigma} \right) \omega^\sigma \wedge \omega_0. \quad (4)$$

By definition of a Lepagean equivalent of  $E$ , one can find using Poincaré lemma local forms  $\rho$ , such that  $\alpha = d\rho$ , and  $\rho$  is an Lepagean equivalent of a Lagrangian for  $E$ . The family of Lepagean equivalents of  $E$  is also called a *Lagrangian system*, and denoted by  $[\alpha]$ . The corresponding Euler–Lagrange equations now take the form

$$J^s \gamma^* i_{J^s \xi} \alpha = 0 \quad \text{for every } \pi - \text{vertical vector field } \xi \text{ on } Y, \quad (5)$$

where  $\alpha$  is any representative of order  $s$  of the class  $[\alpha]$ . A (single) Lepagean equivalent  $\alpha$  of  $E$  on  $J^s Y$  is also called a *Hamiltonian system of order  $s$*  and the equations

$$\delta^* i_\xi \alpha = 0 \quad \text{for every } \pi_s - \text{vertical vector field } \xi \text{ on } J^s Y \quad (6)$$

are called *Hamilton equations*. They represent equations for integral sections  $\delta$  (called *Hamilton extremals*) of the *Hamiltonian ideal*, generated by the system  $\mathcal{D}_\alpha^s$  of  $n$ -forms  $i_\xi \alpha$ , where  $\xi$  runs over  $\pi_s$ -vertical vector fields on  $J^s Y$ . Also, considering  $\pi_{s+1}$ -vertical vector fields on  $J^{s+1} Y$ , one has the ideal  $\mathcal{D}_\alpha^{s+1}$  of  $n$ -forms  $i_\xi \hat{\alpha}$  on  $J^{s+1} Y$ , where  $\hat{\alpha}$  (called *principal part* of  $\alpha$ ) denotes the at most 2-contact part of  $\alpha$ . Its integral sections which moreover annihilate all at least 2-contact forms, are called *Dedecker–Hamilton extremals*. It holds that if  $\gamma$  is an extremal then its  $s$ -prolongation (resp.  $(s+1)$ -prolongation) is a Hamilton (resp. Dedecker–Hamilton) extremal, and (up to projection) every Dedecker–Hamilton extremal is a Hamilton extremal.

Denote by  $r_0$  the minimal order of Lagrangians corresponding to  $E$ . A Hamiltonian system  $\alpha$  on  $J^s Y$ ,  $s \geq 1$ , associated with  $E$  is called *regular* if the system of local generators of  $\mathcal{D}_\alpha^{s+1}$  contains all the  $n$ -forms

$$\omega^\sigma \wedge \omega_i, \omega_{(j_1}^\sigma \wedge \omega_i), \dots, \omega_{(j_1 \dots j_{r_0-1}}^\sigma \wedge \omega_i), \quad (7)$$

where  $(\dots)$  denotes symmetrization in the indicated indices. If  $\alpha$  is regular then every Dedecker–Hamilton extremal is holonomic up to the order  $r_0$ , and its projection is an extremal. (In case of

first order Hamiltonian systems there is an bijection between extremals and Dedecker–Hamilton extremals).  $\alpha$  is called *strongly regular* if the above correspondence holds between extremals and Hamilton extremals. It can be proved that every strongly regular Hamiltonian system is regular, and it is clear that if  $\alpha$  is regular and such that  $\alpha = \hat{\alpha}$  then it is strongly regular. A Lagrangian system is called *regular* (resp. *strongly regular*) if it has a regular (resp. strongly regular) associated Hamiltonian system.

## 2 LAGRANGIANS AFFINE IN SECOND DERIVATIVES

In a fiber chart, second order Lagrangian  $\lambda = L\omega_0$  affine in the variables  $y_{ij}^\sigma$  is expressed by formula

$$L = L_0 + L_\sigma^{ij} y_{ij}^\sigma, \quad L_\sigma^{ij} = L_\sigma^{ji} \quad (8)$$

where functions  $L_0, L_\sigma^{ij}$  do not depend on the variables  $y_{kl}^\nu$ .

We shall consider above Lagrangians and their Lepagean forms (1), (2) satisfying  $\rho = \theta_\lambda + d\phi + \tilde{\mu}$ , where  $\phi = 0$  and  $\tilde{\mu} = \sum_{i=2}^n p_i(\beta)$  and  $\beta$  is defined on  $J^1Y$ .

In general case, the Poincaré–Cartan forms of second order Lagrangian is defined on  $J^3Y$ , but for Lagrangians of the forms (8) the form  $\theta_\lambda$  is projectable onto  $J^2Y$ . Our choice of Lepagean form of the Lagrangian (8) conserves the above Lepagean form defined on  $J^2Y$ .

In fibred chart, we can rewrite the above Lepagean form by following formula

$$\begin{aligned} \rho = & (L_0 + L_\nu^{kl} y_{kl}^\nu) \omega_0 + \left( \frac{\partial L_0}{\partial y_j^\sigma} + \frac{\partial L_\nu^{kl}}{\partial y_j^\sigma} y_{kl}^\nu - d_k L_\sigma^{jk} \right) \omega^\sigma \wedge \omega_j \\ & + L_\sigma^{ij} \omega_i^\sigma \wedge \omega_j + a_{\sigma\nu}^{ij} \omega^\sigma \wedge \omega^\nu \wedge \omega_{ij} + b_{\sigma\nu}^{kij} \omega^\sigma \wedge \omega_k^\nu \wedge \omega_{ij} \\ & + c_{\sigma\nu}^{kl ij} \omega_k^\sigma \wedge \omega_l^\nu \wedge \omega_{ij} + \mu, \end{aligned} \quad (9)$$

where  $\mu$  is at least 3-contact (i.e.,  $\mu = \sum_{i=3}^n p_i(\beta)$  for  $\beta$  defined on  $J^1Y$ ) and funtions  $a_{\sigma\nu}^{ij}, b_{\sigma\nu}^{kij}, c_{\sigma\nu}^{kl ij}$  do not depend on the variables  $y_{pq}^\kappa$  and satisfy the conditions

$$\begin{aligned} a_{\sigma\nu}^{ij} &= -a_{\nu\sigma}^{ij}, \quad a_{\sigma\nu}^{ij} = -a_{\sigma\nu}^{ji}, \quad a_{\sigma\nu}^{ij} = a_{\nu\sigma}^{ji}, \\ b_{\sigma\nu}^{kij} &= -b_{\sigma\nu}^{kji}, \\ c_{\sigma\nu}^{kl ij} &= -c_{\nu\sigma}^{kl ij}, \quad c_{\sigma\nu}^{kl ij} = -c_{\sigma\nu}^{kl ji}. \end{aligned} \quad (10)$$

**Theorem 1** *Let  $\dim X \geq 2$ . Let  $\lambda = L\omega_0$  be a second order Lagrangian (8), and  $\alpha = d\rho$  with  $\rho$  of the form (9), (10), be Lepagean equivalent of Euler–Lagrange form of above Lagrangian. Assume that the matrix*

$$(B_{\nu\sigma}^{klj} \mid C_{\nu\kappa}^{klpq}), \quad (11)$$

*with  $mn^2$  rows (resp.  $mn + mn(n+1)/2$  columns) labelled by  $\nu, k, l$  (resp.  $\sigma, j, \kappa, p, q$ ), where*

$$B_{\nu\sigma}^{klj} = \left( \frac{\partial L_\nu^{kl}}{\partial y_j^\sigma} - \frac{1}{2} \left( \frac{\partial L_\sigma^{jk}}{\partial y_j^\nu} + \frac{\partial L_\sigma^{jl}}{\partial y_k^\nu} \right) - b_{\sigma\nu}^{kij} - b_{\sigma\nu}^{ljk} \right), \quad (12)$$



and

$$C_{\nu\kappa}^{klpq} = (c_{\nu\kappa}^{kpql} + c_{\nu\kappa}^{lpqk}), \quad (13)$$

has maximal rank equal to  $mn(n+3)/2$ .

Then the Hamiltonian system  $\alpha = d\rho$  is regular (i.e. every Dedecker–Hamilton extremal is of the form  $\delta_D = J^2\gamma$ , where  $\gamma$  is an extremal of  $\lambda$ ).

If moreover  $\mu$  is closed then the Hamiltonian system  $\alpha = d\rho$  is strongly regular (i.e. every Hamilton extremal is of the form  $\delta = J^2\gamma$ , where  $\gamma$  is an extremal of  $\lambda$ ).

**Proof** of the regularity of the Hamiltonian system follows from explicit computation  $\alpha = d\rho$ ,  $\hat{\alpha} = p_1(\alpha) + p_2(\alpha)$  and generators of ideal  $\mathcal{D}_{\hat{\alpha}}^3$ .

Expressing the generators of the ideal  $\mathcal{D}_{\hat{\alpha}}^3$  we get

$$\begin{aligned} i_{\frac{\partial}{\partial y^\nu}} \hat{\alpha} &= E_\nu \omega_0 + \left( \frac{\partial^2 L_0}{\partial y_j^\sigma \partial y^\nu} + \frac{\partial^2 L_\kappa^{pq}}{\partial y_j^\sigma \partial y^\nu} y_{pq}^\kappa - \frac{\partial^2 L_0}{\partial y^\sigma \partial y_j^\nu} - \frac{\partial^2 L_\kappa^{pq}}{\partial y^\sigma \partial y_j^\nu} y_{pq}^\kappa \right. \\ &\quad - \frac{\partial}{\partial y^\nu} d_k L_\sigma^{jk} + \frac{\partial}{\partial y^\sigma} d_k L_\nu^{jk} - 2d_k a_{\sigma\nu}^{kj} \Big) \omega^\sigma \wedge \omega_j + \omega_k^\sigma \wedge \omega_j \\ &\quad \times \left( \frac{\partial L_\sigma^{kj}}{\partial y^\nu} - \frac{\partial^2 L_0}{\partial y_k^\nu \partial y_j^\sigma} - \frac{\partial^2 L_\kappa^{pq}}{\partial y_j^\sigma \partial y_k^\nu} y_{pq}^\kappa + \frac{\partial}{\partial y_k^\sigma} d_p L_\nu^{jp} + 4a_{\nu\sigma}^{jk} - 2d_i b_{\nu\sigma}^{kij} \right) \\ &\quad + \left( \frac{\partial L_\sigma^{kl}}{\partial y_j^\nu} - \frac{1}{2} \left( \frac{\partial L_\nu^{jk}}{\partial y_l^\sigma} + \frac{\partial L_\nu^{jl}}{\partial y_k^\sigma} \right) - b_{\nu\sigma}^{kjl} - b_{\nu\sigma}^{kjl} \right) \omega_{kl}^\sigma \wedge \omega_j \\ &\quad + 2 \left( \frac{\partial a_{\sigma\nu}^{ij}}{\partial y^\kappa} + \frac{\partial a_{\kappa\sigma}^{ij}}{\partial y^\nu} + \frac{\partial a_{\nu\kappa}^{ij}}{\partial y^\sigma} \right) \omega^\kappa \wedge \omega^\sigma \wedge \omega_{ij} \\ &\quad + 2 \left( 2 \frac{\partial a_{\sigma\nu}^{ij}}{\partial y_k^\kappa} + \frac{\partial b_{\nu\kappa}^{kij}}{\partial y^\sigma} - \frac{\partial b_{\sigma\kappa}^{kij}}{\partial y^\nu} \right) \omega_k^\kappa \wedge \omega^\sigma \wedge \omega_{ij} \\ &\quad + 2 \left( 2 \frac{\partial c_{\kappa\sigma}^{lkij}}{\partial y^\nu} + \frac{\partial b_{\nu\kappa}^{ij}}{\partial y_k^\sigma} - \frac{\partial b_{\nu\sigma}^{kij}}{\partial y_l^\kappa} \right) \omega_l^\kappa \wedge \omega_k^\sigma \wedge \omega_{ij}, \end{aligned} \quad (14)$$

$$\begin{aligned} i_{\frac{\partial}{\partial y_k^\nu}} \hat{\alpha} &= \left( \frac{\partial^2 L_0}{\partial y_i^\sigma \partial y_k^\nu} + \frac{\partial^2 L_\kappa^{pq}}{\partial y_i^\sigma \partial y_k^\nu} y_{pq}^\kappa - \frac{\partial L_\nu^{kj}}{\partial y^\sigma} - \frac{\partial}{\partial y_k^\nu} d_p L_\sigma^{jp} + 4a_{\nu\sigma}^{ik} \right. \\ &\quad - 2d_j b_{\sigma\nu}^{kij} \Big) \omega^\sigma \wedge \omega_i + \omega_j^\sigma \wedge \omega_i \\ &\quad \times \left( \frac{\partial L_\sigma^{ij}}{\partial y_k^\nu} - \frac{\partial L_\nu^{kj}}{\partial y_i^\sigma} + 2b_{\sigma\nu}^{kij} - 2b_{\nu\sigma}^{ikj} - 4d_l c_{\nu\sigma}^{kilj} \right) \\ &\quad + 2C_{\nu\sigma}^{ikjl} \omega_{jl}^\sigma \wedge \omega_i + 2 \left( 2 \frac{\partial a_{\sigma\kappa}^{ij}}{\partial y_k^\nu} + \frac{\partial b_{\kappa\nu}^{kij}}{\partial y^\sigma} - \frac{\partial b_{\sigma\nu}^{kij}}{\partial y^\kappa} \right) \omega^\sigma \wedge \omega^\kappa \wedge \omega_{ij} \\ &\quad + 2 \left( 2 \frac{\partial c_{\kappa\nu}^{lkij}}{\partial y^\sigma} + \frac{\partial b_{\sigma\kappa}^{ij}}{\partial y_k^\nu} - \frac{\partial b_{\sigma\nu}^{kij}}{\partial y_l^\kappa} \right) \omega^\sigma \wedge \omega_l^\kappa \wedge \omega_{ij} \\ &\quad + 2 \left( 2 \frac{\partial c_{\sigma\nu}^{lkij}}{\partial y_p^\kappa} + \frac{\partial c_{\nu\kappa}^{kpji}}{\partial y_l^\sigma} + \frac{\partial c_{\kappa\sigma}^{plij}}{\partial y_k^\nu} \right) \omega_p^\kappa \wedge \omega_l^\sigma \wedge \omega_{ij}, \end{aligned} \quad (15)$$

$$i_{\frac{\partial}{\partial y_{kl}^\nu}} \hat{\alpha} = B_{\nu\sigma}^{klj} \omega^\sigma \wedge \omega_i + C_{\sigma\nu}^{ijkl} \omega_j^\sigma \wedge \omega_i. \quad (16)$$

Since the ranks of the matrix  $(B_{\nu\sigma}^{klj} | C_{\kappa\nu}^{pqkl})$  is maximal then the  $\omega^\sigma \wedge \omega_i$  and  $\omega_{(k}^\kappa \wedge \omega_{l)}$  are generators of ideal  $\mathcal{D}_\alpha^3$ . We obtain for Dedecker–Hamilton extremals  $\delta_D = J^2\gamma$ , where  $\gamma$  is a section of  $\pi$ .

Substituting this into (6), (14) we get

$$\delta_D^* i_{\frac{\partial}{\partial y^\sigma}} \hat{\alpha} = E_\sigma \circ J^2\gamma$$

for 2nd order Euler–Lagrange form (4) and  $\gamma$  is an extremal of  $\lambda$ .

Let us prove strong regularity: We have to show that under our assumptions, for every section  $\delta$  satisfying Hamilton equations, one has  $\delta = J^2\gamma$ , where  $\gamma$  is a solution of the Euler–Lagrange equations of the Lagrangian  $\lambda$ .

Assuming  $d\mu = 0$  (c.f. (16)), we obtain:

$$\delta^*(i_{\partial/\partial y_{kl}^\sigma} \alpha) = \delta^* \left( (B_{\nu\sigma}^{klj} | C_{\kappa\nu}^{pqkl}) (\omega^\sigma \wedge \omega_i | \omega_{(k}^\kappa \wedge \omega_{l)})^T \right) = 0,$$

i.e.  $\delta^*\omega^\sigma = 0$  and  $\delta^*\omega_j^\sigma = 0$  by the rank condition on the matrix  $(B_{\nu\sigma}^{klj} | C_{\nu\kappa}^{klpq})$ , i.e.  $\partial y^\sigma / \partial x^i = y_i^\sigma$  and  $\partial y_i^\sigma / \partial x^j = y_{ij}^\sigma$  along  $\delta$ . Hence,  $\delta^*(i_{\partial/\partial y_k^\nu} \alpha) = 0$ .

The above obtained conditions on  $\delta$  mean that every solution of Hamilton equations is holonomic up to the second order, i.e., we can write  $\delta = J^2\gamma$ , where  $\gamma$  is a section of  $\pi$ .

Now, the equations  $J^2\delta^*(i_{\partial/\partial y_k^\sigma} \alpha) = 0$  are satisfied identically, and the last set of Hamilton equations, i.e.,  $J^2\delta^*(i_{\partial/\partial y^\sigma} \alpha) = 0$  take the form  $E_\sigma \circ J^2\gamma = 0$  proving that  $\gamma$  is an extremal of  $\lambda$ . This completes the proof.

### 3 LEGENDRE TRANSFORMATION

In this section the Hamiltonian systems admitting Legendre transformation are studied. By the Legendre transformation we understand the coordinates transformation onto  $J^2Y$ .

Writing the Lepagean equivalent  $\rho$  (9), (10) in the form of a noninvariant decomposition we get

$$\begin{aligned} \rho &= -H\omega_0 + p_\sigma^j dy^\sigma \wedge \omega_j + p_\sigma^{ij} dy_i^\sigma \wedge \omega_j \\ &+ a_{\sigma\nu}^{ij} dy^\sigma \wedge dy^\nu \wedge \omega_{ij} + b_{\sigma\nu}^{kij} dy^\sigma \wedge dy_k^\nu \wedge \omega_{ij} \\ &+ c_{\sigma\nu}^{klj} dy_k^\sigma \wedge dy_l^\nu \wedge \omega_{ij} + \mu, \end{aligned} \quad (17)$$

where

$$\begin{aligned} H &= -L + \left( \frac{\partial L}{\partial y_i^\sigma} - d_j L_\sigma^{ij} \right) y_i^\sigma + L_\sigma^{ij} y_{ij}^\sigma + 2a_{\sigma\nu}^{ij} y_i^\sigma y_j^\nu \\ &- (b_{\sigma\nu}^{kij} + b_{\sigma\nu}^{jik}) y_i^\sigma y_{kj}^\nu - \frac{1}{2} (c_{\sigma\nu}^{klj} + c_{\sigma\nu}^{ilkj} + c_{\sigma\nu}^{kjl} + c_{\sigma\nu}^{ijkl}) y_{ik}^\sigma y_{jl}^\nu, \\ p_\sigma^j &= \frac{\partial L}{\partial y_j^\sigma} - d_i L_\sigma^{ij} + 4a_{\sigma\nu}^{ij} y_i^\nu - (b_{\sigma\nu}^{kij} + b_{\sigma\nu}^{jik}) y_{jk}^\nu, \\ p_\sigma^{ij} &= L_\sigma^{ij} + (b_{\nu\sigma}^{ikj} + b_{\nu\sigma}^{jki}) y_k^\nu - 2(c_{\nu\sigma}^{kilj} + c_{\nu\sigma}^{likj}) y_{kl}^\nu. \end{aligned} \quad (18)$$

**Remark 1** In general, the functions  $p_\sigma^{ij}$  are not symmetric in the indices  $i, j$ .

**Theorem 2** Let  $\dim X \geq 2$ . Let  $\lambda = L\omega_0$  be a second order Lagrangian (8), and let  $\rho$  of the form (9), (10) be the Lepagean equivalent of above Lagrangian with noninvariant decomposition (18). Assume that

$$\begin{pmatrix} \frac{\partial p_\sigma^i}{\partial y_k^\nu} & \frac{\partial p_\sigma^i}{\partial y_{kl}^\nu} \\ \frac{\partial p_\sigma^{ij}}{\partial y_k^\nu} & \frac{\partial p_\sigma^{ij}}{\partial y_{kl}^\nu} \end{pmatrix} \quad (19)$$

is regular. Then transformation

$$\psi_2 = (x^k, y^\nu, y_k^\nu, y_{kl}^\nu) \rightarrow (x^i, y^\sigma, p_\sigma^i, p_\sigma^{ij}) = \chi, \quad (20)$$

where  $m(n^2 + 1)/2$  of  $p_\sigma^{ij}$ 's are independent, is coordinate transformation on open set  $U \subset V_2$ . If moreover functions  $c_{\nu\sigma}^{kilj}$  satisfies conditions

$$c_{\nu\sigma}^{kilj} + c_{\nu\sigma}^{likj} = c_{\nu\sigma}^{kjl i} + c_{\nu\sigma}^{ljki} \quad (21)$$

then the functions  $p_\sigma^{ij}$  are symmetric in the indices  $i, j$  (i.e.,  $p_\sigma^{ij} = p_\sigma^{ji}$ ).

**Proof** of above theorem follows from explicit calculation of the Jacobi matrix of transformation (20). If the submatrix (19) of Jacobi matrix is regular, then the transformation (20) is coordinate transformation. Similarly, from explicit calculation we can easily see that (21) are necessary conditions for  $p_\sigma^{ij} = p_\sigma^{ji}$ . This completes the proof.

**Definition** The transformation (20) is called *generalized non-symmetric Legendre transformation*. The transformation (20) with condions  $p_\sigma^{ij} = p_\sigma^{ji}$  is called *generalized Legendre transformation*.

**Remark 2** For first order field Lagrangians the regularity condition and condition for existence Legendre (resp. generalized Legendre transformation) are identical. This fact contrasts with situation of second order field Lagrangians. The Legendre transformation (in classical field theories) and generalized Legendre transformation for second order Lagrangians in field theory do not coincide with regularity (resp. strongly regularity) conditions.

In these generalized Legendre coordinates the Hamilton equations (6) take a rather complicated form.

### Two interesting cases of Hamilton equations.

**a)** The lepagean equivalent (9), (10), (17), (18) of the second order Lagrangians (8) satisfies condition  $\mu$  is closed (i.e.,  $d\mu = 0$ ). In generalized “symetric” Legendre coordinates (i.e.,  $p_\sigma^{ij} = p_\sigma^{ji}$ ) the explicite computation of Hamilton equations reads

$$\begin{aligned} \frac{\partial H}{\partial y^\sigma} &= -\frac{\partial p_\sigma^j}{\partial x^j} + 4\frac{\partial a_{\sigma\nu}^{ij}}{\partial x^j} \frac{\partial y^\nu}{\partial x^i} + 2\left(\frac{\partial a_{\kappa\nu}^{ij}}{\partial y^\sigma} + \frac{\partial a_{\kappa\sigma}^{ij}}{\partial y^\nu} + \frac{\partial a_{\nu\kappa}^{ij}}{\partial y^\sigma}\right) \frac{\partial y^\kappa}{\partial x^i} \frac{\partial y^\nu}{\partial x^j} \\ &- 4\frac{\partial a_{\sigma\nu}^{ij}}{\partial p_\kappa^k} \frac{\partial p_\kappa^k}{\partial x^i} \frac{\partial y^\nu}{\partial x^j} + \frac{\partial b_{\sigma\nu}^{kij}}{\partial x^j} \frac{\partial y_k^\nu}{\partial x^i} + 2\left(\frac{\partial b_{\kappa\nu}^{kij}}{\partial y^\sigma} - \frac{\partial b_{\sigma\nu}^{kij}}{\partial y^\kappa}\right) \frac{\partial y^\kappa}{\partial x^i} \frac{\partial y_k^\nu}{\partial x^j} \\ &- 2\frac{\partial b_{\sigma\nu}^{kij}}{\partial p_\kappa^l} \frac{\partial p_\kappa^k}{\partial x^i} \frac{\partial y_k^\nu}{\partial x^j} + 2\frac{\partial c_{\kappa\nu}^{klij}}{\partial y^\sigma} \frac{\partial y_k^\kappa}{\partial x^i} \frac{\partial y_l^\nu}{\partial x^j}, \\ \frac{\partial H}{\partial p_\sigma^i} &= \frac{\partial y^\sigma}{\partial x^i} + 2\frac{\partial a_{\kappa\nu}^{jk}}{\partial p_\sigma^i} \frac{\partial y^\kappa}{\partial x^j} \frac{\partial y^\nu}{\partial x^k} + 2\frac{\partial b_{\kappa\nu}^{kjl}}{\partial p_\sigma^i} \frac{\partial y^\kappa}{\partial x^j} \frac{\partial y_l^\nu}{\partial x^k} + 2\frac{\partial c_{\kappa\nu}^{kljm}}{\partial p_\sigma^i} \frac{\partial y_k^\kappa}{\partial x^j} \frac{\partial y_l^\nu}{\partial x^m}, \\ \frac{\partial H}{\partial p_\sigma^{ij}} &= \frac{1}{2}\left(\frac{\partial y_i^\sigma}{\partial x^j} + \frac{\partial y_j^\sigma}{\partial x^i}\right). \end{aligned}$$

**b)** The lepagean equivalent (9), (10), (17), (18) of the second order Lagrangians (8) satisfies condition  $\mu$  is closed (i.e.,  $d\mu = 0$ ) and  $p_\sigma^{ij} = p_\sigma^{ji}$ . If moreover  $d\eta = 0$ , where

$$\eta = a_{\sigma\nu}^{ij} dy^\sigma \wedge dy^\nu \wedge \omega_{ij} + b_{\sigma\nu}^{kij} dy^\sigma \wedge dy_k^\nu \wedge \omega_{ij} + c_{\sigma\nu}^{klj} dy_k^\sigma \wedge dy_l^\nu \wedge \omega_{ij}$$

then the Hamilton equations (6) have the following form

$$\frac{\partial H}{\partial y^\sigma} = -\frac{\partial p_\sigma^j}{\partial x^j}, \quad \frac{\partial H}{\partial p_\sigma^i} = \frac{\partial y^\sigma}{\partial x^i}, \quad \frac{\partial H}{\partial p_\sigma^{ij}} = \frac{1}{2} \left( \frac{\partial y_i^\sigma}{\partial x^j} + \frac{\partial y_j^\sigma}{\partial x^i} \right).$$

## CONCLUSION

The paper is generalization of classical Hamiltonian field theory on fibred manifold. The regularization procedure of the first order Lagrangians proposed by Krupková and Smetanová [6] is applied to case of the second order Lagrangians affine in second derivatives. Hamilton equations are created from the Lepagean equivalent whose order of contactness is more than 2-contact (c.f. Hamilton  $p_2$ -equations in [11]). The generalized Legendre transformation is studied. The generalized momenta  $p_\sigma^{ij}$  with  $p_\sigma^{ij} \neq p_\sigma^{ji}$  are found.

Contrary to the Hamilton–De Donder theory the regularity conditions of the Lepagean form (9), (10) and the conditions of existence of the generalized Legendre transformation (20) do not coincide. The regularity conditions do not guarantee the existence of the Legendre transformation. In other hand, the existence of the Legendre transformation does not guarantee the regularity.

## References

- [1] Cantrijn, F., Ibort, A., De León, M. On the geometry of multisymplectic manifolds *Journal of the Australian Mathematical Society* 66 (3), 1999. p. 303–330.
- [2] Dedecker, P. On the generalization of symplectic geometry to multiple integrals in the calculus of variations, *Lecture Notes in Math.* 570 (Springer, Berlin, 1977) p. 395–456.
- [3] Hořava, P. On a covariant Hamilton-Jacobi framework for the Einstein-Maxwell theory, *Classical and Quantum Gravity* 8(11), 1991, p. 2069–2084.
- [4] Krupka, D. Some geometric aspects of variational problems in fibred manifolds, *Folia Fac. Sci. Nat. UJEP Brunensis* 14, 1973, p. 1–65.
- [5] Krupka, D., Štěpánková, O. On the Hamilton form in second order calculus of variations, in: *Geometry and Physics, Proc. Int. Meeting*, Florence, Italy, 1982, M. Modugno, ed. (Pitagora Ed., Bologna, 1983) p. 85–101
- [6] Krupková, O. field theory, *J. Geom. Phys.* 43, 2002, p. 93–132.
- [7] Krupková, O., Smetanová, D. Legendre transformation for regularizable Lagrangians in field theory, *Letters in Math. Phys.* 58, 2001, p. 189–204.
- [8] Krupková, O., Volný, P. Euler-Lagrange and Hamilton equations for non-holonomic systems in field theory, *Journal of Physics A: Mathematical and General* 40(7), 2005, p. 8715–8745.
- [9] Prieto - Martínez, P.D., Román - Roy, N. A new multisymplectic unified formalism for second order classical field theories, *Journal of Geometric Mechanics*, 7(2), 2015. p. 203–253.
- [10] Saunders, D.J. *The Geometry of Jets Bundles*, Cambridge University Press, Cambridge, 1989.

- [11] Smetanová, D. On Hamilton  $p_2$ -equations in second-order field theory, in: *Steps in Differential Geometry, Proc. of the Coll. on Diff. Geom., Debrecen 2000* (University of Debrecen, Debrecen, 2001). p. 329–341.
- [12] Smetanová, D. The regularization of second order Lagrangians in example, *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica*, 58, 2016, p. 158–165.
- [13] Swaczyna, M., Volný, P. Uniform motions in central fields, *Journal of Geometric Mechanics*, 9(1), 2017, p. 91 - 130.
- [14] Swaczyna, M., Volný, P. Geometric concept of isokinetic constraint for a system of particles, *Miskolc Mathematical Notes*, 14(2), 2013, p. 697–704.
- [15] Vey, D. Multisymplectic formulation of vielbein gravity: I. De Donder-Weyl formulation, Hamiltonian  $(n - 1)$ -forms . *Classical and Quantum Gravity*, 32 (9), 2015, 50 pp.

# ENGINEERING EDUCATION AND SCIENCE & TECHNOLOGY POPULARIZATION AMONG YOUNGSTERS SUPPORTED BY IT

Lubica Stuchlíková<sup>1</sup>, Peter Benko<sup>1</sup>, Frantisek Janicek<sup>1</sup>, Ondrej Pohorelec<sup>1</sup>,  
Jiri Hrbacek<sup>2</sup>

<sup>1</sup>Slovak University of Technology in Bratislava, Faculty of Electrical Engineering and  
Information Technology, Ilkovicova 3, 812 19 Bratislava 1, Slovak Republic,  
lubica.stuchlikova@stuba.sk, peter\_benko@stuba.sk,  
frantisek.janicek@stuba.sk, xpohoreleco@stuba.sk

<sup>2</sup>Masaryk University, Faculty of Education, Department of Technical Education and  
Information Science, Porici 623/7, 603 00 Brno, Czech Republic,  
hrbacek@ped.muni.cz

**Abstract:** *Electrical engineering has a unique position and extraordinary strategic importance world-wide. This field of technology is considered to be the moving force of today's modern technical civilization. At the forefront is the need for highly qualified graduates of the electrical engineering fields capable of contributing to the development of the new advanced technologies and their real-world implementation. The education quality of the young generation, also known as Digital Natives, is becoming paramount. In this article, practical experiences gained during expert preparation of the professionals in this field with the IT support are presented. The article is focused on e-learning projects used not only as a tool for knowledge, research results and newest developments transfer into the education process, but also for increasing the interest of youth in science and technology.*

**Keywords:** electrical engineering, digital natives, e-learning, education, science & technology popularization.

## INTRODUCTION

Electrical engineering, as a stand-alone discipline, that examines energy, electrical effects and properties, started to form in the 19th century [1]. The discovery of electrical current, electrical laws and electrical devices all contributed to its advancement. Electrical engineering is currently one of the key branches of the technology that deals with generation, distribution and consumption of electrical energy as well as the devices used for these purposes that is widely implemented across many parts of human lives. Importance of all the branches of electrical engineering, that use progressive technologies of today for improving the quality of life, for nature and technology symbiosis in the form of alternative sources of energy or ecological elements in traffic, for development of “smart” cities and networks, for decreasing the impact of the technical production on the environment, for improving the communication and signal transmission possibilities, for improving the personal security, etc. is still growing [2].

Because of this reason, the praxis has an extremely large interest in highly qualified graduates of the electrical engineering fields, capable of quickly integrating themselves into the production and development process in top-level companies and research centers. Preparation of such an graduates rests in quality of technical education of the young generation [3].

Another important fact to be considered is that today's university students are the new generation, so called digital natives [4]. These students, born after 1982, also called as Millennials, Generation Y, Net Generation, Digital Generation or iGeneration have grown up in the world of the new information and communication technologies [5]. These students expect the same technologies they use every day to be present in education – computers, internet, internet applications, social networks, web 2.0, mobile phones, tablets, videogames, etc. They are interested in connectivity. They prefer interactivity over passivity, for example internet over television. Every day, they are in online contact with dozens of their friends. They have access to a large amount of information from expert lectures from renowned universities' professors, to complete ballast.

Modern information technologies, that formed this generation, also opened new possibilities in the field of education. Multimedia education, cooperative education [6], remote and virtual laboratories [7], mobile education, micro-education, study supports with internal intelligence [8], 3D virtual worlds [9], MOOC (Massive Open Online Course) [10], simulations [11], educational games are gradually becoming a common part of education. These possibilities can mutually interact and complement, so that a new quality is created, that allows more effective goals completion in education process. This computer assisted education, often called summarily e-learning, is becoming a very popular form of education. It has a potential to become not only as an excellent tool for knowledge, research results and newest developments transfer into the education process, but also for increasing the interest of youth in science and technology [12]. It is, of course, important to remember, that virtual world and theoretical knowledge without the ability to apply them in real world situation is not enough. The education should be very closely linked to practical knowledge and skills.

Our own practical experiences gained during expert preparation of the electrical engineering professionals in education and popularization of science and technology with the IT support since 2004 are presented in this article. It is focused on the original e-learning projects available at the educational portals eLearn central.

## **1 CHALLENGES IN THE ELECTRICAL ENGINEERING EDUCATION**

While presenting the challenges in the electrical engineering education, we can look at the challenges that the whole technical education in Slovakia is trying to solve.

Young people's interest in studying technology and natural sciences is decreasing. The usage of "technical devices" is considered to be very interesting and enjoyable, but the study in the technical fields is generally unattractive for young people across the entire world. Social prestige of the technical and crafting professions is very low in Slovakia, while at the same time a lot of top students leave to study at the foreign universities. Young people assume that all the news and development comes from abroad. Therefore, they are thinking, that there is no place for such experts in Slovakia.

The number of pupils at secondary technical schools is decreasing, because of the expansion of the secondary grammar schools (oriented mostly at humanitarian sciences). This causes the decrease in the knowledge level of pupils – knowledge of the average and worse students is higher in classes with larger number of talented pupils. Because of this, the total level of knowledge and skills is lower for the graduates of the elementary schools and, consequently, of the technical secondary schools.

Study of electrical engineering is considered to be among the more difficult by the general populace. It requires good mathematics and physics knowledge, the subjects based on logical relations and abstract thinking. The level of knowledge in mathematics and physics is decreasing. The reason for this lower score is lower lessons dotation as well as the cancellation of the compulsory mathematics leaving examination, not enough young expert teachers of these subjects and almost total suppression of the technical education at the elementary and secondary schools.

For several years we are experiencing the demography curve decline. The outcome is the decline in university students. At the same time, the number of students that do not finish the technical universities is rising. One of the reason is that students from secondary grammar schools ( $\frac{2}{3}$  of the students) have low level of technical knowledge and wrong idea of the study contents. If the study does not fulfil their ideas, they are losing interest in studying.

Education is directly influenced by the development of the science and technology in the 21st century. While the volume of the knowledge is increasing, the capacities and time schedules of the education are not. There is an ever-increasing difference between the progress in the praxis and the education. Schoolbooks and the scripts are becoming obsolete very quickly and their contents do not correspond with the current requirements. Basics must be taught, but we cannot forget the news and requirements from the praxis, if we are to reveal the whole picture. Universities are confronted with major differences in knowledge of the accepted students, that are mostly evident at technical subjects. Our students are young people, that grew up surrounded by technologies, that are erasing the boundary between reality and science fiction – digital natives.

Our basic goal and at the same time the greatest challenge is to lead the students, so that they will be able to think critically, to be highly adaptable and flexible, to support their individuality and creativity as well as their ability to work effectively in a team. Very important task is to affect them so that their internal motivation for gaining new knowledge and skills is strengthened. In spite of our main duty to provide university education, we also have to work on the problems of elementary and secondary education, science and technology popularization for children and general public. Elementary and secondary schools are our partners at fulfilling our common goal – training of professionals in the field of electrical engineering for the future. Children that are going to schools today, should be employed as highly qualified experts by 2034. Thanks to the rapid advancement of the science and technology it is not possible to exactly determine what would be important knowledge and skills for successful involvement in the praxis. Because the future cannot be foreseen, it is necessary to maintain theoretical and practical knowledge of the graduates on the highest possible level, update educational process so that the graduates leaving to join the praxis are best prepared for this change.

## **2 SOLUTIONS IN THE ELECTRICAL ENGINEERING EDUCATION**

Some of the aforementioned challenges are possible to solve by an effective implementation of the communication and information technologies in the education process along with the new pedagogical approaches [13]. One of the very interesting possibilities is e-learning. Electronic education has an immense potential to become the source of motivation and creativity, as well as the carrier of knowledge, and it is necessary to count it as a partner in the



whole education process from the first steps of life, elementary schools, secondary schools to universities, and it will accompany us during the lifelong education.

Our answer to electrical engineering education was the creation of the alternative sources of information – educational portals eLearn central on the educational platform MOODLE, that has the role to present interactive educational materials, courses and projects to our students as a support for the standard face-to-face education. At the beginning of the development of the support interactive e-learning materials, we focused on raising the quality of education of the subject Electronic devices and circuits (Fig. 1).

The figure displays two screenshots of a Moodle course page for 'Electronic Devices and Circuits'. The left screenshot shows the course overview with a navigation menu on the left, a 'Latest news' section, a 'Search forums' box, and a 'Weekly outline' section titled 'Electronic Devices and Circuits' ZS 2017/2018. The right screenshot shows a detailed view of 'Lecture no. 6: Bipolar Transistors' and 'Practice No. 6: Unipolar transistor (experimental)'. It includes an 'Assignment' section, 'Interactive flash animations' with links to various topics, and 'Self-tests' with checkboxes for different tests. At the bottom, there is a 'Credit written test' announcement.

**Fig. 1.** Course Electronic devices and circuits: Exercise instructions; interactive map, materials and information.

Source: own

This subject deals with basic principles of operation and electrical properties of electrical devices and circuits. Emphasis is put on diodes, transistors, operational amplifiers and digital circuits. This subject is taught in the form of standard lectures, laboratory exercises and complex e-learning support, where the contents are continually optimized and updated. Students also have printed scripts at their disposal as well as exercises in the form of exercise sheets in the pdf format. Practical laboratory exercises from this subject allows students to validate theoretical knowledge with practical measurements of electrical characteristics of electrical devices and circuits. Complex e-learning support consists of standard interactive www course Electronic devices and circuits and informational www course Electronic devices and circuits – exercise instructions (Fig. 2).

E405 Topic 1: Measuring of the resistance of a resistor and voltage divider

Name and surname: ..... Rating: .....  
 Workplace: ..... Date of measurement: ..... Exercise time: ..... Signature of trainer: .....

**Measuring of the resistance of a resistor and voltage divider**

Acquaintance with measuring instruments. Measurement of resistance volt-ampere characteristics and voltage divider design.

**Assignment:**

1. Measure of the resistance and V-A characteristic of the resistors  $R_1$  and  $R_2$ . Compare the results.
2. Design and connect the voltage divider from the resistors  $R_1$  and  $R_2$ . To set input voltage on 10 V and measure the value of the output voltage of the voltage divider. Verify the value of the output voltage by calculation.
3. Get familiar with the basic operation of the oscilloscope and the functional generator.

---

**Practical part:** Documents for the implementation of assignment tasks

**To point 1:**

Note: Use the VOLCRAFT VC6508T as multimeter and a source triple DC RIGOL DP832 to measure the resistance of resistors and V-A characteristic of resistors and connect the components to the breadboard (Fig. 1.1).





Fig. 1.1. The breadboard - green line shown by

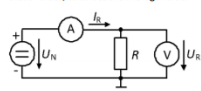
Interesting URL: How to Use a Breadboard - <https://learn.sparkfun.com/tutorials/how-to-use-a-breadboard>



3

E405 Topic 1: Measuring of the resistance of a resistor and voltage divider

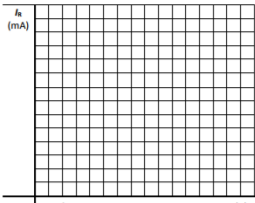
- Measure the resistance of the resistors  $R_1$  and  $R_2$ . Use the VOLCRAFT VC6508T multimeter and the breadboard, write the measured values into Tab. 1.1.
- Connect the electronic circuit for VACH measurement of resistors according to the circuit diagram presented in Fig. 1.2. Measure both resistors. Write the measured values into the Tab. 1.1 and draw the chart into Graph 1.1.
- For each measured voltage and current value, calculate the resistance value of the measured resistor. From these values, calculate the average value.



Obr. 1.2. The circuit diagram for VACH measurement of resistors

Tab. 1.1 Table of measured and calculated values

$R_1$			$R_2$		
Measured			Measured		
$I_R$ (mA)	$V_R$ (V)	Calc. $R_1$ (Ω)	$I_R$ (mA)	$V_R$ (V)	Calc. $R_2$ (Ω)
0,5			0,5		
1			1		
3			3		
5			5		
7			7		
9			9		
12			12		
Calculated average value			Calculated average value		



Graph. 1.1 VACH resistors

**To point 2:**

- Design a voltage divider with the resistors  $R_1$  and  $R_2$ . Connect the voltage divider circuit (Fig. 1.3) to the breadboard. Measure the value of the output voltage of divider at the input voltage 10 V.
- Check the accuracy of the measured value of the output voltage of the divider at the input voltage of 10 V by calculation using the relation  $V_O = V_N \frac{R_2}{R_1 + R_2}$

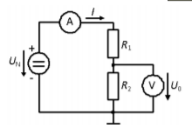


Fig. 1.3. The circuit diagram of the voltage divider

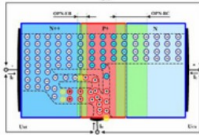
Measured values	$R_1$ =	$R_2$ =
$V_{O_{meas}}$ - measured values (V)		
$V_{O_{calc}}$ - calculated value (V)		
Relative error $\delta$ (%)		
$\delta = ((V_{O_{meas}} - V_{O_{calc}}) / V_{O_{calc}}) \cdot 100$		

4

**Fig. 2.** Course Electronic devices and circuits: Exercise instructions  
Source: own

Basic concepts and physical principles of electronic devices and circuits are available to the users through the interactive course Electronic devices and circuits [14]. It has 10 lessons and is available for free at the eLearn central open portal (<http://uef.fei.stuba.sk/moodleopen>). Lessons are complemented with one or two types of tests (registration is necessary for taking the tests, so that the results can be saved). Dictionary of the terms is also present. Original interactive animations located at the Interactive animations in electronics (Fig. 3), that is available for free are also part of the course.

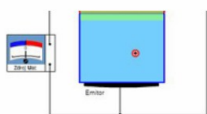
f.4.3 Bipolárny tranzistor v zapojení so spoločnou bázou



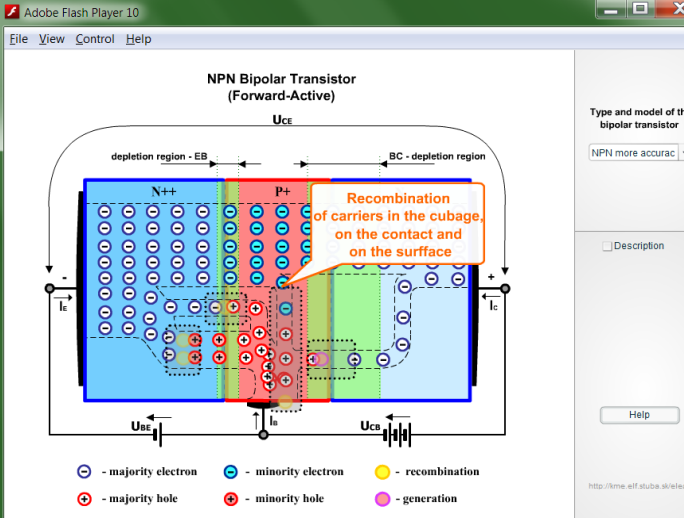
f.4.3 Bipolárny tranzistor v zapojení so spoločnou bázou

---

f.4.4 Bipolárny tranzistor v zapojení so spoločným emitorom



f.4.4 Bipolárny tranzistor v zapojení so spoločným emitorom



NPN Bipolar Transistor (Forward-Active)

U<sub>CE</sub>

depletion region - EB      BC - depletion region

N++      P+

Recombination of carriers in the cubage, on the contact and on the surface

U<sub>BE</sub>      U<sub>CB</sub>

I<sub>E</sub>      I<sub>B</sub>      I<sub>C</sub>

⊖ - majority electron    ⊕ - minority electron    ● - recombination  
⊖ - majority hole      ⊕ - minority hole      ● - generation

Type and model of the bipolar transistor  
NPN more accurac

Description

Help

<http://time.fei.stuba.sk/elearn>

**Fig. 3.** Course Interactive flash animation (designed for Slovak students - in Slovak language), bipolar transistor animation view (in English language).  
Source: own

These animations were created with the goal of showing the student in an interactive and intuitive form the basic internal physical processes in electronic devices as well as operation principles of the electronic circuits. Passive devices, diodes, transistors, examples of device fabrication using planar technology, NAND, TTL, CMOS, imaging devices, etc, are present. Interactive animation are used in all developed courses of the eLearn central, that are related to the animated problems. Course “Electronic devices and circuits – exercises instructions” is a course intended only for the students of the second year of the bachelor study. It has a weekly format, in accordance with weeks in semester. Course contains information about the subject, successful completion requirements, lectures in pdf, exercise sheets and materials for exercises, exam questions, discussion forums, tests and announcements (Fig. 1). From the students results as well as their feedback, we can conclude that this approach (lectures + practical exercises preferring experimental work in pairs + complex e-learning support) is working as a possibility of students’ motivation growth, as well as effective tool for increasing the quality of technical education.

Our experiences were successfully used and are still being used in development of e-learning support for our students for another subjects, individual and team projects, as well as in creating e-learning materials for a wide target group: bachelors, elementary and secondary school pupils and general public in the form of popularization of science and technology.

### 3 SOLUTIONS IN THE SCIENCE & TECHNOLOGY POPULARIZATION

An example of popularization of science and technology is offline professional interactive monograph on DVD “The Mysterious World of Energy” [15] and online e-learning project „Power Engineering dictionary“. Project “The Mysterious World of Energy” (Fig. 4) is our reaction to a necessity of creating an enviro-awareness between children and youth in the field of power engineering.



**Fig. 4.** Interactive monograph „The Mysterious World of Energy“ (in Slovak)  
Source: own

Its goal is to contribute to popularization of the topics of sources, generation and transfer of energy, negative effects on environment as well as possibilities of individual effect on improving the situation, for example saving energy. “The Mysterious World of Energy” is divided into 5 parts. 3 parts are aimed at the students of secondary schools and bachelors (Energy sources, Energy conversion, Power Engineering dictionary) and 2 parts are for elementary school pupils (Why can anybody be a power plant and Games about the world of energy). Belonging to this monography is the Power Engineering dictionary with more than 750 terms (Fig. 5), it is available for free in online version at the eLearn central open portal (<http://uef.fe.i.stuba.sk/moodleopen>). The ambition of the authors was the creation of the dictionary with extensive database of terms from the field of power engineering and to cover most frequent terms of generation, distribution, consumption and price of energy. All this with a goal to help the children, youth and general public to clarify some questions from the field of power engineering, provide answers and catch their interest.



**Fig. 5.** Power Engineering dictionary (it is designed for Slovak children and youth - explanation of the terms are in Slovak language)

Source: own

## CONCLUSION

Electrical engineering is actively partaking in improving working conditions of people, as basic instruments of economy, social, and culture growth. It is the reason for the interest from the praxis for the highly qualified graduates of the electrical engineering fields. It is necessary to secure the high quality education of young generation, digital natives, to prepare such graduates. Work with these people require implementation of modern and effective approaches to education associated with practical knowledge and skills, with high motivation of the students supported by the information and communication technologies.

Examples of the original interactive e-learning projects available at two portals eLearn central <http://uef.fe.i.stuba.sk/moodle/> and <http://uef.fe.i.stuba.sk/moodleopen/> are presented in this article. Portals eLearn central have been used as a support to a standard face-to-face education at the STU in Bratislava since 2004, and for popularization of science and technology between children, youth and general public since 2009. Our interactive online/offline projects are full of interactivity, multimedia elements, animations, illustrations, tests and discussion forums. We prefer many graphical schemes and funny images showing basic properties before

difficult explanation, while we are creating our popularization e-learning projects. Based on the feedback from pupils and students, we can confirm that e-learning is a great tool for knowledge, research results and new advancements transfer to education process, as well as a support tool for increasing the interest in young people for science and technology.

## References

- [1] Lundgreen, P. Engineering education in europe and the u.S.a., 1750-1930: The rise to dominance of school culture and the engineering professions, 1990, *Annals of Science*, 47 (1), pp. 33-75 ISSN: 00033790.
- [2] Sarkar, A. N. Eco-Innovations in Designing Ecocity, Ecotown and Aerotropolis. 2016, *Journal of Architectural Engineering Technology*, 5: 161. ISSN: 2168-9717.
- [3] Plaza, I., Medrano, C. T. Continuous improvement in electronic engineering education, 2007, *IEEE Transactions on Education*, 50 (3), pp. 259-265. ISSN: 00189359.
- [4] Prensky, M. Digital Natives, Digital Immigrants. In *On the Horizon*, 2001, [online] MCB University Press, Vol. 9, No. 5, Bradford, West Yorkshire, UK [cit. 2017-06-06] 6 screens. <http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf>.
- [5] Bennett, S., Maton, K., Kervin, L. The 'digital natives' debate: A critical review of the evidence, *British Journal of Educational Technology*, 2008, 39 (5), pp. 775-786. ISSN: 00071013.
- [6] Paulson, D. R. – Faust, J. L. Active and cooperative learning - Background & Definitions. In *California State University (USA)*. 2002. [online], 2002-05-18, [cit. 2017-06-06]. <http://www.calstatela.edu/dept/chem/chem2/Active/index.htm>.
- [7] Žáková, K. WEB-Based Control Education in Matlab. In: *Web-Based Control and Robotics Education*. 2009. Dordrecht: Springer Verlag, pp. 83-102. ISBN 978-90-481-2504-3.
- [8] Hrbáček, J. Study supports with internal intelligence, In *Technology of education. Science teacher magazine*, 2011. Nitra , Slovakia, SLOVDIDAC, Slovakia, vol. 19, no. 2, pp. 11 - 17. ISSN 1335 -003X.
- [9] Tüzün, H., Özdiñç, F. The effects of 3D multi-user virtual environments on freshmen university students' conceptual and spatial learning and presence in departmental orientation. *Computers and Education*. Vol. 94, 1 March 2016, pp. 228-240. ISSN: 03601315
- [10] Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J. From massive access to cooperation: lessons learned and proven results of a hybrid xMOOC/cMOOC pedagogical approach to MOOCs. *International Journal of Educational Technology in Higher Education*, Vol. 13, Issue 1, 1 December 2016, No. 24. ISSN: 23659440.
- [11] Pribytný, P., Donoval, D., Chvála, A., Marek, J., Molnár, M. Electro-Thermal Analysis and Optimization of Edge Termination of Power Diode Supported by 2D Numerical Modeling and Simulation. In: *Microelectronics Reliability*. 2012. vol. 52, pp. 463-468. ISSN 0026-2714.
- [12] Stuchlikova, L., Benkovska, J. e-Learn central - the Journey to e-Learning. *Proceedings of the 14th International Conference on Interactive Collaborative Learning and 11th International Conference Virtual University*. Piešťany, Slovakia, September 21-23, 2011. Piscataway : IEEE, 2011, pp.16-23. ISBN 978-1-4577-1746-8.
- [13] Stuchlikova, L. Challenges of education in the 21st century. In *ICETA 2016, 14th IEEE International conference on emerging elearning technologies and applications*, 2016.



- Starý Smokovec, Slovakia. November 24 - 25, 2016. Danvers : IEEE, pp. 335-340. ISBN 978-1-5090-4701-7.
- [14] Stuchlikova, L, Gron, M., Csabay, O., Helbich, M., Radobicky, J., Beno, J., Mondocko, P., Vacek, F., Hulený, L., Kinder, R., Rovanova, L., Redhammer, R., Nemcok, P., Bednar, M., Števo, M., Lazistan, F., Benkovic, M., Svrcek, P., Benko, M., Nagy, A. Interactive www Course "Electronic Devices and Circuits". In *Virtual University VU '06: 7th International Conference*, Bratislava, Slovak Republic, 14.-15.12.2006. Bratislava : STU v Bratislave, 2006, s.119-124. ISBN 80-227-2542-0.
- [15] Janicek, F., Stuchlikova, L., Farkas Smitkova, M., Holjencik, J., Cerman, A., Zuscak, J., Halan, I. 2014. *The Mysterious World of Energy*. 1.vyd. Bratislava, Nakladateľstvo STU, ISBN 978-80-227-4281-8.

## Acknowledgement

The work presented in this paper has been supported by the agency KEGA the Ministry of Education, Science, Research and Sport of the Slovak Republic for under Grant 020STU-4/2015 and by the Slovak Research and Development Agency (APVV) under the Contract No. APVV-15-0326.

# LINEAR DIFFERENCE WEAKLY DELAYED SYSTEMS, THE CASE OF COMPLEX CONJUGATE EIGENVALUES OF THE MATRIX OF NON-DELAYED TERMS

Jan Šafařík, Josef Diblík

Faculty of Civil Engineering, Faculty of Electrical Engineering and Communication,  
Brno University of Technology, Brno, Czech Republic.  
Technická 3058/10, Žabovřesky, 61600, Brno, Czech republic.  
xsafar19@stud.feec.vutbr.cz, diblik@feec.vutbr.cz

**Abstract:** *A linear weakly delayed discrete system with single delay*

$$x(k+1) = Ax(k) + Bx(k-m), \quad k = 0, 1, \dots,$$

*in  $\mathbb{R}^3$  is considered, where  $A$  and  $B$  are  $3 \times 3$  matrices and  $m \geq 1$  is an integer. Assuming that the characteristic equation of the matrix  $A$  has a pair of complex conjugate roots, the general solution of the given system is constructed.*

**Keywords:** Discrete system, weakly delayed system, linear system, initial problem, single delay.

## INTRODUCTION

Consider a linear system of difference equations with delay

$$x(k+1) = Ax(k) + Bx(k-m), \quad k = 0, 1, \dots \quad (1)$$

where  $A = (a_{ij})_{i,j=1}^3$ ,  $B = (b_{ij})_{i,j=1}^3$  are  $3 \times 3$  real constant matrices and  $m \geq 1$  is a natural number.

In the sequel, it is assumed that the system (1) is weakly delayed as defined below (compare [5] and [1], [2]).

**Definition 1** *System (1) is called weakly delayed if the characteristic equations for (1) and for the system without delay  $x(k+1) = Ax(k)$  have identical roots, that is, if, for every  $\lambda \in \mathbb{C} \setminus \{0\}$ ,*

$$\det(A + \lambda^{-m}B - \lambda E) = \det(A - \lambda E),$$

*where  $E$  is a  $3 \times 3$  unit matrix.*

The below lemma is used.

**Lemma 1** *If the system (1) is weakly delayed, then its arbitrary linear nonsingular transformation again leads to a weakly delayed system.*

For the proof, we refer to [5], Lemma 1.2.

**Theorem 1 ([3])** *Let  $l = 3$  in (1). Then, (1) is a weakly delayed system if and only if conditions (2)–(7) below hold:*

$$b_{11} + b_{22} + b_{33} = 0, \quad (2)$$

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} = 0, \quad (3)$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0, \quad (4)$$

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} = 0, \quad (5)$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0, \quad (6)$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \\ + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0. \quad (7)$$

For the proof, we refer to [3], Theorem 1.3.

Assume that the matrix  $A$  has one real eigenvalue  $\lambda_1 = \lambda$  and two eigenvalues are complex conjugate, i.e.,  $\lambda_{2,3} = p \pm iq$ , with  $q \neq 0$ . Then, the Jordan form  $\Lambda$  assigned to  $A$  is

$$A : \Lambda = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & p & q \\ 0 & -q & p \end{pmatrix}. \quad (8)$$

For the considered case, it is easy to give a coefficient criterion for the system to be weakly delayed.

**Theorem 2** *If matrix  $A$  has the above eigenvalues, then system (1) is weakly delayed if and only if conditions (9) – (14) below hold:*

$$b_{11} = 0, \quad (9)$$

$$b_{22} + b_{33} = 0, \quad (10)$$

$$b_{23} - b_{32} = 0, \quad (11)$$

$$b_{22}b_{33} - b_{12}b_{21} - b_{13}b_{31} - b_{23}b_{32} = 0, \quad (12)$$



$$(\lambda - p)(b_{12}b_{21} + b_{13}b_{31}) + q(b_{12}b_{31} - b_{13}b_{21}) = 0, \quad (13)$$

$$b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} = 0. \quad (14)$$

*Proof.* Although the proof is given in [9], we give here an improved version to fill some gaps in the original version.

It is possible to simplify conditions (4), (6) and (7). From (4), we get

$$\begin{aligned} \lambda(b_{22}b_{33} - b_{23}b_{32}) + p(b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31}) \\ + q(b_{11}b_{23} + b_{12}b_{31} - b_{11}b_{32} - b_{13}b_{21}) = 0 \end{aligned} \quad (15)$$

because

$$\begin{aligned} & \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \\ &= \begin{vmatrix} \lambda & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & p & q \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & -q & p \end{vmatrix} = \\ &= \lambda(b_{22}b_{33} - b_{23}b_{32}) + p(b_{11}b_{33} - b_{13}b_{31}) - q(b_{11}b_{32} - b_{12}b_{31}) \\ &\quad + q(b_{11}b_{23} - b_{13}b_{21}) + p(b_{11}b_{22} - b_{12}b_{21}) = \\ &= \lambda(b_{22}b_{33} - b_{23}b_{32}) + p(b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31}) \\ &\quad + q(b_{11}b_{23} + b_{12}b_{31} - b_{11}b_{32} - b_{13}b_{21}) = 0. \end{aligned}$$

From (6) we get

$$\lambda(p(b_{22} + b_{33}) + q(b_{23} - b_{32})) + b_{11}(p^2 + q^2) = 0 \quad (16)$$

since

$$\begin{aligned} & \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \\ &= \begin{vmatrix} \lambda & 0 & 0 \\ 0 & p & q \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} \lambda & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ 0 & -q & p \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & p & q \\ 0 & -q & p \end{vmatrix} = \\ &= \lambda(pb_{33} - qb_{32}) + \lambda(pb_{22} + qb_{23}) + b_{11}(p^2 + q^2) = \\ &= \lambda(p(b_{22} + b_{33}) + q(b_{23} - b_{32})) + b_{11}(p^2 + q^2) = 0. \end{aligned}$$

From (7) we get

$$\lambda(b_{22} + b_{33}) + p(2b_{11} + b_{22} + b_{33}) + q(b_{23} - b_{32}) = 0 \quad (17)$$

since

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix}$$

$$\begin{aligned}
& + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \\
& = \begin{vmatrix} \lambda & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ 0 & p & q \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \\
& + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & p & q \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 \\ 0 & -q & p \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ 0 & -q & p \end{vmatrix} = \\
& = \lambda b_{22} + \lambda b_{33} + p b_{33} - q b_{32} + p b_{11} + p b_{11} + p b_{22} + q b_{23} = \\
& = \lambda(b_{22} + b_{33}) + p(2b_{11} + b_{22} + b_{33}) + q(b_{23} - b_{32}) = 0.
\end{aligned}$$

From (2), we have  $b_{22} + b_{33} = -b_{11}$ . Step by step, expression (17) yields

$$\begin{aligned}
\lambda(b_{22} + b_{33}) + p(2b_{11} + b_{22} + b_{33}) + q(b_{23} - b_{32}) &= 0, \\
\lambda(-b_{11}) + p b_{11} + q(b_{23} - b_{32}) &= 0, \\
-b_{11}(\lambda - p) + q(b_{23} - b_{32}) &= 0.
\end{aligned}$$

From the last expression, we have  $q(b_{23} - b_{32}) = (\lambda - p)b_{11}$ . A substitution into (16) yields

$$\begin{aligned}
\lambda(p(b_{22} + b_{33}) + q(b_{23} - b_{32})) + b_{11}(p^2 + q^2) &= 0, \\
\lambda p(-b_{11}) + \lambda b_{11}(\lambda - p) + b_{11}(p^2 + q^2) &= 0, \\
b_{11}(\lambda^2 - 2\lambda p + p^2 + q^2) &= 0, \\
b_{11}((\lambda - p)^2 + q^2) &= 0.
\end{aligned}$$

Since  $(\lambda - p)^2 + q^2 \neq 0$ , we get

$$b_{11} = 0. \tag{18}$$

i.e. (9) holds and, moreover, (5) reduces to (12). From (2), utilizing (18), we derive

$$b_{22} + b_{33} = 0 \tag{19}$$

and (10) is valid. Substituting (18) and (19) into (17), we have

$$b_{23} - b_{32} = 0,$$

so (11) holds. Simplifying (5) leads to

$$\begin{aligned}
(b_{11}b_{22} - b_{12}b_{21}) + (b_{11}b_{33} - b_{13}b_{31}) + (b_{22}b_{33} - b_{23}b_{32}) &= 0, \\
-b_{12}b_{21} - b_{13}b_{31} + b_{22}b_{33} - b_{23}b_{32} &= 0.
\end{aligned}$$

Then, from the last expression, we get

$$b_{22}b_{33} - b_{23}b_{32} = b_{12}b_{21} + b_{13}b_{31}.$$

Substituting this together with (18) into (15), we obtain (13):

$$\begin{aligned}\lambda(b_{12}b_{21} + b_{13}b_{31}) + p(-b_{12}b_{21} - b_{13}b_{31}) + q(b_{12}b_{31} - b_{13}b_{21}) &= 0, \\ (b_{12}b_{21} + b_{13}b_{31})(\lambda - p) + q(b_{12}b_{31} - b_{13}b_{21}) &= 0.\end{aligned}$$

Condition (3) can be simplified to (14). □

**Example 1** *Let system (1) be of the form*

$$\begin{aligned}x_1(k+1) &= 2x_1(k) && +x_2(k-m) && +x_3(k-m), \\ x_2(k+1) &= 2x_2(k) + 3x_3(k) - x_1(k-m) && -\sqrt{2}x_2(k-m), \\ x_3(k+1) &= -3x_2(k) + 2x_3(k) - x_1(k-m) && +\sqrt{2}x_3(k-m)\end{aligned}$$

where  $k \in \mathbb{Z}_0^\infty$ . In this case

$$A = \Lambda = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & -3 & 2 \end{pmatrix}, \quad (20)$$

$\lambda = 2, p = 2, q = 3$  and

$$B = \begin{pmatrix} 0 & 1 & 1 \\ -1 & -\sqrt{2} & 0 \\ -1 & 0 & \sqrt{2} \end{pmatrix}. \quad (21)$$

*It is easy to verify that conditions (9)–(14) are valid and system (1) is weakly delayed.*

In the paper, we consider a solution of initial Cauchy problem (1), (22) where

$$x(0) = x_0 = \begin{pmatrix} x_{0,1} \\ x_{0,2} \\ x_{0,3} \end{pmatrix}, \dots, x(-m) = x_{-m} = \begin{pmatrix} x_{-m,1} \\ x_{-m,2} \\ x_{-m,3} \end{pmatrix} \quad (22)$$

and  $x_{i,j}, i = 0, -1, \dots, -m, j = 1, 2, 3$  are real constants.

## 1 RESULT

Assuming, without loss of generality, that the matrix  $A$  in (1) is in its Jordan form  $\Lambda$  (this is possible due to Lemma 1), we will investigate a system

$$x(k+1) = \Lambda x(k) + Bx(k-m) \quad (23)$$

together with the initial data as given by (22). Let us transform (23) into a higher-dimensional system without delay. Let  $z^1, \dots, z^m$  be the new dependent 3-dimensional vector variables defined by the formulas

$$z^1(k) = x(k-1) \quad \Rightarrow \quad z^1(k+1) = x(k),$$

$$\begin{aligned}
z^2(k) = x(k-2) & \Rightarrow z^2(k+1) = x(k-1), \\
\vdots & \\
z^m(k) = x(k-m) & \Rightarrow z^m(k+1) = x(k-(m-1)).
\end{aligned}$$

Then, an equivalent system without delay is

$$\begin{aligned}
x(k+1) &= \Lambda x(k) & + Bz^m(k), \\
z^1(k+1) &= x(k), \\
z^2(k+1) &= z^1(k), \\
z^3(k+1) &= z^2(k), \\
\vdots & \quad \ddots \\
z^m(k+1) &= z^{m-1}(k).
\end{aligned}$$

Below, we rename the dependent variables as

$$\begin{aligned}
y_i(k) &:= x_i(k), \quad i = 1, 2, 3, \\
y_{j+3}(k) &:= z_j^1(k), \quad j = 1, 2, 3, \\
y_{j+6}(k) &:= z_j^2(k), \quad j = 1, 2, 3, \\
&\vdots \\
y_{j+3m}(k) &:= z_j^m(k), \quad j = 1, 2, 3
\end{aligned}$$

and, instead of (23), we will consider a system of  $3m+3$  equations

$$y(k+1) = \mathcal{A}y(k), \quad k \geq 0 \quad (24)$$

where

$$\mathcal{A} = \begin{pmatrix} \Lambda & \Theta & \dots & \Theta & B \\ E & \Theta & \dots & \Theta & \Theta \\ \Theta & E & \dots & \Theta & \Theta \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Theta & \Theta & \dots & E & \Theta \end{pmatrix}$$

is a  $(3m+3) \times (3m+3)$  matrix,  $\Theta$  is a  $3 \times 3$  zero matrix and  $y(k) = (y_1(k), \dots, y_{3m+3}(k))^T$ .

### 1.1 Solving the system (24)

The initial conditions for (24), as can be seen from (22) and from the performed transformation, are

$$y(0) = (y_1(0), y_2(0), \dots, y_{3m+3}(0))^T = (x(0), x(-1), \dots, x(-m))^T. \quad (25)$$

Let  $y(k) = Sw(k)$  where  $S$  is a regular transient matrix transforming  $\mathcal{A}$  to a Jordan form. Then, by (24),

$$Sw(k+1) = \mathcal{A}Sw(k)$$

and

$$w(k+1) = \gamma w(k) \quad (26)$$

where

$$\gamma = S^{-1}AS.$$

System (26) is  $(3m+3)$ -dimensional. The initial Cauchy problem for (26) derived from (25), is

$$w(0) = S^{-1}y(0), \quad (27)$$

and the solution of (26) is given by the formula (see, e.g. [7])

$$w(k) = \gamma^k w(0), \quad k = 1, 2, 3, \dots \quad (28)$$

Below, we will need the following auxiliary result.

**Theorem 3** *Let a matrix  $A$  be of the type (8) and let the entries of a matrix  $B$  satisfy (9)–(14). Then, the eigenvalues  $\mu_i, i = 1, \dots, 3m+3$  of the matrix  $\mathcal{A}$  are  $\mu_1 = \lambda, \mu_2 = p + qi, \mu_3 = p - qi, \mu_4 = \mu_5 = \dots = \mu_{3m+3} = 0$ .*

*Proof.* Computing  $\det(\mathcal{A} - \mu I)$ , where  $I$  is a  $3m+3$  by  $3m+3$  unit matrix, we get (performed computations are indicated)

$$\Delta = \det(\mathcal{A} - \mu I) = \begin{vmatrix} \Lambda - \mu E & \Theta & \Theta & \dots & \Theta & \Theta & B \\ E & -\mu E & \Theta & \dots & \Theta & \Theta & \Theta \\ \Theta & E & -\mu E & \dots & \Theta & \Theta & \Theta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \Theta & \Theta & \Theta & \dots & E & -\mu E & \Theta \\ \Theta & \Theta & \Theta & \dots & \Theta & E & -\mu E \end{vmatrix}.$$

$\cdot \mu \underbrace{\hspace{1.5cm}}_{+}$

Multiplying the first column of the matrix by  $\mu$  and adding it to the second column, we obtain:

$$\Delta = \begin{vmatrix} \Lambda - \mu E & \mu(\Lambda - \mu E) & \Theta & \dots & \Theta & \Theta & B \\ E & \Theta & \Theta & \dots & \Theta & \Theta & \Theta \\ \Theta & E & -\mu E & \dots & \Theta & \Theta & \Theta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \Theta & \Theta & \Theta & \dots & E & -\mu E & \Theta \\ \Theta & \Theta & \Theta & \dots & \Theta & E & -\mu E \end{vmatrix}.$$

$\cdot \mu \underbrace{\hspace{1.5cm}}_{+}$

Further, we multiply the second column  $\mu$  and add it to the third column and to get:

$$\Delta = \begin{vmatrix} \Lambda - \mu E & \mu(\Lambda - \mu E) & \mu^2(\Lambda - \mu E) & \dots & \Theta & \Theta & B \\ E & \Theta & \Theta & \dots & \Theta & \Theta & \Theta \\ \Theta & E & \Theta & \dots & \Theta & \Theta & \Theta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \Theta & \Theta & \Theta & \dots & E & -\mu E & \Theta \\ \Theta & \Theta & \Theta & \dots & \Theta & E & -\mu E \end{vmatrix}.$$

We repeat this until we multiply the  $m$ -th column by  $\mu$  and add it to the  $(m + 1)$ -st column finally getting the determinant:

$$\Delta = \begin{vmatrix} \Lambda - \mu E & \mu(\Lambda - \mu E) & \dots & \mu^{m-1}(\Lambda - \mu E) & \mu^m(\Lambda - \mu E) + B \\ E & \Theta & \dots & \Theta & \Theta \\ \Theta & E & \dots & \Theta & \Theta \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Theta & \Theta & \dots & \Theta & \Theta \\ \Theta & \Theta & \dots & E & \Theta \end{vmatrix}.$$

By the Laplace expansion with respect to the last column, we have:

$$\begin{aligned} \Delta &= (-1)^m \det(\mu^m(\Lambda - \mu E) + B) \\ &= (-1)^m \begin{vmatrix} \mu^m(\lambda - \mu) + b_{11} & b_{12} & b_{13} \\ b_{21} & \mu^m(p - \mu) + b_{22} & q\mu^m + b_{23} \\ b_{31} & -q\mu^m + b_{32} & \mu^m(p - \mu) + b_{33} \end{vmatrix}. \end{aligned}$$

Now, direct computation leads to:

$$\begin{aligned} \Delta &= (-1)^m [-\mu^{3m+3} + (\lambda + 2p)\mu^{3m+2} + (-2\lambda p - p^2 - q^2)\mu^{3m+1} + (\lambda p^2 + \lambda q^2)\mu^{3m} \\ &\quad + (b_{11} + b_{22} + b_{33})\mu^{2m+2} \\ &\quad + (-2b_{11}p - b_{22}\lambda - b_{22}p - b_{23}q + b_{32}q - b_{33}\lambda - b_{33}p)\mu^{2m+1} \\ &\quad + (b_{11}p^2 + b_{11}q^2 + b_{22}\lambda p + b_{23}\lambda q - b_{32}\lambda q + b_{33}\lambda p)\mu^{2m} \\ &\quad + (-b_{11}b_{22} - b_{11}b_{33} - b_{22}b_{33} + b_{12}b_{21} + b_{13}b_{31} + b_{23}b_{32})\mu^{m+1} \\ &\quad + (b_{11}b_{22}p + b_{11}b_{23}q - b_{11}b_{32}q + b_{11}b_{33}p - b_{12}b_{21}p + b_{12}b_{31}q - b_{13}b_{21}q \\ &\quad - b_{13}b_{31}p + b_{22}b_{33}\lambda - b_{23}b_{32}\lambda)\mu^m \\ &\quad + b_{11}b_{22}b_{33} - b_{11}b_{23}b_{32} - b_{12}b_{21}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31}]. \end{aligned}$$

Since (9)–(14) hold, further simplification of  $\Delta$  gives:

$$\begin{aligned} \Delta &= (-1)^m [-\mu^{3m+3} + (\lambda + 2p)\mu^{3m+2} + (-2\lambda p - p^2 - q^2)\mu^{3m+1} + (\lambda p^2 + \lambda q^2)\mu^{3m}] \\ &= (-1)^{m+1} \mu^{3m} (\mu - \lambda)(\mu^2 - 2\mu p + p^2 + q^2) \\ &= (-1)^{m+1} \mu^{3m} (\mu - \lambda)(\mu - (p + qi))(\mu - (p - qi)). \end{aligned}$$

Now it is easy to see that the roots of the equation  $\det(\mathcal{A} - \mu I) = 0$  are as formulated in the theorem.  $\square$

The following example illustrates the validity of Theorem 3 by using mathematical software.

**Example 2** Let matrices  $A, B$  be defined by formulas (20), (21). Then

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 2 & 3 & 0 & 0 & 0 & -1 & -\sqrt{2} & 0 \\ 0 & -3 & 2 & 0 & 0 & 0 & -1 & 0 & \sqrt{2} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

It is easy to verify that the eigenvalues of  $A$  are

$$\lambda_1 = 1, \lambda_2 = 2 + 3i, \lambda_3 = 2 - 3i,$$

and the eigenvalues of  $B$  are

$$\lambda_4 = \lambda_5 = \lambda_6 = 0.$$

The eigenvalues of  $\mathcal{A}$  (calculated by WolframAlpha software) are

$$\mu_1 = 1, \mu_2 = 2 + 3i, \mu_3 = 2 - 3i, \mu_4 = \dots = \mu_9 = 0.$$

Eigenvalues  $\lambda_i, i = 1, \dots, 6$  (derived by Theorem 3) are the same as eigenvalues  $\mu_j, j = 1, \dots, 6$ .

When using formula (28), it is necessary to compute powers of the matrix  $\gamma$ . The computations depend on the geometrical multiplicity of the zero eigenvalue of matrix  $B$ . Below,  $\Theta^*$  denotes a  $3m \times 3m$  zero matrix.

### 1.1.1 Case I - the geometrical multiplicity of $B$ equals 1

Due to Theorem 3, we can assume that the transition matrix  $S$  is such that

$$\gamma := \gamma_1 = \left( \begin{array}{c|ccc} \Lambda & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & & \\ \Theta & & & \end{array} \begin{array}{c} \\ \\ G_1 \\ \end{array} \right),$$

where

$$G_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix},$$

is a  $3m \times 3m$  matrix.

Then,

$$\gamma^k := \gamma_1^k = \left( \begin{array}{c|ccc} \Lambda^k & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & & \\ \Theta & & & \end{array} \begin{array}{c} \\ \\ G_k \\ \end{array} \right), \quad 1 \leq k < 3m$$

where

$$G_k = G_1^k = \begin{pmatrix} \overbrace{0 \dots 0}^k & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 \end{pmatrix}$$

and

$$\gamma^k := \gamma_1^k = \left( \begin{array}{c|ccc} \Lambda^k & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & \Theta^* & \\ \Theta & & & \end{array} \right), \quad k \geq 3m$$

where (the following formula holds for  $k \geq 1$ )

$$\Lambda^k = \begin{pmatrix} \lambda^k & 0 & 0 \\ 0 & \operatorname{Re}(p+iq)^k & \operatorname{Im}(p+iq)^k \\ 0 & -\operatorname{Im}(p+iq)^k & \operatorname{Re}(p+iq)^k \end{pmatrix}. \quad (29)$$

In (29),

$$\begin{aligned} \operatorname{Re}(p+iq)^k &= \sum_{s=0}^{\lfloor k/2 \rfloor} (-1)^s \binom{k}{2s} p^{k-2s} q^{2s}, \\ \operatorname{Im}(p+iq)^k &= \sum_{s=0}^{\lfloor k/2 \rfloor} (-1)^s \binom{k}{2s+1} p^{k-2s-1} q^{2s+1}, \end{aligned}$$

$\lfloor \cdot \rfloor$  is the floor function and, for the whole numbers  $k, \ell$ ,

$$\binom{k}{\ell} := \begin{cases} \frac{k!}{\ell!(k-\ell)!} & \text{if } k \geq \ell \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Assume that the roots  $\lambda_2, \lambda_3$  are given in the exponential form

$$\lambda_2 = re^{i\varphi}, \quad \lambda_3 = re^{-i\varphi},$$

where  $r > 0$  and  $\varphi \in (0, \pi)$ . Then,

$$\operatorname{Re} \lambda_2^k = \operatorname{Re}(re^{i\varphi})^k = \operatorname{Re} r^k e^{ki\varphi} = r^k \cos k\varphi,$$



$$\begin{aligned}
\operatorname{Im}\lambda_2^k &= \operatorname{Im}(re^{i\varphi})^k = \operatorname{Im}r^k e^{ki\varphi} = r^k \sin k\varphi, \\
\operatorname{Re}\lambda_3^k &= \operatorname{Re}\lambda_2, \\
\operatorname{Im}\lambda_3^k &= -\operatorname{Im}\lambda_2.
\end{aligned}$$

Now (29) can be written as

$$\Lambda^k = \begin{pmatrix} \lambda^k & 0 & 0 \\ 0 & r^k \cos k\varphi & r^k \sin k\varphi \\ 0 & -r^k \sin k\varphi & r^k \cos k\varphi \end{pmatrix}. \quad (30)$$

### 1.1.2 Case II - the geometrical multiplicity of $B$ equals 2

Due to Theorem 3, we can assume that the transition matrix  $S$  is such that

$$\gamma := \gamma_2 = \left( \begin{array}{c|ccc} \Lambda & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & & \\ \Theta & & H_1 & \end{array} \right),$$

where

$$H_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix},$$

is a  $3m \times 3m$  matrix.

Then,

$$\gamma^k := \gamma_2^k = \left( \begin{array}{c|ccc} \Lambda^k & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & & \\ \Theta & & H_k & \end{array} \right), \quad 1 \leq k < 3m - 1$$

where

$$H_k = H_1^k = \begin{pmatrix} \overbrace{0 \dots 0}^{k+1} & 0 & 0 & \dots & 0 \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix}$$

and

$$\gamma^k := \gamma_2^k = \left( \begin{array}{c|ccc} \Lambda^k & \Theta & \dots & \Theta \\ \hline \Theta & & & \\ \vdots & & & \\ \Theta & & \Theta^* & \end{array} \right), \quad k \geq 3m - 1$$

where powers  $\Lambda^k$  are given by (29) or (30).

## 1.2 Solution of the problem (23), (22)

The solution of system (24) is given by the formula

$$y(k) = Sw(k) = S\gamma_i^k w(0), \quad k = 1, 2, 3, \dots$$

where  $i = 1$  if the geometrical multiplicity of the zero eigenvalue of  $B$  equals 1 and  $i = 2$  if the geometrical multiplicity of the zero eigenvalue of  $B$  equals 2. Using an auxiliary matrix

$$Q = (E, \underbrace{\Theta, \dots, \Theta}_m),$$

we can write the solution of the initial problem (23), (22) in the form

$$x(k) = QS\gamma_i^k w(0), \quad i = 1, 2, \quad k = 1, 2, 3, \dots, \quad (31)$$

where (by (25) and (27))

$$w(0) = S^{-1}y(0) = S^{-1}(x(0), x(-1), \dots, x(-m))^T. \quad (32)$$

Therefore, the following theorem holds.

**Theorem 4** *Let the matrix  $A$  have the form (8) with one real eigenvalue  $\lambda_1 = \lambda$  and two complex conjugate eigenvalues  $\lambda_{2,3} = p \pm iq$ , let the elements of the matrix  $B$  satisfy (9)–(14). Then, the solution of the initial problem (1), (22) is given by formula (31) where  $i = 1$  if the geometrical multiplicity of the zero eigenvalue of  $B$  equals 1 and  $i = 2$  if the geometrical multiplicity of the zero eigenvalue of  $B$  equals 2 and  $w(0)$  is given by (32).*

## CONCLUSION

The paper is concerned with weakly delayed systems (1). Assuming that the Jordan form assigned to the matrix  $A$  is given by (8), i.e., the matrix  $A$  has one real eigenvalue  $\lambda_1 = \lambda$  and two eigenvalues  $\lambda_{2,3} = p \pm iq$  are complex conjugate with  $q \neq 0$ , a criterion is given for (1) to be weakly delayed. To solve system (1) (or equivalent system (23) where  $A$  is replaced by (8)), system (23) is transformed into a higher-dimensional system without delay (24). The solution of system (1), depending on the geometrical multiplicity of the zero eigenvalue of  $B$ , and satisfying initial data (22), is given by formula (31).

The present investigation extends the previous analysis of weakly delayed systems in [1]–[6], [8], [9]

## Reference

- [1] Diblík J., Halfarová H.: *Explicit general solution of planar linear discrete systems with constant coefficients and weak delays*. Adv. Difference Equ. 2013, Art. number: 50, doi:10.1186/1687-1847-2013-50, 1–29. Available at: <<https://link.springer.com/article/10.1186/1687-1847-2013-50>>.
- [2] Diblík J., Halfarová H.: *General explicit solution of planar weakly delayed linear discrete systems and pasting its solutions*. Abstr. Appl. Anal. 2014, doi:10.1155/2014/627295, 1–37. Available at: <<https://www.hindawi.com/journals/aaa/2014/627295/>>.
- [3] Diblík J., Halfarová H.: *Discrete systems of linear equations with weak delay*. In *XXIX International Colloquium on the Management of Educational Process*. Brno: 2011. s. 1-4. ISBN: 978-80-7231-779- 0.
- [4] Diblík, J., Halfarová, H., Šafařík, J.: *Conditional Stability and Asymptotic Behavior of Solutions of Weakly Delayed Linear Discrete Systems in  $\mathbb{R}^2$* . Discrete Dynamics in Nature and Society, Volume 2017 (2017), Article ID 6028078, p. 1-10. ISSN: 1607-887X. DOI 10.1155/2017/6028078. Available at: <<https://www.hindawi.com/journals/ddns/2017/6028078/>>.
- [5] Diblík J., Khusainov D. Ya., Šmarda Z.: *Construction of the general solution of planar linear discrete systems with constant coefficients and weak delay*. Adv. Difference Equ. 2009, Art. ID 784935, 18 pp. Available at: <<https://link.springer.com/article/10.1155/2009/784935>>.
- [6] Diblík, J., Šafařík, J.: *Solution of weakly delayed linear discrete systems in  $\mathbb{R}^3$* . In Aplimat 2017, 16th Conference on Applied Mathematics, Proceedings. First Edition. Bratislava: Slovak University of Technology, 2017. s. 454-460. ISBN: 978-80-227-4650- 2. Available at: <<http://toc.proceedings.com/33721webtoc.pdf>>.
- [7] Elaydi, S. N.: *An Introduction to Difference Equations*, Third Edition, Springer, 2005.
- [8] Šafařík, J.: *Solution of a Weakly Delayed Difference System*. In Proceedings of the 22nd Conference STUDENT EEICT 2016. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2016. s. 763-767. ISBN: 978-80-214-5350- 0.
- [9] Šafařík, J., Diblík, J., Halfarová, H.: *Weakly Delayed Systems of Linear Discrete Equations in  $\mathbb{R}^3$* . In MITAV 2015 (Matematika, informační technologie a aplikované vědy), Post-conference proceedings of extended versions of selected papers. Brno: Univerzita obrany v Brně, 2015. s. 105-121. ISBN: 978-80-7231-436-2. Available at: <<http://mitav.unob.cz/data/MITAV2015Proceedings.pdf>>.

## Acknowledgement

The authors were supported by the Grant FEKT-S-17-4225 of Faculty of Electrical Engineering and Communication, BUT.

# HOMOTHETY CURVATURE HOMOGENEITY

Alena Vanžurová

Institute of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic. Email:  
vanzurova.a@fce.vutbr.cz; alena.vanzurova@upol.cz

**Abstract:** *Curvature homogeneous manifolds are Riemannian or pseudo-Riemannian spaces whose curvature tensor of type (0,4) is “the same” in all points. Connected locally homogeneous manifolds are trivial examples. I.M. Singer [21] introduced also curvature homogeneity of higher order. We study here a natural modification of this concept, namely homothety curvature homogeneity and homothety  $r$ -curvature homogeneity [13], [14].*

**Keywords:** Riemannian space, curvature tensor, curvature operator, locally homogeneous space, curvature homogeneous space of order  $r$ , homothety curvature homogeneous space of order  $r$ .

## INTRODUCTION

Curvature homogeneous spaces in the classical setting have been studied by many authors since 60', starting from the pioneering paper by I.M. Singer [21]. Curvature homogeneous spaces are Riemannian or pseudo-Riemannian manifolds whose curvature tensor of type (0,4) is “the same” in all points. (In local coordinates it means that around two points, components of the curvature tensor are the same in the “corresponding” local maps in small neighborhoods.) Connected locally homogeneous manifolds are trivial examples. Recall that a Riemannian manifold is locally homogeneous if the pseudogroup of local isometries acts transitively on it, and if it is moreover complete then the manifold is locally isometric to a homogeneous space.

In [21], Singer introduced also curvature homogeneity of higher order. It means that not only the curvature tensors, but also their covariant derivatives (with respect to the Levi-Civita connection of the metric) coincide up to a suitable order.

We generalise here the concept of curvature homogeneity, introduced by Singer, to homothety  $r$ -curvature homogeneity (called originally curvature homogeneity of order  $r$  of type (1,3) in [13], [14]).

## HOMOTHETY CURVATURE HOMOGENEOUS SPACES

Let  $(M, g)$  denote a smooth Riemannian manifold equipped with a positive metric  $g$  where  $M$  is a smooth  $m$ -dimensional manifold,  $m \geq 2$ .

Denote by  $\mathcal{R}$  the curvature operator (the curvature tensor of type (1,3)) of  $(M, g)$  given by  $\mathcal{R}(X, Y)Z = [D_X, D_Y]Z - D_{[X, Y]}Z$  for vector fields  $X, Y, Z$  on  $M$ , where  $D$  is the Riemannian (Levi-Civita) connection of  $(M, g)$ , while  $R$  denotes here the curvature tensor (of the type (0,4)) of  $(M, g)$ . These objects are related by the identity  $R(X, Y, Z, W) = g(\mathcal{R}(X, Y)Z, W)$  for vector fields  $X, Y, Z, W$  on  $M$ .  $R_p$  or  $\mathcal{R}_p$  is the value of the corresponding tensor in the point  $p \in M$ .

**Definition 1** A Riemannian manifold  $(M, g)$  is said to be *homothety curvature homogeneous* (or (1,3)-curvature homogeneous, [13]) if given two points  $p, q \in M$ , there is a curvature-preserving linear homothety  $f: T_p M \rightarrow T_q M$ , i.e. such that  $f^*(\mathcal{R}_q) = \mathcal{R}_p$  where  $\mathcal{R}_p$  and  $\mathcal{R}_q$  are the (1,3)-curvature tensors in the points  $p$  and  $q$ , respectively.

The following was proved in [13]:

**Theorem 1** Let  $(M, g)$  be a smooth Riemannian manifold and let  $\mathcal{R}$  or  $R$  denote its curvature tensor field of type (1,3), or of type (0,4), respectively,  $p \in M$ . Then the following conditions are equivalent:

- (i) For each  $q \in M$ , there is a linear homothety  $f_q: T_p M \rightarrow T_q M$  such that  $\mathcal{R}_p = f_q^*(\mathcal{R}_q)$ .
- (ii) There is a smooth function  $\varphi$  on  $M$  such that  $\varphi(p) = 0$  and for each  $q \in M$ ,  $R_p = e^{2\varphi(q)} F_q^*(R_q)$  where  $F_q: T_p M \rightarrow T_q M$  is a linear isometry.

If one of the conditions (i), (ii) is satisfied (for  $p \in M$  fixed) then the space  $(M, g)$  is homothety curvature homogeneous.

In what follows we show that we are able to construct examples of homothety curvature homogeneous spaces in arbitrary dimensions, [13], by generalising a metric from an example by K. Sekigawa [18].

## GENERALISED CURVATURE HOMOGENEOUS SPACES

I.M. Singer introduced the following condition, [21]

$P(r)$ : For every  $p, q \in (M, g)$  there exists a linear isometry  $F: T_p M \rightarrow T_q M$  such that  $F^*((D^k \mathcal{R})_q) = (D^k \mathcal{R})_p$  for  $k = 0, 1, \dots, r$ .

A space  $(M, g)$  with such a property is said to be *curvature homogeneous up to order  $r$* . All standard first order ( $r = 1$ ) curvature homogeneous Riemannian manifolds of dimension 3 are automatically locally homogeneous. Singer proved for Riemannian spaces that a connected locally homogeneous space is curvature homogeneous of all orders, and there is always a finite number  $s \leq m(m-1)/2$  ( $m = \dim M$ ) such that, if the Riemannian manifold  $(M, g)$  is curvature homogeneous up to order  $s$ , then it is automatically locally homogeneous, i.e. "too nice" in a sense.

We introduce the following condition (for each integer  $k \geq 0$  separately):

$Q(k)$ : For every  $p, q \in (M, g)$  there exists a linear homothety  $h: T_p M \rightarrow T_q M$  such that  $h^*((D^k R)_q) = (D^k R)_p$ .

**Definition 2** A space  $(M, g)$  satisfying the conditions  $Q(0), \dots, Q(r)$  is called to be *curvature homogeneous up to order  $r$  and of type (1,3), briefly homothety  $r$ -curvature homogeneous*.

Note that in our definition the linear homotheties above are in general completely independent for different integers  $k$ . The following technical result (an analogy for higher order of Theorem 1) can be checked, [14], and used in examples:

**Theorem 2** The following conditions for a smooth Riemannian manifold  $(M, g)$  are equivalent:

- (i)  $(M, g)$  satisfies the condition  $Q(k)$ , i.e., for every  $p, q \in (M, g)$  there exists a linear homothety  $h: T_p M \rightarrow T_q M$  such that  $h^*((D^k R)_q) = (D^k R)_p$ .
- (ii) There is a smooth function  $\varphi$  on  $M$  such that  $\varphi(p) = 0$  for a fixed  $p \in M$  and  $(D^k \mathcal{R})_p = e^{(k+2)\varphi(q)} F^*((D^k \mathcal{R})_q)$  for each  $q \in M$ , where  $F: T_p M \rightarrow T_q M$  is a linear isometry.

We proved [14] that for a 3-dimensional Riemannian manifold of Sekigawa type [18] the conditions  $Q(0)$  and  $Q(1)$  are satisfied, therefore  $(M, g)$  is homothety 1-curvature homogeneous,  $Q(2)$  is not satisfied, hence  $(M, g)$  is not homothety 2-curvature homogeneous, and  $(M, g)$  is not locally homogeneous:

**Example 1** Recall that an example by K. Sekigawa on  $R^3[w, x, y]$ , [18], [14], has a metric expressed with respect to the particular orthonormal co-frame  $\omega^0 = f(x)dw$ ,  $\omega^1 = dx - ydw$ ,  $\omega^2 = dy + xdw$  by the formula  $g = \sum_{i=0}^2 (\omega^i)^2$  where  $a, b$  are positive real numbers and  $f(x) = ae^x + be^{-x}$ . It is known that the space  $(R^3, g)$  is simply connected, complete, irreducible and it satisfies the condition  $P(0)$ , i.e. it is 0-curvature homogeneous. But it is not locally homogeneous, hence it does not satisfy the condition  $P(1)$ . In our new (1,3)-setting, it can be checked that the conditions  $Q(0)$ ,  $Q(1)$  hold, but the condition (ii) from Theorem 2 is not satisfied, hence  $Q(2)$  is not valid. Together, we verified that the space  $(R^3, g)$  is 0-curvature homogeneous but is not 1-curvature homogeneous, it is homothety 1-curvature homogeneous but is not homothety 2-curvature homogeneous, and  $(R^3, g)$  is not locally homogeneous, i.e. is not "too nice".

## GENERALISATION OF K. SEKIGAWA'S EXAMPLE FOR ARBITRARY DIMENSIONS

The above example can be generalised for an arbitrary dimension  $m = n + 1$  as follows. Consider  $R^{n+1}$  with standard coordinates  $(w, x^1, \dots, x^n)$ . Take an open subset  $U$  of  $R^2[w, x^1]$ , a non-vanishing (no-where zero) smooth function  $f: U \rightarrow R$  on  $U$ , a skew-symmetric smooth  $(n \times n)$ -matrix function  $A(w) = (A_j^i(w))$  of one variable, and define the metric  $g_{f,A(w)}$  (on an open subset  $\tilde{U}$  of  $R^{n+1}$ ) by

$$g_{f,A(w)} = \sum_{j=0}^{n+1} \omega^j \otimes \omega^j$$

with respect to a special orthonormal co-frame (as it can be easily checked) introduced by

$$\omega^0 = f(w, x^1)dw, \quad \omega^i = dx^i + \sum_{j=1}^n A_j^i(w)x^j dw, \quad i = 1, \dots, n.$$

Let  $\langle X_0, X_1, \dots, X_n \rangle$  be the corresponding orthonormal basis of vector fields. According to [12, pp. 51-52] (see also [4] and [5]), the above metric is a generalisation of an example by K. Sekigawa [18]. By standard evaluation we can verify that the Riemannian (0,4)-curvature tensor is given by the formula ([12])

$$R = -4f^{-1}f''_{x^1x^1} \omega^0 \wedge \omega^1 \otimes \omega^0 \wedge \omega^1. \quad (1)$$

Hence  $R_{0110} = -R_{0101} = -R_{1010} = R_{1001} = f^{-1}f''_{x^1x^1}$  and all other components  $R_{ijkl}$  vanish ( $f^{-1} = 1/f$ ).

Now, the metric  $g_{f,A(w)}$  above is nonflat and curvature homogeneous, in the classical sense, if and only if the function  $f$  satisfies  $f^{-1}f''_{x^1x^1} = k$  where  $k$  is a non-zero constant. Equivalently, if and only if the function  $f$  is a solution of the second order homogeneous differential equation with constant coefficients  $f''_{x^1x^1} - kf = 0$ . Solving the equation we get that  $f$  must take the form

$$\begin{aligned} f(w, x^1) &= a(w) \exp(\sqrt{k}x^1) + b(w) \exp(-\sqrt{k}x^1) \text{ if } k > 0, \\ f(w, x^1) &= a(w) \cos(\sqrt{-k}x^1) + b(w) \sin(\sqrt{-k}x^1) \text{ if } k < 0 \end{aligned} \quad (2)$$

where  $a(w)$  and  $b(w)$  are differentiable real functions such that  $f(w, x^1) > 0$  in  $U$ . (Here  $U$  can be the whole plane in the case  $k < 0$  and an open strip in the plane for  $k > 0$ ). Recall that this class of spaces is remarkable because it includes all irreducible curvature homogeneous spaces which are not locally homogeneous and whose curvature tensor  $R$  “is the same” as that of a Riemannian symmetric space (so-called “non-homogeneous relatives of symmetric spaces”, see [11]).

For the following, we need a technical Lemma:

**Lemma 1** *Let  $(M, g)$  be a Riemannian manifold and let  $\langle E_1, \dots, E_n \rangle$  be an orthonormal moving frame on a domain  $U \subset M$ . Fix a point  $p \in U$ . Suppose that, with respect to this moving frame,  $R_{ijkl}(q) = \phi(q)R_{ijkl}(p)$  for each point  $q \in U$  and for all choices of indices, where  $\phi(q)$  is a smooth and positive function on  $U$ . Then there is a smooth function  $\varphi(q)$  such that  $\varphi(p) = 0$  and, for each point  $q$ ,  $\mathcal{R}_p = e^{2\varphi(q)}F_q^*(\mathcal{R}_q)$  where  $F_q: T_pM \rightarrow T_qM$  is a linear isometry.*

*Proof.* The condition above means that

$$\mathcal{R}_q(E_{iq}, E_{jq}, E_{kq}, E_{\ell q}) = \phi(q)\mathcal{R}_p(E_{ip}, E_{jp}, E_{kp}, E_{\ell p})$$

for each  $q$  and all indices. The map  $F = F_q: T_pM \rightarrow T_qM$  which sends the orthonormal frame  $\langle E_{1p}, \dots, E_{np} \rangle$  at  $p$  onto the orthonormal frame  $\langle E_{1q}, \dots, E_{nq} \rangle$  at  $q$  is a linear isometry. Then we can write

$$\begin{aligned} \mathcal{R}_q(E_{iq}, E_{jq}, E_{kq}, E_{\ell q}) &= \mathcal{R}_q(FE_{ip}, FE_{jp}, FE_{kp}, FE_{\ell p}) \\ &= (F^*\mathcal{R}_p)(E_{ip}, E_{jp}, E_{kp}, E_{\ell p}). \end{aligned}$$

This is valid for every choice of the indices and hence  $F^*(\mathcal{R}_p) = \phi(q)\mathcal{R}_p$ . Because  $\phi$  is smooth and positive we get  $\mathcal{R}_p = 1/\phi(q)F^*(\mathcal{R}_p)$ ,  $e^{2\varphi(q)} = 1/\phi(q)$ , and  $\varphi(q) = 1/2 \ln(1/\phi(q))$ .

Let now  $f$  be an arbitrary smooth function on  $R^2$  such that  $f$  and  $f^{-1}f''_{x^1x^1}$  are nonzero in all points and such that  $f''_{x^1x^1}/f$  is never a constant in an open domain of  $R^2$ . Then the corresponding metric  $\bar{g} = g_{f,A(w)}$  defined on  $R^{n+1}$  has the curvature components as in the formula (1). We can see that, for these curvature components  $\bar{R}_{ijkl}$ , we have

$$\bar{R}_{ijkl}(q) = (f^{-1}(q)f''_{x^1x^1}(q))/(f^{-1}(p)f''_{x^1x^1}(p))\bar{R}_{ijkl}(p)$$

for any pair of points  $p, q \in R^{n+1}$  and all indices  $i, j, k, \ell$ . Let now the point  $p$  be fixed. Then the assumptions of the Lemma 1 are satisfied, where the corresponding function  $\phi(q)$  is defined as

$$\phi(q) = f^{-1}(q)f''_{x^1x^1}(q)/(f^{-1}(p)f''_{x^1x^1}(p))$$

and hence positive. From Theorem 1 and our special assumptions we deduce that the space  $(R^{n+1}, \bar{g})$  is (1,3)-curvature homogeneous but not (0,4)-curvature homogeneous.

### HOMOTHETY CURVATURE HOMOGENEITY IN DIMENSION 3

In dimension  $m = 3$  we are able to prove more ( $m = 4$  was treated in [19], [20]). We show that actually, the class of curvature homogeneous 3-dimensional analytic Riemannian manifolds depends on 3 real analytic functions of 2 variables. In comparison the class of homothety curvature homogeneous 3-dimensional analytic Riemannian manifolds depends on 1 analytic function of 3 variables and 3 analytic functions of 2 variables, consequently is much bigger.

In the classical setting, a three-dimensional Riemannian manifold  $(M, g)$  is curvature homogeneous if and only if the Ricci eigenvalues  $\varrho_1, \varrho_2, \varrho_3$  are constant at all points. Indeed, the curvature tensor  $\mathcal{R}$  is uniquely determined by the corresponding Ricci tensor  $\varrho$  and the metric  $g$ , see formula (3) below. In what follows metrics and functions are supposed to be real analytic. The following results can be proved by means of the Cauchy-Kowalewski theorem:

**Theorem 3** ([6]) *All real analytic Riemannian manifolds with the prescribed constant Ricci eigenvalues  $\varrho_1 = \varrho_2 \neq \varrho_3$  depend, up to a local isometry, on two arbitrary (real analytic) functions of one variable.*

The case of distinct constant eigenvalues was discussed in [7] and classified in [9]:

**Theorem 4** ([9]) *All real analytic Riemannian manifolds with the prescribed distinct constant Ricci eigenvalues  $\varrho_1 > \varrho_2 > \varrho_3$  depend, up to a local isometry, on three arbitrary (real analytic) functions of two variables.*

The classification of all triplets of distinct real numbers which can be realized as Ricci eigenvalues on a 3-dimensional locally homogeneous space was made in [8]. From these results it follows that the spaces  $(M, g)$  with prescribed constant Ricci eigenvalues are, with rare exceptions, not locally homogeneous, and on an open subset of  $R^3$ , the prescribed triplets of constant Ricci eigenvalues can be realized only on spaces which are not locally homogeneous.

Theorem 4 was later generalized in

**Theorem 5** ([10]) *All Riemannian metrics defined in a domain  $U \subset R^3[x, y, z]$  with the prescribed distinct real analytic Ricci eigenvalues  $\varrho_1(x, y, z) > \varrho_2(x, y, z) > \varrho_3(x, y, z)$  depend, up to a local isometry, on three arbitrary real analytic functions of two variables. Every solution of the problem is defined at least locally, i.e. in a neighborhood  $U' \subset U$  of a fixed point  $p \in U$ .*

In a domain  $U \subset R^3[x, y, z]$ , fix a point  $p$  and choose a real analytic function  $\varphi(x, y, z)$  on  $U$  vanishing at  $p$ . According to Theorem 5 let us construct a (local) Riemannian metric  $g$  about  $p$  such that their Ricci eigenvalues are of the form  $\varrho_i = e^{2\varphi}\lambda_i$ ,  $i = 1, 2, 3$  where  $\lambda_1 > \lambda_2 > \lambda_3$  are nonzero constants. Then  $\varrho_1 > \varrho_2 > \varrho_3$  at each point as required. Denote by  $g$  such a local metric. Choose a Ricci adapted orthonormal moving frame  $\langle E_1, E_2, E_3 \rangle$  in a neighborhood of  $p$ . Then we get  $\varrho_{ij} = \varrho_i\delta_{ij} = \varrho_j\delta_{ij} = e^{2\varphi}\lambda_i\delta_{ij} = e^{2\varphi}\lambda_j\delta_{ij}$  for  $i, j = 1, 2, 3$ .

Now the formula for components of the curvature tensor which is valid in the 3-dimensional case (cf. [1], [2], [3])

$$R_{ijkl} = \frac{1}{n-2}(g_{ik}\varrho_{jl} - g_{il}\varrho_{jk} + g_{jl}\varrho_{ik} - g_{jk}\varrho_{il}) + \frac{\tau}{(n-1)(n-2)}(g_{il}g_{jk} - g_{ik}g_{jl}), \quad (3)$$



(where  $R_{ijk\ell}$  denote the components of  $\mathcal{R}$ ,  $\varrho_{ij}$  the components of the Ricci tensor and  $\tau$  the scalar curvature, with respect to any local moving frame) is reduced to

$$R_{ijk\ell} = \frac{e^{2\varphi}}{n-2}(\lambda_j(\delta_{ik}\delta_{j\ell} - \delta_{i\ell}\delta_{jk}) + \lambda_i(\delta_{j\ell}\delta_{ik} - \delta_{jk}\delta_{i\ell}) \\ + \frac{e^{2\varphi}(\lambda_1 + \lambda_2 + \lambda_3)}{(n-1)(n-2)}(\delta_{i\ell}\delta_{jk} - \delta_{ik}\delta_{j\ell}).$$

for  $i, j, k, \ell = 1, 2, 3$ . In particular, we get

$$R_{ijk\ell}(p) = \frac{1}{n-2}(\lambda_j(\delta_{ik}\delta_{j\ell} - \delta_{i\ell}\delta_{jk}) + \lambda_i(\delta_{j\ell}\delta_{ik} - \delta_{jk}\delta_{i\ell}) \\ + \frac{(\lambda_1 + \lambda_2 + \lambda_3)}{(n-1)(n-2)}(\delta_{i\ell}\delta_{jk} - \delta_{ik}\delta_{j\ell}).$$

Now, the assumption of Lemma 1 is satisfied and hence Theorem 1 can be used. Therefore in general, the corresponding metric  $g$  is homothety curvature homogeneous and not curvature homogeneous.

A Riemannian manifold is called *generic* if the Ricci eigenvalues are distinct in all points. Due to Theorem 1, we see easily that all generic 3-dimensional homothety curvature homogeneous Riemannian manifolds are constructed just in the way described above. Hence we get the following

**Theorem 6** *All generic real analytic homothety curvature homogeneous three-dimensional Riemannian manifolds are locally parametrised, up to a local isometry, by one arbitrary real analytic function of three variables and three arbitrary real analytic functions of two variables.*

## THE PSEUDO-RIEMANNIAN CASE

In pseudo-Riemannian case the situation is a bit different. Three-dimensional Lorentzian manifolds were examined in [16], [17]. Pseudo-Riemannian curvature homogeneous spaces, of arbitrary dimension, signature and order, were examined in [22], [23].

In [24] the authors constructed irreducible pseudo-Riemannian manifolds of arbitrary signature  $(p, q)$  with the same curvature tensor as a pseudo-Riemannian symmetric space which is a direct product of a two-dimensional Riemannian space form  $M_2(c)$  and a pseudo-Euclidean space of the signature either  $(p, q-2)$  or  $(p-2, q)$ . Their examples are again inspired by three-dimensional examples of Sekigawa type but the construction of metrics is modified in comparison with the Riemannian case. They consider  $R^{n+1}$  with standard coordinates  $(w, x^1, \dots, x^n)$ , a sequence  $(\epsilon_0, \epsilon_1, \dots, \epsilon_n)$  of prescribed signatures  $\epsilon_i = \pm 1$ , and a family of positive functions of  $w$ ,  $\lambda_1(w), \dots, \lambda_{n-1}(w)$ . Further, the matrix function  $A = A_j^i(w)$  has the only non-zero entries just  $A_i^{i+1}(w) = \lambda_i(w)$ ,  $A_{i+1}^i(w) = -\epsilon_i \epsilon_{i+1} \lambda_i(w)$ ,  $i = 1, \dots, n-1$ . Now we can use the same definition for 1-forms as in the Riemannian case to arrive to a pseudo-orthonormal co-frame, and introduce the metric in an analogous way. Conditions under which the space is curvature homogeneous are settled in [24].

Let us also mention here that our results were an inspiration for P. Gilkey and his co-workers who started to develop our theory for pseudo-Riemannian spaces and constructed new examples, [15] and the references therein. Among others, the following was proved in [15]:

**Lemma 2** *The following conditions are equivalent for a (pseudo-)Riemannian manifold  $(M, g)$  (and the manifold is said to be homothety  $k$ -curvature homogeneous if any of them is satisfied):*

- (i) *Given any two points  $p, q \in M$  there is a linear homothety  $\Phi = \Phi_{p,q}$  from  $T_p M$  to  $T_q M$  so that if  $0 \leq \ell \leq k$ , then  $\Phi^*(D^\ell \mathcal{R})_q = (D^\ell \mathcal{R})_p$ .*
- (ii) *Given any two points  $p, q \in M$  there is a linear isometry  $\phi = \phi_{p,q}$  from  $T_p M$  to  $T_q M$  and there exists  $0 \neq \lambda = \lambda_{p,q}$  so that if  $0 \leq \ell \leq k$ , then  $\phi^*(D^\ell \mathcal{R})_q = \lambda^{-\ell-2}(D^\ell \mathcal{R})_p$ .*
- (iii) *There exist constants  $\epsilon_{ij}, c_{i_1 \dots i_{\ell+4}}$  such that for all  $q \in M$  there exists a basis  $\{e_1^q, \dots, e_m^q\}$  for  $T_q M$  and there exists a real number  $\lambda_q \neq 0$  so that if  $0 \leq \ell \leq k$ , then for all indices  $i_1, i_2, \dots$ , the value  $g_q(e_{i_1}^q, e_{i_2}^q) = \epsilon_{ij}$  and  $(D^\ell \mathcal{R})_q(e_{i_1}^q, e_{i_2}^q, \dots, e_{i_{\ell+4}}^q) = \lambda_q^{-\ell-2} c_{i_1 \dots i_{\ell+4}}$ .*

## CONCLUSION

We propose investigation of a new topic, namely homothety  $r$ -curvature homogeneity (curvature homogeneity of type (1,3) and order  $r$ ) for Riemannian spaces. The class of 1-curvature homogeneous Riemannian spaces of type (1,3) is much wider than the class of 1-curvature homogeneous spaces of type (0,4). We proved that for a 3-dimensional Riemannian manifold  $(M, g)$  of Sekigawa type [18] the conditions  $Q(0)$  and  $Q(1)$  are satisfied, therefore  $(M, g)$  is homothety 1-curvature homogeneous,  $Q(2)$  is not satisfied, hence  $(M, g)$  is not homothety 2-curvature homogeneous, and  $(M, g)$  is not locally homogeneous. So we bring an example of a space that is homothety 1-curvature homogeneous but not homothety 2-curvature homogeneous, and not locally homogeneous. Among others, our results were an inspiration for P. Gilkey and his co-workers [15] who started to develop the theory for pseudo-Riemannian spaces and constructed new examples.

## References

- [1] Favard, J.: *Cours de Géométrie Différentielle Locale*. Gauthier-Villars, Paris 1957.
- [2] Wey, H.: Reine Infinitesimalgeometrie, *Math. Zeitschrift* **2** (1918), 384-411.
- [3] Yano, K.: *Integral Formulas in Riemannian Geometry*. Marcel Dekker, Inc., 1970.
- [4] Kowalski, O., Tricerri, F., Vanhecke, L.: *New examples of non-homogeneous Riemannian manifolds whose cubature tensor is that of a Riemannian symmetric space*. C. R. Acad. Sci. Paris, **311**, Série I, (1990), 355-360.
- [5] Kowalski, O., Tricerri, F., Vanhecke, L.: *Curvature homogeneous Riemannian manifolds*. J. Math. Pures Appl., **71**, 1992, 471-501.
- [6] Kowalski, O.: *A classification of Riemannian 3-manifolds with constant principal Ricci curvatures  $\varrho_1 = \varrho_2 \neq \varrho_3$* . Nagoya Math. J. **132** (1993), 1-36.
- [7] Kowalski, O., Prüfer, F.: *On Riemannian 3-manifolds with distinct constant Ricci eigenvalues*. Math. Ann. **300** (1994), 17-28.
- [8] Kowalski, O., Nikčević, S.Ž.: *On Ricci eigenvalues of locally homogeneous Riemannian manifolds*. Geometriae Dedicata **62** (1996), 65-72.
- [9] Kowalski, O., Vlášek, Z.: *Classification of Riemannian 3-manifolds with distinct constant principal Ricci curvatures*. Bulletin of the Belgian Mathematical Society-Simon Stevin **5** (1998), 59-68.

- [10] Kowalski, O., Vlášek, Z.: *On 3D-manifolds with prescribed Ricci eigenvalues*. In: Complex, Contact and Symmetric Manifolds-In Honor of L. Vanhecke. Progress in Mathematics, Vol. **234**, Birkhäuser Boston-Basel-Berlin, pp. 187-208 (2005).
- [11] E. Boeckx, O. Kowalski, L. Vanhecke: Non-homogeneous relatives of symmetric spaces. *Diff. Geom. and Appl.* **4** (1994), 45-69.
- [12] Boeckx, E., Kowalski, O., Vanhecke, L.: *Riemannian Manifolds of Conullity two*. World Scientific, 1996.
- [13] Kowalski O., Vanžurová, A.: *On curvature-homogeneous spaces of type (1,3)*. Math. Nachr. 284, No. 17-18, 2127-2132 (2011).
- [14] Kowalski O., Vanžurová, A.: *On a Generalization of Curvature Homogeneous Spaces*. Results in Mathematics: Vol. 63, Issue 1 (2013), pp. 129-134.
- [15] García-Río, E., Gilkey, P., Nikčević, S.: *Homothety curvature homogeneity and homothety homogeneity*. To appear.
- [16] P. Bueken, P.: *On curvature homogeneous three-dimensional Lorentzian manifolds*. J. Geom. Phys. 22(1997), 349-362.
- [17] Bueken, P. and Djorić, M.: *Three-dimensional Lorentz metrics and curvature homogeneity of order one*. Ann. Glob. Anal. Geom. 18 (2000), 85-103.
- [18] Sekigawa, K.: *On some 3-dimensional Riemannian manifolds*. Hokkaido Math. J. 2 (1973), 259-270.
- [19] Sekigawa, K., Suga, H. and Vanhecke, L.: *Four-dimensional curvature homogeneous spaces*. Comment. Math. Univ. Carolinae **33** (1992), 261-268.
- [20] Sekigawa, K., Suga, H. and Vanhecke, L.: *Curvature homogeneity for four-dimensional manifolds*. J. Korean Math. Soc **32** (1995), 93-101.
- [21] Singer, I.M.: *Infinitesimally homogeneous spaces*. Comm. Pure Appl. Math. 13(1960), 685-697.
- [22] Gilkey, P.J.: *The Geometry of Curvature Homogeneous Pseudo-Riemannian Manifolds*. ICP Advanced Texts in Mathematics - Vol. 2. Imperial College Press, 2007.
- [23] Brozos-Vázquez, M., Gilkey, P. and Nikčević, S.: *Geometric realizations of curvature*. Imperial College Press (in preparation).
- [24] Kowalski O., Dušek, Z.: *Pseudo-Riemannian spaces modelled on symmetric spaces*. Monatsh. Math. Vol. 165, Issue 3-4 (2012), 319-326.

## Acknowledgement

The author was supported by the project of specific university research of the Brno University of Technology No. FAST-S-16-3385.

# LIMITATION OF SEQUENCES OF BANACH SPACE THROUGH INFINITE MATRIX

Tomáš Visnyai

Faculty of Chemical and Food Technology STU in Bratislava,  
Radlinského 9, 812 37 Bratislava 1, Slovak Republic.

Email: visnyai@stuba.sk

**Abstract:** *The aim of the paper is to discuss some properties of the matrix transformation of sequences of elements of Banach space.*

**Keywords:** matrix transformation, sequences of elements of Banach space, convergence.

## INTRODUCTION

In this article we will investigate some matrix methods of summability of sequences of elements of an arbitrary Banach space  $(X, \|\cdot\|)$ . Let  $A = (a_{nk})$  is an infinite matrix with real numbers and let  $\alpha = (\alpha_k)$  is a sequence of elements of the space  $(X, \|\cdot\|)$ . A transformed sequence  $\beta = (\beta_n)$  is defined as  $\beta_n = \sum_{k=1}^{\infty} a_{nk} \alpha_k$  provided that the series on the right side converges. We will present the necessary and sufficient condition to existence of the transformed sequence  $\beta = (\beta_n)$  for all sequences  $\alpha = (\alpha_k)$  which are converge to the zero element of the space  $X$ . Next, that the  $\beta = (\beta_n)$  to be bounded for all bounded sequences  $\alpha = (\alpha_k)$ . Finally, that the  $\beta = (\beta_n)$  to converges to the same element as the sequence  $\alpha = (\alpha_k)$ . These conditions will relate to the infinite matrix  $A = (a_{nk})$ .

## 1. PRELIMINARIES

In this section we will investigate the sequences of elements of Banach space. Let as first we show some properties about these sequences. The notion Banach space we will considered as well-known notion. The elements of Banach space we denote as  $\alpha, \beta, \dots$ . The zero element as  $\Theta$  and the unit element as  $\varepsilon$ .

**Definition 1.** The sequence  $(\alpha_n)_1^{\infty}$  converges to  $\alpha$  if

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \forall n \in \mathbb{N} : n \geq n_0 \Rightarrow \|\alpha_n - \alpha\| < \varepsilon.$$

Let  $A = (a_{mn})$  is an infinite matrix with real numbers. The linear transformation defined by this matrix, transforms each sequence  $(\alpha_n)_1^{\infty}$  to  $(\beta_m)_1^{\infty}$ , if the series

$$\beta_m = \sum_{n=1}^{\infty} a_{mn} \alpha_n \quad (1)$$

converges for all  $m = 1, 2, \dots$ . Note that the expression (1) we can interpret as a series in Banach space.

**Proposition 2.** The series  $\sum_{n=1}^{\infty} \xi_n$  converges if and only if

$$\forall \varepsilon > 0 \quad \exists n_0 \in N \quad \forall n, m \in N : m > n \geq n_0 \Rightarrow \|\xi_{n+1} + \dots + \xi_m\| < \varepsilon.$$

Then provided that the series (1) converges for all  $m=1,2,\dots$  we have the sequence  $(\beta_m)_1^\infty$ . Now we show which properties must have the matrix  $A=(a_{mn})$  to existence the sequence  $(\beta_m)_1^\infty$  for  $\alpha=(\alpha_n)_1^\infty, \alpha_n \rightarrow \Theta$ . Next that the  $(\beta_m)_1^\infty$  to be bounded for all bounded sequences  $(\alpha_n)_1^\infty$ . Finally that the  $(\beta_m)_1^\infty$  converges to the same element as the sequence  $(\alpha_n)_1^\infty$ .

## 2. MAIN RESULTS

Now we introduce some results which generalize the statements from [2] and [3] for the sequences of elements of Banach space.

**Theorem 3.** Let  $A=(a_{mn})$  is an infinite matrix with real numbers.

- a) Then  $A\alpha$  exists for all bounded sequences if and only if  $A\alpha$  exists for all sequences  $\alpha=(\alpha_n)_1^\infty, \alpha_n \rightarrow \Theta$ .
- b) A necessary and sufficient condition for  $A\alpha$  to exists for all sequences  $\alpha=(\alpha_n)_1^\infty, \alpha_n \rightarrow \Theta$  is that

$$\sum_{n=1}^{\infty} |a_{mn}| < \infty, \text{ for all } m=1,2,\dots \quad (2)$$

*Proof.* b) The condition (2) is enough to existence of  $A\alpha$  for an arbitrary bounded sequence  $\alpha=(\alpha_n)_1^\infty$  i.e., the series  $\sum_{n=1}^{\infty} a_{mn}\alpha_n$  converges. This follows from Theorem 3 a). The sequence  $\alpha=(\alpha_n)_1^\infty$  is bounded i.e.,  $\|\alpha_n\| \leq M, M \in R$ . The series  $\sum_{n=1}^{\infty} a_{mn}\alpha_n$  converges, because

$$\|a_{m+1}\alpha_{n+1} + \dots + a_{mp}\alpha_p\| \leq M \cdot (|a_{m+1}| + \dots + |a_{mp}|) < M \cdot \frac{\varepsilon}{M} = \varepsilon$$

for all  $m=1,2,\dots$ . Suppose that the (2) is not true. For example  $m=1, \sum_{n=1}^{\infty} |a_{1n}| = \infty$ . Then there exists a sequence of non-negative integers  $0 = n_1 < n_2 < \dots < n_j < \dots$ , such that

$$\sum_{k=n_j+1}^{n_{j+1}} |a_{1k}| > 1, j=1,2,\dots$$

Put  $\alpha_k = \frac{\varepsilon}{j} \text{sign} a_{1k}$ ,  $\|\varepsilon\|=1$  for  $n_j+1 \leq k \leq n_{j+1}, j=1,2,\dots$ , then clearly  $\alpha_k \rightarrow \Theta$  and

$$\left\| \sum_{k=n_j+1}^{n_{j+1}} a_{1k} \alpha_k \right\| = \frac{\|\varepsilon\|}{j} \cdot \sum_{k=n_j+1}^{n_{j+1}} |a_{1k}| > \frac{\|\varepsilon\|}{j}, \text{ hence } \left\| \sum_{k=1}^{n_{j+1}} a_{1k} \alpha_k \right\| = \left\| \sum_{i=1}^j \sum_{k=n_i+1}^{n_{i+1}} a_{1k} \alpha_k \right\| > \sum_{i=1}^j \frac{1}{i}.$$

We proved that the series  $\sum_{k=1}^{\infty} a_{1k} \alpha_k$  diverges, hence  $A\alpha$  does not exist. Therefore the condition (2) is equivalent to the existence of  $A\alpha$  for all sequences  $\alpha=(\alpha_n)_1^\infty$  if and only if  $\alpha_n \rightarrow \Theta$ .

a) If  $A\alpha$  exists for all bounded sequences  $\alpha$ , then it exists for every sequence which converges to  $\Theta$ , because such a sequence is bounded. Now if  $A\alpha$  does not exist for some

bounded sequence  $\alpha$  then the (2) is not true and from the first part of the proof we have, that there exists such a sequence converges to  $\Theta$ . Hence  $A\alpha$  does not exist.  $\square$

**Example 4.** Now we show that if  $\alpha$  is a bounded sequence of real numbers and the condition of Theorem 3 is true for the matrix  $A = (a_{mn})$ , then the sequence  $A\alpha$  may not be bounded. Denote  $\beta = A\alpha$ . The condition of Theorem 3 guarantees the existence of the sequence  $\beta$ , but it is not enough to be bounded sequence  $\beta$ . Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 1 & 2 & 0 & 0 & 0 & \dots \\ 1 & 2 & 3 & 0 & 0 & \dots \\ 1 & 2 & 3 & 4 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \text{ and } \alpha = \begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \vdots \end{pmatrix}, \text{ then } A\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 1 & 2 & 0 & 0 & 0 & \dots \\ 1 & 2 & 3 & 0 & 0 & \dots \\ 1 & 2 & 3 & 4 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \vdots \end{pmatrix} = (1, 2, 3, \dots) = \beta.$$

It is easy to see that the sequence  $\beta$  is not bounded sequence of real numbers. We need more conditions for  $\sum_{n=1}^{\infty} |a_{mn}|$  to be bounded sequence  $\beta$ .

**Theorem 5.** Let  $A = (a_{mn})$  is an infinite matrix with real numbers. A sufficient and necessary condition for  $A$  to transform all bounded sequences  $\alpha = (\alpha_n)_1^{\infty}$  to bounded sequence  $\beta = A\alpha$  is that there exists a constant  $M > 0$  such that

$$\sum_{n=1}^{\infty} |a_{mn}| \leq M \quad (3)$$

for all  $m = 1, 2, \dots$

*Proof.* The condition (3) is sufficient. It can prove same as in Theorem 3. Let

$$\limsup_{m \rightarrow \infty} \sum_{n=1}^{\infty} |a_{mn}| = \infty.$$

Then we have two cases:

- a) there is an  $j$  such that  $\limsup_{m \rightarrow \infty} |a_{mj}| = \infty$ ,
- b) the case a) is not true, i.e.  $\limsup_{m \rightarrow \infty} |a_{mn}| < \infty$  for all  $n = 1, 2, \dots$

In case a) put

$$\alpha_j = \varepsilon \text{ for } \|\varepsilon\| = 1 \quad \text{and} \quad \alpha_l = \Theta \text{ for } l \neq j.$$

Therefore  $\beta_m = \sum_{n=1}^{\infty} a_{mn} \alpha_n = a_{mj} \varepsilon$  for all  $m = 1, 2, \dots$  and

$$\limsup_{m \rightarrow \infty} \|\beta_m\| = \limsup_{m \rightarrow \infty} |a_{mj} \varepsilon| = \limsup_{m \rightarrow \infty} |a_{mj}| \|\varepsilon\| = \limsup_{m \rightarrow \infty} |a_{mj}| = \infty.$$

Hence for the sequence  $\alpha = (\alpha_l)_1^{\infty}$  the sequence  $\beta = (\beta_m)_1^{\infty}$  is not bounded, while  $\alpha_l \rightarrow \Theta$ .

In case b)  $\limsup_{m \rightarrow \infty} |a_{mn}| < \infty$  for all  $n = 1, 2, \dots$ . Then for any  $i = 1, 2, \dots$  there exists  $K_i$  such that  $|a_{mi}| \leq K_i$   $m = 1, 2, \dots$ . Now put  $M(r) = K_1 + \dots + K_r$  and from this

$$\sum_{j=1}^r |a_{mj}| \leq M(r) \text{ for all } m = 1, 2, \dots \quad (4)$$

Since  $\sum_{n=1}^{\infty} |a_{mn}| < \infty$  for  $m=1,2,\dots$  (see Theorem 3) and  $\limsup_{m \rightarrow \infty} \sum_{n=1}^{\infty} |a_{mn}| = \infty$  we can choose two increasing sequences of non-negative integers  $(m_k)_1^{\infty}, (n_k)_1^{\infty}$  such that their terms satisfy the following conditions. Chose  $m_1=1$  and  $n_1$  such that  $\sum_{n=n_1+1}^{\infty} |a_{1n}| < 1$ . Let we have  $m_{k-1}$  and  $n_{k-1}$ . The other terms we chose as  $m_k > m_{k-1}$  such that

$$\sum_{n=1}^{\infty} |a_{m_k n}| > (k+1)M(n_{k-1}) + k^2 + 1 \quad (5)$$

and  $n_k > n_{k-1}$  such that

$$\sum_{n=n_k+1}^{\infty} |a_{m_k n}| < 1 \quad (6)$$

Then from (5) and (6) we have  $\sum_{n=1}^{n_k} |a_{m_k n}| > (k+1)M(n_{k-1}) + k^2$  and from this and (4) we get

$$\sum_{n=n_{k-1}+1}^{n_k} |a_{m_k n}| > k.M(n_{k-1}) + k^2 \quad (7)$$

Now put

$$\alpha_j = \varepsilon, \|\varepsilon\| = 1, j = 1, 2, \dots, n_1 \quad \text{and} \quad \alpha_n = \frac{\varepsilon}{k} \text{sign} a_{m_k n} \text{ for } n_{k-1} < n \leq n_k, k = 2, 3, \dots$$

Then  $\alpha_n \rightarrow \Theta$  ( $n \rightarrow \infty$ ), hence  $\|\alpha_n\| = \left\| \frac{\varepsilon}{k} \text{sign} a_{m_k n} \right\| = \|\varepsilon\| \cdot \left| \frac{1}{k} \text{sign} a_{m_k n} \right| = \left| \frac{1}{k} \right|$ . Therefore  $\|\alpha_n\| \rightarrow 0$ ,

because  $\frac{1}{k} \rightarrow 0$ . From (4), (6) and (7) we have:

$$\begin{aligned} \|\beta_{m_k}\| &= \left\| \sum_{n=1}^{\infty} a_{m_k n} \alpha_n \right\| \geq \left\| \sum_{n=n_{k-1}+1}^{n_k} a_{m_k n} \alpha_n \right\| - \left\| \sum_{n=1}^{n_{k-1}} a_{m_k n} \alpha_n \right\| - \left\| \sum_{n=n_k+1}^{\infty} a_{m_k n} \alpha_n \right\| \\ &\geq \frac{1}{k} \cdot \sum_{n=n_{k-1}+1}^{n_k} |a_{m_k n}| \cdot \|\varepsilon\| - \sum_{n=1}^{n_{k-1}} |a_{m_k n}| \cdot \|\varepsilon\| - \sum_{n=n_k+1}^{\infty} |a_{m_k n}| \cdot \|\varepsilon\| \\ &> \frac{1}{k} \cdot (k.M(n_{k-1}) + k^2) - M(n_{k-1}) - 1 \\ &= k - 1. \end{aligned}$$

Therefore the sequence  $\beta_m$  is not bounded. □

The method of summation defined by the infinite matrix  $A = (a_{mn})$  is called regular if it transforms the convergent sequence to convergent sequence with the same limit (see e.g. [1], [4]). Now we show that this is also true for the sequences of elements of Banach space.

**Theorem 6.** Let  $A = (a_{mn})$  is an infinite matrix with real numbers. The sequence  $\beta_m = \sum_{n=1}^{\infty} a_{mn} \alpha_n$  converges to  $\alpha$  for  $m \rightarrow \infty$  and  $\alpha_n \rightarrow \alpha$  if and only if the following conditions hold:

- a)  $\exists M > 0, \forall m=1,2,\dots, \sum_{n=1}^{\infty} |a_{mn}| \leq M$ ,
- b)  $\forall n=1,2,\dots, \lim_{m \rightarrow \infty} a_{mn} = 0$ ,
- c)  $\lim_{m \rightarrow \infty} \sum_{n=1}^{\infty} a_{mn} = 1$ .

*Proof.* 1) Suppose  $A = (a_{mn})$  satisfies the three conditions. Let  $\alpha = (\alpha_n)_1^\infty, \alpha_n \rightarrow \alpha$  is a sequence of elements of Banach space. Since  $\alpha_n \rightarrow \alpha$ ,  $(\alpha_n)_1^\infty$  is the bounded sequence. Since the condition a) holds, from Theorem 5 there exists a sequence  $\beta_m = \sum_{n=1}^\infty a_{mn} \alpha_n$ ,  $m=1,2,\dots$ , such is also bounded. We show that  $\beta_m \rightarrow \alpha$ .

Let  $\varepsilon > 0$ . The  $\alpha_n \rightarrow \alpha$  i.e.  $\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \forall n \in \mathbb{N} : n \geq n_0 \Rightarrow \|\alpha_n - \alpha\| < \varepsilon$ , and by c) there is an integer  $m_0$ , such that

$$\left| 1 - \sum_{n=1}^\infty a_{mn} \right| < \varepsilon$$

for all  $m > m_0$ . Using also a) we then have for  $m > m_0$

$$\begin{aligned} \|\beta_m - \alpha\| &= \left\| \sum_{n=1}^\infty a_{mn} \alpha_n - \alpha \right\| = \\ &= \left\| \sum_{n=1}^\infty a_{mn} (\alpha_n - \alpha) + \alpha \left( \sum_{n=1}^\infty a_{mn} - 1 \right) \right\| \leq \left\| \sum_{n=1}^\infty a_{mn} (\alpha_n - \alpha) \right\| + \left\| \alpha \left( \sum_{n=1}^\infty a_{mn} - 1 \right) \right\| \leq \\ &\leq \sum_{n=1}^\infty |a_{mn}| \|\alpha_n - \alpha\| + \|\alpha\| \left| \sum_{n=1}^\infty a_{mn} - 1 \right| < \sum_{n=1}^{n_0} |a_{mn}| \|\alpha_n - \alpha\| + \varepsilon \sum_{n=n_0+1}^\infty |a_{mn}| + \varepsilon \|\alpha\| \leq \\ &\leq \sum_{n=1}^{n_0} |a_{mn}| \|\alpha_n - \alpha\| + \varepsilon M + \varepsilon \|\alpha\|. \end{aligned}$$

But by assumption b) there exists  $m_1 > m_0$ , such that for all  $m > m_1$

$$|a_{mn}| < \frac{\varepsilon}{(L+1)n_0},$$

where  $n=1,2,\dots,n_0$  and  $L = \max \{\|\alpha_1 - \alpha\|, \dots, \|\alpha_{n_0} - \alpha\|\}$ . Then from the previous for all  $m > m_0$  we have  $\|\beta_m - \alpha\| < \varepsilon(1 + M + \|\alpha\|)$ . The sequence  $\beta_m = \sum_{n=1}^\infty a_{mn} \alpha_n$ ,  $m=1,2,\dots$  converges to  $\alpha$ .

2) Let  $\beta_m = \sum_{n=1}^\infty a_{mn} \alpha_n$ ,  $m=1,2,\dots$  converges to  $\alpha$ , where  $\alpha_n$  is a sequence of elements of Banach space,  $\alpha_n \rightarrow \alpha$ . The necessity of condition a) has already been proved in Theorem 5.

For every  $k=1,2,\dots$  define the sequence  $\varepsilon^{(k)} = (\varepsilon_n^{(k)})_1^\infty$  by

$$\varepsilon_n^{(k)} = \varepsilon \text{ if } n = k, \|\varepsilon\| = 1 \quad \text{and} \quad \varepsilon_n^{(k)} = \Theta \text{ if } n \neq k.$$

Then  $\beta_m^{(k)} = \sum_{n=1}^\infty a_{mn} \varepsilon_n^{(k)} = a_{mk} \varepsilon$ . Since  $\lim_{n \rightarrow \infty} \varepsilon_n^{(k)} = \Theta$ , for all  $k=1,2,\dots$  follows that  $\lim_{m \rightarrow \infty} \beta_m^{(k)} = \lim_{m \rightarrow \infty} a_{mk} \varepsilon = 0$ . It is true if and only if  $\lim_{m \rightarrow \infty} a_{mk} = 0$  for all  $k=1,2,\dots$  i.e. b) holds.

Next consider the sequence  $(\alpha_n)_1^\infty = (\varepsilon, \varepsilon, \dots), \|\varepsilon\| = 1$  which converges to  $\varepsilon$ . Therefore the sequence  $\beta_m = \sum_{n=1}^\infty a_{mn} \alpha_n = \sum_{n=1}^\infty a_{mn} \varepsilon$  converges to  $\varepsilon$  only if  $\sum_{n=1}^\infty a_{mn}$  converges to 1 for  $m \rightarrow \infty$ . Then  $\beta_m = \sum_{n=1}^\infty a_{mn} \varepsilon = a_{m1} \varepsilon + \dots + a_{mn} \varepsilon + \dots = \varepsilon \cdot \sum_{n=1}^\infty a_{mn}$ ,  $\beta_m \rightarrow \varepsilon$ , thus necessarily  $\lim_{m \rightarrow \infty} \sum_{n=1}^\infty a_{mn} = 1$ , i.e. c) holds.  $\square$

We showed that the results for the regular matrix method of real sequences can be applied for the sequences of elements of Banach space. Another results could be find in [1].



### 3. EXAMPLES

**Example 7.** Define the sequence of continuous functions on  $\langle 0,1 \rangle$  by

$$f_n(x) = \frac{1-nx}{n}.$$

It is clear that if  $n \rightarrow \infty$ , then  $f_n(x) \rightarrow f(x) = -x$  where  $x \in \langle 0,1 \rangle$  according to the norm  $\|f\| = \max_{x \in \langle 0,1 \rangle} |f(x)|$ . Let  $Z = (a_{mn})$  is a regular matrix defined as follows:

$$a_{mm} = a_{mm+1} = \frac{1}{2}, \quad m = 1, 2, \dots \quad \text{and} \quad a_{mn} = 0 \quad \text{if } m < n, n > m+1.$$

Then the sequence  $g_m(x) = \sum_{n=1}^{\infty} a_{mn} f_n(x)$ ,  $x \in \langle 0,1 \rangle$  can be written as  $g_m(x) = \frac{1}{2} \left( \frac{1}{m} + \frac{1}{m+1} \right) - x$ .

Clearly  $g_m \rightarrow g \equiv 0$  according to the norm  $\|f\|$  in the space  $(C(\langle 0,1 \rangle), \|\cdot\|)$ .

**Example 8.** Let  $\alpha^{(n)} = (\alpha_i^{(n)})$  is a sequence of elements in the space  $l^2$  defined as follows

$$\alpha^{(1)} = (1, 0, 0, \dots), \quad \alpha^{(2)} = (0, \frac{1}{2}, 0, \dots), \quad \dots \quad \alpha^{(n)} = (0, \dots, 0, \frac{1}{n}, 0, \dots).$$

Clearly  $\alpha^{(n)} \rightarrow (0, 0, 0, \dots)$  according to the norm  $\|x\| = \left( \sum_{i=1}^{\infty} x_i^2 \right)^{\frac{1}{2}}$ . Let  $C = (c_{mn})$  is a matrix defined as:

$$c_{mn} = \frac{1}{m} \quad \text{if } n \leq m \quad \text{and} \quad c_{mn} = 0 \quad \text{if } n > m.$$

Create a sequence  $\beta^{(m)} = \sum_{n=1}^{\infty} \alpha^{(n)} c_{mn} = C \alpha^{(n)}$ . Then  $\beta^{(m)} = \left( \frac{1}{m}, \frac{1}{2m}, \dots, \frac{1}{m^2}, 0, \dots \right)$ . Therefore  $\beta^{(m)} \rightarrow (0, 0, 0, \dots)$  because

$$\|\beta^{(m)} - 0\| = \sqrt{\frac{1}{m^2} + \frac{1}{(2m)^2} + \dots + \frac{1}{(m^2)^2}} = \sqrt{\frac{1}{m^2} \left( 1 + \frac{1}{4} + \dots + \frac{1}{m^2} \right)} \leq \frac{1}{m} \cdot \frac{\pi}{\sqrt{6}} \rightarrow 0 \quad \text{for } m = 1, 2, \dots$$

### CONCLUSION

In this article we have generalized the notion of convergence through regular matrix. In monographies [2], [3] and [1] are mentioned conditions for infinite matrix  $A = (a_{mn})$ , which transforms the sequence of real numbers to another sequence of real numbers and preserves the boundedness and convergence to the same limit. In [4] is stated a theorem without proof which says about the transformation of sequences of elements of Banach space. We give the proofs of three theorems which say about the form of regular matrix, transforms a (bounded) convergent sequence of elements of Banach space to (bounded) convergent sequence. There are listed examples of transformation of sequences of different spaces.

### Acknowledgement

The financial support from the Cultural and Educational Grant Agency under the grant KEGA No. 047STU-4/2016 is gratefully acknowledged.

### LITERATURE

- [1] KOSTYRKO, P. Convergence fields of regular matrix transformation, *Tatra Mountains Math. Publ.*, **28** (2004), p. 153-157. ISSN 1210-3195.
- [2] PETERSEN, G. M. *Regular matrix transformations*. London: McGraw-Hill, 1966.

- [3] ŠALÁT, T. *Infinite series* [in Slovak]. Praha: Academia, 1974.
- [4] VISNYAI, T. Convergence fields of regular matrix transformations of sequences of elements of Banach spaces, *Miskolc Mathematical Notes*, **7** (2006), p. 101-108. ISSN 1787-2413.