Mathematics,
Information Technologies
and
Applied Sciences
2016

post-conference proceedings of extended versions of selected papers

Editors:

Jaromír Baštinec and Miroslav Hrubý

Brno, Czech Republic, 2016

ISBN: 978-80-7231-400-3

Aims and target group of the conference:

The conference **MITAV 2016** 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 16 and 17, 2016.

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 Lengyelfalusy, 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
Karel Lepka	Masaryk University in Brno, Faculty of Education, Department of Mathematics
Pavlína Račková	University of Defence in Brno, Faculty of Military Technology, Department of Mathematics and Physics
Jan Vondra	Masaryk University in Brno, Faculty of Science, Department of Mathematics and Statistics

Programme of the conference:

Thursday, June 16, 2016

12:00-13:45	Registration of the participants
13:45-14:00	Opening of the conference
14:00-14:50	Keynote lecture No. 1 (Jiří Krtička, Czech republic)
14:50-15:10	Break
15:10-16:00	Keynote lecture No. 2 (Jaromír Šimša, Czech republic)
16:00-16:30	Break
16:35-18:00	Presentations of papers
18:00-19:15	Conference dinner
19:30-22:00	Social event (University of Defence Club – performance by folklore ensemble Lučina with cimbalom)

Friday, June 17, 2016

09:00-09:45	Keynote lecture No. 3 (Václav Talhofer, Czech republic)
09:45-10:00	Break
10:00-11:30	Presentations of papers
11:30-11:45	Break
11:45-13:15	Presentations of papers
13:15	Closing

Each MITAV 2016 participant received printed collection of abstracts **MITAV 2016** with ISBN 978-80-7231-464-5. CD supplement of this printed volume contains all the accepted contributions of the conference.

Now, in autumn 2016, this **post-conference CD** was published, containing extended versions of selected MITAV 2016 contributions. The proceedings are published in English and contain extended versions of 8 selected conference papers. Published articles have been chosen from 34 conference papers and every article was reviewed by two reviewers.

Webpage of the MITAV conference:

http://mitav.unob.cz

Content:

List of reviewers:

doc. RNDr. Jaromír Baštinec, CSc.
doc. RNDr. Jaroslav Beránek, Ph.D.
doc. RNDr. Edita Kolářová, Ph.D.
RNDr. Karel Lepka, Dr
Prof. RNDr. Miroslava Růžičková, Ph.D.
Prof. Dr. Hab. Ewa Schmeidel, Ph.D.
RNDr. Zdeněk Svoboda, CSc.

Stability of the Zero Solution of Stochastic Differential Systems with Four-Dimensional Brownian Motion

Jaromír Baštinec, Marie Klimešová

Department of Mathematics, Faculty of Electrical Engineering and
Communication Brno University of Technology,
Technická 2848/8, Žabovřesky, 61600, Brno, Czech republic.
Email: bastinec@feec.vutbr.cz,
xklime01@stud.feec.vutbr.cz

Abstract: The natural world is influenced by stochasticity therefore stochastic models are used to test various situations because only the stochastic model can approximate the real model. For example, the stochastic model is used in population, epidemic and genetic simulations in medicine and biology, for simulations in physical and technical sciences, for analysis in economy, financial mathematics, etc. The crucial characteristic of the stochastic model is its stability. Stability of stochastic differential equations (SDEs) has become a very popular theme of recent research in mathematics and its applications. This article studies the fundamental theory of the stochastic stability. There is investigated the stability of the solution of stochastic differential equations and systems of SDEs. The article begins with a summary of the stochastic theory. Then, there are inferred conditions for the asymptotic mean square stability of the zero solution of stochastic system with Brownian motion. There is used a Lyapunov function for proofs of main results. The method of Lyapunov functions for the analysis of qualitative behavior of SDEs provides some very useful information in the study of stability properties for concrete stochastic dynamical systems, conditions of existence the stationary solutions of SDEs and related problems. There are proved conditions for the stability (asymptotic, stochastic asymptotic). The results are illustrated by trivial examples for special types of matrices.

Keywords: Brownian motion, stochastic differential equation, Lyapunov function, stochastic Lyapunov function, stability, stochastic stability.

Introduction

Stochastic modeling has come to play an important role in many branches of science and industry where more and more people have encountered stochastic differential equations. Stochastic model can be used to solve problem which evinces by accident, noise, etc. Definition of probability spaces, stochastic process, stochastic differential equation and an existence and uniqueness of solution of these equations, were mentioned in [11], [12], [14]. It was taken from B. Øksendal [19], E. Kolářová [15], B. Maslowski [17], S. Ditlevsen [4], M. Navara [18] and J. Staněk [20] and others. In this paper we focus on the description of the stochastic stability. Stability is studied both for difference equations and systems [6], and for differential equations and systems [1], [3], [5], [7] or [8]. In this paper we use definitions of the stability theory of the stochastic system defined by R. Z. Khasminskii [10]. The general principles of various types of stochastic systems are described for example X.Mao [16].
In the paper we study the linear matrix systems. We derived sufficient conditions of stochastic stability for general system of the zero solution of the stochastic differential equation using Lyapunov function. The same method can also be used for constant matrix. Stochastic models may find the use in the optimization.

Definition 1 [*19*] *If Ω is a given set, then a σ-algebra $\mathcal{F}$ on Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:*

(i) $\emptyset \in \mathcal{F}$

(ii) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$*, where* $F^C = \Omega \setminus \mathcal{F}$ *is the complement of* $\mathcal{F}$ *in* Ω

(iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A := \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}.$

The pair $(\Omega, \mathcal{F})$ *is called a measurable space.*

Definition 2 [*19*] *A probability measure* P *on a measurable space* $(\Omega, \mathcal{F})$ *is a function* $P : \mathcal{F} \longrightarrow [0, 1]$ *such that*

(a) $P(\emptyset) = 0, P(\Omega) = 1.$

(b) if $A_1, A_2, \dots \in \mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$) then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space. It is called a complete probability space if $\mathcal{F}$ contains all subsets G of Ω with P-outer measure zero, i.e. with

$$P^*(G) := \inf\{P(F); F \in \mathcal{F}, G \subset F\} = 0.$$

Definition 3 *The stochastic process B_t is called Brownian motion or Wiener process if the process has some basic properties:*

(i) $B_0 = 0$

(ii) $B_t - B_s$ *has the distribution* $N(0, t-s)$ *for* $t \geq s \geq 0$

(iii) B_t *has independent increments, i.e.*

$$B_{t_1},\ B_{t_2} - B_{t_1},\ \dots,\ B_{t_k} - B_{t_{k-1}}$$

are independent for all $0 \leq t_1 < t_2 \cdots < t_k$.

Note. The unconditional probability density function at a fixed time t

$$f_{B_t}(X_t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{X_t^2}{2t}\right).$$

The expectation is zero; $E\left[B_t\right] = 0$ for $t > 0$. The variance is t; $E\left[B_t^2\right] = t$.

Theorem 1 *Let B_t be Brownian motion. Then*

$$E\left[B_t B_s\right] = min\left\{t, s\right\} \ \ for\ t \geq 0,\ s \geq 0.$$

Proof: [15], pp. 14.

Definition 4 *Let $B_i(t), t = 1, 2, \dots, m$, be a Brownian motion. Then $B(t) = (B_1(t), ..., B_m(t))$ denote m-dimensional Brownian motion.*

Definition 5 *[19] Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $B_t = (B_1(t), ..., B_m(t))$ be m-dimensional Brownian motion and $b : [0,T] \times R^n \to R^n$, $\sigma : [0,T] \times R^n \to R^{n \times m}$ be measurable functions. Then the process $X_t = (X_1(t), ..., X_m(t)), X_t \equiv X(t)$, $t \in [0,T]$ is the solution of the stochastic differential equation*

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t, \tag{1}$$

$b(t, X_t) \in R \times R^n,\ \sigma(t, X_t)B_t \in R \times R^n$. After the integration of equation (1) we give the different form of the solution of the SDE

$$X_t = X_0 + \int\limits_0^t b(s, X_s)\mathrm{d}s + \int\limits_0^t \sigma(s, X_s)\mathrm{d}B_s.$$

Assume that for every initial value $X(t_0) = X_0 \in R^n$ equation (1) has a unique global solution that is denoted by $X(t; t_0, X_0)$. We know that the solution has continuous sample paths and its every moment is finite.

In the following text we will study the stability of various types of stability of solutions of the system (1) and we will suppose that $b(t, o)\mathrm{d}t + \sigma(t, o)\mathrm{d}B_t = o$, where o is a zero vector . So equation (1) has **the trivial solution** o corresponding to the initial value $X(t_0) = 0$.

1 Stability of Stochastic Differential Equations

In 1892 A.M. Lyapunov developed a methods for determining stability without solving the equation. We are used the second Lyapunov method:
Let K denote the family of all continuous nondecreasing functions $\mu : R_+ \to R_+$ such that $\mu(0) = 0$ and $\mu(r) > 0$ if $r > 0$. Let $S_h = \{X_t \in R^n : |X_t| < h\}$ for $h > 0$. A continuous function $V(X_t, t)$ defined on $S_h \times [t_0, \infty)$ is said to be **positive-definite** (in the sense of Lyapunov) if $V(0, t) \equiv 0$ and, for some $\mu \in K$,

$$V(X_t, t) \geq \mu(|x|) \quad \text{for all } (X_t, t) \in S_h \times [t_0, \infty).$$

A function $V(X_t, t)$ is said to be **negative-definite** if $(-V(X_t, t))$ is positive-definite. A continuous non-negative function $V(X_t, t)$ is said to be **decrescent** (i.e. to have an arbitrarily small upper bound) if for some $\mu \in K$,

$$V(X_t, t) \leq \mu(|X_t|) \quad \text{for all } (X_t, t) \in S_h \times [t_0, \infty).$$

A function $V(X_t, t)$ defined on $R^n \times [t_0, \infty)$ is said to be **radially unbounded** if

$$\lim_{|x|\to\infty} \left(\inf_{t \geq t_0} V(X_t, t) \right) = \infty.$$

Let $C^{1,1}(S_h \times [t_0, \infty), R_+)$ denote the family of all continuous functions $V(X_t, t)$ from $S_h \times [t_0, \infty)$ to R_+ with continuous first partial derivatives with respect to every component of X_t and to t. Then $V(t) = V(t, X_t)$ represents a function of t with the derivative

$$\dot{V}(t) = V_t(t, X_t) + V_{X_t}(t, X_t)b(t, X_t) = \frac{\partial V(t, X_t)}{\partial t} + \sum_{i=1}^{n} \frac{\partial V(t, X_t)b_i(t, X_t)}{\partial X_i}.$$

If $\dot{V}(t) \leq 0$, then $V(t)$ will not increase so the distance of X_t from the equilibrium point measured by $V(t, X_t)$ does not increase. If $\dot{V}(t) < 0$, then $V(t)$ will decrease to zero so the distance will decrease to zero, that is $X_t \to 0$. [16]

1.1 Stability in probability

Theorem 2 (***Lyapunov theorem***) *[16] If there exists a positive-definite function $V(X_t, t) \in C^{1,1}(S_h \times [t_0, \infty), R_+)$ such that*

$$\dot{V}(X_t, t) := V_t(t, X_t) + V_{X_t}(t, X_t)b(t, X_t) \leq 0$$

for all $(X_t, t) \in S_h \times [t_0, \infty)$, then the trivial solution is stable. If there exists a positive-definite decrescent function $V(X_t, t) \in C^{1,1}(S_h \times [t_0, \infty), R_+)$ such that $\dot{V}(X_t, t)$ is negative-definite, then trivial solution of the system is asymptotically stable.

Suppose one would like to let the initial value be a random variable. It should also be pointed out that when $\sigma(t, X_t) = 0$, these definitions reduce to the corresponding deterministic ones. We now extend the Lyapunov Theorem 2 to the stochastic case. Let $0 < h \leq \infty$. Denote by $C^{2,1}(S_h \times R_+, R_+)$ the family of all nonnegative functions $V(X_t, t)$ defined on $S_h \times R_+$ such that they are continuously twice differentiable in X_t and once in t. Define the differential operator LV associated with equation (1) by

$$LV = \frac{\partial V}{\partial t} + \sum_{i=1}^{n} \frac{\partial V(t, X_t)b_i(X_t, t)}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 V \left[\sigma(X_t, t)\sigma^T(X_t, t)\right]_{ij}}{\partial X_i \partial X_j}.$$

The inequality $\dot{V}(X_t, t) \leq 0$ will be replaced by $LV(X_t, t) \leq 0$ in order to get the stochastic stability assertions.

Definition 6 [*16*] *The trivial solution of equation* (1) *is stochastically* ***stable*** *if there exists a positive-definite function* $V(X_t, t) \in C^{2,1}(S_h \times [t_0, \infty), R_+)$ *such that*

$$LV(X_t, t) \leq 0$$

for all $(X_t, t) \in S_h \times [t_0, \infty)$.

If there exists a positive-definite decrescent function $V(X_t, t) \in C^{2,1}(S_h \times [t_0, \infty), R_+)$ *such that* $LV(X_t, t)$ *is negative-definite, then the trivial solution of equation* (1) *is stochastically* ***asymptotically stable****.*

If there exists a positive-definite decrescent radially unbounded function $V(X_t, t) \in C^{2,1}(R^n \times [t_0, \infty), R_+)$ *such that* $LV(X_t, t)$ *is negative-definite, then the trivial solution of equation* (1) *is stochastically* ***asymptotically stable in the large.***

Proof: [16], pp. 111.

1.2 Almost sure exponential stability

Definition 7 [*16*] *The trivial solution of equation* (1) *is said to be almost surely exponentially stable if*

$$\lim_{t \to \infty} \sup \frac{1}{t} \log |X(t, t_0, X_0)| < 0 \quad a.s. \tag{2}$$

for all $X_0 \in R^n$*. The left-hand side of (2) is called the sample Lyapunov exponents of the solution of the stochastic system.*

The trivial solution is almost surely exponentially stable if and only if the sample Lyapunov exponents are negative. The almost sure exponential stability means that almost all sample paths of the solution will tend to the equilibrium position $X_t = 0$ exponentially fast.

Lemma 1 [*16*] *For all* $X_{t_0} \neq 0$ *in* R^n

$$P\left\{X(t, t_0, X_0) \neq 0 \quad on \;\; t \geq t_0\right\} = 1,$$

where $P\left\{X(t, t_0, X_0)\right\}$ *is the probability that the occurrence* X_t *is based on point* $X_0(t_0)$*. That is, almost all the sample path of any solution starting from a non-zero state will never reach the origin.*

Definition 8 *[16] Assume that there exists a function* $V \in C^{2,1}(R^n \times [t_0, \infty), R_+)$ *and constants* $p > 0, c_1 > 0, c_2 \in R, c_3 \geq 0,$ *such that for all* $X_0 \neq 0$ *and* $t \geq t_0,$

(i) $c_1 |x|^p \leq V(X_t, t),$

(ii) $LV(X_t, t) \leq c_2 V(X_t, t)$

(iii) $|V_{X_t}(X_t, t)\sigma(X_t, t)|^2 \geq c_3 V^2(X_t, t)$

Then

$$\lim_{t\to\infty} \sup \frac{1}{t} \log |X(t, t_0, X_0)| \leq -\frac{c_3 - 2c_2}{2p} \quad a.s.$$

for all $X_0 \in R^n$*. In particular, if* $c_3 > 2c_2$*, the trivial solution of equation* (1) *is* ***almost surely exponentially stable****.*

Proof: [16], pp. 121.

Definition 9 *[16] Assume that there exists a function* $V \in C^{2,1}(R^n \times [t_0, \infty), R_+)$ *and constants* $p > 0, c_1 > 0, c_2 \in R, c_3 \geq 0,$ *such that for all* $X_0 \neq 0$ *and* $t \geq t_0,$

(i) $c_1 |x|^p \geq V(X_t, t) > 0,$

(ii) $LV(X_t, t) \geq c_2 V(X_t, t)$

(iii) $|V_{X_t}(X_t, t)\sigma(X_t, t)|^2 \leq c_3 V^2(X_t, t)$

Then

$$\lim_{t\to\infty} \inf \frac{1}{t} \log |X(t, t_0, X_0)| \geq -\frac{2c_2 - c_3}{2p} \quad a.s.$$

for all $X_0 \in R^n$*. In particular, if* $2c_2 > c_3$*, then almost all the sample paths of* $|X(t, t_0, X_0)|$ *will tend to infinity, and we say in this case that the trivial solution of equation* (1) *is* ***almost surely exponentially unstable****.*

Proof: [16], pp. 121.

1.3 Moment exponential stability

Definition 10 $[16]$ *The trivial solution of equation* (1) *is said to be* $p-$*th moment exponentially stable if there is a pair of positive constants* λ *and* C *such that*

$$E\left|X(t,t_0,X_0)\right|^p \leq C\left|X_0\right|^p exp\left(-\lambda(t-t_0)\right) \quad on \quad t \geq t_0$$

for all $X_0 \in R^n$. *When* $p=2$, *it is usually said to be exponentially stable in mean square. It also follows that*

$$\lim_{t\to\infty} \sup \frac{1}{t}\log\left(E\left|X(t,t_0,X_0)\right|^p\right) < 0. \tag{3}$$

The $p-$th moment exponential stability means that the $p-$th moment of the solution will tend to 0 exponentially fast. The left-hand side of (3) is called the $p-$th moment Lyapunov exponent of the solution.

Theorem 3 $[16]$ *Assume that there is a positive constant* K *such that*

$$X_t^T b(X_t,t) \vee \left|\sigma(X_t,t)\right|^2 \leq K\left|X_t\right|^2 \quad for \;\; all \quad (X_t,t) \in R^n \times [t_0,\infty).$$

Then the p*th moment exponential stability of the trivial solution of equation* (1) *implies the almost sure exponential stability.*

Proof: [16], pp. 128.

Theorem 4 $[16]$ *Assume that there is a function* $V(X_t,t) \in C^{2,1}(R^n \times [t_0,\infty), R_+)$ *and positive constants* $c_1, c_2, c_3,$ *such that*

$$c_1\left|X_t\right|^p \leq V(X_t,t) \leq c_2\left|X_t\right|^p \quad and \quad LV(X_t,t) \leq -c_3 V(X_t,t)$$

for all $(X_t,t) \in R^n \times [t_0,\infty)$. *Then*

$$E\left|X(t,t_0,X_0)\right|^p \leq \frac{c_2}{c_1}\left|X_0\right|^p exp\left(-c_3(t-t_0)\right) \quad on \;\; t \geq t_0 \tag{4}$$

for all $X_0 \in R^n$. *In other words, the trivial solution of equation* (1) *is* $p-$*th moment exponentially stable and the* $p-$*th moment Lyapunov exponent should not be greater than* $-c_3$.

Proof: [16], pp. 130.

Theorem 5 *[16] Let $q > 0$. Assume that there is a function $V(X_t, t) \in C^{2,1}(R^n \times [t_0, \infty), R_+)$ and positive constants c_1, c_2, c_3, such that*

$$c_1 |X_t|^q \leq V(X_t, t) \leq c_2 |X_t|^q \quad and \quad LV(X_t, t) \geq c_3 V(X_t, t)$$

for all $(X_t, t) \in R^n \times [t_0, \infty)$. Then

$$E |X(t, t_0, X_0)|^q \geq \frac{c_1}{c_2} |X_0|^q \, exp\left(c_3(t - t_0)\right) \quad on \;\; t \geq t_0 \tag{5}$$

for all $X_0 \in R^n$, and we say in this case that the trivial solution of equation (1) is $q-$th moment exponentially unstable.

Proof: [16], pp. 131.

1.4 Stochastic Stability and Nonstability

It is not surprising that noise can destabilize a stable system. And the noise can stabilized the unstable system. In this section we shall establish a general theory of stochastic stabilization and destabilization for a given nonlinear system. Suppose that the given system is described by a nonlinear ordinary differential equation

$$\dot{y}(t) = f(y(t)) \;\; on \;\; t \geq t_0, y(t_0) = X_0 \in R^d.$$

Here $f : R^d \times R_+ \to R^d$ is a locally Lipschitz continuous function and particularly, for some $K > 0$,

$$|f(X_t, t)| \leq K |X_t| \quad for \;\; all \;\; (X_t, t) \in R^d \times R_+. \tag{6}$$

We now use the m-dimensional Brownian motion $B(t) = (B_1(t), \ldots, B_m(t))^T$ as the source of noise to perturb the given system. For simplicity, suppose the stochastic perturbation is of a linear form, that is the stochastically perturbed system is described by the semilinear Itô equation

$$\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sum_{i=1}^{m} G_i X_t \mathrm{d}B_i(t) \;\; on \;\; t \geq t_0, X(t_0) = X_0 \in R^d, \tag{7}$$

where all $G_i, 1 \leq i \leq m$ are $d \times d$ matrices. Clearly, equation (7) has a unique solution denoted by $X(t; t_0, X_0)$ again and, moreover, it admits a trivial solution $X_t \equiv 0$.

Theorem 6 $[16]$ *Let (6) hold. Assume that there are two constants $\lambda > 0$ and $\rho \geq 0$ such that*

$$\sum_{i=1}^{m} \left| G_i X_t^2 \right| \leq \lambda \left| X_t \right|^2 \quad and \quad \sum_{i=1}^{m} \left| X_t^T G_i X_t^2 \right| \geq \rho \left| X_t \right|^4 \tag{8}$$

for all $X_t \in R^d$. Then

$$\lim_{t \to \infty} sup \frac{1}{t} \log \left| X(t; t_0, X_0) \right| \leq -\left(\rho - K - \frac{\lambda}{2} \right) a.s. \tag{9}$$

for all $X_0 \in R^d$. In particular, if $\rho > K + \frac{1}{2}\lambda$, then the trivial solution of equation (7) is almost surely exponentially stable.

Proof: [16], pp. 137.

2 Main results

2.1 Four-Dimensional Brownian Motion

We have a matrix linear stochastic differential equation

$$\mathrm{d}X_t = AX_t\mathrm{d}t + G\mathrm{d}B_t, \tag{10}$$

where $X_t = \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix}$, $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$, $G = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix}$,
$B_t = \begin{pmatrix} B_1(t) \\ B_2(t) \\ B_3(t) \\ B_4(t) \end{pmatrix}$, a_{ij}, g_{ij} for $i, j = 1, 2, 3, 4$ are constants.

Definition 11 *Lyapunov quadratic function V is given*

$$V(X_t) = X_t^T Q \, X_t,$$

where Q is a symmetric positive-definite matrix.

The Euclidean matrix norm of the matrix $A = (a_{ij}), i = 1, \ldots, n; j = 1, \ldots, n$ on the space R^n can be define as

$$\|A\|_E := \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2}.$$

2.2 Results for the general matrix Q

Definition 12 *Lyapunov quadratic function V is given*

$$V(X_t) = X_t^T Q \; X_t,$$

where $Q = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix}$ *is a symmetric positive-definite matrix. Positive-definite matrix is verified by the Sylvester's criterion. There have to apply these conditions together*

$$\Delta_1 = q_1 > 0,$$

$$\Delta_2 = q_1^2 - q_2^2 > 0,$$

$$\Delta_3 = q_1^3 + 2q_2^2 q_3 - q_1 q_3^2 - 2q_1 q_2^2 > 0,$$

$$\begin{gathered} \Delta_4 = q_1 q_2^3 + q_1 q_2 q_3^2 + q_1^3 q_4 - q_1 q_2^2 q_4 - 2q_1^2 q_2 q_3 - q_1^2 q_2^2 - 2q_2^2 q_3^2 - q_2^3 q_4 + q_2^4 + q_3^4 \\ +2q_1 q_2^2 q_3 + 4q_1 q_2 q_3 q_4 + q_2^2 q_4^2 - 2q_2 q_3^2 q_4 - q_1^2 q_3^2 - q_2^3 q_4 - q_1^2 q_4^2 > 0. \end{gathered}$$

Theorem 7 *Zero solution of equation* (10) *is stochastically stable if holds*

$$LV < 0,$$

where

$$\begin{aligned} LV \;\; = \;\; & 2\left(a_{11}q_1 + a_{21}q_2 + a_{31}q_3 + a_{41}q_4\right) X_1^2(t) + 2\left(a_{12}q_2 + a_{22}q_1 + a_{32}q_2\right. \\ + \;\; & \left. a_{42}q_3\right) X_2^2(t) + 2\left(a_{13}q_3 + a_{23}q_2 + a_{33}q_1 + a_{43}q_2\right) X_3^2(t) + 2\left(a_{14}q_4\right. \\ + \;\; & \left. a_{24}q_3 + a_{34}q_2 + a_{44}q_1\right) X_4^2(t) + 2\left(a_{12}q_1 + a_{11}q_2 + a_{22}q_2 + a_{21}q_1 + a_{32}q_3\right. \\ + \;\; & \left. a_{31}q_2 + a_{42}q_4 + a_{41}q_3\right) X_1(t)X_2(t) + 2\left(a_{13}q_1 + a_{11}q_3 + a_{23}q_2 + a_{23}q_1\right. \\ + \;\; & \left. a_{21}q_2 + a_{33}q_3 + a_{31}q_1 + a_{43}q_4 + a_{41}q_2\right) X_1(t)X_3(t) + 2\left(a_{14}q_1 + a_{11}q_4\right. \\ + \;\; & \left. a_{24}q_2 + a_{21}q_3 + a_{34}q_3 + a_{31}q_2 + a_{44}q_4 + a_{41}q_1\right) X_1(t)X_4(t) + 2\left(a_{13}q_2\right. \end{aligned}$$

$$
\begin{aligned}
+ & \quad a_{12}q_3 + a_{22}q_2 + a_{33}q_2 + a_{32}q_1 + a_{43}q_3 + a_{42}q_2\big) X_2(t)X_3(t) + 2\left(a_{14}q_2\right. \\
+ & \quad a_{24}q_1 + a_{22}q_3 + a_{34}q_2 + a_{32}q_2 + a_{44}q_3 + a_{42}q_1\big) X_2(t)X_4(t) + 2\left(a_{14}q_3\right. \\
+ & \quad a_{24}q_2 + a_{23}q_3 + a_{34}q_1 + a_{33}q_2 + a_{44}q_2 + a_{43}q_1\big) X_3(t)X_4(t) + q_1\left(g_{11}^2\right. \\
+ & \quad g_{12}^2 + g_{13}^2 + g_{14}^2 + g_{21}^2 + g_{22}^2 + g_{23}^2 + g_{24}^2 + g_{31}^2 + g_{32}^2 + g_{33}^2 + g_{34}^2 + g_{41}^2 \\
+ & \quad g_{42}^2 + g_{43}^2 + g_{44}^2\big) + 2q_2\left(g_{11}g_{21} + g_{12}g_{22} + g_{13}g_{23} + g_{14}g_{24} + g_{21}g_{31} + g_{22}g_{32}\right. \\
+ & \quad g_{23}g_{33} + g_{24}g_{34} + g_{31}g_{41} + g_{32}g_{42} + g_{33}g_{43} + g_{34}g_{44}\big) + 2q_3\left(g_{11}g_{31}\right. \\
+ & \quad g_{12}g_{32} + g_{13}g_{33} + g_{14}g_{34} + g_{21}g_{41} + g_{22}g_{42} + g_{23}g_{43} + g_{24}g_{44}\big) + 2q_4 \\
\times & \quad \left(g_{11}g_{41} + g_{12}g_{42} + g_{13}g_{43} + g_{14}g_{44}\right).
\end{aligned}
$$

Proof:
We compute derivation of Lyapunov function of equation (10). We get the equation

$$
\begin{aligned}
\mathrm{d}V(X_t) &= X_t^TQAX_t\mathrm{d}t + X_t^TQG\mathrm{d}B_t + X_t^TA^T\mathrm{d}tQX_t + \mathrm{d}B_t^TG^TQX_t \\
&+ \mathrm{d}B_t^TG^TQG\mathrm{d}B_t.
\end{aligned}
$$

In matrix form

$$
\begin{aligned}
& \mathrm{d}V\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix} \\
= & \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix}^T \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix} \mathrm{d}t \\
+ & \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix}^T \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix} \begin{pmatrix} \mathrm{d}B_1(t) \\ \mathrm{d}B_2(t) \\ \mathrm{d}B_3(t) \\ \mathrm{d}B_4(t) \end{pmatrix} \\
+ & \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix}^T \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}^T \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix} \mathrm{d}t
\end{aligned}
$$

$$
\begin{aligned}
+ \quad & \begin{pmatrix} \mathrm{d}B_1(t) \\ \mathrm{d}B_2(t) \\ \mathrm{d}B_3(t) \\ \mathrm{d}B_4(t) \end{pmatrix}^T \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix}^T \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix} \\
+ \quad & \begin{pmatrix} \mathrm{d}B_1(t) \\ \mathrm{d}B_2(t) \\ \mathrm{d}B_3(t) \end{pmatrix}^T \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix}^T \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_1 & q_2 & q_3 \\ q_3 & q_2 & q_1 & q_2 \\ q_4 & q_3 & q_2 & q_1 \end{pmatrix} \\
\times \quad & \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix} \begin{pmatrix} \mathrm{d}B_1(t) \\ \mathrm{d}B_2(t) \\ \mathrm{d}B_3(t) \\ \mathrm{d}B_4(t) \end{pmatrix}.
\end{aligned}
$$

We get

$$
\begin{aligned}
& \mathrm{d}V(X_t) \\
= \quad & 2\left(a_{11}q_1 + a_{21}q_2 + a_{31}q_3 + a_{41}q_4\right) X_1^2(t)\mathrm{d}t + 2\left(a_{12}q_2 + a_{22}q_1 + a_{32}q_2 + a_{42}q_3\right) \\
\times \quad & X_2^2(t)\mathrm{d}t + 2\left(a_{13}q_3 + a_{23}q_2 + a_{33}q_1 + a_{43}q_2\right) X_3^2(t)\mathrm{d}t + 2\left(a_{14}q_4 + a_{24}q_3\right. \\
+ \quad & \left. a_{34}q_2 + a_{44}q_1\right) X_4^2(t)\mathrm{d}t + 2\left(a_{12}q_1 + a_{11}q_2 + a_{22}q_2 + a_{21}q_1 + a_{32}q_3 + a_{31}q_2\right. \\
+ \quad & \left. a_{42}q_4 + a_{41}q_3\right) X_1(t)X_2(t)\mathrm{d}t + 2\left(a_{13}q_1 + a_{11}q_3 + a_{23}q_2 + a_{23}q_1 + a_{21}q_2 + a_{33}q_3\right. \\
+ \quad & \left. a_{31}q_1 + a_{43}q_4 + a_{41}q_2\right) X_1(t)X_3(t)\mathrm{d}t + 2\left(a_{14}q_1 + a_{11}q_4 + a_{24}q_2 + a_{21}q_3 + a_{34}q_3\right. \\
+ \quad & \left. a_{31}q_2 + a_{44}q_4 + a_{41}q_1\right) X_1(t)X_4(t)\mathrm{d}t + 2\left(a_{13}q_2 + a_{12}q_3 + a_{22}q_2 + a_{33}q_2 + a_{32}q_1\right. \\
+ \quad & \left. a_{43}q_3 + a_{42}q_2\right) X_2(t)X_3(t)\mathrm{d}t + 2\left(a_{14}q_2 + a_{24}q_1 + a_{22}q_3 + a_{34}q_2 + a_{32}q_2 + a_{44}q_3\right. \\
+ \quad & \left. a_{42}q_1\right) X_2(t)X_4(t)\mathrm{d}t + 2\left(a_{14}q_3 + a_{24}q_2 + a_{23}q_3 + a_{34}q_1 + a_{33}q_2 + a_{44}q_2 + a_{43}q_1\right) \\
\times \quad & X_3(t)X_4(t)\mathrm{d}t + q_1\left(g_{11}^2 + g_{12}^2 + g_{13}^2 + g_{14}^2 + g_{21}^2 + g_{22}^2 + g_{23}^2 + g_{24}^2 + g_{31}^2 + g_{32}^2\right. \\
+ \quad & \left. g_{33}^2 + g_{34}^2 + g_{41}^2 + g_{42}^2 + g_{43}^2 + g_{44}^2\right)\mathrm{d}t + 2q_2\left(g_{11}g_{21} + g_{12}g_{22} + g_{13}g_{23} + g_{14}g_{24}\right. \\
+ \quad & \left. g_{21}g_{31} + g_{22}g_{32} + g_{23}g_{33} + g_{24}g_{34} + g_{31}g_{41} + g_{32}g_{42} + g_{33}g_{43} + g_{34}g_{44}\right)\mathrm{d}t \\
+ \quad & 2q_3\left(g_{11}g_{31} + g_{12}g_{32} + g_{13}g_{33} + g_{14}g_{34} + g_{21}g_{41} + g_{22}g_{42} + g_{23}g_{43} + g_{24}g_{44}\right)\mathrm{d}t \\
+ \quad & 2q_4\left(g_{11}g_{41} + g_{12}g_{42} + g_{13}g_{43} + g_{14}g_{44}\right)\mathrm{d}t + 2\left[\left(q_1X_1(t) + q_2X_2(t) + q_3X_3(t)\right.\right. \\
+ \quad & \left. q_4X_4(t)\right)\left(g_{11}\mathrm{d}B_1(t) + g_{12}\mathrm{d}B_2(t) + g_{13}\mathrm{d}B_3(t) + g_{14}\mathrm{d}B_4(t)\right) + \left(q_2X_1(t)\right. \\
+ \quad & \left. q_1X_2(t) + q_2X_3(t) + q_3X_4(t)\right)\left(g_{21}\mathrm{d}B_1(t) + g_{22}\mathrm{d}B_2(t) + g_{23}\mathrm{d}B_3(t) + g_{24}\mathrm{d}B_4(t)\right) \\
+ \quad & \left(q_3X_1(t) + q_2X_2(t) + q_1X_3(t) + q_2X_4(t)\right)\left(g_{31}\mathrm{d}B_1(t) + g_{32}\mathrm{d}B_2(t) + g_{33}\mathrm{d}B_3(t)\right.
\end{aligned}
$$

$$\begin{aligned} &+ \quad g_{34}\mathrm{d}B_4(t)) + (q_4X_1(t) + q_3X_2(t) + q_2X_3(t) + q_1X_4(t))\,(g_{41}\mathrm{d}B_1(t) + g_{42}\mathrm{d}B_2(t) \\ &+ \quad g_{43}\mathrm{d}B_3(t) + g_{44}\mathrm{d}B_4(t))] \end{aligned}$$

We apply expectation $E\left\{\mathrm{d}V(X_t)\right\}$

$$\begin{aligned} E\left\{\mathrm{d}V(X_t)\right\} &= 2\left(a_{11}q_1 + a_{21}q_2 + a_{31}q_3 + a_{41}q_4\right)X_1^2(t) + 2\left(a_{12}q_2 + a_{22}q_1 + a_{32}q_2\right. \\ &+ \left. a_{42}q_3\right)X_2^2(t) + 2\left(a_{13}q_3 + a_{23}q_2 + a_{33}q_1 + a_{43}q_2\right)X_3^2(t) + 2\left(a_{14}q_4\right. \\ &+ \left. a_{24}q_3 + a_{34}q_2 + a_{44}q_1\right)X_4^2(t) + 2\left(a_{12}q_1 + a_{11}q_2 + a_{22}q_2 + a_{21}q_1\right. \\ &+ \left. a_{32}q_3 + a_{31}q_2 + a_{42}q_4 + a_{41}q_3\right)X_1(t)X_2(t) + 2\left(a_{13}q_1 + a_{11}q_3\right. \\ &+ \left. a_{23}q_2 + a_{23}q_1 + a_{21}q_2 + a_{33}q_3 + a_{31}q_1 + a_{43}q_4 + a_{41}q_2\right)X_1(t)X_3(t) \\ &+ 2\left(a_{14}q_1 + a_{11}q_4 + a_{24}q_2 + a_{21}q_3 + a_{34}q_3 + a_{31}q_2 + a_{44}q_4 + a_{41}q_1\right) \\ &\times X_1(t)X_4(t) + 2\left(a_{13}q_2 + a_{12}q_3 + a_{22}q_2 + a_{33}q_2 + a_{32}q_1 + a_{43}q_3\right. \\ &+ \left. a_{42}q_2\right)X_2(t)X_3(t) + 2\left(a_{14}q_2 + a_{24}q_1 + a_{22}q_3 + a_{34}q_2 + a_{32}q_2\right. \\ &+ \left. a_{44}q_3 + a_{42}q_1\right)X_2(t)X_4(t) + 2\left(a_{14}q_3 + a_{24}q_2 + a_{23}q_3 + a_{34}q_1\right. \\ &+ \left. a_{33}q_2 + a_{44}q_2 + a_{43}q_1\right)X_3(t)X_4(t) + q_1\left(g_{11}^2 + g_{12}^2 + g_{13}^2 + g_{14}^2 + g_{21}^2\right. \\ &+ \left. g_{22}^2 + g_{23}^2 + g_{24}^2 + g_{31}^2 + g_{32}^2 + g_{33}^2 + g_{34}^2 + g_{41}^2 + g_{42}^2 + g_{43}^2 + g_{44}^2\right) \\ &+ 2q_2\left(g_{11}g_{21} + g_{12}g_{22} + g_{13}g_{23} + g_{14}g_{24} + g_{21}g_{31} + g_{22}g_{32} + g_{23}g_{33}\right. \\ &+ \left. g_{24}g_{34} + g_{31}g_{41} + g_{32}g_{42} + g_{33}g_{43} + g_{34}g_{44}\right) + 2q_3\left(g_{11}g_{31} + g_{12}g_{32}\right. \\ &+ \left. g_{13}g_{33} + g_{14}g_{34} + g_{21}g_{41} + g_{22}g_{42} + g_{23}g_{43} + g_{24}g_{44}\right) + 2q_4\left(g_{11}g_{41}\right. \\ &+ \left. g_{12}g_{42} + g_{13}g_{43} + g_{14}g_{44}\right) = LV\mathrm{d}t \end{aligned}$$

.

2.3 Results for the unit matrix Q

For $Q = I$, where I is a unit matrix, we get

$$\begin{aligned} LV &= 2a_{11}X_1^2(t) + 2a_{22}X_2^2(t) + 2a_{33}X_3^2(t) + 2a_{44}X_4^2(t) + 2\left(a_{12} + a_{21}\right)X_1(t)X_2(t) \\ &+ 2\left(a_{13} + a_{23} + a_{31}\right)X_1(t)X_3(t) + 2\left(a_{14} + a_{41}\right)X_1(t)X_4(t) + 2a_{32}X_2(t)X_3(t) \\ &+ 2\left(a_{24} + a_{42}\right)X_2(t)X_4(t) + 2\left(a_{34} + a_{43}\right)X_3(t)X_4(t) + \left(g_{11}^2 + g_{12}^2 + g_{13}^2 + g_{14}^2\right. \\ &+ \left. g_{21}^2 + g_{22}^2 + g_{23}^2 + g_{24}^2 + g_{31}^2 + g_{32}^2 + g_{33}^2 + g_{34}^2 + g_{41}^2 + g_{42}^2 + g_{43}^2 + g_{44}^2\right) \end{aligned}$$

Now we can find conditions of a stability system. The system will be stable if the

Lyapunov function LV is negative definite, so

$$\begin{aligned} & 2a_{11}X_1^2(t)+2a_{22}X_2^2(t)+2a_{33}X_3^2(t)+2a_{44}X_4^2(t)+2\left(a_{12}+a_{21}\right)X_1(t)X_2(t) \\ + & 2\left(a_{13}+a_{23}+a_{31}\right)X_1(t)X_3(t)+2\left(a_{14}+a_{41}\right)X_1(t)X_4(t)+2a_{32}X_2(t)X_3(t) \\ + & 2\left(a_{24}+a_{42}\right)X_2(t)X_4(t)+2\left(a_{34}+a_{43}\right)X_3(t)X_4(t)+\|G\|^2\leq 0. \end{aligned}$$

Remark: Because $\|G\|^2 \geq 0$, therefore the matrix A must be sufficiently negative, to obtain a negative definite function. We will demonstrate that the matrix A must be more dominant than the matrix G for the stability of the stochastic system,

$$\|A\| \gg \|G\|\,.$$

3 Examples

3.1 Example 1

We consider matrices A and G in the form

$$A=\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix}, G=\begin{pmatrix} \frac{a}{10} & 0 & 0 & 0 \\ 0 & \frac{a}{10} & 0 & 0 \\ 0 & 0 & \frac{a}{10} & 0 \\ 0 & 0 & 0 & \frac{a}{10} \end{pmatrix}.$$

3.1.1 Conditions for the existence of solutions

The matrix A will be negative definite under following conditions:

$$\begin{aligned} & D_1=a<0, \\ & D_2=a^2>0,\ D_2 \text{ follows from } D_1, \\ & D_3=a^3<0 \Leftrightarrow a<0 \wedge a^2>0, D_3 \text{ follows from } D_1, D_2, \\ & D_4=a^4>0 \Leftrightarrow a^2>0,\ D_4 \text{ follows from } D_2. \end{aligned}$$

From these conditions it is evident that $a<0$ or the first condition D_1.

3.1.2 Solution of the differential system A

We find eigenvalues of matrix A as the solution of the characteristic equation

$$\begin{vmatrix} a-\lambda & 0 & 0 & 0 \\ 0 & a-\lambda & 0 & 0 \\ 0 & 0 & a-\lambda & 0 \\ 0 & 0 & 0 & a-\lambda \end{vmatrix} = 0,$$

$$(a-\lambda)^4 = 0 \Rightarrow \lambda_{1,2,3,4} = a.$$

Then

$$\begin{aligned} X_1(t) &= e^{at}, \\ X_2(t) &= te^{at}, \\ X_3(t) &= t^2e^{at}, \\ X_4(t) &= t^3e^{at}. \end{aligned}$$

The general solution is given by a linear combination $X_t = C_1X_1(t) + C_2X_2(t) + C_3X_3(t) + C_4X_4(t)$ with arbitrary constants C_1, C_2, C_3, C_4, so

$$X_t = C_1e^{at} + C_2te^{at} + C_3t^2e^{at} + C_4t^3e^{at}, t \in \mathbb{R},$$

and because $a < 0$, then this solution is stable.

3.1.3 Solution of the stochastic system

We determine stability of solution for $Q = I$

$$\begin{aligned} \mathrm{d}V(X_t) &= 2\left(aX_1^2(t) + aX_2^2(t) + aX_3^2(t) + aX_4^2(t) + \frac{a^2}{50}\right)\mathrm{d}t + \frac{a}{5}X_1(t)\mathrm{d}B_1(t) \\ &+ \frac{a}{5}X_2(t)\mathrm{d}B_2(t) + \frac{a}{5}X_3(t)\mathrm{d}B_3(t) + \frac{a}{5}X_4(t)\mathrm{d}B_4(t). \end{aligned}$$

$$E\left\{\mathrm{d}V(X_t)\right\} = 2\left(aX_1^2(t) + aX_2^2(t) + aX_3^2(t) + aX_4^2(t) + \frac{a^2}{50}\right)\mathrm{d}t = LV\mathrm{d}t.$$

If holds the inequality $LV \leq 0$, thus

$$a\left\|X(t)\right\|^2 \leq -\frac{a^2}{50},$$

for $X_t = C_1e^{at} + C_2te^{at} + C_3t^2e^{at} + C_4t^3e^{at}, t \in \mathbb{R}$, then the system is stochastic stable.

3.2 Example 2

We consider matrices A and G in the form

$$A = \begin{pmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \\ 0 & 0 & a_3 & 0 \\ 0 & 0 & 0 & a_4 \end{pmatrix}, G = \begin{pmatrix} \frac{a_1}{10} & 0 & 0 & 0 \\ 0 & \frac{a_2}{10} & 0 & 0 \\ 0 & 0 & \frac{a_3}{10} & 0 \\ 0 & 0 & 0 & \frac{a_4}{10} \end{pmatrix},$$

where $a_i \neq a_j$ for $i \neq j; i, j = 1, 2, 3, 4$.

3.2.1 Conditions for the existence of solutions

The matrix A will be negative definite under following conditions:

$$\begin{aligned} &D_1 = a_1 < 0, \\ &D_2 = a_1a_2 > 0 \Leftrightarrow a_2 < 0,\ D_2 \text{ follows from } D_1, \\ &D_3 = a_1a_2a_3 < 0 \Leftrightarrow a_3 < 0, D_3 \text{ follows from } D_2, \\ &D_4 = a_1a_2a_3a_4 > 0 \Leftrightarrow a_4 < 0,\ D_4 \text{ follows from } D_3. \end{aligned}$$

From these conditions it is evident that $a_i < 0, i = 1, 2, 3, 4$.

3.2.2 Solution of the differential system A

We find eigenvalues of matrix A as the solution of the characteristic equation

$$\begin{vmatrix} a_1 - \lambda & 0 & 0 & 0 \\ 0 & a_2 - \lambda & 0 & 0 \\ 0 & 0 & a_3 - \lambda & 0 \\ 0 & 0 & 0 & a_4 - \lambda \end{vmatrix} = 0,$$

$$(a_1 - \lambda)(a_2 - \lambda)(a_3 - \lambda)(a_4 - \lambda) = 0.$$

Then

$$\lambda_i = a_i \Rightarrow X_i(t) \quad = \quad e^{a_it}.$$

The general solution with arbitrary constants C_1, C_2, C_3, C_4 is given by

$$X_t = \sum_{i=1}^{4} C_i e^{a_it}, t \in \mathbb{R},$$

and because $a_i < 0$, then this solution is stable.

3.2.3 Solution of the stochastic system

We determine stability of solution for $Q = I$

$$\begin{aligned}
\mathrm{d}V(X_t) &= 2\left(a_1X_1^2(t) + a_2X_2^2(t) + a_3X_3^2(t) + a_4X_4^2(t) + \frac{a_1^2 + a_2^2 + a_3^2 + a_4^2}{200}\right)\mathrm{d}t \\
&+ \frac{a_1}{5}X_1(t)\mathrm{d}B_1(t) + \frac{a_2}{5}X_2(t)\mathrm{d}B_2(t) + \frac{a_3}{5}X_3(t)\mathrm{d}B_3(t) + \frac{a_4}{5}X_4(t)\mathrm{d}B_4(t).
\end{aligned}$$

$$\begin{aligned}
E\left\{\mathrm{d}V(X_t)\right\} &= 2\left(a_1X_1^2(t) + a_2X_2^2(t) + a_3X_3^2(t) + a_4X_4^2(t) + \frac{a_1^2 + a_2^2 + a_3^2 + a_4^2}{200}\right)\mathrm{d}t \\
&= LV\mathrm{d}t.
\end{aligned}$$

If holds the inequality $LV \leq 0$, thus

$$a_1X_1^2(t) + a_2X_2^2(t) + a_3X_3^2(t) + a_4X_4^2(t) \leq -\frac{a_1^2 + a_2^2 + a_3^2 + a_4^2}{100},$$

for $X_t = \sum_{i=1}^{4} C_i e^{a_i t}, t \in \mathbb{R}$, then the system is stochastic stable.

3.3 Example 3

We consider matrices A and G in the form

$$A = \begin{pmatrix} a_1 & 1 & 1 & 1 \\ 0 & a_2 & 1 & 1 \\ 0 & 0 & a_3 & 1 \\ 0 & 0 & 0 & a_4 \end{pmatrix}, G = \begin{pmatrix} \frac{a_1}{10} & 1 & 1 & 1 \\ 0 & \frac{a_2}{10} & 1 & 1 \\ 0 & 0 & \frac{a_3}{10} & 1 \\ 0 & 0 & 0 & \frac{a_4}{10} \end{pmatrix}.$$

3.3.1 Conditions for the existence of solutions

The matrix A will be negative definite under following conditions:

$$\begin{aligned}
&D_1 = a_1 < 0, \\
&D_2 = a_1a_2 > 0 \Leftrightarrow a_2 < 0,\ D_2 \text{ follows from } D_1, \\
&D_3 = a_1a_2a_3 < 0 \Leftrightarrow a_3 < 0, D_3 \text{ follows from } D_2, \\
&D_4 = a_1a_2a_3a_4 > 0 \Leftrightarrow a_4 < 0,\ D_4 \text{ follows from } D_3.
\end{aligned}$$

From these conditions it is evident that $a_i < 0, i = 1, 2, 3, 4$.

3.3.2 Solution of the differential system A

We find eigenvalues of matrix A as the solution of the characteristic equation

$$\begin{vmatrix} a_1-\lambda & 1 & 1 & 1 \\ 0 & a_2-\lambda & 1 & 1 \\ 0 & 0 & a_3-\lambda & 1 \\ 0 & 0 & 0 & a_4-\lambda \end{vmatrix} = 0,$$

$$(a_1-\lambda)(a_2-\lambda)(a_3-\lambda)(a_4-\lambda) = 0.$$

According to previous example the general solution with arbitrary constants C_1, C_2, C_3, C_4 is given by

$$X_t = C_1e^{a_1t} + C_2e^{a_2t} + C_3e^{a_3t} + C_4e^{a_4t}, t \in \mathbb{R}.$$

We can write for a general matrix H

$$H = \begin{pmatrix} a_1 & \alpha & \beta & \gamma \\ 0 & a_2 & \delta & \epsilon \\ 0 & 0 & a_3 & \kappa \\ 0 & 0 & 0 & a_4 \end{pmatrix},$$

where $\alpha, \beta, \gamma, \delta, \epsilon, \kappa \in \mathbb{R}$, the general solution is

$$X_t = C_1e^{a_1t} + C_2e^{a_2t} + C_3e^{a_3t} + C_4e^{a_4t}, t \in \mathbb{R},$$

where C_1, C_2, C_3, C_4 are constants.

3.3.3 Solution of the stochastic system

We determine stability of solution for $Q = I$.

$$\begin{aligned} \mathrm{d}V(X_t) &= 2(a_1X_1^2(t) + a_2X_2^2(t) + a_3X_3^2(t) + a_4X_4^2(t) + X_1(t)X_2(t) + 2X_1(t)X_3(t) \\ &+ X_1(t)X_4(t) + X_2(t)X_4(t) + X_3(t)X_4(t) + \frac{a_1^2+a_2^2+a_3^2+a_4^2}{200} + 3)\mathrm{d}t \\ &+ 2X_1(t)\left(\frac{a_1}{10}\mathrm{d}B_1(t) + \mathrm{d}B_2(t) + \mathrm{d}B_3(t) + \mathrm{d}B_4(t)\right) + 2X_4(t)\left(\frac{a_4}{10}\mathrm{d}B_4(t)\right) \\ &+ 2X_2(t)\left(\frac{a_2}{10}\mathrm{d}B_2(t) + \mathrm{d}B_3(t) + \mathrm{d}B_4(t)\right) + 2X_3(t)\left(\frac{a_3}{10}\mathrm{d}B_3(t) + \mathrm{d}B_4(t)\right). \end{aligned}$$

$$\begin{aligned} E\left\{\mathrm{d}V(X_t)\right\} &= 2(a_1X_1^2(t)+a_2X_2^2(t)+a_3X_3^2(t)+a_4X_4^2(t)+X_1(t)X_2(t)+2X_1(t)X_3(t) \\ &+ X_1(t)X_4(t)+X_2(t)X_4(t)+X_3(t)X_4(t)+\frac{a_1^2+a_2^2+a_3^2+a_4^2}{200}+3)\mathrm{d}t \\ &= LV\mathrm{d}t. \end{aligned}$$

If holds the inequality $LV \leq 0$, thus

$$\begin{aligned} & a_1X_1^2(t)+a_2X_2^2(t)+a_3X_3^2(t)+a_4X_4^2(t)+X_1(t)X_2(t)+2X_1(t)X_3(t)+X_1(t)X_4(t) \\ + \;& X_2(t)X_4(t)+X_3(t)X_4(t) \leq -\frac{a_1^2+a_2^2+a_3^2+a_4^2}{100}-6, \end{aligned}$$

for $X_t = C_1e^{a_1t}+C_2e^{a_2t}+C_3e^{a_3t}+C_4e^{a_4t}, t \in \mathbb{R}$, then the system is stochastic stable.

3.4 Example 4

We consider matrices A and G in the form

$$A = \begin{pmatrix} a_1 & 0 & 0 & a_2 \\ 0 & a_1 & a_2 & 0 \\ 0 & a_2 & a_1 & 0 \\ a_2 & 0 & 0 & a_1 \end{pmatrix}, G = \begin{pmatrix} \frac{a_1}{10} & 0 & 0 & \frac{a_2}{10} \\ 0 & \frac{a_1}{10} & \frac{a_2}{10} & 0 \\ 0 & \frac{a_2}{10} & \frac{a_1}{10} & 0 \\ \frac{a_2}{10} & 0 & 0 & \frac{a_1}{10} \end{pmatrix}.$$

3.4.1 Conditions for the existence of solutions

The matrix A will be negative definite under following conditions:

$D_1 = a_1 < 0,$
$D_2 = a_1^2 > 0,\; D_2$ follows from $D_1,$
$D_3 = a_1^3 - a_1a_2^2 < 0 \Leftrightarrow a_1 < 0 \wedge a_1^2 - a_2^2 > 0 \Rightarrow |a_2| < |a_1|.$
$D_4 = a_1^4 - 2a_1^2a_2^2 + a_2^4 > 0 \Leftrightarrow (a_1^2 - a_2^2)^2 > 0,\; D_4$ holds for arbitrary $|a_1| \neq |a_2|.$

From these conditions it is evident that $a_1 < 0$ and $|a_2| < |a_1|$.

3.4.2 Solution of the differential system A

We find eigenvalues of matrix A as the solution of the characteristic equation

$$\begin{vmatrix} a_1-\lambda & 0 & 0 & a_2 \\ 0 & a_1-\lambda & a_2 & 0 \\ 0 & a_2 & a_1-\lambda & 0 \\ a_2 & 0 & 0 & a_1-\lambda \end{vmatrix} = 0,$$

$$\begin{aligned}[(a_1-\lambda)^2-a_2^2]^2 &= 0,\\ |a_1-\lambda| &= |a_2|\,.\end{aligned}$$

Then according to Example (3.2) in paper [2] we get

$$\begin{array}{llll}\text{for} & a_2>0 & \text{is} & X_{1,2}(t)=(1,1)^Te^{(-a_1+a_2)t},\\ \text{for} & a_2<0 & \text{is} & X_{1,2}(t)=(-1,1)^Te^{(-a_1+a_2)t},\\ \text{for} & a_2<0 & \text{is} & X_{3,4}(t)=(1,1)^Te^{(-a_1-a_2)t},\\ \text{for} & a_2>0 & \text{is} & X_{3,4}(t)=(1,-1)^Te^{(-a_1-a_2)t}.\end{array}$$

The general solution is given by a linear combination $X_t = C_1X_1(t)+C_2X_2(t)+C_3X_3(t)+C_4X_4(t)$, with arbitrary constants C_1, C_2, C_3, C_4.

3.4.3 Solution of the stochastic system

We determine stability of solution for $Q = I$.

$$\begin{aligned}\mathrm{d}V(X_t) &= 2\left[a_1(X_1^2(t)+X_2^2(t)+X_3^2(t)+X_4^2(t))+a_2(X_1(t)X_3(t)+2X_1(t)X_4(t)\right.\\ &+ \left.X_2(t)X_3(t))+\frac{a_1^2}{50}+\frac{a_2^2}{50}\right]\mathrm{d}t+2X_1(t)\left(\frac{a_1}{10}\mathrm{d}B_1(t)+\frac{a_2}{10}\mathrm{d}B_4(t)\right)\\ &+ 2X_2(t)\left(\frac{a_1}{10}\mathrm{d}B_2(t)+\frac{a_2}{10}\mathrm{d}B_3(t)\right)+2X_3(t)\left(\frac{a_2}{10}\mathrm{d}B_2(t)+\frac{a_1}{10}\mathrm{d}B_3(t)\right)\\ &+ 2X_4(t)\left(\frac{a_2}{10}\mathrm{d}B_1(t)+\frac{a_1}{10}\mathrm{d}B_4(t)\right).\end{aligned}$$

$$\begin{aligned}E\left\{\mathrm{d}V(X_t)\right\} &= 2\left[a_1(X_1^2(t)+X_2^2(t)+X_3^2(t)+X_4^2(t))+a_2(X_1(t)X_3(t)+2X_1(t)X_4(t)\right.\\ &+ \left.X_2(t)X_3(t))+\frac{a_1^2}{50}+\frac{a_2^2}{50}\right]\mathrm{d}t\\ &= LV\mathrm{d}t.\end{aligned}$$

If holds the inequality $LV \leq 0$, thus

$$a_1\left\|X(t)\right\|^2+a_2(X_1(t)X_3(t)+2X_1(t)X_4(t)+X_2(t)X_3(t))\leq-\frac{a_1^2+a_2^2}{50},$$

for $X_t = C_1X_1(t)+C_2X_2(t)+C_3X_3(t)+C_4X_4(t), t\in\mathbb{R}$, then the system is stochastic stable.

4 Conclusion

There was defined stability and stochastic stability of the stochastic differential system. Conditions for stochastic stability were established on the model of the stochastic differential system with four-dimensional Brownian motion by using Lyapunov theorem. Results were illustrated on trivial examples. Such type of equations can be used also in biomedical engineering, in meteorology, epidemic modeling, predicting economics, etc.

Acknowledgement

This research was supported by Grant FEKT-S-11-2-921 of Faculty of Electrical Engineering and Communication, Brno University of Technology.

Reference

[1] BAŠTINEC, 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. p. 23 - 29. ISBN 978-80-7231-815-5.

[2] BAŠTINEC, J.; KLIMEŠOVÁ, M. *Stability of the Zero Solution of Stochastic Differential System with Three- dimensional Brownian motion.* In Matematika, Informační technologie a aplikované vědy. Brno: UNOB, 2016. s. 1-8. ISBN: 978-80-7231-464- 5.

[3] DIBLÍK, J., KHUSAINOV, D.Y., BAŠTINEC, J., RYVOLOVÁ, A.: *Exponential stability and estimation of solutions of linear differential systems with constant delay of neutral type.* In 6. konference o matematice a fyzice na vysokých školách technických s mezinárodní účastí. Brno, UNOB Brno. 2009. p. 139 - 146. ISBN 978-80-7231-667-0.

[4] DITLEVSEN, S., BATZEL, J., BACHAR, M.: *Stochastic biomathematical models*, Heidelberg: Springer, 2013, 206 p.

[5] DURRETT, R.: *Probability: theory and examples*, 3. ed. Belmont, CA: Thomson Brooks/Cole, 2005, 497 s. ISBN 05-344-2441-4.

[6] IGNATYEV, A. O; IGNATYEV, O. *Quadratic forms as Lyapunov functions in the study of stability of solutions to difference equations.* Electronic Journal of Differential Equations. 2011, (19): 1-21. ISSN 1072-6691. Available from: http://ejde.math.txstate.edu

[7] DZHALLADOVA, I.A.: *Optimization of stochastic systems*, Kiev, KNEU Press, 2005. ISBN 966-574-774-6.

[8] DZHALLADOVA, I.; BAŠTINEC, J.; DIBLÍK, J.; KHUSAINOV, D.: *Estimates of exponential stability for solutions of stochastic control systems with delay.* Abstract and Applied Analysis. 2011. 2011(1). p. 1 - 14. ISSN 1085-3375. (IF=1,318).

[9] GILBERT, G. T. *Positive Definite Matrices and Sylvester's Criterion.* The American Mathematical Monthly 98.1 (1991): 44-46. DOI:10.2307/2324036.

[10] KHASMINSKII, R.: *Stochastic stability of differential equations.* New York: Springer Berlin Heidelberg, 2011, 358 s. ISBN 978-3-642-23279-4.

[11] KLIMEŠOVÁ, M.: *Stochastic Differential Equations*, Student EEICT. Brno: LITERA, 2014. s. 150-154. ISBN: 978-80-214-4924-4.

[12] KLIMEŠOVÁ, M.: *Stability of the Stochastic Differential Equations*, Student EEICT. Brno: BUT, 2015. s. 526-530. ISBN: 978-80-214-214-5148-3.

[13] KLIMEŠOVÁ, M.; BAŠTINEC, J. *Application of Stochastic Differential Equations.* MITAV. Brno: UO, 2014. s. 1-6. ISBN: 978-80-7231-961-9.

[14] KLIMEŠOVÁ, M.; BAŠTINEC, J. *Stability of the Zero Solution of Stochastic Differential Systems with Two-dimensional Brownian motion.* Mathematics, Information Technologies, and Applied Science 2015. Brno: UNOB, 2015. s. 8-20. ISBN: 978-80-7231-436- 2.

[15] KOLÁŘOVÁ, E.: *Stochastické diferenciální rovnice v elektrotechnice.* Thesis. Brno: BUT, 2006, 26 s. ISBN 80-214-3330-2.

[16] MAO, X.: *Stochastic differential equations and applications.* Chichester: Horwood Pub., 2008, 422 s. ISBN 978-1-904275-34-3.

[17] MASLOWSKI, B., MILOTA, J.: *Proceedings of Seminar in Differential Equations.* Plzeň: Vydavatelský servis, 2006, 118 s. ISBN 80-86843-14-9.

[18] NAVARA, M.: *Pravděpodobnost a matematická statistika.* Skriptum. Praha: FEL ČBUT, 2007, 240 s. ISBN 978-80-01-03795-9.

[19] ØKSENDAL, B.: Stochastic Differential Equations. *An Introduction with Applications*, Springer-Verlag, 1995.

[20] STANĚK, J.: *Stochastické diferenciální rovnice*, KDM - MFF UK, 2011.

[21] STEPANOV, V. V.: *Course of differential equations (review)*, Uspekhi Mat. Nauk, 1939, no. 6, 288-289.

HYPERBOLIC SINE AND COSINE FROM THE ITERATION THEORY POINT OF VIEW

Jaroslav Beránek
Pedagogická fakulta MU
Poříčí 7, 603 00 Brno, Česká Republika
beranek@ped.muni.cz

Abstract: *The article is devoted to an example of a discrete representation of functions from the point of view of an iteration theory. On the basis of two real functions, the hyperbolic sine and cosine are represented through their vertex graphs and the existence of iterative roots is solved. The article includes also basic information and examples of isomorphic mono-unary algebras. In the conclusion of the article there is given a discrete description of a function $f(x) = \cosh x - 1$, where a formal description of its second iterative roots is demonstrated.*

Keywords: Hyperbolic functions; theory of iteration; iterative roots; vertex graph; mono-unary algebra.

INTRODUCTION

Currently, the subject matter and methods of mathematics teaching in the world are undergoing a phase of substantial changes which have resulted from the development of mathematics and its applications. Mathematics teaching has to lead to its applicability. Mathematics should not be taught as a theoretical discipline, but as a tool for solving practical problems. The need of the conscious acquisition of mathematics requires the subject matter to form an integrated system where all pieces of knowledge are brought into accord through multiple mutual relations, links and connections which finally contribute to their understanding, retention and usage. Such a synthesizing attitude should determine all knowledge of students.

Especially, the dual "approach" to elementary real functions of one variable (from the continuous and discrete points of view) is of great significance. Within the scope of school mathematics at all types and stages of schools, the continuous approach is preferred, where functions are represented through their Cartesian graphs. The other approach, the discrete one, where functions are represented through their vertex graphs, is nearly unknown to the students.

This article shows an example of the discrete approach to functions through the description of vertex graphs of two hyperbolic functions – the sine and cosine – including the use of their graphs for solving the problem of iterative roots of these functions. In the final part there is given one "modification" of the hyperbolic cosine which shows further remarkable applications for solving functional equations. The reason for the choice of these two hyperbolic functions is, among others, the fact that although these functions are relatively simple in the form and they have quite a lot of practical applications, they are unjustly neglected at school mathematics.

1. HYPERBOLIC FUNCTIONS

Let $x \in \boldsymbol{R}$ be arbitrary, then $sinh\ x = \dfrac{e^x - e^{-x}}{2}$, $cosh\ x = \dfrac{e^x + e^{-x}}{2}$. The Cartesian graphs are represented in Figure 1.

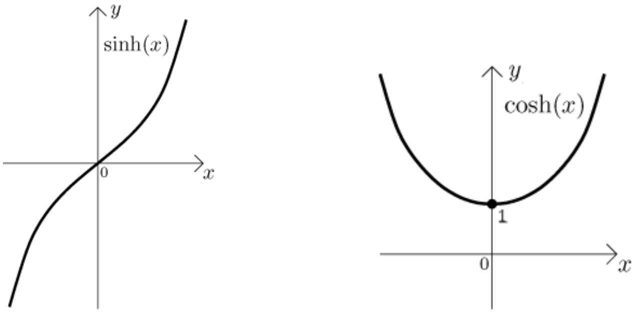

Fig. 1

From the definitions and Cartesian graphs it is evident that both functions are defined on the whole real axis, the range of function *sinh x* is the set ***R***, the range of function *cosh x* is an interval $\langle 1, \infty)$. The hyperbolic sine is an odd function, while the hyperbolic cosine is an even one.

Now let us show why the given functions are the hyperbolic ones. From the definitions we can derive the formula $cosh^2\ x - sinh^2\ x = 1$; it means the point in the plain the Cartesian coordinates of which are *[a cosh t, b sinh t]*, where *a, b* are positive real numbers, *t* is a real parameter, lies on the hyperbola $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$. Through equations $x = a\ cosh\ t,\ y = b\ sinh\ t$ it is possible to parametrize the hyperbola in the same way as the ellipse through equations $x = a\cos t, y = b\sin t$.

There arises a natural question: Do hyperbolic and goniometric functions have similar properties? The answer is a positive one: similarly to trigonometry, it is possible to define the hyperbolic tangent and hyperbolic cotangent, and to derive formulas similar to the ones which apply for goniometric functions. Let us state some of them:

$$tgh\ x = \frac{sinh\ x}{cosh\ x},\ \ cotgh\ x = \frac{cosh\ x}{sinh\ x}, x \neq 0,\ \ sinh\ 2x = 2\ sinh\ x\ cosh\ x,$$

$$cosh\ 2x = sinh^2 x + cosh^2 x,\ \ sinh\ (x \pm y) = sinh\ x\ cosh\ y \pm cosh\ x\ sinh\ y, \text{ etc.}$$

Hyperbolic functions have a wide range of applications in mathematical analysis and physics. E.g. it is possible to find the solution of the integral of the type $\int \dfrac{dx}{\sqrt{x^2 + a^2}}$ through the substitution $x = a\ sinh\ t$, the integral of the type $\int \dfrac{dx}{\sqrt{x^2 - a^2}}$ could be solved through the substitution $x = a\ cosh\ t$. The expansion of both basic hyperbolic functions to the Taylor series is also widely used:

$$sinh\ x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}, \qquad cosh\ x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}.$$

If we consider the Cartesian graphs of hyperbolic sine and cosine, we can notice their striking "resemblance" to the Cartesian graphs of polynomial functions; the function *coshx* resembles the function $f(x) = x^{2k}+1,\ k \in \boldsymbol{N}$, the function *sinh x* resembles the function

$f(x) = x^{2k+1}$, $k \in \boldsymbol{N}$**.** There remains the question if hyperbolic functions have some property similar to the ones of the polynomials. Again, the answer is the positive one; in the discrete concept (in the iteration theory), functions hyperbolic sine and cosine behave like polynomials. First of all we need to define necessary terms and state theorems. Let us restrain ourselves to a brief description. Details could be found in [10], [16].

2. HYPERBOLIC FUNCTIONS FROM THE ITERATION THEORY POINT OF VIEW

Now we will deal with the discrete description of both above mentioned hyperbolic functions, their representation with the help of vertex graphs and its application. Let $\boldsymbol{N}$ be the set of all positive integers, $\boldsymbol{N}_0 = \boldsymbol{N} \cup \{0\}$. The non-empty mapping f of the set A $(A \neq \phi)$ into itself will be called the *transformation* of the given set. For $n \in \boldsymbol{N}_0$ let us define the n-th iteration of the transformation f as follows: $f^{\,0}(x) = x$, $f^{\,1}(x) = f(x)$, $f^{\,n}(x) = (f \, o \, f^{\,n-1})(x)$ for any $x \in A$. Every transformation f of the set A determines the equivalence $\sim_f$ on A as follows: $x \sim_f y$, if and only if there exists a pair of positive integers m, n such that $f^{\,m}(x) = f^{\,n}(y)$. The blocks of the decomposition of the set A determined by the equivalence $\sim_f$ are called *orbits* of the transformation f, in short f-orbits. The *vertex graph* of the transformation (function) f will be plotted in the following way: the elements of the set A will be plotted as points in the plain. We will join the element x to an element y with an arrow if and only if $y = f(x)$. If $f(x) = x$ (x is a fixed point of the function f), we will draw a loop around the point x. If the orbit contains k elements $x_1, \dots x_k$ with the property $f(x_1) = x_2$, $f(x_2) = x_3$, ..., $f(x_{k-1}) = x_k$, $f(x_k) = x_1$, then we say that it is k-cyclic and the given k elements form the *k-cycle*. From the point of view of the graph theory, the vertex graph of the function is the oriented graph, the orbits of the given function f are its weakly connected subgraphs.

Let f, g be functions defined on the same set A. Let for every $x \in A$ be $g^n = f$ for any $n \in \boldsymbol{N}$, $n \geq 2$. Then the function g is called the *n-th iterative root* of the function f. The problem of the existence and construction of the iterative root of the given function f is solved through the analysis of the vertex graph of this function f. It has been proved (see [16]), that the vertex graph of the iterative root g of the function f could be obtained by mating of the orbits of the function f. Precisely, the n-th iterative root of the function f exists if and only if the set of f-orbits can be decomposed to such blocks that the number of the orbits in every block is the divisor of the number n and the orbits in every block are n-mateable.

The ordered pair (A, f) will be called the *mono-unary algebra* (the symbols A, f have the above mentioned meaning). If (A, f), (B, g) are two mono-unary algebras, then we say that they are isomorphic and we write $(A, f) \cong (B, g)$, if and only if there exists the bijective mapping h of the set A to the set B with the property $h \, o \, f = g \, o \, h$. The symbol o denotes the operation composition of mappings. Both functions f, g are in this case called the *conjugated* ones. The vertex graphs of functions of two isomorphic mono-unary algebras are "the same" from the point of view of the graph theory (if we do not take into account the labelling of the elements in the graphs). *The component* of the mono-unary algebra (A, f) will be called the pair, where the carrier set is the orbit of the function f, and the function is the restriction f on this orbit. Obviously, the components are minimal sub-algebras (with respect to the inclusion) of the given mono-unary algebra. Let us give an example of two real functions $f(x) = 2x(1-x)$, $g(x) = x^2$, for which there exists the bijection $h\colon \boldsymbol{R} \to \boldsymbol{R}$, $h(x) = 1-2x$ with the property $h \, o \, f = g \, o \, h$; so there holds $(\boldsymbol{R}, f) \cong (\boldsymbol{R}, g)$.

Let us get back to the hyperbolic functions. The function $f(x) = \sinh x$ is bijective, while for the only real number $x = 0$ there holds $f(x) = x$. The vertex graph of the function $\sinh x$ contains innumerably many two-sidedly infinite chains (details can be found in [3], [5], [10]) and one fixed point $x = 0$; the vertex graph of the polynomial function $f(x) = x^3$, whose

Cartesian graph is analogical to the one of the function *sinh x*, contains also innumerably many two-sidedly infinite chains and three fixed points $x_1 = 0, x_2 = -1, x_3 = 1$. The vertex graph of the function $f(x)= \sinh x$ is in Figure 2:

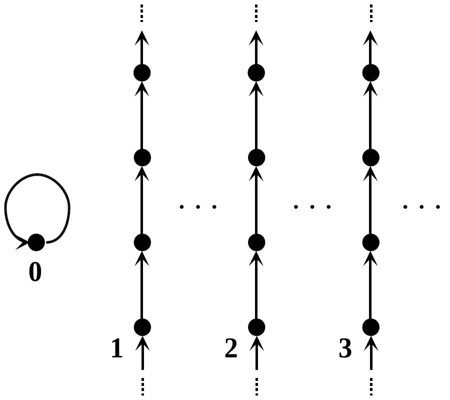

Fig. 2

Evidently, these functions are not conjugated due to the different number of the fixed points. However, if we add two fixed points to the vertex graph of the function *sinh*, then the newly defined function will be conjugated with the function x^3. We will perform such modification through extending the real axis by adding two improper points $+\infty, -\infty$, so we will define the set $\boldsymbol{R}_N = \boldsymbol{R} \cup \{+\infty, -\infty\}$. On the set $\boldsymbol{R}_N$ we will define the function *Sinh x* as follows: $Sinh(-\infty) = -\infty$, $Sinh(+\infty) = +\infty$; for all $x \in \boldsymbol{R}$ there holds $Sinh\, x = \sinh x$. Then the both considered mono-unary algebras are isomorphic, and we write $(\boldsymbol{R}_N, Sinh) \cong (\boldsymbol{R}, x^3)$. If we examine the iterative roots of the function *sinh x*, then obviously for every $n \in \boldsymbol{N}$, $n \geq 2$ there exists the iterative root of the order *n*. For the fixed point $x = 0$ there always applies $f^n(0) = 0$, the other two-sidedly infinite chains can be mated for every chosen *n*. For $n = 3$, the mating is shown in Figure 3:

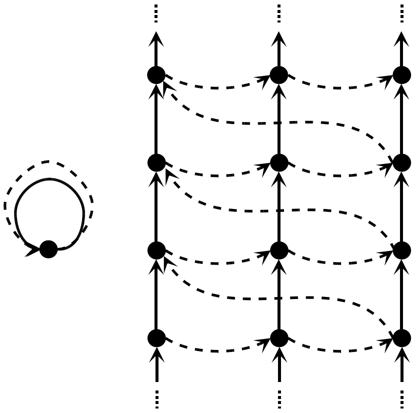

Fig. 3

Now, let us consider the function *cosh x* and the analogical polynomial function $x^2 + 1$. Neither of these functions has any fixed point. Both of them are even functions, their vertex graphs are the same (only the labelling of the vertices differs). They contain one orbit in the shape of the chain with the minimal element *0*, its successor is number *1*, etc. and because the functions are even, on every positive element of the chain beginning with the vertex *1* there is mapped the vertex labelled with the opposite negative number. Further, they contain the vertex graphs of innumerably many orbits of the "similar" shape as the first orbit, whose minimal elements are pairs $\pm k$, where $k \in (-1, 0) \cup (0, 1)$.The vertex graph is in Figure 4:

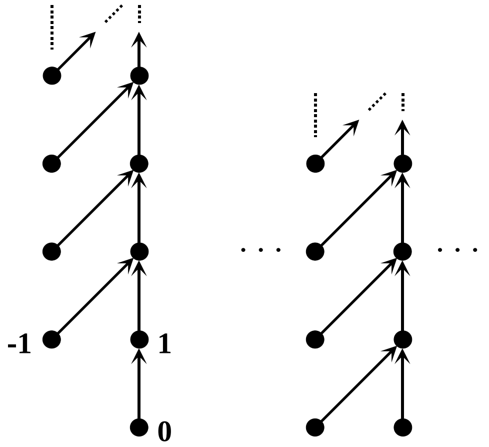

Fig. 4

Mono-unary algebras formed of both functions are isomorphic, so we can write $(\boldsymbol{R}, cosh) \cong (\boldsymbol{R}, x^2+1)$. If we want to solve the problem of the existence and construction of the iterative roots, we can state that the function *cosh x* has iterative roots of all orders. One orbit containing the zero is mateable with other orbits for every permissible *n*, other orbits are mateable as well (because they are the same). For $n = 3$, the mating is illustrated by Figure 5:

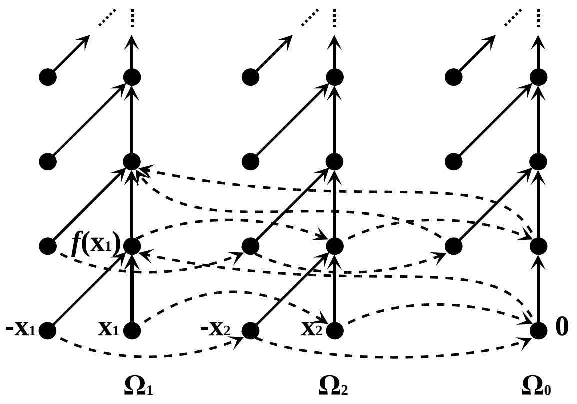

Fig. 5

In order not to restrict ourselves only on the intuitive representation of the iterative roots of the function $f(x) = cosh\ x$ with the help of the picture, we will show the formal description of mating of one orbit containing the zero with $m-1$ other orbits for an arbitrary iterative root of the order *m*. Let us show that for any permissible *m* these orbits are *m*-mateable. We will denote the orbits: the orbit containing the zero will be denoted as Ω_0, other orbits $\Omega_1, ..., \Omega_{m-1}$. The minimal non-negative element of the orbit Ω_0 is 0, the minimal positive elements of other orbits will be labelled as $x_1, x_2,...., x_{m-1}$. Let us define the function *g* as follows: $g(x_i) = x_{i+1}$, $g(-x_i) = -x_{i+1}$ for $i = 1,..., m\text{-}2$, $g(x_{m-1}) = g(-x_{m-1}) = 0$, $g(0) = f(x_1)$, further for $p \in \boldsymbol{N}$ let us define $g[f^p(x_i)] = f^p(x_{i+1})$, $g[-f^p(x_i)] = -f^p(x_{i+1})$ for $i = 1,..., m\text{-}1$, $g[f^p(0)] = g[-f^p 0)] = f^{p+1}(x_1)$. For the function *g* there applies $g^m = f$, so the orbit Ω_0 is *m*-mateable with orbits $\Omega_1, ..., \Omega_{m-1}$ and the function $f = cosh\ x$ has iterative roots of all orders. Another possible description of iterative roots will be shown in the next section.

3. ITERATIVE ROOTS OF ORDER TWO

We have already shown that from the point of view of the iterative theory, the function hyperbolic cosine is analogic to the quadratic function $x^2 + 1$. We will further "extend" this analogy as follows: the simplest quadratic function is $f(x) = x^2$. Under such analogy, the corresponding function should be the function $q(x) = \cosh x - 1$. We will show the correctness of such consideration, and with the help of the function $q(x)$ we will show next interesting courses of our exploration, more suitable for university students due to its formal demands.

The orbit structure of the function $q(x) = \cosh x - 1$ is the same as the one at the function $f(x) = x^2$. We will present the description of the quadratic real function $f(x) = x^2$ using its vertex graph. The mono-unary algebra ($\boldsymbol{R}, f$) has just two finite components with carrier sets $K_0 = \{0\}$, $K_1 = \{-1, 1\}$ with 1-element cycles $\{0\}$, $\{1\}$ in the given order. Further, it is formed of uncountable many enumerable 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. It is easy to show (see [8]) that this vertex graph is the same as at the mono-unary algebra ($\mathbf{Z}, \mu$), where μ: $\mathbf{Z} \to \mathbf{Z}$, $\mu(z) = z +2$ for an odd z, $\mu(z) = z +1$ for an even z.

Now let us proceed to the function $q(x) = \cosh x - 1$, whose vertex graph is analogic to the above mentioned simplest quadratic function. The equation $\cosh x - 1 = x$ has an evident solution $x_0 = 0$, and also an approximate solution $x_1 = 1{,}616$. These are two fixed points of the function q; considering that the function q is the even one, there also holds that $f(-1{,}616) = 1{,}616$. Let us denote the number $-1{,}616$ as x_2. The orbital structure of the function $q(x) = \cosh x - 1$ is as follows: one orbit (let us name it K_0) is a 1-element one and contains the vertex 0, the second orbit is a 2-element one (let us name it K_1), containing two vertices x_1, x_2 and is finished with the loop in the vertex x_1. Further, the vertex graph contains uncountable many orbits, the base of which are two-sidedly infinite chains, where on each positive vertex there is mapped one vertex denoted by the opposite number (denoted analogically as K_t, $t \in (0,1)$). The vertex graph is shown in Figure 6.

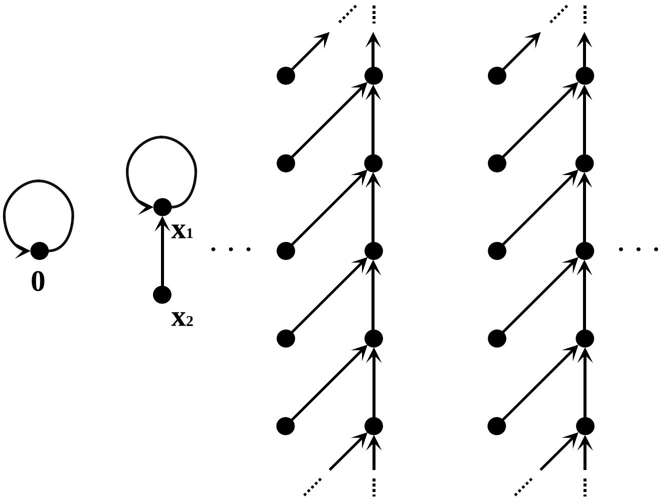

Fig. 6

In the further text we will use the monoid of the function´s endomorphism $q(x) = \cosh x -1$, denoted as $End(\boldsymbol{R},q)$. According to the familiar definition, there holds $End(\boldsymbol{R},q) = \{h:\boldsymbol{R} \to \boldsymbol{R};\ h \circ q = q \circ h\}$; sometimes we also call it the centralizer of the transformation q in the full transformational monoid $T(\boldsymbol{R})$. Now let us prove that the function $q(x) = \cosh x - 1$ has iterative roots of all orders, even in the set $End(\boldsymbol{R}, q)$. The set of all second iterative roots of the function q will be denoted $\sqrt{q}^{\,*}$.

Theorem 1: (See [8]) Let q: $\boldsymbol{R} \to \boldsymbol{R}$ be the function defined by the formula $q(x) = \cosh x - 1$ for every $x \in \boldsymbol{R}$. The equation $f^n = q$ has a solution in the monoid $End(\boldsymbol{R},q)$ for every $n \in \boldsymbol{N}$.

Proof: From the description of the vertex graph of function g there applies (Fig.6) that it contains two finite components K_0, K_1 and uncountable many infinite countable components K_t, $t \in (0,1)$. Let us consider the mono-unary algebra $(\boldsymbol{R},q)$ in the form $(\boldsymbol{R},q) = \sum_{t\in\langle 0,1\rangle}(K_t, q_t)$, where $q_t = q \,|\, K_t$ for $t \in \langle 0,1\rangle$, $K_0 = \{0\}$, $K_1 = \{x_1, x_2\}$. Let γ be the decomposition of the set system $\{K_t;\ t\in(0,1)\}$ to such blocks that every block of the decomposition γ contains just n elements. Let us consider an arbitrary block of the decomposition γ and let us denote its elements K_{t_1} ,..., K_{t_n}. For every pair of the indexes $i, j \in \{1,..., n\}$ there holds that (K_{t_i}, q_{t_i}), (K_{t_j}, q_{t_j}) are isomorphic continuous mono-unary algebras. Let $f_{t_i} :(K_{t_i}, q_{t_i}) \to (K_{t_{i+1}}, q_{t_{i+1}})$ be the relevant isomorphism (a firmly chosen one from many isomorphisms of these components) for $i = 1, 2,..., n-1$. Let $f_{t_n} :(K_{t_n}, q_{t_n}) \to (K_{t_1}, q_{t_1})$ be the homomorphism defined by the relation

$f_{t_n}(x) = \cosh y - 1$, where $y = (f_{t_{n-1}} o f_{t_{n-2}} o \dots o f_{t_1})^{-1}(x) = (f_{t_1}^{-1} o \dots o f_{t_{n-1}}^{-1})(x)$.

The mappings f_{t_1} ,..., $f_{t_{n-1}}$ are bijections, so their composition is the bijective one as well. The existence of the homomorphism f_{t_n} follows from [12]. Now let us denote $f\colon \boldsymbol{R} \to \boldsymbol{R}$ as the function: $f(0) = 0$, $f(x_2) = f(x_1) = x_1$, $f(K_{t_i}) = f_{t_i}$ for every $t_i \in (0,1)$. From the definition of the function f it is evident that $f(\cosh x - 1) = \cosh f(x) - 1$ for every $x \in \boldsymbol{R}$, i.e. $f \in End(\boldsymbol{R},q)$. Now let us show that $f^{\,n} = q$. Let $x \in \boldsymbol{R}$ be an arbitrary number. If $x \in \{x_1, x_2, 0\}$, then $f^{\,n}(x) = \cosh x - 1 = q(x)$. Let $x \in \boldsymbol{R} - \{x_1, x_2, 0\}$. Then there exists a block $\{K_{t_i};\ i = 1,..., n\}$ of the decomposition γ, for which $x \in \bigcup_{i=1}^{n} K_{t_i}$,i.e. $x \in K_{t_k}$ for some suitable natural number k, $1 \le k \le n$. At first, let us assume that $k = 1$; then $f^{n}(x) = f_{t_n}[(f_{t_{n-1}} o \dots o f_{t_1})(x)] = \cosh y - 1$, where $y = (f_{t_{n-1}} o \dots o f_{t_1})^{-1}\ o\ (f_{t_{n-1}} o \dots o f_{t_1})(x) = (f_{t_1}^{-1} o \dots o f_{t_{n-1}}^{-1} o f_{t_{n-1}} o \dots o f_{t_1})(x) = x$, so $f^{n}(x) = \cosh x - 1$. Let $1 < k \le n$. Then $f^{n}(x) = (f_{t_{k-1}} o \dots o f_{t_1} o f_{t_n} o f_{t_{n-1}} o \dots o f_{t_k})(x) \overset{(*)}{=} (f_{t_{k-1}} o \dots o f_{t_1})(f_{t_n}(u))$, where $u = (f_{t_{n-1}} o \dots o f_{t_k})(x)$. According to the definition of f_{t_n} there holds $f_{t_n}(u) = \cosh[(f_{t_1}^{-1} o \dots o f_{t_{n-1}}^{-1})(u)] - 1 \overset{(**)}{=} \cosh[(f_{t_1}^{-1} o \dots o f_{t_{k-1}}^{-1})(x)] - 1$. As f_{t_m} is the isomorphism of the mono-unary algebra (K_{t_m}, q_{t_m}) on the mono-unary algebra $(K_{t_{m+1}}, q_{t_{m+1}})$ for every $m \in \{1, 2,..., n-1\}$, the mapping $\varphi = f_{t_{k-1}} o \dots o f_{t_1}$ is the isomorphism of the mono-unary algebra (K_{t_1}, q_{t_1}) on the algebra (K_{t_k}, q_{t_k}). With respect to the equalities $(*)$, $(**)$, we will get $f^{n}(x) = \varphi(f_{t_n}(u)) = \varphi(\cosh[\varphi^{-1}(x)] - 1) = \cosh(\varphi\ o\ [\varphi^{-1}(x)]) - 1 = \cosh x - 1 = q(x)$, because also the function φ is interchangeable with the function q. Thus the proof is finished.

In Theorem 1 we proved that for every $n \in \boldsymbol{N}$, $n \ge 2$ there exists the n-th iterative root of the function $q(x) = \cosh x - 1$ interchangeable with this function. In the case of the iterative roots of the second order we will prove even more; we will prove that every function $f\colon \boldsymbol{R} \to \boldsymbol{R}$, which is the solution of the equation $f^{2} = q$, commutes with the function q. Firstly, let us give the auxiliary statements, the details see [8].

Lemma 1: Let $f \in \sqrt{q}^{\,*}$ .Then there holds: $f\,(K_0 \cup K_1) \subset K_0 \cup K_1$, $f\,(\bigcup_{t\in(0,1)} K_t\,) \subset \bigcup_{t\in(0,1)} K_t$.

Proof: From the shape of the orbits of the function q and from the formula $f^2 = q = \cosh x - 1$ we immediately realize that for any function $f \in \sqrt{q}^{\,*}$ either $f(0) = 0$ (and then $f(x_2) = f(x_1) = x_1$) or $f(0) = x_1$ (which requires $f(x_2) = f(x_1) = 0$). This also implies that for $x \in \boldsymbol{R} - \{ x_1, x_2, 0\}$ there holds $f(x) \in \boldsymbol{R} - \{ x_1, x_2, 0\}$.

Lemma 2: Let $f \in \sqrt{q}^{\,*}$. Then for every $t \in (0,1)$ and every $x \in K_t$ there holds $f(x) \in \boldsymbol{R} - K_t$.

Proof: The Lemma immediately results from the general theory of mating of orbits (see [16]). None of the considered orbits contains a cycle, so according to the general theory none of them can be self-mateable for any natural number n, $n \geq 2$. Let us show the proof without using this theory.

We have already mentioned the fact that the vertex graph of the function $q(x) = \cosh x - 1$ is the same as the one of the function $p(x) = x^2$ (only the labelling of the elements is different). Therefore we can write $(\boldsymbol{R}, q) \cong (\boldsymbol{R}, p)$. Now we will use this isomorphism and perform the proof for the function p instead for the function q. This will make the notation substantially easier. Without detriment to universality there holds $f \in \sqrt{p}^{\,*}$

Let us assume that there exists a number $x_0 \in K_t$ (for some $t \in (0, 1)$) such that $f(x_0) \in K_t$. Without detriment to universality we can assume that $x_0 > 0$. Then there exists $n \in \boldsymbol{N}$, $n \geq 2$ with the property $f(x_0) = x_0^{2^n}$. The case when $x_0 = [f(x_0)]^{2^m}$ for a suitable $m > 1$ implies $f[f(x_0)] = x_0^2 = [f(x_0)]^{2^{m+1}}$, so it is sufficient to consider the case when $f(x_0)$ is above x_0 in the ordering $\leq_q$, i.e. $f(x_0) = x_0^{2^n}$. Then $f(x_0^{2^n}) = x_0^2$ and there further holds $f(x_0^2) = x_0^{2^{n+1}}$, so then $f(x_0^{2^{2n-1}}) = x_0^{2^n}$. Then we will get $f^2(x_0^{2^{2n-1}}) = f(x_0^{2^n}) = x_0^2 \neq x_0^{2^{2n}} = q(x_0^{2^{2n-1}})$, which is the contradiction with the premise $x_0 \in K_t$, $f(x_0) \in K_t$. So there holds $f(x_0) \notin K_t$.

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

Proof: Let $f \in \sqrt{q}^{\,*}$, $x_0 \in \boldsymbol{R}$ be arbitrary. Evidently, for $x_0 \in \{x_1, x_2, 0\}$ with respect to Lemma 1 there holds the relation $\cosh[f(x_0)] - 1 = f(\cosh x_0 - 1)$. Now let us consider the pair $s, t \in (0,1)$, for which $x_0 \in K_t$, $f(x_0) \in K_s$; according to Lemma 2 $s \neq t$. There applies $f^2(x_0) = \cosh x_0 - 1$, so $\cosh[f(x_0)] - 1 = f^2[f(x_0)] = f[f^2(x_0)] = f(\cosh x_0 - 1)$, therefore $f(\cosh x - 1) = \cosh f(x) - 1$ for every number $x \in \boldsymbol{R}$.

Theorem 2 can be expressed in the following way: Every solution f: $\boldsymbol{R} \rightarrow \boldsymbol{R}$ of the functional equation $f^2(x) = \cosh x - 1$ is the solution of the functional equation $f(\cosh x - 1) = \cosh f(x) - 1$ ·

Now let us describe formally the structure of functions $f \in \sqrt{q}^{\,*}$ .We will show that this structure can only be of two types. The corresponding Theorem 3 will be given without the proof owing to the length of the article (the proof can be found in [8]). However, from the didactic point of view it is interesting that although the whole situation is absolutely evident from diagrams, it requires a quite complicated formal notation. This is one of the difficult aspects of the iteration theory when it is possible to decide relatively easily if the roots exist, but their formal notation can be considerably complicated.

Definition: Let for two components K_i, K_j of the mono-unary algebra $(\boldsymbol{R}, q)$ and for some function $f \in \sqrt{q}^{\,*}$ there hold: $f(x) \in K_j$ a $f(y) \in K_i$ for every pair $(x,y) \in K_i \times K_j$. This pair of components (K_i, K_j) will then be called the *f-pair of components* of the mono-unary algebra $(\boldsymbol{R},q)$ corresponding to the function f.

Let us remark that such situation with two components is enforced because none of the components K_t, $t \in (0,1)$ is a cyclic one, so according to the general theory of the iterative roots none of the components can be self-mateable (the relevant theory is again in [16]). As we discuss the second iterative roots, the only alternative is to mate the infinite orbits in pairs.

Definition: Let (K_i, K_j) be the f-pair of components of the mono-unary algebra $(\boldsymbol{R}, q)$ corresponding to the function f, $f \in \sqrt{q}^{\,*}$. Then the function $f\big| K_i = f_{(i,j)}$: $K_i \to K_j$ and the function $f\big| K_j = f_{(j,i)}$: $K_j \to K_i$ will be called *connective functions* with respect to the f-pair of components (K_i, K_j).

It is evident that any component (K, f_K) of the mono-unary algebra $(\boldsymbol{R}, f)$ can be constructed as follows: $K = K_i \cup K_j$, where (K_i, K_j) is the f-pair of components corresponding to the function $f \in \sqrt{q}^{\,*}$, $f_K = f_{(j,i)} \cup f_{(i,j)}$, where $f_{(i,j)}$, $f_{(j,i)}$ are connective functions with respect to the f-pair of components (K_i, K_j).

Theorem 3: Let us denote $P = \boldsymbol{R} - \{x_1, x_2, 0\}$ (the elements of the set $\{x_1, x_2, 0\}$ are of the same meaning as in Theorem 1; they denote the elements of finite components of the vertex graph of the function $q(x) = \cosh x - 1$). Let $f \in \sqrt{q}^{\,*}$. Let (K, f_K) be any component of the mono-unary algebra $(\boldsymbol{R}, f)$, $K \subset P$. Then there exits the f-pair (K_i, K_j) of the components of the mono-unary algebra $(\boldsymbol{R},q)$, for which either f_K is an even non-negative function or $f_K = f_{(i,j)} \cup f_{(j,i)}$ (under a convenient choice of indexes i, j), where $f_{(i,j)}$ is an odd connective function and $f_{(j,i)}$ is an even connective function with respect to the f-pair of components (K_i, K_j).

Proof: Can be found in [8].

Remark: Both possible cases of the construction of the second iterative roots of the function q for infinite components K_t are depicted in the following Figure 7. On the left there is the case when f_K is an even 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.

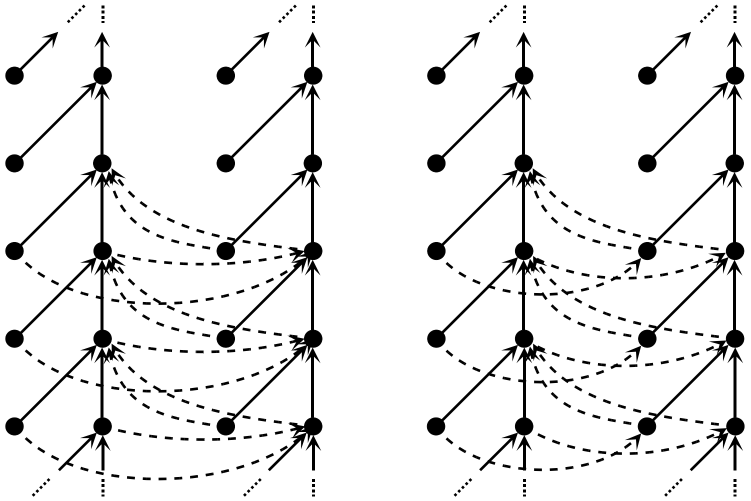

Fig. 7

CONCLUSION

The article gives the discrete description of two hyperbolic functions: the sine and the cosine. This non-traditional approach to these functions with the usage of their vertex graphs enables both the efficient formal description of the construction of their iterative roots, and the formal description of the second iterative root of the hyperbolic cosine. The text points out the fact that, while using the discrete approach to functions, it is possible to solve problems which, while using the classical continuous approach, would be solved with great difficulties (e.g. some types of equations of one variable). The choice of hyperbolic functions was not random as well; it demonstrates that these two functions, neglected while teaching mathematics at universities, offer a wide spectrum of topics for students´ individual projects and a great number of unexpected connections. Thus, they enable the necessary synthesizing approach to mathematics, which has already been pointed out in the introduction. In the conclusion, let us mention that the iterative theory of functions itself is by far more extensive and contains many more and deeper results and applications to the practice (e.g. in the area of numerical methods). Those interested in this theory could find further information in e.g. [10], [14], [15], [16], [17], [18], [19], [20].

REFERENCES

[1] Bartsch, H. J.: *Matematické vzorce*. SNTL, Praha 1987.

[2] Beránek. J., Chvalina, J.: *On Tabor`s problem concerning a certain quasi-ordering of iterative roots of functions*. Aequ. Math. 39 (1990), pp. 1–5.

[3] Beránek, J.: *O iterativních kořenech jisté polynomické funkce.* In: Sborník prací Pedagogické fakulty MU, no. 133, MU Brno, Brno 1993, pp. 5–15.

[4] Beránek, J.: *Přístup k vysokoškolské výuce základních matematických poznatků z hlediska teorie orientovaných grafů a speciálních algeber.* In: Řízení osvojovacího procesu XII. Proceedings of Abstracts from the XIIth Scientific Colloquium, VVŠ PV, Vyškov 1994, pp. 73–80.

[5] Beránek, J.: *K problematice výuky hyperbolických funkcí*. In: Řízení osvojovacího procesu XIII. Proceedings of Abstracts from the XIIIth Scientific Colloquium, VVŠ PV, Vyškov 1995, pp. 48–57.

[6] Coufalová, Y., Francová, M., Chvalina, J.: *Konjugace zobrazení a funkcí.* In: Sborník prací Pedagogické fakulty MU, MU Brno, Brno 1993, pp. 31–47.

[7] Hejný, M. et al.: *Teória vyučovania matematiky 2.* SPN, Bratislava 1990.

[8] Chvalina, J.,Beránek, J.: *O iteračních odmocninách kvadratické funkce.* In: Sborník prací pedagogické fakulty UJEP, UJEP Brno, Brno 1990, pp.7–19.

[9] Chvalina, J.:*Diskrétní orbitální struktura zobrazení a funkcí.* In:Diskrétní matematika. Sborník VIII. brněnské konference o vyučování matematice, 1992. JČMF, Brno 1992, pp. 23–28.

[10] Chvalina, J.: *Funkcionální grafy, kvaziuspořádané množiny a komutativní hypergrupy.* Masarykova Univerzita, Brno 1995, pg. 205.

[11] Isaacs, R.: *Iterates of fractional order*. Canad. J. Math. 2 (1950), pp. 409-416.

[12] Novotný, M.: *Sur un problème de la théorie des applications.* Publ. Fac. Sci. Univ. Masaryk 344 (1953), pp. 53–64.

[13] Skornjakov, L.A.: *Unars*. In: Colloq. Math. Soc. János Bolyai 29, Univ.Algebra, Esztergom 1977, pp. 735-743.

[14] Smítal, J.: *O funkciách a funkcionálnych rovniciach.* Alfa, Bratislava 1984.

[15] Snowden, M., Howie, J. M.: *Square roots in finite full transformation semigroups*. Glasgow Math. J. **23** (1982), no. 2, pp. 137–149.

[16] Targonski, G.: *Topics in Iteration Theory.* Vandenhoeck et Ruprecht, Göttingen and Zürich 1981

[17] Targoński, G.: *New directions and open problems in iteration theory*. Berichte [Reports], 229. Forschungszentrum Graz, Mathematisch-Statistische Sektion, Graz, 1984. 51 pp.

[18] Targoński, G.: *Iteration theory and its functional equations.* Rend. Sem. Mat. Fis. Milano 58 (1988), pp. 207–220 (1990).

[19] Zdun, M., C.: *The structure of iteration groups of continuous functions*. Aequationes Math. 46 (1993), no. 1-2, pp. 19–37.

[20] Zdun, M., C., Solarz, P.: *Recent results on iteration theory: iteration groups and semigroups in the real case*. Aequationes Math. 87 (2014), no. 3, pp. 201–245.

Formula for explicit solutions of a class of linear discrete equations with delay

Author J. Diblík, K. Mencáková

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

Abstract: In the paper there is given explicit formula for solutions of linear systems of discrete equations with delay. Used method is based on a transformation of given system to a system without delay.

Keywords: Discrete system, delay, linear system.

Introduction

We consider a system of discrete equations with delay of the form

$$\Delta x(k) = Bx(k-1), \tag{1}$$

where $k \geq 0$, $B = (b_{ij})_{i,j=1}^{2}$ is a constant matrix and $x(k) = (x_1(k), x_2(k))^T$ is an unknown vector.

We suggest a method of solution based on a transformation of system (1) to a system without delay. Systems of this form are often used in the theory of digital filters ([1] – [3]).

1 Transformation of the system

Consider a system with delay (1). Since $\Delta x(k) = x(k+1) - x(k)$, $k \geq 0$, the system (1) is equivalent with system

$$x(k+1) = x(k) + Bx(k-1). \tag{2}$$

Define a new unknown vectors $v_1(k)$, $v_2(k)$, $k \geq 0$, by formulas

$$v_1(k) = x(k-1), \tag{3}$$

$$v_2(k) = x(k). \tag{4}$$

Then $v_1(k) = v_2(k-1)$, $v_2(k+1) = v_2(k) + Bv_1(k)$.

Moreover, define a 4-dimensional vector

$$v(k) = (v_1(k), v_2(k))^T. \tag{5}$$

Then

$$v(k+1) = (v_1(k+1), v_2(k+1))^T = (v_2(k), v_2(k) + Bv_1(k))^T$$

and system (2) can be transformed into a system without delay

$$v(k+1) = \begin{pmatrix} \Theta & E \\ B & E \end{pmatrix} v(k), \quad k \geq 0 \tag{6}$$

where E is a 2×2 unit matrix and Θ is a 2×2 null matrix.

Let a matrix A be defined as

$$A = \begin{pmatrix} \Theta & E \\ B & E \end{pmatrix}. \tag{7}$$

Then system (6) can be written as

$$v(k+1) = Av(k), \quad k \geq 0. \tag{8}$$

Consider an initial condition of system (8)

$$v(0) = v_0,$$

where v_0 is an initial value for dependent vector v. Then, by the well-known formula for the solution of non-delayed linear systems with constant matrices [4], we have

$$v(k) = A^k v(0), \quad k \geq 0. \tag{9}$$

2 Formula for A^k, $k \geq 0$

It is important to give a recommendation to find powers of matrices A in system (9). The following theorem gives formulas to compute povers A^k through powers of the matrix B.

Theorem 1. *For powers of the matrix A, given by (7) it holds*

$$A^k = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix}, \quad k \geq 0, \tag{10}$$

where a_k, b_k, c_k, d_k are 2×2 matrices, $a_0 = E$ and

$$a_k = \frac{B}{2^{k-2}} \sum_{i=0}^{\lfloor (k-2)/2 \rfloor} \binom{k-1}{2i+1} (E+4B)^i, \quad k \geq 1, \tag{11}$$

$$b_k = \frac{1}{2^{k-1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i, \quad k \geq 0, \tag{12}$$

$$c_k = \frac{B}{2^{k-1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i, \quad k \geq 0, \tag{13}$$

$$d_k = \frac{1}{2^{k}} \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i, \quad k \geq 0. \tag{14}$$

where $\lfloor \cdot \rfloor$ is the floor function.

Proof. We use method of mathematical induction.

I. Let $k = 0$. Then

$A^0 = \begin{pmatrix} a_0 & b_0 \\ c_0 & d_0 \end{pmatrix} =$

$$= \begin{pmatrix} a_0 & \frac{1}{2^{-1}} \sum_{i=0}^{-1} \binom{0}{2i+1} (E+4B)^i \\ \frac{B}{2^{-1}} \sum_{i=0}^{-1} \binom{0}{2i+1} (E+4B)^i & \frac{1}{2^{0}} \sum_{i=0}^{0} \binom{1}{2i+1} (E+4B)^i \end{pmatrix} = \begin{pmatrix} E & \Theta \\ \Theta & E \end{pmatrix}.$$

II. Suppose the formula (10) holds for some $k \geq 0$. We show then it holds also for $k+1$.
We use the relation

$$A^{k+1} = A^k \cdot A = A \cdot A^k,$$

i.e.

$$\begin{pmatrix} a_{k+1} & b_{k+1} \\ c_{k+1} & d_{k+1} \end{pmatrix} = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix} \cdot \begin{pmatrix} \Theta & E \\ B & E \end{pmatrix}.$$

The rest of the proof is divided into four parts by formulas (11 – 14) for terms a_k, b_k, c_k, d_k of matrix A^k. Then

II. a) $a_{k+1} = b_k \cdot B = B \cdot b_k = \dfrac{B}{2^{k-1}} \displaystyle\sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i,$

so the formula (11) with $k+1$ holds.

II. b) $b_{k+1} = a_k + b_k$,

we substitute formulas (11), (12) for expressions on each sides and in the following we will modify the whole equation by simple operations without mentioning. So we obtain

$$\begin{aligned}
&\frac{1}{2^k}\sum_{i=0}^{\lfloor k/2\rfloor}\binom{k+1}{2i+1}(E+4B)^i \\
&\quad= \frac{B}{2^{k-2}}\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^i + \frac{1}{2^{k-1}}\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i,
\end{aligned}$$

$$\begin{aligned}
\frac{1}{2^k}\sum_{i=0}^{\lfloor k/2\rfloor}\binom{k+1}{2i+1}(E+4B)^i &= \frac{(4B+E)-E}{4\cdot 2^{k-2}}\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^i \\
&\quad+\frac{1}{2^{k-1}}\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i,
\end{aligned}$$

$$\begin{aligned}
\frac{1}{2^k}\sum_{i=0}^{\lfloor k/2\rfloor}\binom{k+1}{2i+1}(E+4B)^i &= \frac{1}{2^k}\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^{i+1} \\
&\quad-\frac{1}{2^k}\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^i + \frac{1}{2^{k-1}}\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i.
\end{aligned}$$

Multiplying the both sides by 2^k we get

$$\begin{aligned}
\sum_{i=0}^{\lfloor k/2\rfloor}\binom{k+1}{2i+1}(E+4B)^i &= \sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^{i+1} \\
&\quad-\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^i + 2\cdot\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i,
\end{aligned}$$

$$\sum_{i=0}^{\lfloor k/2\rfloor}\binom{k+1}{2i+1}(E+4B)^i-\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i$$
$$=\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^{i+1}+\sum_{i=0}^{\lfloor (k-1)/2\rfloor}\binom{k}{2i+1}(E+4B)^i$$
$$-\sum_{i=0}^{\lfloor (k-2)/2\rfloor}\binom{k-1}{2i+1}(E+4B)^i. \quad (15)$$

To verify the last formula, we consider two cases: either k is even nor odd.

For even k we get from (15)

$$\sum_{i=0}^{k/2}\binom{k+1}{2i+1}(E+4B)^i-\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^i$$
$$=\sum_{i=0}^{k/2-1}\binom{k-1}{2i+1}(E+4B)^{i+1}+\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^i$$
$$-\sum_{i=0}^{k/2-1}\binom{k-1}{2i+1}(E+4B)^i,$$

$$\sum_{i=0}^{k/2-1}\binom{k+1}{2i+1}(E+4B)^i+\binom{k+1}{k+1}(E+4B)^{k/2}$$
$$-\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^i=\sum_{i=0}^{k/2-1}\binom{k-1}{2i+1}(E+4B)^{i+1}$$
$$+\sum_{i=0}^{k/2-1}\left[\binom{k}{2i+1}-\binom{k-1}{2i+1}\right](E+4B)^i.$$

Utilising formula

$$\binom{m+1}{l}-\binom{m}{l}=\binom{m}{l-1} \quad (16)$$

we can transform the last expression into

$$\sum_{i=0}^{k/2-1}\left[\binom{k+1}{2i+1}-\binom{k}{2i+1}\right](E+4B)^i+(E+4B)^{k/2}$$

$$= \sum_{i=0}^{k/2-1} \binom{k-1}{2i+1}(E+4B)^{i+1} + \sum_{i=0}^{k/2-1} \binom{k-1}{2i}(E+4B)^{i},$$

$$\sum_{i=0}^{k/2-1} \binom{k}{2i}(E+4B)^{i} - \sum_{i=0}^{k/2-1} \binom{k-1}{2i}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=0}^{k/2-1} \binom{k-1}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{k/2-1} \left[\binom{k}{2i} - \binom{k-1}{2i} \right] (E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=0}^{k/2-1} \binom{k-1}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=0}^{k/2-1} \binom{k-1}{2i+1}(E+4B)^{i+1}.$$

If $i = 0$ the first combinative number on the left-hand side equals zero. Therefore we can write

$$\sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=0}^{k/2-1} \binom{k-1}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=1}^{k/2} \binom{k-1}{2i-1}(E+4B)^{i},$$

$$\sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + \binom{k-1}{k-1}(E+4B)^{k/2}$$

and finally the following parity obviously holds

$$\sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2} = \sum_{i=1}^{k/2-1} \binom{k-1}{2i-1}(E+4B)^{i} + (E+4B)^{k/2}.$$

For k odd we have from (15)

$$\begin{aligned}\sum_{i=0}^{(k-1)/2}\binom{k+1}{2i+1}(E+4B)^i-\sum_{i=0}^{(k-1)/2}\binom{k}{2i+1}(E+4B)^i&\\ =\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^{i+1}+\sum_{i=0}^{(k-1)/2}\binom{k}{2i+1}(E+4B)^i&\\ -\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^i,&\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^{(k-1)/2}\left[\binom{k+1}{2i+1}-\binom{k}{2i+1}\right](E+4B)^i&\\ =\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^{i+1}+\sum_{i=0}^{(k-3)/2}\binom{k}{2i+1}(E+4B)^i&\\ +\binom{k}{k}(E+4B)^{(k-1)/2}-\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^i,&\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^{(k-1)/2}\binom{k}{2i}(E+4B)^i=\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^{i+1}&\\ +\sum_{i=0}^{(k-3)/2}\left[\binom{k}{2i+1}-\binom{k-1}{2i+1}\right](E+4B)^i+(E+4B)^{(k-1)/2},&\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^{(k-1)/2}\binom{k}{2i}(E+4B)^i=\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^{i+1}&\\ +\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i}(E+4B)^i+(E+4B)^{(k-1)/2},&\end{aligned}$$

$$\begin{aligned}\sum_{i=0}^{(k-3)/2}\binom{k}{2i}(E+4B)^i+\binom{k}{k-1}(E+4B)^{(k-1)/2}-\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i}(E+4B)^i&\\ -(E+4B)^{(k-1)/2}=\sum_{i=0}^{(k-3)/2}\binom{k-1}{2i+1}(E+4B)^{i+1},&\end{aligned}$$

$$\sum_{i=0}^{(k-3)/2}\left[\binom{k}{2i}-\binom{k-1}{2i}\right](E+4B)^i+k\cdot(E+4B)^{(k-1)/2}$$

$$-(E+4B)^{(k-1)/2} = \sum_{i=0}^{(k-3)/2} \binom{k-1}{2i+1}(E+4B)^{i+1},$$

and finally

$$\sum_{i=0}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2}$$
$$= \sum_{i=0}^{(k-3)/2} \binom{k-1}{2i+1}(E+4B)^{i+1}.$$

If $i = 0$ the first combinative number on the left-hand side is again equal to zero. Therefore

$$\sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2}$$
$$= \sum_{i=0}^{(k-3)/2} \binom{k-1}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2}$$
$$= \sum_{i=1}^{(k-1)/2} \binom{k-1}{2i-1}(E+4B)^{i},$$

$$\sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2}$$
$$= \sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + \binom{k-1}{k-2}(E+4B)^{(k-1)/2},$$

$$\sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2}$$
$$= \sum_{i=1}^{(k-3)/2} \binom{k-1}{2i-1}(E+4B)^{i} + (k-1)(E+4B)^{(k-1)/2},$$

and the formula (12) with $k+1$ holds.

II. c) $c_{k+1} = d_k \cdot B = B \cdot d_k = \dfrac{B}{2^k} \displaystyle\sum_{i=0}^{\lfloor (k)/2 \rfloor} \binom{k+1}{2i+1}(E+4B)^{i},$

so the formula (13) with $k+1$ holds.

II. d) $d_{k+1} = c_k + d_k$,

we substitute formulas (13), (14) and we will again modify the whole equation without mentioning of simple steps.

$$\frac{1}{2^{k+1}} \sum_{i=0}^{\lfloor (k+1)/2 \rfloor} \binom{k+2}{2i+1} (E+4B)^i$$
$$= \frac{B}{2^{k-1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i + \frac{1}{2^k} \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i,$$

$$\frac{1}{2^{k+1}} \sum_{i=0}^{\lfloor (k+1)/2 \rfloor} \binom{k+2}{2i+1} (E+4B)^i$$
$$= \frac{(4B+E)-E}{4 \cdot 2^{k-1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i + \frac{1}{2^k} \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i,$$

$$\frac{1}{2^{k+1}} \sum_{i=0}^{\lfloor (k+1)/2 \rfloor} \binom{k+2}{2i+1} (E+4B)^i = \frac{1}{2^{k+1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^{i+1}$$
$$- \frac{1}{2^{k+1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i + \frac{1}{2^k} \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i.$$

Multiplying the both sides by 2^{k+1} we get

$$\sum_{i=0}^{\lfloor (k+1)/2 \rfloor} \binom{k+2}{2i+1} (E+4B)^i = \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^{i+1}$$
$$- \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i + 2 \cdot \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i,$$

$$\sum_{i=0}^{\lfloor (k+1)/2 \rfloor} \binom{k+2}{2i+1}(E+4B)^i - \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1}(E+4B)^i$$
$$= \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1}(E+4B)^{i+1} + \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1}(E+4B)^i$$
$$- \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1}(E+4B)^i. \quad (17)$$

For even k the previous parity (17) is

$$\sum_{i=0}^{k/2} \binom{k+2}{2i+1}(E+4B)^i - \sum_{i=0}^{k/2} \binom{k+1}{2i+1}(E+4B)^i$$
$$= \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^{i+1} + \sum_{i=0}^{k/2} \binom{k+1}{2i+1}(E+4B)^i$$
$$- \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^i,$$

$$\sum_{i=0}^{k/2} \left[\binom{k+2}{2i+1} - \binom{k+1}{2i+1} \right] (E+4B)^i = \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^{i+1}$$
$$+ \sum_{i=0}^{k/2} \binom{k+1}{2i+1}(E+4B)^i - \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^i.$$

Utilising the form (16) we get

$$\sum_{i=0}^{k/2} \binom{k+1}{2i}(E+4B)^i = \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^{i+1}$$
$$+ \sum_{i=0}^{k/2-1} \binom{k+1}{2i+1}(E+4B)^i + \binom{k+1}{k+1}(E+4B)^{k/2}$$
$$- \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^i,$$

$$\sum_{i=0}^{k/2} \binom{k+1}{2i}(E+4B)^i = \sum_{i=0}^{k/2-1} \binom{k}{2i+1}(E+4B)^{i+1}$$

$$+\sum_{i=0}^{k/2-1}\left[\binom{k+1}{2i+1}-\binom{k}{2i+1}\right](E+4B)^{i}+1\cdot(E+4B)^{k/2},$$

$$\sum_{i=0}^{k/2}\binom{k+1}{2i}(E+4B)^{i}$$
$$=\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^{i+1}+\sum_{i=0}^{k/2-1}\binom{k}{2i}(E+4B)^{i}+(E+4B)^{k/2},$$

$$\sum_{i=0}^{k/2-1}\binom{k+1}{2i}(E+4B)^{i}+\binom{k+1}{k}(E+4B)^{k/2}-\sum_{i=0}^{k/2-1}\binom{k}{2i}(E+4B)^{i}$$
$$-(E+4B)^{k/2}=\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{k/2-1}\left[\binom{k+1}{2i}-\binom{k}{2i}\right](E+4B)^{i}+(k+1)(E+4B)^{k/2}$$
$$-(E+4B)^{k/2}=\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+k\cdot(E+4B)^{k/2}=\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^{i+1}.$$

If $i=0$ the first combinative number on the left-hand side is again equal to zero.

$$\sum_{i=1}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+k\cdot(E+4B)^{k/2}=\sum_{i=0}^{k/2-1}\binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=1}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+k\cdot(E+4B)^{k/2}=\sum_{i=1}^{k/2}\binom{k}{2i-1}(E+4B)^{i},$$

$$\sum_{i=1}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+k\cdot(E+4B)^{k/2}$$
$$=\sum_{i=1}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+\binom{k}{k-1}(E+4B)^{k/2},$$

$$\sum_{i=1}^{k/2-1}\binom{k}{2i-1}(E+4B)^{i}+k\cdot(E+4B)^{k/2}$$

$$= \sum_{i=1}^{k/2-1} \binom{k}{2i-1}(E+4B)^i + k \cdot (E+4B)^{k/2}$$

and it holds.

For odd k the parity (17) has form

$$\sum_{i=0}^{(k+1)/2} \binom{k+2}{2i+1}(E+4B)^i - \sum_{i=0}^{(k-1)/2} \binom{k+1}{2i+1}(E+4B)^i$$
$$= \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^{i+1} + \sum_{i=0}^{(k-1)/2} \binom{k+1}{2i+1}(E+4B)^i$$
$$- \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^i,$$

$$\sum_{i=0}^{(k-1)/2} \binom{k+2}{2i+1}(E+4B)^i + \binom{k+2}{k+2}(E+4B)^{(k+1)/2}$$
$$- \sum_{i=0}^{(k-1)/2} \binom{k+1}{2i+1}(E+4B)^i = \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^{i+1}$$
$$+ \sum_{i=0}^{(k-1)/2} \left[\binom{k+1}{2i+1} - \binom{k}{2i+1} \right] (E+4B)^i,$$

$$\sum_{i=0}^{(k-1)/2} \left[\binom{k+2}{2i+1} - \binom{k+1}{2i+1} \right] (E+4B)^i + 1 \cdot (E+4B)^{(k+1)/2}$$
$$= \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^{i+1} + \sum_{i=0}^{(k-1)/2} \binom{k}{2i}(E+4B)^i,$$

$$\sum_{i=0}^{(k-1)/2} \binom{k+1}{2i}(E+4B)^i - \sum_{i=0}^{(k-1)/2} \binom{k}{2i}(E+4B)^i + (E+4B)^{(k+1)/2}$$
$$= \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{(k-1)/2} \left[\binom{k+1}{2i} - \binom{k}{2i} \right] (E+4B)^i + (E+4B)^{(k+1)/2}$$
$$= \sum_{i=0}^{(k-1)/2} \binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=0}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2}=\sum_{i=0}^{(k-1)/2}\binom{k}{2i+1}(E+4B)^{i+1}.$$

If $i=0$ the first combinative number on the left-hand side is again equal to zero.

$$\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2}=\sum_{i=0}^{(k-1)/2}\binom{k}{2i+1}(E+4B)^{i+1},$$

$$\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2}=\sum_{i=1}^{(k+1)/2}\binom{k}{2i-1}(E+4B)^{i},$$

$$\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2}$$
$$=\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+\binom{k}{k}(E+4B)^{(k+1)/2},$$

$$\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2}$$
$$=\sum_{i=1}^{(k-1)/2}\binom{k}{2i-1}(E+4B)^i+(E+4B)^{(k+1)/2},$$

the formula (13) with $k+1$ holds and the theorem is proved. □

3 Solution of the system (1)

Let an initial value of system (1) be given:

$$x(0)=x_0, x(-1)=x_{-1}, \tag{18}$$

where $x_0=(x_{01},x_{02})^T$, $x_{-1}=(x_{-10},x_{-11})^T$ are 2-dimensional constant vectors.

Then the relevant initial value of transformed system (8) is, by (3) – (5),

$$v(0)=v_0=(x_{-1},x_0)^T=(x_{-10},x_{-11},x_{01},x_{02})^T. \tag{19}$$

The solution of initial problem (8), (19) is given by formula (9)

$$v(k)=A^kv_0=A^k(x_{-1},x_0)^T, \quad k\geq 0. \tag{20}$$

Since

$$v(k)=(x_1(k-1),x_2(k-1),x_1(k),x_2(k))^T, \quad k\geq 0,$$

the solution $x(k) = (x_1(k), x_2(k))^T$ of system (1) satisfying initial data (18) can be derived by separating the last two rows from (20).

It is easy to see from (20) that

$$x(k) = c_k x_{-1} + d_k x_0, \quad k \geq 0,$$

and the following theorem holds.

Theorem 2. *Solution of the initial-value problem* (1)*,* (18) *is given by formula*

$$\begin{aligned} x(k) = & \left(\frac{B}{2^{k-1}} \sum_{i=0}^{\lfloor (k-1)/2 \rfloor} \binom{k}{2i+1} (E+4B)^i \right) \cdot x_{-1} \\ + & \left(\frac{1}{2^k} \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k+1}{2i+1} (E+4B)^i \right) \cdot x_0, \quad k \geq 0. \end{aligned}$$

4 Conclusion

In the paper we solved an initial problem of system (1) with a single delay. The original system is transformed into a system without delay. The solution was found by the well-known formula, but we derived exact formulas for powers of the defined matrix of linear terms. Thanks to this, it was possible to express explicitely the solution of the initial problem.

Acknowledgement

The authors were supported by Grant FEKT-S-14-2200 of Faculty of Electrical Engineering and Communication, BUT.

References

[1] Vích, R., Smékal, Z.: Číslicové filtry, Praha, Academia 2000, ISBN 80-200-0761-X.

[2] Proakis, J. G., Manolakis, D. G.: Digital Signal Processing, Fourth edition, New Jersey, Prentice Hall 2006, ISBN 978-0131873742.

[3] Oppenheim A. V., Schafer R. W., Buck J. R.: Discrete-Time Signal Processing, New Jersey, Prentice Hall 2009, ISBN 0-13-754920-2.

[4] Elaydi, S. N.: An Introduction to Difference Equations, Third edition, Springer 2005, ISBN 978-0-387-23059-7.

Properties of counting function of pseudoprimes

Alexander Maťašovský, Tomáš Visnyai

Faculty of Chemical and Food Technology STU in Bratislava,
Radlinského 9, 812 37 Bratislava 1, Slovak Republic.
Email: matasovsky@stuba.sk, visnyai@stuba.sk

Abstract: The aim of this article is present some sufficient and necessary conditions to convergence of subseries of the harmonic series along any subset of positive integers.

Keywords: convergence of series, counting function, pseudoprimes

Introduction

In this article we will investigate the counting functions. Denote $A(x)$ the number of elements $n \in A \subseteq \mathbb{N}$, where $n \leq x$. We show the join between some properties of $A(x)$ and the series $\sum_{n\in A} n^{-1}$. Then we show the relationship between $A(x)$ and asymptotic density of the set A. Finally we apply the results for counting function of the set of all pseudoprime numbers to base 2.

1 Basic Notions and Definitions

Let $A = \{a_1 < a_2 < \cdots < a_n < \cdots\} \subseteq \mathbb{N} = \{1, 2, \dots\}$. Let $A(x)$ denote the number of elements of the set A less then or equal to $x \in \mathbb{N}$ i.e.

$$A(x) = |\{n \in \mathbb{N} \; : \; n \leq x\}|.$$

The lower asymptotic density of the set $A \subseteq \mathbb{N}$ is defined by $\underline{d}(A) = \liminf_{x\to\infty} \frac{A(x)}{x}$ and the upper asymptotic density is defined by $\overline{d}(A) = \limsup_{x\to\infty} \frac{A(x)}{x}$. If $\underline{d}(A) = \overline{d}(A) = d(A)$, then $d(A)$ is called the asymptotic density of the set A. It is clear that $d(A) \in \langle 0, 1\rangle$. For example $d(\mathbb{N}) = 1$, $d(\mathbb{P}) = 0$, $d(S) = \frac{6}{\pi^2}$ and $d(\mathbb{N}^2) = 0$, where $\mathbb{P}$ denotes the set of all prime numbers, S denotes the set of all square free numbers and $\mathbb{N}^2 = \{1^1, 2^2, 3^2, \dots\}$ (see [1],[6] and [7]).

Recall the definition of symbol O. Let f, g are functions defined on $\mathbb{R}$. We have

$$f(x) = O(g(x)) \Leftrightarrow \exists M \in \mathbb{R} \; : \; \frac{f(x)}{g(x)} \leq M$$

or

$$\limsup_{x\to\infty} \frac{f(x)}{g(x)} < +\infty.$$

Next we denote

$$f(x) \sim g(x) \Leftrightarrow \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

The notion of pseudoprime number is defined as follows. The composite number n is called pseudoprime number on base a if

$$a^n \equiv a \pmod{n}.$$

For example the integer 341 is preudoprime number on base 2 (see [2]), because $341 = 31 \cdot 11$ and $a = 2$ then

$$\begin{aligned} 2^{10} &\equiv 1 \pmod{341} \\ 2^{340} &\equiv 1 \pmod{341} \\ 2^{341} &\equiv 2 \pmod{341} \end{aligned}$$

This number was found in 1819. It can be shown that the set of all pseudoprime numbers is infinite. Paul Erdős in 1950 for the counting function proved the following inequality (see [3])

$$P_2(x) < 2x \exp\left\{-\frac{1}{3}\log^{\frac{1}{4}} x\right\}.$$

2 Main Results

Let us recall the relationship between the convergence of series of inverse values from A and asymptotic density of the set A. We need the following Lemma.

Lemma 1. *Let* $a_1 \geq a_2 \geq \cdots a_n \geq \cdots$ *is a sequence of real numbers. Let* $\lim_{n\to\infty} a_n = 0$ *and* α_n $(n = 1, 2, \dots)$ *are non-negative real numbers. If the series* $\sum_{n=1}^{\infty} \alpha_n a_n$ *converges then* $\lim_{n\to\infty}(\alpha_1 + \alpha_2 + \cdots + \alpha_n)a_n = 0$.

Proof. By using the Cauchy-Bolzano convergence criterion i.e. for all $\varepsilon > 0$ there exists $m \in \mathbb{N}$ that for all $n > m$ the inequality $\alpha_{m+1}a_{m+1} + \cdots + \alpha_n a_n < \varepsilon$ holds. From monotonicity of $(a_n)^{\infty}$ we have

$$a_n(\alpha_{m+1} + \cdots + \alpha_n) \leq \alpha_{m+1}a_{m+1} + \cdots + \alpha_n a_n < \varepsilon$$

for $n > m$. Hence $\lim_{n\to\infty} a_n(\alpha_{m+1} + \cdots + \alpha_n) = 0$ for fixed m. Therefore $\lim_{n\to\infty}(\alpha_1 + \cdots + \alpha_n)a_n = 0$. □

Now we present the sufficient condition to convergence of subseries of harmonic series along a set A.

Theorem 2. *Let* $A = \{n_1 < n_2 < \cdots < n_k < \cdots\}$. *If* $\sum_{k=1}^{\infty} n_k^{-1} < \infty$ *then* $d(A) = 0$.

Proof. Set the $\alpha_{n_k} = 1$ for $k = 1, 2, \dots$ and $\alpha_m = 0$ for $m \neq n_k$. On the base of Lemma 1 $\lim_{n\to\infty} \frac{\alpha_1+\alpha_2+\cdots+\alpha_n}{n} = 0$ and $A(n) = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. Then we have $\lim_{n\to\infty} \frac{A(n)}{n} = 0$ i.e. $d(A) = 0$. □

Remark 3. The converse of Theorem 2 is not true. For instance the set of all prime numbers has asymptotic density equals to 0, but $\sum_{k=1}^{\infty} p_k^{-1} = +\infty$, where $\mathbb{P} = \{p_1 < p_2 < \cdots < p_k < \cdots\}$.

Theorem 4. *If* $A(x) = O\left(\frac{x}{\log^{\alpha} x}\right)$, $\alpha > 1$, *then* $\sum_{a\in A} a^{-1} < +\infty$.

Proof. Let $A = \{a_1 < a_2 < \cdots < a_n < \cdots\}$. By assumption there exist $M > 0$ such that $A(x) \leq M \cdot \frac{x}{\log^{\alpha} x}$. Put $x = a_n$ then $A(x) = n \leq M \cdot \frac{x}{\log^{\alpha} x}$. Hence $\frac{1}{a_n} \leq M \cdot \frac{1}{n \log^{\alpha} a_n}$. Because $a_n \geq n$ $(n = 1, 2, \dots)$ is $\log^{\alpha} a_n \geq \log^{\alpha} n$ for $n > n_0$, therefore

$$\frac{1}{a_n} \leq M \cdot \frac{1}{n \log^{\alpha} n}.$$

The sequence $\sum_{n=2}^{\infty} \frac{1}{n \log^{\alpha} n}$ converges for $\alpha > 1$. Hence the series $\sum_{n=1}^{\infty} a_n^{-1}$ also converges. □

From Theorems 2 and 4 we have the following corollary.

Corollary 5. *If* $A(x) = O\left(\frac{x}{\log^{\alpha} x}\right)$, $\alpha > 1$, *then* $d(A) = 0$, *where* $A = \{a_1 < \cdots < a_n < \cdots\}$.

We show that the converse of Theorem 4 is not true.

Theorem 6. *Let* $\alpha > 1$ *then there exists a set* $A = \{a_1 < \cdots < a_n < \cdots\}$ *such that* $\sum_{n=1}^{\infty} a_n^{-1} < +\infty$ *and* $A(x) \neq O\left(\frac{x}{\log^{\alpha} x}\right)$.

Proof. Let $\beta \in \mathbb{R}^+$, such that $1 < \beta < \alpha$. Create a set A in the following way: $A = \cup_{k=1}^{\infty} A_k$ where

$$A_k = \{2^k + 1, 2^k + 2, \dots, 2^k + [t_k 2^k]\}$$

and $t_k = \frac{1}{k^\beta}$ for $k = 1, 2, \dots$. Since $\beta > 1$ and

$$\sum_{j \in A_k} j^{-1} \leq \frac{1}{2^k}[t_k 2^k] \leq t_k = \frac{1}{k^\beta}$$

the series $\sum_{j \in A} j^{-1}$ converges. Next we show that $A(x) \neq O\left(\frac{x}{\log^\alpha x}\right)$. Suppose that $A(x) = O\left(\frac{x}{\log^\alpha x}\right)$. Then there exists $M > 0$ such that $A(x) \leq M \cdot \frac{x}{\log^\alpha x}$. Specially for $x = 2^k + [t_k 2^k]$ we have

$$\sum_{j=1}^{k} [t_j 2^j] \leq M \cdot \frac{2^k + [t_k 2^k]}{\log^\alpha (2^k + [t_k 2^k])}$$

therefore

$$\sum_{j=1}^{k} [t_j 2^j] \leq M \cdot \frac{2^k + [t_k 2^k]}{\log^\alpha (2^k(1 + t_k))}.$$

The left side is greater then $t_k 2^k \log^\alpha(2^k(1 + t_k))$ hence

$$\begin{aligned} k^\alpha (\log^\alpha 2) t_k 2^k &\leq M 2^k (1 + t_k) \\ (\log^\alpha 2)\frac{k^\alpha}{k^\beta} &\leq M(1 + \frac{1}{k^\beta}) \\ (\log^\alpha 2) k^\alpha &\leq M(k^\beta + 1). \end{aligned}$$

The last inequality does not hold for $k \to \infty$, because by assumption $\alpha > \beta$. □

In the next we will investigate the counting function on base 2. The Erdős's equality says that (cf. [3])

$$P_2(x) = O\left(x e^{-\frac{1}{3}\sqrt[4]{\log x}}\right), \tag{1}$$

where $P_2(x)$ denotes the number of all pseudoprimes to base 2 less then or equal to x. We will show that

$$P_2(x) = O\left(\frac{x}{\log^\alpha x}\right), \quad \alpha > 0. \tag{2}$$

From (1) exists $K > 0$ such that $P_2(x) \leq K \cdot \frac{x}{e^{\frac{1}{3}\sqrt[4]{\log x}}}$.

Let we show that the inequality

$$\frac{x}{e^{\frac{1}{3}\sqrt[4]{\log x}}} \leq \frac{x}{\log^{\alpha} x}, \quad \alpha > 0.$$

It is clear that the $\log^{\alpha} x \leq e^{\frac{1}{3}\sqrt[4]{\log x}}$, $\alpha > 0$. It is enough to represent the right side of inequality with Taylor series. Otherwise the following inequalities hold:

$$3\alpha \log\log x \leq (\log x)^{\frac{1}{4}}, \quad \alpha > 0$$
$$\log\log^{\alpha} x \leq \frac{1}{3}\sqrt[4]{\log x}\log e$$
$$\log^{\alpha} x \leq e^{\frac{1}{3}\sqrt[4]{\log x}}.$$

The last inequality holds for sufficiently large x. Hence the (2) holds. From Theorem 4 and Corollary 5 we have the following theorem.

Theorem 7. *a) The series of inverse values of pseudoprimes to base* 2 *converges.*

b) The asymptotic density of the set of all pseudoprimes to base 2 *is equal to* 0.

The condition b) of theorem above is proven in [4] by another way. Finally we show an important property of $P_2(x)$.

Theorem 8. *We have*

$$\lim_{x\to\infty} \frac{P_2(x)}{\frac{x}{\log x}} = 0.$$

Proof. Because there exist $K > 0$ and $\varepsilon > 0$, from (2) we have

$$\frac{P_2(x)\log x}{x} \leq \frac{Kx}{(\log x)^{1+\varepsilon}}.\frac{\log x}{x} = \frac{K}{(\log x)^{\varepsilon}}.$$

Since $\lim_{x\to\infty} \frac{K}{(\log x)^{\varepsilon}} = 0$ therefore $\lim_{x\to\infty} \frac{P_2(x)\log x}{x} = 0$. □

The Theorem 8 gives an important information of $P_2(x)$ i.e. the function $P_2(x)$ grows more slowly with $x \to \infty$ than $\frac{x}{\log x}$. We see that the analogy with prime number theorem does not hold. It says that $\lim_{x\to\infty} \frac{\pi(x)\log x}{x} = 1$, where $\pi(x)$ is the prime-counting function.

Easier problems regarding this issue can be an appropriate topic in aiming the course N422S1_4B Semestral Project I as a research activity for students attending the 4th semester.

Conclusion

We showed that if the counting function $A(x)$ of the set A i.e. $A(x) = \sum_{n \leq x} 1$, $n \in A$ has some propreties (see Theorem 4), then the series $\sum_{n \in A} a_n^{-1}$ converges. This result we applied to the counting function of the set of all pseudoprime numbers to base 2.

Acknowledgement

Tomáš Visnyai would acknowledge the financial support under the grant No. KEGA 047STU-4/2016.

References

[1] Šalát T.: *Nekonečné rady.* Praha, Academia, 1974.

[2] Křížek M. - Somer L.: *Pseudoprvočísla.* Pokroky matematiky, fyziky a astronómie, **48**/2003, p. 143-151.

[3] Erdős P.: *On almost primes.* AMM, **57**/1950, p. 404-407.

[4] Pomerance C.: *A new lower bound for the the pseudoprime counting function.* Illinois journal of Math., **26**/1982, No. 1, p. 4-9.

[5] Szymiczek K.: *On Pseudoprimes Which are Products of Distinct Primes.* AMM, **74**/1967, No. 1, p. 35-37

[6] Powell B. J. - Šalát T.: *Convergence of subseries of the harmonic series and asymptotic densities of sets of integers.*, Publ. Inst. Math. (Beograd) **50**(64)/1991, p. 60-70.

[7] Visnyai T.: *Convergence of series along the sets from ideals.* In Proceedings of 15th Conference on Applied Mathematics APLIMAT2016, STU Bratislava, 2016, p. 1105-1109.

YIELD POINT PHENOMENA IN METALS AND ALLOYS

Vladislav Navrátil

Department of Physics, Chemistry and Vocational Education
Faculty of Education, Masaryk University
Poříčí 7, 603 00 Brno, Czech Republic
navratil@ped.muni.cz

Abstract: *In previous papers [4,5] the plastic deformation of Cd single crystals has been investigated and in this way both the thermally activated and macroscopic parameters have been obtained. The present paper deals with the creep behavior of Cd + 0.44 at. % Zn alloy single crystals. Contrary to the Cd single crystals discontinuous creep curves have been observed, which do not occur at 78 K. The low – temperature limit of this effect is 130K approximately. The characteristics of these discontinuous creep curves are discussed on the assumption that the Portevin – Le Chatelier effect is acting in the Cd + 0.44 at. % Zn Alloy.*

Keywords: metals, deformation, yield point, repeated creep curve, Portevin – Le Chatelier effect.

INTRODUCTION

When certain materials such as mild steel or duralumin are deformed in tension, it is found that the stress – strain curve is not smooth, but shows marked irregularities, with negative slopes occurring at or near the initial yield on the curve. The actual shape of the stress - strain curve is dependent, to some extent, on the type and characteristics of the tensile testing machine used; nevertheless one may include all cases where $d\sigma/d\varepsilon$ is negative as example of yield point effects deserving attention [1]. Tensile machines are divided into two types, the so – called "soft" and "hard" machines. The effects of machine rigidity may be simply illustrated by reference to Fig 1a. Here, the tensile specimen shown is imagined to have a Young´s modulus *E*, while the machine and supporting members have an effective spring constant *K*. Thus, under a load *L*, the extension of the system is *L/K + L.l/(S.E)* where *l* is the specimen length and *S* its cross section. If the specimen extends by an amount dl the overall extension is constant; the load measured changes by *dL* so that

$$dL\left(\frac{1}{K}+\frac{l}{S.E}\right)+\frac{L.dl}{S.E}=0 \qquad (1)$$

so that

$$\frac{dL}{L}=-\frac{dL}{\frac{S.E}{K}+l} \qquad (2)$$

For certain dl further follows from this relationship that when *K* → 0 (very soft machine), is $\frac{dL}{L}\to 0$.

The spring constant *K* of the machine may be determined quite simply by a dial gauge using a heavy specimen or by determining of the shape of the elastic region of the load – time curve, which, as the above analysis shows, will always be less than the normally accepted modulus of elasticity

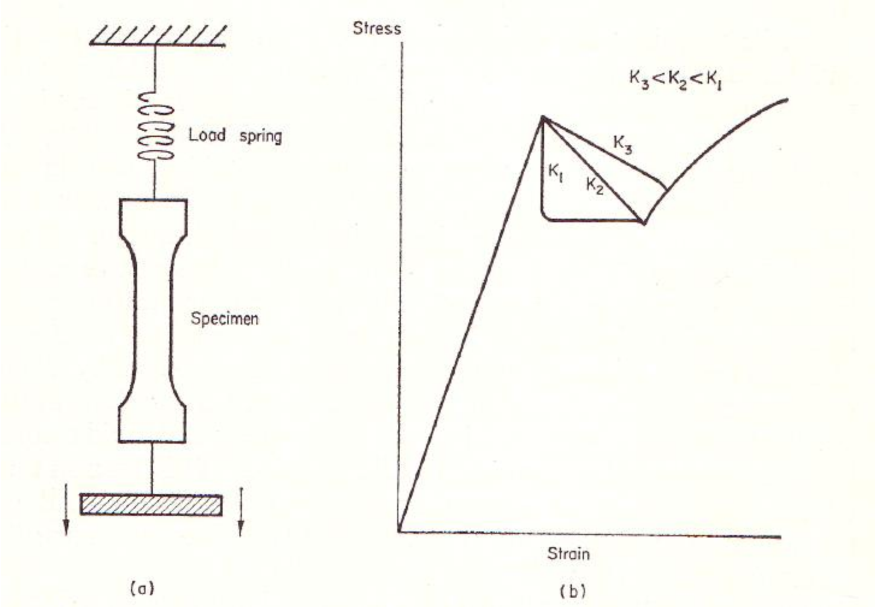

Fig.1 (a) Elastic elements of a tensile machine.
(b) Effect of the spring constant K on the stress – strain curve [1].

The elastic parameters of the machine will also affect the magnitude of the yield point drop. As the effective stiffness of the machine decreases the load relaxation decreases and will become less abrupt, until, as shown in Fig. 1(b), only a rounded yield is seen. Here the stress barely falls below the nucleation stress, and the Luders band will be forced through the specimen at much higher velocities [1].

The yield points observed with hard machines are found to take many different forms of instability, dependent on material and testing temperature. Fig. 2 shows a series of successive yield points obtained in mild steel at elevated temperatures, at about 500 K. The multiple yield points seen here as deformation begins are the result of interrupted motion of the Luders band along the specimen [2]. The movement of dislocations near the band front becomes locked – a phenomenon known as strain hardening – and as result the stress has to rise to release the band from again. The ductility is thereby reduced – a phenomenon known as blue brittleness – a result of simultaneous straining and ageing.

Fig 3 shows a case of austenitic stainless steel at high temperatures [3]. Here the general stress level continues to rise as deformation proceeds – and in certain cases the curves are smooth at the commencement of yield. As the strain increases, serrations build up slowly and reach their maximum at ultimate tensile strength. This mode, characteristic of duralumin and bronzes, nickel – hydrogen and even some magnesium – base alloys is properly called the Portevin – Le Chatelier effect after its discoverers (1923) who first noted it in duralumin.

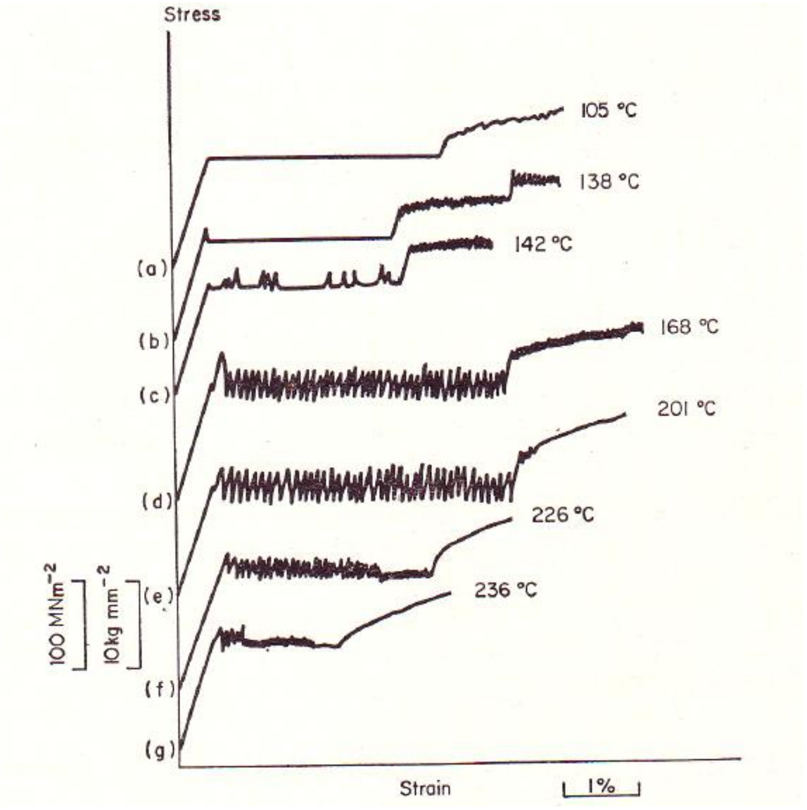

Fig. 2. Stress – strain curves for polycrystalline mild steel at elevated temperature [2]

The name of the effect is often applied to curves such as Fig. 2 and 3, although, as we shall see, the mode of locking may differ [1].

The purpose of the present work is to show that similar phenomenon can be observed also in creep deformation of Cd-Zn single crystal alloys.

1. EXPERIMENTAL PROCEDURE

To measure the characteristics of the transient creep of Cd + 0.44 at. % Zn alloy single crystals equivalent incrementally loading method (the sample is gradually loaded with increments in constant time interval 10 minutes) and experimental equipment as for measurements of pure Cd single crystals have been used [4,5].

The single crystals of that alloy were the same orientation $\lambda_0 = 49^o$ and $\chi_0 = 45^o$. The samples of the cylindrical form (3.9 mm of diameter and about 35 mm of length) were prepared at the Department of Solid State Physics of the Charles University, Prague.

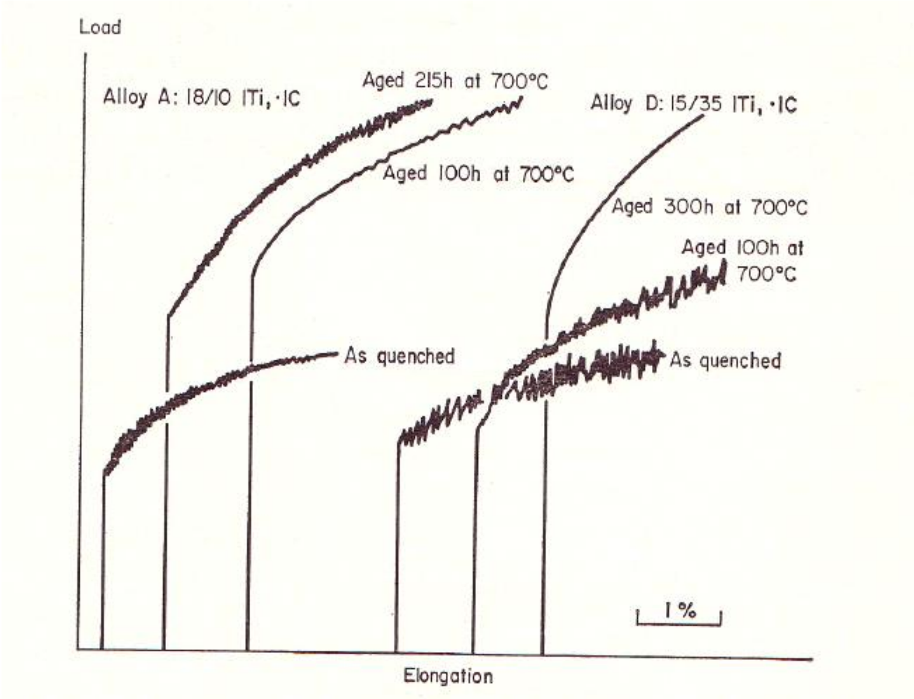

Fig. 3. Stress – strain curves for austenitic stainless steels at 770 K [3]

2. MACROSCOPIC CHARACTER OF TRANSIENT CREEP

Contrary to the Cd single crystals a considerable change in creep behavior of Cd + 0.44 at.% Zn alloy has been observed. The typical creep curve for pure Cd single crystals is illustrated in Fig.4. The change of he creep curves is shown by a special discontinuous “step like” shape of the creep curves (Fig. 5). As shown in the figure, an abrupt elongation of the specimen about the Δs length always occurs at a constant resolved shear stress value after a certain time interval Δt, where the length of he given specimen keeps at a constant value till the further abrupt elongation..

A similar behavior of these discontinuous flows in creep was observed at 238 K and 202 K. At temperature 78 K the “step like” curve shape of creep curves however has not appeared at all. On the basis of this fact we can conclude on the occurrence of the Portevin –Le Chatelier effect in this alloy. Since this effect occurs as a rule in a certain region of temperatures and shear stresses only we tried to determine the lower temperature level of this effect for the alloy mentioned at least approximately. Decreasing continually the temperature (by means of a petroleum ether bath, cooled by liquid nitrogen), the original “step like” shape of the creep curves has become continuous at temperature T ~ 130 K and at resolved shear stress τ = 200 p.mm^{-2} (Fig. 6)

Then the bath was heated again, until the continuous shape of creep curves changes into “step like” one. This change occurs at T ~ 230 K. The increase in the “transition” temperature is probably due to increase of the resolved shear stress to the value τ = 420 p.mm^{-2} (Fig. 7).

The “step like” curves can be characterized by the length of the “step” Δs and the time between two “steps” Δt (Fig. 5). At room temperature the length of the “step” is practically independent of time in one creep segment. The value of Δs however depends on the resolved shear stress. Typical examples of time dependence of the Δs are shown in Fig. 8. It is evident, that during one creep segment Δs is approximately constant. Fig.9 shows the stress dependence of the average length of "step" $\overline{\Delta s}$. It is obvious that $\overline{\Delta s}$ has a maximum value for a certain resolved shear stress ($\tau \sim 90$ p.mm^{-2}).

The time interval between two "steps" Δt during one creep segment increases in dependence on time, except the region about $\tau \sim 90$ p.mm^{-2}, where the course of the dependence $\Delta t = \Delta t(t)$ is not clear (Fig.10).

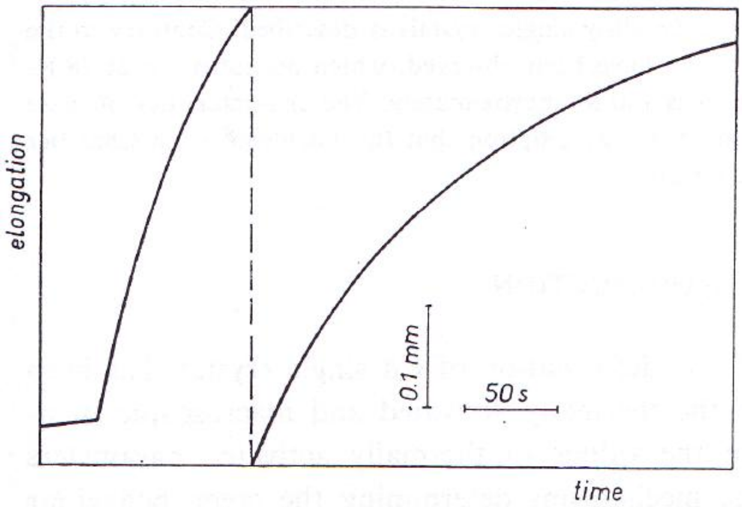

Fig.4. A typical creep curve for pure Cd single crystals (transient creep, T = 295 K, τ = 776 p.mm^{-2}, registered by means of recorder with zero suppresion [4].

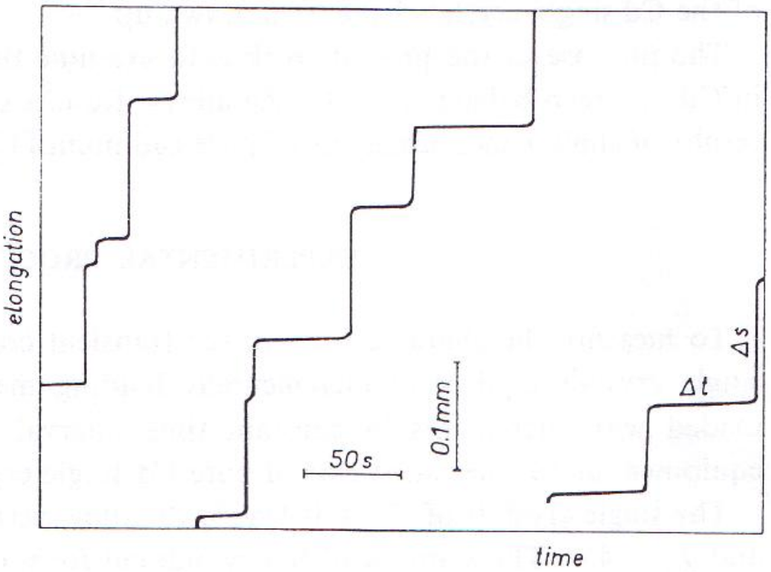

Fig.5. Discontinuous creep curve for Cd + 0.44 at. % Zn alloy single crystals. (T = 295 K, τ = 745 p.mm^{-2} registered by means of recorder with zero suppression)

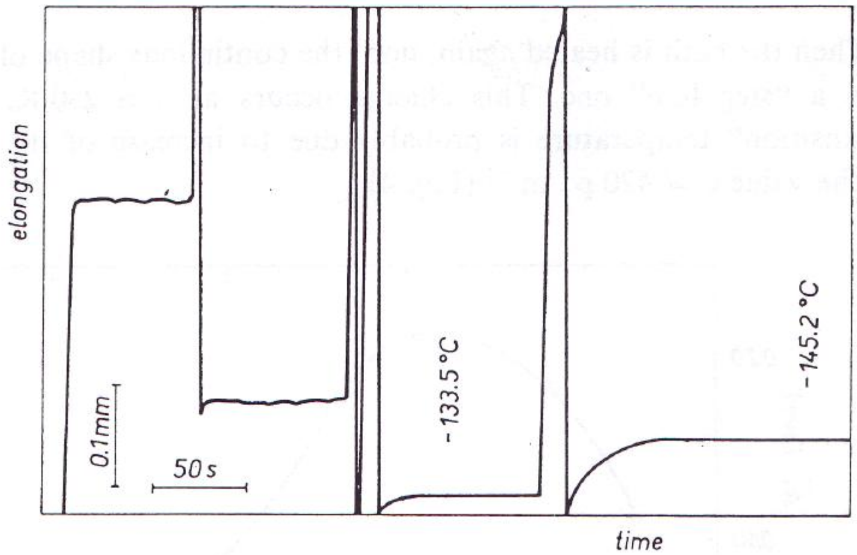

Fig.6. Transition from the „step like“ creep curves to the continuous ones For the Cd + 0.44 at. % Zn alloy single crystals (T = 130 K, τ = 200 p.mm^{-2}, registered by means of recorder with zero suppresion)

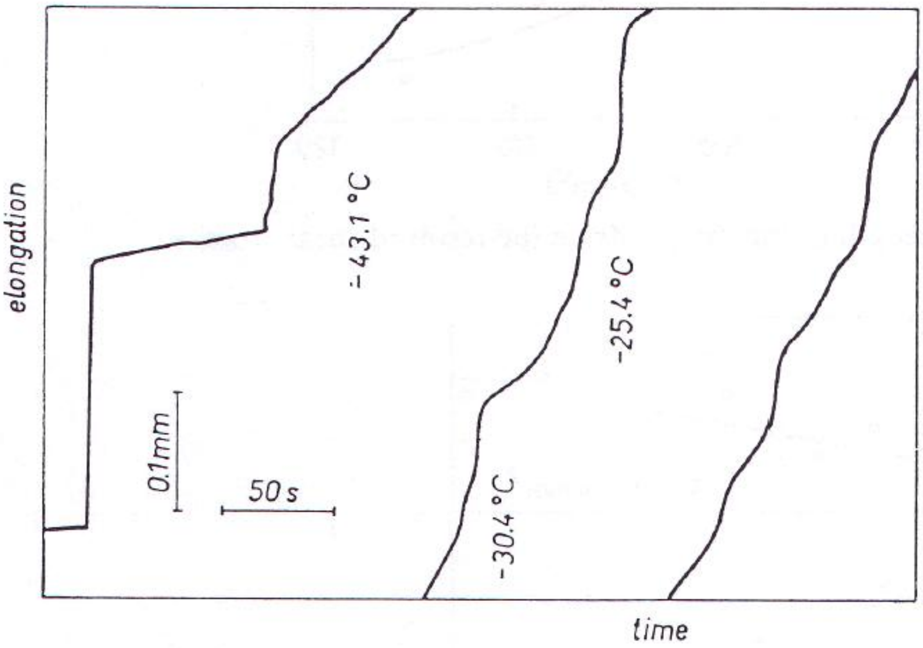

Fig. 7. Transition from the continuous creep curves ton the “step like” ones for the Cd + 0.44 at % Zn alloy single crystals (T = 230 K, τ = 420 p.mm^{-2}, registered by means of recorder with zero suppresion)

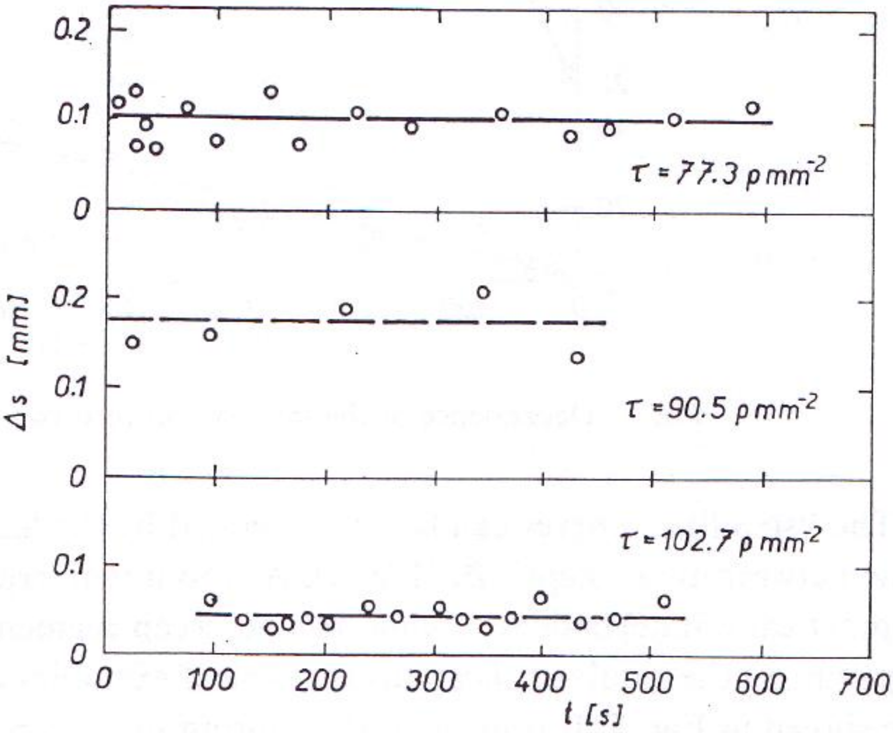

Fig.8. Dependence of the length of the „step“ *Δs* on time

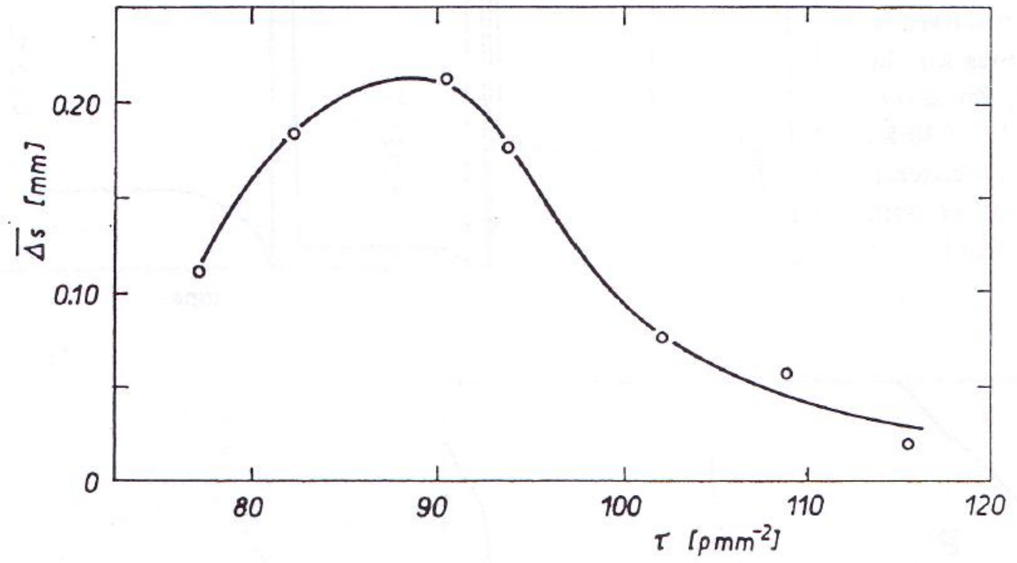

Fig.9. Dependence of the average length of the “step” $\overline{\Delta s}$ on the resolved shear stress *τ*.

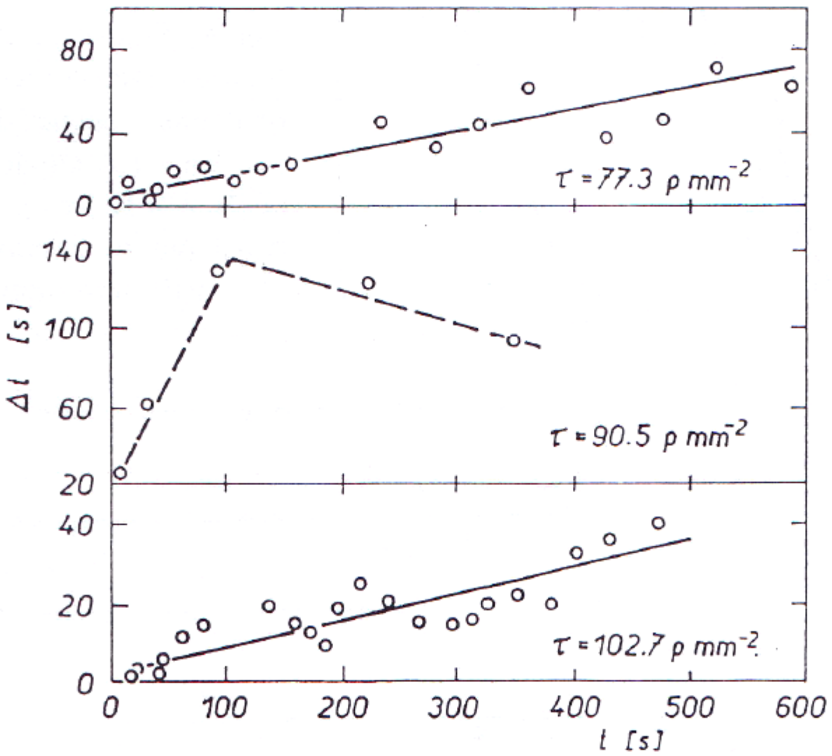

Fig.10. Dependence of the interval between two steps *Δt* on time *t*.

CONCLUSION

From the results of our measurements it follows that a new effect appears which is characterised by a typical "step like" shape of the creep curves. As shown by measurements, this effect appears in a certain interval of temperatures and resolved shear stresses. Hence we conclude on the Portevin – Le Chatelier effect occurring in the alloy.

The Portevin –Le Chatelier effect mostly observed by tensile tests on many interstitial and substitutional alloys becomes evident so that the stress – strain curve in a certain region of the resolved shear stress and temperature exhibits a characteristic jerky shape. Cottrell [6] assumes this effect to be due to an interaction of solute atoms and vacancies with dislocations. For the case of the creep it has not been investigated up to now, and for Cd – Zn alloys single crystals has not been investigated neither in tensile tests [7,8].

The average length of the "step" can be caused by the blocking dislocations by great number of solute atoms (Zn). This number of atoms depends on concentration of vacancies, which is proportional to the strain ε ($c_v \sim 10^{-4}\ \varepsilon$) and/or to the resolved shear stress. Hence it follows that with increasing τ, $\overline{\Delta s}$ ought to increase too. This effect has been observed up to a certain value of the resolved shear stress τ = 90 p.mm^{-2} only. Besides the increasing number of vacancies, however, we must also take into account the increase of the dislocation density with increasing resolved shear stress. At a greater number of dislocations (and at a constant density of solute atoms), it belongs to one dislocation a smaller number of blocking atoms, the blocking of dislocations is reduced and therefore $\overline{\Delta s}$ decrease. In consequence of this decrease the creep curve becomes a continuous shape.

LITERATURE

[1] HALL, E.O.: *Yield Point Phenomena in Metals and Alloys*. Plenum Press New York SBN 306 – 30490 - 2

[2] BLAKEMORE, J.S., HALL, E.O.: *J. Iron Steel Inst*.,1966, **204**, 817 pp

[3] HARDING, H.J., HONEYCOMBE, R.W.K.: *J. Iron Steel Inst*., 1966, **204**, 259 pp.

[4] HAMERSKÝ, M., LUKÁČ, P.: *Czech. J. Phys*. 1973, **B23**, 1345 pp.

[5] HAMERSKÝ, M., *Czech. J. Phys*., 1970, **B20**, 1327 pp.

[6] COTTRELL, A.H., *Dislocations and Plastic Flow*. Oxford University Press, Oxford 1953.

[7] MARTIN, J.L., CAILLARD, D: Thermally Activated Mechanism in Crystal Plasticity. Pergamon Press, 2003.

[8] NAVRATIL, V., NOVOTNA, J.: Aplimat 6th. Int. Conf. 2007, 125 – 130 pp.

A formula for the sum of the series of reciprocals of the polynomial of degree two with different positive integer roots

Radovan Potůček

Department of Mathematics and Physics, Faculty of Military Technology,
University of Defence, Kounicova 65, 662 10 Brno. Email:
Radovan.Potucek@unob.cz

Abstract: This contribution, which is a follow-up to author's paper [1] and [2] deals with the series of reciprocals of the quadratic polynomials with different positive integer roots. We derive the formula for the sum of this series and verify it by some examples evaluated using the basic programming language of the computer algebra system Maple 16. This contribution can be an inspiration for teachers of mathematics whose are teaching the topic Infinite series or as a subject matter for work with talented students.

Keywords: Sequence of partial sums, telescoping series, harmonic number, computer algebra system Maple.

Introduction and basic notions

Let us recall the basic terms. For any sequence $\{a_k\}$ of numbers the associated *series* is defined as the sum

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \cdots \ .$$

The *sequence of partial sums* $\{s_n\}$ associated to a series $\sum\limits_{k=1}^{\infty} a_k$ is defined for each n as the sum of the sequence $\{a_k\}$ from a_1 to a_n, i.e.

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \,.$$

The series $\sum\limits_{k=1}^{\infty} a_k$ *converges* to a limit s if and only if the sequence of partial sums $\{s_n\}$ converges to s, i.e. $\lim\limits_{n\to\infty} s_n = s$. We say that the series $\sum\limits_{k=1}^{\infty} a_k$ has a *sum* s and write $\sum\limits_{k=1}^{\infty} a_k = s$.

The *telescoping series* is any series where nearly every term cancels with a preceding or following term, so its partial sums eventually only have a fixed number of terms after cancellation. Telescoping series are not very common in mathematics but are interesting to study. The method of changing series whose terms are rational functions into telescoping series is that of transforming the rational functions by the method of partial fractions.

For example, the series $\sum\limits_{k=1}^{\infty} \dfrac{1}{(k-1)(k-2)}$, where obviously the summational index $k \neq 1, 2$, has the general kth term $a_k = \dfrac{1}{(k-1)(k-2)} = \dfrac{A}{k-1} + \dfrac{B}{k-2}$. After removing the fractions we get the equation $1 = A(k-2) + B(k-1)$. For $k = 1$ we get $A = -1$ and for $k = 2$ we obtain $B = 1$, so we have $a_k = -1/(k-1) + 1/(k-2) = 1/(k-2) - 1/(k-1)$. After that we arrange the terms of the nth partial sum $s_n = a_3 + a_4 + \cdots + a_n$ in a form where can be seen what is cancelling. Then we find the limit of the sequence of the partial sums s_n in order to find the sum s of the infinite telescoping series as $s = \lim\limits_{n\to\infty} s_n$. In our case we get

$$s_n = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-3} - \frac{1}{n-2}\right) + \left(\frac{1}{n-2} - \frac{1}{n-1}\right) = 1 - \frac{1}{n-1}.$$

So we have $s = \lim\limits_{n\to\infty}\left(1 - \dfrac{1}{n-1}\right) = 1$.

1 The sum of the series of reciprocals of the quadratic polynomial with different positive integer roots

Let us consider the series of reciprocals of the normalized quadratic polynomials of the form $k^2 - (a+b)k + ab = (k-a)(k-b)$ with two different integer roots

$0 < a < b$, i.e. the series

$$\sum_{\substack{k=1 \\ k \neq a,b}}^{\infty} \frac{1}{(k-a)(k-b)}\,, \tag{1}$$

and let us determine its sum $s(a, b)$.

This series can split into three parts – two finite series and the infinite one, so we have

$$s(a,b) = \sum_{k=1}^{a-1} \frac{1}{(k-a)(k-b)} + \sum_{k=a+1}^{b-1} \frac{1}{(k-a)(k-b)} + \sum_{k=b+1}^{\infty} \frac{1}{(k-a)(k-b)}\,. \tag{2}$$

We differentiate four following cases:

1. If $a = 1$ and $b = 2$, then both finite parts of the series (2) are not defined and we have
$$s(1,2) = \sum_{k=3}^{\infty} \frac{1}{(k-1)(k-2)} = 1\,,$$
as we derived in the example mentioned in Introduction.

2. If $a = 1$ and $b \geq 3$, then the first finite part of the series (2) is not defined and we get
$$s(1,b) = \sum_{k=2}^{b-1} \frac{1}{(k-1)(k-b)} + \sum_{k=b+1}^{\infty} \frac{1}{(k-1)(k-b)}\,.$$

3. If $a \geq 2$ and $b = a + 1$, then the second finite part of the series (2) is not defined and we obtain
$$s(a,a+1) = \sum_{k=1}^{a-1} \frac{1}{(k-a)(k-a-1)} + \sum_{k=a+2}^{\infty} \frac{1}{(k-a)(k-a-1)}\,.$$

4. If $a \geq 2$ and $b \geq a+2$, then there remain all three parts of the series (2) and we have
$$s(a,b) = \sum_{k=1}^{a-1} \frac{1}{(k-a)(k-b)} + \sum_{k=a+1}^{b-1} \frac{1}{(k-a)(k-b)} + \sum_{k=b+1}^{\infty} \frac{1}{(k-a)(k-b)}\,.$$

We concentrate on the last form for expressing the sum $s(a, b)$, where $a \geq 2$, $b \geq a + 2$, and determine this sum using the equality

$$\frac{1}{(k-a)(k-b)} = \frac{1}{a-b}\left(\frac{1}{k-a} - \frac{1}{k-b}\right).$$

The sums $s(1, b)$ and $s(a, a+1)$ corresponding with items 2. and 3. we give further in Corollary 1

The sum s' of the first finite part of the series (2) is

$$\begin{aligned} s' = \frac{1}{a-b}\biggl[&\left(\frac{1}{1-a} - \frac{1}{1-b}\right) + \left(\frac{1}{2-a} - \frac{1}{2-b}\right) + \left(\frac{1}{3-a} - \frac{1}{3-b}\right) + \cdots \\ &\cdots + \left(\frac{1}{-3} - \frac{1}{a-b-3}\right) + \left(\frac{1}{-2} - \frac{1}{a-b-2}\right) + \left(\frac{1}{-1} - \frac{1}{a-b-1}\right)\biggr] = \\ &= \frac{1}{a-b}\biggl[-\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{a-3} + \frac{1}{a-2} + \frac{1}{a-1}\right) + \\ &+ \left(\frac{1}{b-a+1} + \frac{1}{b-a+2} + \frac{1}{b-a+3} + \cdots + \frac{1}{b-3} + \frac{1}{b-2} + \frac{1}{b-1}\right)\biggr] = \\ &= \frac{1}{b-a}\bigl[H_{a-1} - (H_{b-1} - H_{b-a})\bigr] = \frac{H_{b-a} + H_{a-1} - H_{b-1}}{b-a}\,, \end{aligned}$$

where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is the nth *harmonic number*, H_0 being defined as 0. First ten values of the harmonic numbers are stated in the following table:

n	1	2	3	4	5	6	7	8	9	10
H_n	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$	$\frac{49}{20}$	$\frac{363}{140}$	$\frac{761}{280}$	$\frac{7\,129}{2\,520}$	$\frac{7\,381}{2\,520}$

Table 1: First ten values of the harmonic numbers

Basic information about harmonic numbers can be found e.g. in the web-site [3] or in [4], interesting information are included e.g. in the paper [5].

Now, let us determine the sum s'' of the second finite part of the series (2). We get

$$\begin{aligned}s'' = \frac{1}{a-b}\Bigg[&\left(\frac{1}{1}-\frac{1}{a-b+1}\right)+\left(\frac{1}{2}-\frac{1}{a-b+2}\right)+\left(\frac{1}{3}-\frac{1}{a-b+3}\right)+\cdots\\ &\cdots+\left(\frac{1}{b-a-3}-\frac{1}{-3}\right)+\left(\frac{1}{b-a-2}-\frac{1}{-2}\right)+\left(\frac{1}{b-a-1}-\frac{1}{-1}\right)\Bigg]=\\ =\frac{1}{a-b}\Bigg[&\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{b-a-3}+\frac{1}{b-a-2}+\frac{1}{b-a-1}\right)+\\ &+\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{b-a-3}+\frac{1}{b-a-2}+\frac{1}{b-a-1}\right)\Bigg]=\frac{2H_{b-a-1}}{a-b}\,.\end{aligned}$$

Finally, let us express the nth partial sum s_n of the infinite part of the series (2). We have

$$\begin{aligned}s_n = \frac{1}{a-b}\Bigg[&\left(\frac{1}{b-a+1}-\frac{1}{1}\right)+\left(\frac{1}{b-a+2}-\frac{1}{2}\right)+\left(\frac{1}{b-a+3}-\frac{1}{3}\right)+\cdots\\ &+\left(\frac{1}{n-a-2}-\frac{1}{n-b-2}\right)+\left(\frac{1}{n-a-1}-\frac{1}{n-b-1}\right)+\left(\frac{1}{n-a}-\frac{1}{n-b}\right)\Bigg].\end{aligned}$$

It is evident that the 1st summand in the 1st parenthesis $\left(\frac{1}{b-a+1}-\frac{1}{1}\right)$ cancels with the 2nd summand in the $(b-a+1)$st parenthesis $\left(\frac{1}{-2a+2b+1}-\frac{1}{-a+b+1}\right)$. Further, the 1st summand in the 2nd parenthesis $\left(\frac{1}{b-a+2}-\frac{1}{2}\right)$ cancels with the 2nd summand in the $(b-a+2)$nd parenthesis $\left(\frac{1}{-2a+2b+2}-\frac{1}{-a+b+2}\right)$, and so forth, up to the 1st summand in the $(b-a)$th parenthesis $\left(\frac{1}{2b-2a}-\frac{1}{b-a}\right)$ cancels with the 2nd summand in the $(2b-2a)$th parenthesis $\left(\frac{1}{3b-3a}-\frac{1}{2b-2a}\right)$. In the $(b-a+1)$st parenthesis $\left(\frac{1}{2b-2a+1}-\frac{1}{b-a+1}\right)$ and in the several following parentheses cancel both summands, so in the beginning of the expression of the nth partial sum s_n remains after cancelling the sum

$$-\frac{1}{1}-\frac{1}{2}-\cdots-\frac{1}{b-a-1}-\frac{1}{b-a}\,.$$

Analogously, in the ending of the nth partial sum s_n, the 2nd summand in the $(n-b)$th parenthesis $\left(\dfrac{1}{n-a}-\dfrac{1}{n-b}\right)$ cancels with the 1st summand in the $(n+a-2b)$th parenthesis $\left(\dfrac{1}{n-b}-\dfrac{1}{n+a-2b}\right)$. Further, the 2nd summand in the $(n-b-1)$st parenthesis $\left(\dfrac{1}{n-a-1}-\dfrac{1}{n-b-1}\right)$ cancels with the 1st summand in the $(n-a-1)$st parenthesis $\left(\dfrac{1}{n-1}-\dfrac{1}{n-a-1}\right)$, and so forth, up to the 2nd summand in the $(n-2b+a+1)$st parenthesis $\left(\dfrac{1}{n-b+1}-\dfrac{1}{n-2b+a+1}\right)$ cancels with the 1st summand in the $(n-3b+2a+1)$st parenthesis $\left(\dfrac{1}{n-2b+a+1}-\dfrac{1}{n-3b+2a+1}\right)$. In the $(n-2b+a)$th parenthesis $\left(\dfrac{1}{n-b}-\dfrac{1}{n-2b+a}\right)$ and in the several preceding parentheses cancel both summands, so in the ending of the expression of the nth partial sum s_n remains after cancelling the sum

$$\frac{1}{n-b+1}+\frac{1}{n-b+2}+\cdots+\frac{1}{n-a-1}+\frac{1}{n-a}\,.$$

The sum of the infinite part of the series (2) is $s''' =$

$$=\frac{1}{a-b}\lim_{n\to\infty}\left(-\frac{1}{1}-\frac{1}{2}-\cdots-\frac{1}{b-a}+\frac{1}{n-b+1}+\cdots+\frac{1}{n-a-1}+\frac{1}{n-a}\right)=$$
$$=\frac{1}{b-a}\left(1+\frac{1}{2}+\cdots+\frac{1}{b-a-1}+\frac{1}{b-a}\right)=\frac{H_{b-a}}{b-a}\,.$$

Altogether, for $a\geq 2$, $b\geq a+2$ we get the sum $s(a,b)$ of the series (1) in the form

$$s(a,b)=s'+s''+s'''=\frac{H_{b-a}+H_{a-1}-H_{b-1}}{b-a}+\frac{2H_{b-a-1}}{a-b}+\frac{H_{b-a}}{b-a}\,,$$

so we derived this statement:

Theorem 1 The series $\sum\limits_{\substack{k=1\\k\neq a,b}}^{\infty} \dfrac{1}{(k-a)(k-b)}$, where $a \geq 2$ and $b \geq a+2$ are positive integers, has the sum

$$s(a,b) = \frac{1}{b-a}\left(H_{a-1} - H_{b-1} + 2H_{b-a} - 2H_{b-a-1}\right), \tag{3}$$

where H_n is the nth harmonic number.

Corollary 1 From Theorem 1 and from the reasoning above it follows:

1. For the sum $s(a,b)$ above it obviously holds: $s(a,b) = s(b,a)$.

2. For $b \geq 3$ it holds: $s(1,b) = \dfrac{1}{b-1}(H_{b-1} - 2H_{b-2})$.

3. For $a \geq 2$ it holds: $s(a,a+1) = H_{a-1} - H_a + 2$, *whence* $\lim\limits_{a\to\infty} s(a,a+1) = 2$.

4. For $a = 1, b = 2$ it holds: $s(1,2) = H_1 = 1$.

Remark 1 *Let us note, that the formula* (3) *includes also the three above special cases. We can so state that for arbitrary two different positive integer roots $a < b$ of the normalized quadratic polynomial $(k-a)(k-b)$ the series* $\sum\limits_{\substack{k=1\\k\neq a,b}}^{\infty} \dfrac{1}{(k-a)(k-b)}$ *has the sum*

$$s(a,b) = \frac{1}{b-a}\left(H_{a-1} - H_{b-1} + 2H_{b-a} - 2H_{b-a-1}\right).$$

Example 1 *Using* **i)** *nth partial sum,* **ii)** *formula* (3) *calculate the sum of the series*

$$\sum_{k=1}^{\infty} \frac{1}{(k-3)(k-8)}.$$

i) *The series* $\sum\limits_{k=1}^{\infty} \dfrac{1}{(k-3)(k-8)}$, *where $a = 3$, $b = 8$, has, using the equality* $\dfrac{1}{(k-a)(k-b)} = \dfrac{1}{a-b}\left(\dfrac{1}{k-a} - \dfrac{1}{k-b}\right)$, *the nth partial sum (where $k \neq 3, 8$,* *i.e.* $k = 1, 2, 4, 5, 6, 7, 9, 10, 11, \ldots, n-2, n-1, n$*)*

$$\begin{aligned}
s_n = {} & \left(\frac{1}{(-2)(-7)} + \frac{1}{(-1)(-6)}\right) + \left(\frac{1}{1(-4)} + \frac{1}{2(-3)} + \frac{1}{3(-2)} + \frac{1}{4(-1)}\right) + \\
& + \left(\frac{1}{6\cdot 1} + \frac{1}{7\cdot 2} + \frac{1}{8\cdot 3} + \cdots + \frac{1}{(n-5)(n-10)} + \frac{1}{(n-4)(n-9)} + \right. \\
& \left. + \frac{1}{(n-3)(n-8)}\right) = \\
& = \left(\frac{1}{14} + \frac{1}{6}\right) + \left(-\frac{1}{4} - \frac{1}{6} - \frac{1}{6} - \frac{1}{4}\right) + \frac{1}{3-8}\left[\left(\frac{1}{6} - \frac{1}{1}\right) + \left(\frac{1}{7} - \frac{1}{2}\right) + \right. \\
& + \left(\frac{1}{8} - \frac{1}{3}\right) + \left(\frac{1}{9} - \frac{1}{4}\right) + \left(\frac{1}{10} - \frac{1}{5}\right) + \left(\frac{1}{11} - \frac{1}{6}\right) + \cdots \\
& \cdots + \left(\frac{1}{n-8} - \frac{1}{n-13}\right) + \left(\frac{1}{n-7} - \frac{1}{n-12}\right) + \left(\frac{1}{n-6} - \frac{1}{n-11}\right) + \\
& \left. + \left(\frac{1}{n-5} - \frac{1}{n-10}\right) + \left(\frac{1}{n-4} - \frac{1}{n-9}\right) + \left(\frac{1}{n-3} - \frac{1}{n-8}\right)\right] = \\
& = \frac{5}{21} - \frac{5}{6} - \frac{1}{5}\left(-\frac{1}{1} - \frac{1}{2} - \frac{1}{3} - \frac{1}{4} - \frac{1}{5} + \frac{1}{n-7} + \frac{1}{n-6} + \frac{1}{n-5} + \right. \\
& \left. + \frac{1}{n-4} + \frac{1}{n-3}\right).
\end{aligned}$$

Because for arbitrary integer c is $\lim\limits_{n\to\infty} \dfrac{1}{n+c} = 0$*, we have*

$$\begin{aligned}
s(3,8) = \lim_{n\to\infty} s_n &= \frac{5}{21} - \frac{5}{6} + \frac{1}{5}\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right) = \\
&= \frac{5}{21} - \frac{5}{6} + \frac{1}{5}\cdot\frac{137}{60} = -\frac{97}{700} = 0.13\overline{857142}.
\end{aligned}$$

ii) *By the formula* (3) *from Theorem 1, using values of the harmonic numbers from the table 1, we get the sum* $s(3,8)$ *more easily:*

$$\begin{aligned}
s(3,8) &= \frac{1}{8-3}\left(H_{3-1} - H_{8-1} + 2H_{8-3} - 2H_{8-3-1}\right) = \\
&= \frac{1}{5}\left(H_2 - H_7 + 2H_5 - 2H_4\right) = \frac{1}{5}\left(\frac{3}{2} - \frac{363}{140} + 2\cdot\frac{137}{60} - 2\cdot\frac{25}{12}\right) = \\
&= \frac{1}{5}\left(\frac{3}{2} - \frac{363}{140} + \frac{137}{30} - \frac{25}{6}\right) = -\frac{97}{700} = 0.13\overline{857142}.
\end{aligned}$$

Example 2 *Using* **i)** n*th partial sum,* **ii)** *formula* (3)*, and* **iii)** *formula for the sum* $s(a, a+1)$ *from Corollary 1 calculate the sum of the series*

$$\sum_{k=1}^{\infty} \frac{1}{(k-3)(k-4)} .$$

i) *The series* $\displaystyle\sum_{k=1}^{\infty} \frac{1}{(k-3)(k-4)}$*, where* $a = 3$*,* $b = 4$*, has, using the equality* $\displaystyle\frac{1}{(k-a)(k-b)} = \frac{1}{a-b}\left(\frac{1}{k-a} - \frac{1}{k-b}\right)$*, the following* n*th partial sum (where* $k = 1, 2, 5, 6, 7, \ldots, n-2, n-1, n$*):*

$$\begin{aligned} s_n &= \left(\frac{1}{(-2)(-3)} + \frac{1}{(-1)(-2)}\right) + \left(\frac{1}{2\cdot 1} + \frac{1}{3\cdot 2} + \frac{1}{4\cdot 3} + \cdots \right.\\ &\quad \left. + \frac{1}{(n-5)(n-6)} + \frac{1}{(n-4)(n-5)} + \frac{1}{(n-3)(n-4)}\right) = \\ &= \left(\frac{1}{6} + \frac{1}{2}\right) + \frac{1}{3-4}\left[\left(\frac{1}{2} - \frac{1}{1}\right) + \left(\frac{1}{3} - \frac{1}{2}\right) + \left(\frac{1}{4} - \frac{1}{3}\right) + \cdots \right.\\ &\quad \left. \cdots + \left(\frac{1}{n-5} - \frac{1}{n-6}\right) + \left(\frac{1}{n-4} - \frac{1}{n-5}\right) + \left(\frac{1}{n-3} - \frac{1}{n-4}\right)\right] = \\ &= \frac{2}{3} - \left(-\frac{1}{1} + \frac{1}{n-3}\right) = \frac{5}{3} - \frac{1}{n-3} . \end{aligned}$$

Because for arbitrary integer c *is* $\displaystyle\lim_{n\to\infty} \frac{1}{n+c} = 0$*, we have*

$$s(3,4) = \lim_{n\to\infty} s_n = \frac{5}{3} - 0 = 1.\overline{6}.$$

ii) *By the formula* (3) *from Theorem 1, using values of the harmonic numbers from the table 1, we get the sum* $s(3, 4)$ *more easily:*

$$\begin{aligned} s(3,4) &= \frac{1}{4-3}\left(H_{3-1} - H_{4-1} + 2H_{4-3} - 2H_{4-3-1}\right) = H_2 - H_3 + 2H_1 - 2H_0 = \\ &= \frac{3}{2} - \frac{11}{6} + 2\cdot 1 - 2\cdot 0 = \frac{5}{3} = 1.\overline{6}. \end{aligned}$$

The identical result we get by formula $s(a, a+1) = H_{a-1} - H_a + 2$*. For* $a = 3$ *we have*

$$s(3,4) = H_2 - H_3 + 2 = \frac{3}{2} - \frac{11}{6} + 2 = \frac{5}{3} = 1.\overline{6}.$$

2 Numerical verification

In this paper we solve the problem to determine the values of the sum $s(a,b) = \sum\limits_{\substack{k=1 \\ k\neq a,b}}^{\infty} \frac{1}{(k-a)(k-b)}$ for $a = 1, 2, \ldots, 9$ and $b = a+1, a+2, \ldots, 10$. We use on the one hand an approximative evaluation of the sum $s(a,b,t) = \sum\limits_{\substack{k=1 \\ k\neq a,b}}^{t} \frac{1}{(k-a)(k-b)}$, where $t = 10^8$, using the basic programming language of the computer algebra system Maple 16, and on the other hand the formula (3) for evaluation the sum $s(a,b)$. We compare $45 = 9+8+\cdots+1$ pairs of these ways obtained sums $s(a,b,10^8)$ and $s(a,b)$ to verify the formula (3). We use following simple procedures `hn`, `rp2abpos` and two **for** statements:

```
hn:=proc(h)
    local i,s; s:=0;
    if h=0 then s:=0 else
    for i from 1 to h do
        s:=s+1/i;
    end do;
    end if;
end proc:

rp2abpos:=proc(a,b,n)
    local k,sab,sabt; sabt:=0;
    sab:=(hn(a-1)-hn(b-1)+2*hn(b-a)-2*hn(b-a-1))/(b-a);
    print("n=",n,"s(",a,b,")=",evalf[20](sab));
    for k from 1 to n do
        if k<>a then
                if k=<>b then sabt:=sabt+1/((k-a)*(k-b))
        else sabt:=sabt+0; end if; end if;
    end do;
    print("sum(",a,b,")=",evalf[20](sabt));
    print("diff=",evalf[20](abs(sabt-sab)));
end proc:

for i from 1 to 9 do
    for j from i+1 to 10 do
        rp2abpos(i,j,100000000);
    end do;
end do;
```

The approximative values of the sums $s(a,b)$ rounded to 4 decimals obtained by these procedures are written into the following table:

$s(a,b)$	$b=2$	$b=3$	$b=4$	$b=5$	$b=6$
$a=1$	1.0000	-0.2500	-0.3889	-0.3958	-0.3767
$a=2$	×	1.5000	0.0833	-0.1389	-0.1958
$a=3$	×	×	1.6667	0.2083	-0.0389
$a=4$	×	×	×	1.7500	0.2750
$a=5$	×	×	×	×	1.8000
$s(a,b)$	$b=7$	$b=8$	$b=9$	$b=10$	×
$a=1$	-0.3528	-0.3296	-0.3085	-0.2896	×
$a=2$	-0.2100	-0.2099	-0.2046	-0.1974	×
$a=3$	-0.1125	-0.1386	-0.1474	-0.1490	×
$a=4$	0.0167	-0.0649	-0.0969	-0.1104	×
$a=5$	0.3167	0.0524	-0.0336	-0.0691	×
$a=6$	1.8333	0.3452	0.0774	-0.0114	×
$a=7$	×	1.8571	0.3661	0.0959	×
$a=8$	×	×	1.8750	0.3819	×
$a=9$	×	×	×	1.8889	×

Table 2: The approximate values of the sums $s(a,b)$ for $a = 1, 2, \ldots, 9$, $b = a+1, a+2, \ldots, 10$.

Computation of 45 couples of the sums $s(a,b,10^8)$ and $s(a,b)$ took over 4 hours. The absolute errors, i.e. the differences $\left|s(a,b) - s(a,b,10^8)\right|$, are about 10^{-8}.

3 Conclusion

As regards the problem to state the sum of the series of reciprocals of the polynomial of degree two, the author found only a mention on the web site [6] regarding the sum of the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{n(n+k)}$, where k be a positive integer. It is stated that the sum of this series is equal to the fraction $\dfrac{H_k}{k}$, where H_k is the kth harmonic number.

Another mention about sum of the special series of reciprocals of the polynomial of degree two concerns the sum of the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^2+a^2}$, where a is arbitrary non-zero real number. It is deduced that the sum of this series equals to the fraction $\frac{\pi a \coth(\pi a) - 1}{2a^2}$.

So we can say that these paper dealing with the sum of the series of reciprocals of the quadratic polynomials with different positive integer roots a and b, i.e. with the series

$$\sum_{\substack{k=1 \\ k\neq a,b}}^{\infty} \frac{1}{(k-a)(k-b)}\,,$$

where $0 < a < b$ are integers, brings new results which are not yet discuss in the literature.

We derived that the sum $s(a,b)$ of this series is given by the following formula using the nth harmonic numbers H_n

$$s(a,b) = \frac{1}{b-a}\left(H_{a-1} - H_{b-1} + 2H_{b-a} - 2H_{b-a-1}\right).$$

We verified this result by computing 45 various sums by using the computer algebra system Maple 16.

We stated four basic properties of the sum $s(a,b)$:

1. $s(a,b) = s(b,a)$, 2. $s(1,b) = \frac{1}{b-1}(H_{b-1} - 2H_{b-2})$ for $b \geq 3$,
3. $s(a,a+1) = H_{a-1} - H_a + 2$ for $a \geq 2$, whence $\lim\limits_{a\to\infty} s(a,a+1) = 2$,
4. $s(1,2) = H_1 = 1$.

The series of reciprocals of the quadratic polynomials with different positive integer roots so belong to special types of infinite series, such as geometric and telescoping ones, which sums are given analytically by means of a simple formula.

Reference

[1] Potůček R.: The sums of reciprocals of some quadratic polynomials. In *Proceedings of AFASES 2010, 12th International Conference "Scientific Research and Education in the Air Force"* (CD-ROM). Brasov, Romania, 2010, p. 1206-1209. ISBN 978-973-8415-76-8.

[2] Potůček R.: The sum of the series of reciprocals of the quadratic polynomials with double non-positive integer root. In *Proceedingd of the 15th Conference on Applied Mathematics APLIMAT 2016*. Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, 2016, p. 919-925. ISBN 978-80-227-4531-4.

[3] Wikipedia contributors: *Harmonic number*. Wikipedia, The Free Encyclopedia, [online], [cit. 2016-09-01]. Available from: https://en.wikipedia.org/wiki/Harmonic_number.

[4] Weisstein, E. W.: *Harmonic Number*. From MathWorld – A Wolfram Web Resource, [online], [cit. 2016-09-01]. Available from: http://mathworld.wolfram.com/HarmonicNumber.html

[5] Benjamin, A. T., Preston, G. O., Quinn, J. J.: *A Stirling Encounter with Harmonic Numbers*. Mathematics Magazine 75 (2), 2002, p. 95 –103, [online], [cit. 2016-09-01]. Available from: https://www.math.hmc.edu/∼benjamin/papers/harmonic.pdf

[6] Wikipedia contributors: *Telescoping series*. Wikipedia, The Free Encyclopedia, [online], [cit. 2016-09-01]. Available from: https://en.wikipedia.org/wiki/Telescoping_series.

[7] Mathematics Stack Exchange – A question and answer website for people studying math. [online], [cit. 2016-09-01]. Available from: http://math.stackexchange.com/questions/208317/show-sum-n-0-infty-frac1a2n2-frac1a-pi-coth-a-pi2a2.

Weakly Delayed Difference Systems in $\mathbb{R}^3$ and their Solution

Jan Šafařík

Faculty of Civil Engineering,
Faculty of Electrical Engineering and Communication,
Brno University of Technology, Brno, Czech Republic.
Email: `xsafar19@stud.feec.vutbr.cz`

Josef Diblík

Faculty of Civil Engineering,
Faculty of Electrical Engineering and Communication,
Brno University of Technology, Brno, Czech Republic.
Email: `diblik.j@fce.vutbr.cz`

Abstract: The paper is concerned with a weakly delayed difference system

$$x(k+1) = Ax(k) + Bx(k-1)$$

where $k = 0, 1, \ldots$ and $A = (a_{ij})_{i,j=1}^{3}$, $B = (b_{ij})_{i,j=1}^{3}$ are constant matrices. It is demonstrated that the initial delayed system can be transformed into a linear system without delay and, moreover, that all the eigenvalues of the matrix of the linear terms of this system can be obtained as the union of all the eigenvalues of matrices A and B.
In such a case, the new linear system without delay can be solved easily, e.g., by utilizing the well-known Putzer algorithm with one of the possible cases being considered in the paper.

Keywords: Discrete system, weak delay, initial problem, Putzer algorithm.

Introduction

The theory of weakly delayed systems is considered, for planar discrete systems, in the papers [1] – [3]. In this paper, we investigate a system of difference equations

$$x(k+1) = Ax(k) + Bx(k-1), \quad k = 0, 1, \ldots \tag{1}$$

where A and B are 3 by 3 constant matrices with elements a_{ij} and b_{ij}, $i, j = 1, 2, 3$.
It is known that, for every matrix A, there exists a nonsingular matrix S transforming it into the corresponding Jordan form A^*. This means that

$$A^* = S^{-1}AS$$

where A^* can have one of the following seven possible forms (denoted below as $A_1, \ldots, A_7$), depending on the roots of the characteristic equation

$$\det\left(A - \lambda I\right) = 0, \tag{2}$$

where I (throughout the paper) is a 3 by 3 unit matrix.
If (2) has three real distinct roots λ_1, λ_2, λ_3, then

$$A_1 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \tag{3}$$

if (2) has one double real root λ_1, $\lambda_2 = \lambda_3$, then

$$A_2 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \tag{4}$$

or

$$A_3 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix}, \tag{5}$$

in the case of one triple real root $\lambda = \lambda_{1,2,3}$, the following forms are possible

$$A_4 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \tag{6}$$

$$A_5 = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \tag{7}$$

$$A_6 = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \tag{8}$$

and, finally, if one root is real and two roots are complex conjugate, i.e. $\lambda_{2,3} = p \pm iq$, with $q \neq 0$, then

$$A_7 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & p & q \\ 0 & -q & p \end{pmatrix}. \tag{9}$$

We assume that (1) is a weakly delayed system in the sense of the following definition.

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 $\lambda \in \mathbb{C} \setminus \{0\}$,

$$\det\left(A + \lambda^{-1}B - \lambda I\right) = \det\left(A - \lambda I\right).$$

Applying Definition 1 to system (1), we get conditions under which the system is weakly delayed. Such conditions are given in the next part.
One way of solving system (1) is transforming (1) into a system without delay. Then, (1) can be written as

$$y(k+1) = \mathcal{A}_i y(k) \tag{10}$$

where

$$\mathcal{A}_i = \left(\begin{array}{c|c} A_i & B \\ \hline I & \Theta \end{array}\right), \quad i = 1, \dots, 7,$$

where Θ is zero matrix and

$$y_j(k) = x_j(k), \;\; j = 1, 2, 3, \quad y_j(k) = x_{j-3}(k-1), \;\; j = 4, 5, 6.$$

To solve (10) by Putzer algorithm, we need all eigenvalues of matrices $\mathcal{A}_i, i = 1, \dots, 7$.

1 Relationship between the eigenvalues of A_i, B, and $\mathcal{A}_i$

The main purpose of this paper is to show that the set of all eigenvalues of matrices $\mathcal{A}_i, i = 1, \dots, 7$ can be written as the union of the sets of all eigenvalues of matrices A_i and relevant matrix B.
In other words we prove the folloving theorem.

Theorem 1 (Main result) *Let system* (1) *be weakly delayed and let* $i \in \{1, \dots, 7\}$ *be fixed. Then the set of all the eigenvalues* $\mu_j^i, j = 1, \dots, 6$ *of the matrix* $\mathcal{A}_i$ *equals to the union of the sets of all the eigenvalues* $\lambda_j, j = 1, 2, 3$ *of the matrix* A_i *and all the eigenvalues* $\lambda_j, j = 4, 5, 6$ *of the matrix* B.

The property mentioned by Theorem 1 is not obvious and does not hold for arbitrary matrices A and B as shown by the following example.

Example 1 *Let*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 & -2 \\ -1 & 0 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

Then

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 & 1 & 2 & -2 \\ 0 & 2 & 0 & -1 & 0 & 2 \\ 0 & 0 & 3 & -2 & 2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

It is easy to verify that the eigenvalues of A *are*

$$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3,$$

and the eigenvalues of B *are*

$$\lambda_4 = 1, \lambda_5 = 3, \lambda_6 = -2.$$

The eigenvalues of $\mathcal{A}$ *(calculated by WolframAlpha software) are*

$$\mu_1 \doteq 3.57695,$$

$$\begin{aligned}\mu_2 &\doteq 1.61803,\\ \mu_3 &\doteq 1.14869 + 0.773951i,\\ \mu_4 &\doteq 1.14869 - 0.773951i,\\ \mu_5 &\doteq -0.874334,\\ \mu_6 &\doteq -0.618034.\end{aligned}$$

The eigenvalues $\lambda_i \neq \mu_j, i, j = 1, \ldots, 6$.

1.1 Proof of Theorem 1 if $i = 1$

The following theorem is proved in [5], Theorem 3.

Theorem 2 *System* (1) *is a weakly delayed system if and only if*

$$b_{11} = b_{22} = b_{33} = 0, \tag{11}$$

$$b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} = 0, \tag{12}$$

$$b_{12}b_{21} + b_{13}b_{31} + b_{23}b_{32} = 0, \tag{13}$$

$$\lambda_3 b_{12}b_{21} + \lambda_2 b_{13}b_{31} + \lambda_1 b_{23}b_{32} = 0. \tag{14}$$

Now we prove that Theorem 1 holds if $i = 1$.

Lemma 1 *Let a matrix A be of type* (3) *and let the entries of a matrix B satisfy* (11)–(14). *Then, the eigenvalues $\mu_i, i = 1, \ldots, 6$ of the matrix*

$$\mathcal{A}_1 = \left(\begin{array}{ccc|ccc} \mu_1 & 0 & 0 & 0 & b_{12} & b_{13} \\ 0 & \mu_2 & 0 & b_{21} & 0 & b_{23} \\ 0 & 0 & \mu_3 & b_{31} & b_{32} & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} A_1 & B \\ \hline I & \Theta \end{array}\right)$$

are $\mu_1 = \lambda_1$, $\mu_2 = \lambda_2$, $\mu_3 = \lambda_3$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Computing $\det(\mathcal{A}_1 - \mu I)$, we get

$$\Delta_1 = \det(\mathcal{A}_1 - \mu I) = \begin{vmatrix} \lambda_1 - \mu & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \lambda_2 - \mu & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \lambda_3 - \mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}.$$

Multiplying the first (the second, the third) column by μ and adding it to the fourth (the fifth, the sixth) column we get:

$$\Delta_1 = \begin{vmatrix} \lambda_1 - \mu & 0 & 0 & \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ 0 & \lambda_2 - \mu & 0 & b_{21} & \mu(\lambda_2 - \mu) + b_{22} & b_{23} \\ 0 & 0 & \lambda_3 - \mu & b_{31} & b_{32} & \mu(\lambda_3 - \mu) + b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}.$$

By Laplace decomposition with respect to the sixth row, we have:

$$\Delta_1 = -\begin{vmatrix} \lambda_1 - \mu & 0 & \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ 0 & \lambda_2 - \mu & b_{21} & \mu(\lambda_2 - \mu) + b_{22} & b_{23} \\ 0 & 0 & b_{31} & b_{32} & \mu(\lambda_3 - \mu) + b_{33} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{vmatrix}.$$

Again, by Laplace decomposition with respect to the last row, we derive:

$$\Delta_1 = \begin{vmatrix} \lambda_1 - \mu & \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ 0 & b_{21} & \mu(\lambda_2 - \mu) + b_{22} & b_{23} \\ 0 & b_{31} & b_{32} & \mu(\lambda_3 - \mu) + b_{33} \\ 1 & 0 & 0 & 0 \end{vmatrix}.$$

Finally, by Laplace decomposition with respect to the last row, we obtain:

$$\Delta_1 = -\begin{vmatrix} \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ b_{21} & \mu(\lambda_2 - \mu) + b_{22} & b_{23} \\ b_{31} & b_{32} & \mu(\lambda_3 - \mu) + b_{33} \end{vmatrix}.$$

Now, direct computation leads to:

$$\Delta_1 = \mu^6 + (-\lambda_1 - \lambda_2 - \lambda_3)\mu^5 + (\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 - b_{11} - b_{22} - b_{33})\mu^4$$

$$\begin{aligned}&+((b_{11}+b_{22})\lambda_3+(b_{11}+b_{33})\lambda_2+(b_{22}+b_{33})\lambda_1-\lambda_1\lambda_2\lambda_3)\mu^3\\&+(-b_{23}b_{32}+b_{22}b_{33}-b_{11}\lambda_2\lambda_3-b_{13}b_{31}-b_{33}\lambda_1\lambda_2-b_{12}b_{21}-b_{22}\lambda_1\lambda_3\\&\quad+b_{11}b_{22}+b_{11}b_{33})\mu^2\\&+(b_{23}b_{32}\lambda_1-b_{11}b_{33}\lambda_2-b_{11}b_{22}\lambda_3+b_{13}b_{31}\lambda_2-b_{22}b_{33}\lambda_1+b_{12}b_{21}\lambda_3)\mu\\&-b_{11}b_{22}b_{33}-b_{12}b_{23}b_{31}-b_{13}b_{21}b_{32}+b_{12}b_{21}b_{33}+b_{13}b_{22}b_{31}+b_{11}b_{23}b_{32}.\end{aligned}$$

Since (11)–(14) hold, further simplification of Δ_1 gives:

$$\begin{aligned}\Delta_1&=\mu^6+(-\lambda_1-\lambda_2-\lambda_3)\mu^5+(\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3)\mu^4+(-\lambda_1\lambda_2\lambda_3)\mu^3\\&=\mu^3(\mu-\lambda_1)(\mu-\lambda_2)(\mu-\lambda_3).\end{aligned}$$

Now it is easy to see that the roots of the equation $\det(\mathcal{A}_1-\mu I)=0$ are as formulated in the lemma.

Example 2 *Let*

$$A=\begin{pmatrix}1&0&0\\0&2&0\\0&0&3\end{pmatrix},B=\begin{pmatrix}0&1&-2\\2&0&2\\2&1&0\end{pmatrix}.$$

Then

$$\mathcal{A}=\begin{pmatrix}1&0&0&0&1&-2\\0&2&0&2&0&2\\0&0&3&2&1&0\\1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}.$$

It is easy to verify that the eigenvalues of A are

$$\lambda_1=1,\lambda_2=2,\lambda_3=3,$$

with the eigenvalues of B being

$$\lambda_4=\lambda_5=\lambda_6=0.$$

The eigenvalues of $\mathcal{A}$ (calculated by WolframAlpha software) are

$$\mu_1=1,\mu_2=2,\mu_3=3,\mu_4=\mu_5=\mu_6=0.$$

Eigenvalues $\lambda_i, i=1,\dots,6$ are the same as eigenvalues $\mu_j, j=1,\dots,6$.

1.2 Proof of Theorem 1 if $i = 2$

The following theorem is proved in [5], Theorem 4.

Theorem 3 *System* (1) *is a weakly delayed system if and only if*

$$\begin{align} b_{11} &= 0, \tag{15}\\ b_{22} + b_{33} &= 0, \tag{16}\\ b_{12}b_{21} + b_{13}b_{31} &= 0, \tag{17}\\ b_{22}b_{33} - b_{23}b_{32} &= 0, \tag{18}\\ b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} &= 0. \tag{19} \end{align}$$

Now we prove that Theorem 1 holds if $i = 2$.

Lemma 2 *Let a matrix A_2 be of type* (4) *and let the entries of a matrix B satisfy* (15)–(19). *Then, the eigenvalues $\mu_i, i = 1, \dots, 6$ of the matrix*

$$\mathcal{A}_2 = \left(\begin{array}{ccc|ccc} \mu_1 & 0 & 0 & b_{11} & b_{12} & b_{13}\\ 0 & \mu_2 & 0 & b_{21} & b_{22} & b_{23}\\ 0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33}\\ \hline 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} A_2 & B\\ \hline I & \Theta \end{array}\right)$$

are $\mu_1 = \lambda_1$, $\mu_2 = \mu_3 = \lambda_2$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Computing $\det(\mathcal{A}_2 - \mu I)$, we get

$$\Delta_2 = \det(\mathcal{A}_2 - \mu I) = \begin{vmatrix} \lambda_1 - \mu & 0 & 0 & b_{11} & b_{12} & b_{13}\\ 0 & \lambda_2 - \mu & 0 & b_{21} & b_{22} & b_{23}\\ 0 & 0 & \lambda_2 - \mu & b_{31} & b_{32} & b_{33}\\ 1 & 0 & 0 & -\mu & 0 & 0\\ 0 & 1 & 0 & 0 & -\mu & 0\\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$
= \begin{vmatrix}
\lambda_1-\mu & 0 & 0 & \mu(\lambda_1-\mu)+b_{11} & b_{12} & b_{13} \\
0 & \lambda_2-\mu & 0 & b_{21} & \mu(\lambda_2-\mu)+b_{22} & b_{23} \\
0 & 0 & \lambda_2-\mu & b_{31} & b_{32} & \mu(\lambda_2-\mu)+b_{33} \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0
\end{vmatrix}
$$

$$
= \cdots = - \begin{vmatrix}
\mu(\lambda_1-\mu)+b_{11} & b_{12} & b_{13} \\
b_{21} & \mu(\lambda_2-\mu)+b_{22} & b_{23} \\
b_{31} & b_{32} & \mu(\lambda_2-\mu)+b_{33}
\end{vmatrix}
$$

$$
\begin{aligned}
&= \mu^6 + (-\lambda_1 - 2\lambda_2)\mu^5 + (2\lambda_1\lambda_2 + \lambda_2^2 - b_{11} - b_{22} - b_{33})\mu^4 \\
&\quad + (-\lambda_1\lambda_2^2 + 2b_{11}\lambda_2 + b_{22}\lambda_1 + b_{22}\lambda_2 + b_{33}\lambda_1 + b_{33}\lambda_2)\mu^3 \\
&\quad + (-b_{11}\lambda_2^2 - b_{33}\lambda_1\lambda_2 - b_{33}\lambda_1\lambda_2 + b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31} + b_{22}b_{33} \\
&\qquad - b_{23}b_{32})\mu^2 \\
&\quad + (-b_{11}b_{22}\lambda_2 - b_{11}b_{33}\lambda_2 + b_{12}b_{21}\lambda_2 + b_{13}b_{31}\lambda_2 - b_{22}b_{33}\lambda_1 + b_{23}b_{32}\lambda_1)\mu \\
&\quad - b_{11}b_{22}b_{33} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33} - b_{12}b_{23}b_{31} - b_{13}b_{21}b_{32} + b_{13}b_{22}b_{31} \\
&= \mu^6 + (-\lambda_1 - 2\lambda_2)\mu^5 + (2\lambda_1\lambda_2 + \lambda_2^2)\mu^4 + (-\lambda_1\lambda_2^2)\mu^3 \\
&= \mu^3(\mu - \lambda_1)(\mu - \lambda_2)^2
\end{aligned}
$$

and the roots of the equation $\det(\mathcal{A}_2 - \mu I) = 0$ are as formulated in the lemma.

1.3 Proof of Theorem 1 if $i = 3$

The following theorem is proved in [5], Theorem 5.

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

$$b_{11} = 0, \tag{20}$$

$$b_{22} + b_{33} = 0, \tag{21}$$

$$b_{32} = 0, \tag{22}$$

$$b_{22}b_{33} - b_{12}b_{21} - b_{13}b_{31} = 0, \tag{23}$$

$$(\lambda_1 - \lambda_2)b_{22}b_{33} + b_{12}b_{31} = 0, \tag{24}$$

$$b_{12}b_{23}b_{31} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} = 0. \tag{25}$$

Now we prove that Theorem 1 holds if $i = 3$.

Lemma 3 *Let a matrix* A_3 *be of type* (5) *and let the entries of a matrix* B *satisfy* (20)–(25)*. Then, the eigenvalues* $\mu_i, i = 1, \dots, 6$ *of the matrix*

$$\mathcal{A}_3 = \left(\begin{array}{ccc|ccc} \mu_1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \mu_2 & 1 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33} \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} A_3 & B \\ \hline I & \Theta \end{array}\right)$$

are $\mu_1 = \lambda_1$, $\mu_2 = \mu_3 = \lambda_2$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Computing $\det(\mathcal{A}_3 - \mu I)$, we get

$$\Delta_3 = \det(\mathcal{A}_3 - \mu I) = \begin{vmatrix} \lambda_1 - \mu & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \lambda_2 - \mu & 1 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \lambda_2 - \mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$= \begin{vmatrix} \lambda_1 - \mu & 0 & 0 & \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ 0 & \lambda_2 - \mu & 1 & b_{21} & \mu(\lambda_2 - \mu) + b_{22} & \mu + b_{23} \\ 0 & 0 & \lambda_2 - \mu & b_{31} & b_{32} & \mu(\lambda_2 - \mu) + b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \cdots = -\begin{vmatrix} \mu(\lambda_1 - \mu) + b_{11} & b_{12} & b_{13} \\ b_{21} & \mu(\lambda_2 - \mu) + b_{22} & \mu + b_{23} \\ b_{31} & b_{32} & \mu(\lambda_2 - \mu) + b_{33} \end{vmatrix}$$

$$\begin{aligned} &= \mu^6 + (-\lambda_1 - 2\lambda_2)\mu^5 + (-b_{22} - b_{33} + 2\lambda_1\lambda_2 + \lambda_2^2 - b_{11})\mu^4 \\ &\quad + (b_{33}\lambda_1 + b_{33}\lambda_1 + b_{33}\lambda_2 + 2b_{11}\lambda_2 - \lambda_1\lambda_2^2 + b_{33}\lambda_2 - b_{32})\mu^3 \\ &\quad + (b_{22}b_{33} - b_{12}b_{21} + \lambda_1 b_{32} - b_{22}\lambda_1\lambda_2 + b_{11}b_{33} - b_{33}\lambda_1\lambda_2 - b_{11}\lambda_2^2 + b_{11}b_{22} \end{aligned}$$

$$
\begin{aligned}
&\quad - b_{23}b_{32} - b_{13}b_{31})\mu^2 \\
&+ (b_{23}b_{32}\lambda_1 + b_{13}b_{31}\lambda_2 + b_{12}b_{21}\lambda_2 + b_{11}b_{32} - b_{22}b_{33}\lambda_1 - b_{11}b_{33}\lambda_2 - b_{11}b_{22}\lambda_2 \\
&\quad - b_{31}b_{12})\mu \\
&+ b_{11}b_{23}b_{32} - b_{13}b_{21}b_{32} - b_{12}b_{23}b_{31} + b_{13}b_{22}b_{31} - b_{11}b_{22}b_{33} + b_{12}b_{21}b_{33} \\
= {}& \mu^6 + (-\lambda_1 - 2\lambda_2)\mu^5 + (2\lambda_1\lambda_2 + \lambda_2^2)\mu^4 + (-\lambda_1\lambda_2^2)\mu^3 \\
= {}& \mu^3(\mu - \lambda_1)(\mu - \lambda_2)^2,
\end{aligned}
$$

i.e. the lemma holds.

1.4 Proof of Theorem 1 if $i = 4$

The following theorem is proved in [5], Theorem 6.

Theorem 5 *System* (1) *is a weakly delayed system if and only if*

$$
\begin{aligned}
b_{11} + b_{22} + b_{33} &= 0, && (26)\\
b_{11}b_{22} + b_{11}b_{33} + b_{22}b_{33} - b_{12}b_{21} - b_{13}b_{31} - b_{23}b_{32} &= 0, && (27)\\
b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} - b_{11}b_{23}b_{32} &= 0. && (28)
\end{aligned}
$$

Now we prove that Theorem 1 holds if $i = 4$.

Lemma 4 *Let a matrix A_4 be of type* (6) *and let the entries of a matrix B satisfy* (26)–(28)*. Then, the eigenvalues $\mu_i, i = 1, \ldots, 6$ of the matrix*

$$
\mathcal{A}_4 = \left(\begin{array}{ccc|ccc}
\mu_1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\
0 & \mu_2 & 0 & b_{21} & b_{22} & b_{23} \\
0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33} \\
\hline
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0
\end{array}\right) = \left(\begin{array}{c|c} A_4 & B \\ \hline I & \Theta \end{array}\right)
$$

are $\mu_1 = \mu_2 = \mu_3 = \lambda$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Similarly as above, we get

$$\Delta_4 = \det(\mathcal{A}_4 - \mu I) = \begin{vmatrix} \lambda-\mu & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \lambda-\mu & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \lambda-\mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$= \begin{vmatrix} \lambda-\mu & 0 & 0 & \mu(\lambda-\mu)+b_{11} & b_{12} & b_{13} \\ 0 & \lambda-\mu & 0 & b_{21} & \mu(\lambda-\mu)+b_{22} & b_{23} \\ 0 & 0 & \lambda-\mu & b_{31} & b_{32} & \mu(\lambda-\mu)+b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \cdots = - \begin{vmatrix} \mu(\lambda-\mu)+b_{11} & b_{12} & b_{13} \\ b_{21} & \mu(\lambda-\mu)+b_{22} & b_{23} \\ b_{31} & b_{32} & \mu(\lambda-\mu)+b_{33} \end{vmatrix}$$

$$\begin{aligned} &= \mu^6 - 3\lambda\mu^5 + (3\lambda^2 - b_{11} - b_{22} - b_{33})\mu^4 + (-\lambda^3 + 2b_{11}\lambda + 2b_{22}\lambda + 2b_{33}\lambda)\mu^3 \\ &\quad + (-b_{11}\lambda^2 - b_{22}\lambda^2 - b_{33}\lambda^2 + b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31} + b_{22}b_{33} \\ &\qquad - b_{23}b_{32})\mu^2 \\ &\quad + (-b_{11}b_{22}\lambda - b_{11}b_{33}\lambda + b_{12}b_{21}\lambda + b_{13}b_{31}\lambda - b_{22}b_{33}\lambda + b_{23}b_{32}\lambda)\mu \\ &\quad - b_{11}b_{22}b_{33} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33} - b_{12}b_{23}b_{31} - b_{13}b_{21}b_{32} + b_{13}b_{22}b_{31} \\ &= \mu^6 + (-3\lambda)\mu^5 + (3\lambda^2)\mu^4 + (-\lambda^3)\mu^3 \\ &= \mu^3(\mu-\lambda)^3. \end{aligned}$$

1.5 Proof of Theorem 1 if $i = 5$

The following theorem is proved in [5], Theorem 7.

Theorem 6 *System* (1) *is a weakly delayed system if and only if*

$$b_{11} + b_{22} + b_{33} = 0, \tag{29}$$

$$b_{21} = 0, \tag{30}$$

$$b_{23}b_{31} = 0, \tag{31}$$

$$b_{11}b_{22} + b_{11}b_{33} + b_{22}b_{33} - b_{13}b_{31} - b_{23}b_{32} = 0, \tag{32}$$

$$b_{11}b_{22}b_{33} - b_{13}b_{22}b_{31} - b_{11}b_{23}b_{32} = 0. \tag{33}$$

Now we prove that Theorem 1 holds if $i = 5$.

Lemma 5 *Let a matrix* A_5 *be of type* (7) *and let the entries of a matrix* B *satisfy* (29)–(33)*. Then, the eigenvalues* $\mu_i, i = 1, \ldots, 6$ *of the matrix*

$$\mathcal{A}_5 = \left(\begin{array}{ccc|ccc} \mu_1 & 1 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \mu_2 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33} \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right) = \left(\begin{array}{c|c} A_5 & B \\ \hline I & \Theta \end{array} \right)$$

are $\mu_1 = \mu_2 = \mu_3 = \lambda$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Obviously

$$\Delta_5 = \det(\mathcal{A}_5 - \mu I) = \begin{vmatrix} \lambda - \mu & 1 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \lambda - \mu & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \lambda - \mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$= \begin{vmatrix} \lambda - \mu & 1 & 0 & \mu(\lambda - \mu) + b_{11} & \mu + b_{12} & b_{13} \\ 0 & \lambda - \mu & 0 & b_{21} & \mu(\lambda - \mu) + b_{22} & b_{23} \\ 0 & 0 & \lambda - \mu & b_{31} & b_{32} & \mu(\lambda - \mu) + b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \cdots = - \begin{vmatrix} \mu(\lambda - \mu) + b_{11} & \mu + b_{12} & b_{13} \\ b_{21} & \mu(\lambda - \mu) + b_{22} & b_{23} \\ b_{31} & b_{32} & \mu(\lambda - \mu) + b_{33} \end{vmatrix}$$

$$
\begin{aligned}
&= \mu^6 - 3\lambda\mu^5 + (3\lambda^2 - b_{11} - b_{22} - b_{33})\mu^4 \\
&\quad + (-\lambda^3 + 2b_{11}\lambda + 2b_{33}\lambda + 2b_{33}\lambda - b_{21})\mu^3 \\
&\quad + (-b_{11}\lambda^2 - b_{22}\lambda^2 - b_{33}\lambda^2 + b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31} + b_{21}\lambda + b_{22}b_{33} \\
&\qquad - b_{23}b_{32})\mu^2 \\
&\quad + (-b_{11}b_{22}\lambda - b_{11}b_{33}\lambda + b_{12}b_{21}\lambda + b_{13}b_{31}\lambda - b_{22}b_{33}\lambda + b_{23}b_{32}\lambda + b_{21}b_{33} \\
&\qquad - b_{23}b_{31})\mu \\
&\quad - b_{11}b_{22}b_{33} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33} - b_{12}b_{23}b_{31} - b_{13}b_{21}b_{32} + b_{13}b_{22}b_{31} \\
&= \mu^6 + (-3\lambda)\mu^5 + (3\lambda^2)\mu^4 + (-\lambda^3)\mu^3 \\
&= \mu^3(\mu - \lambda)^3.
\end{aligned}
$$

1.6 Proof of Theorem 1 if $i = 6$

The following theorem is proved in [5], Theorem 8.

Theorem 7 *System* (1) *is a weakly delayed system if and only if*

$$
\begin{aligned}
b_{11} + b_{22} + b_{33} &= 0, && (34)\\
b_{21} + b_{32} &= 0, && (35)\\
b_{31} &= 0, && (36)\\
b_{21}b_{33} + b_{11}b_{32} &= 0, && (37)\\
b_{11}b_{22} + b_{11}b_{33} + b_{22}b_{33} - b_{12}b_{21} - b_{23}b_{32} &= 0, && (38)\\
b_{11}b_{22}b_{33} + b_{13}b_{21}b_{32} - b_{12}b_{21}b_{33} - b_{11}b_{23}b_{32} &= 0. && (39)
\end{aligned}
$$

Now we prove that Theorem 1 holds if $i = 6$.

Lemma 6 *Let a matrix* A_6 *be of type* (8) *and let the entries of a matrix* B *satisfy* (34)–(39). *Then, the eigenvalues* $\mu_i, i = 1, \dots, 6$ *of the matrix*

$$
\mathcal{A}_6 = \left(\begin{array}{ccc|ccc}
\mu_1 & 1 & 0 & b_{11} & b_{12} & b_{13} \\
0 & \mu_2 & 1 & b_{21} & b_{22} & b_{23} \\
0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33} \\
\hline
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0
\end{array}\right) = \left(\begin{array}{c|c} A_6 & B \\ \hline I & \Theta \end{array}\right)
$$

are $\mu_1 = \mu_2 = \mu_3 = \lambda$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Computing $\det(\mathcal{A}_6 - \mu I)$, we get

$$\Delta_6 = \det(\mathcal{A}_6 - \mu I) = \begin{vmatrix} \lambda-\mu & 1 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \lambda-\mu & 1 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \lambda-\mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$= \begin{vmatrix} \lambda-\mu & 1 & 0 & \mu(\lambda-\mu)+b_{11} & \mu+b_{12} & b_{13} \\ 0 & \lambda-\mu & 1 & b_{21} & \mu(\lambda-\mu)+b_{22} & \mu+b_{23} \\ 0 & 0 & \lambda-\mu & b_{31} & b_{32} & \mu(\lambda-\mu)+b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \cdots = - \begin{vmatrix} \mu(\lambda-\mu)+b_{11} & \mu+b_{12} & b_{13} \\ b_{21} & \mu(\lambda-\mu)+b_{22} & \mu+b_{23} \\ b_{31} & b_{32} & \mu(\lambda-\mu)+b_{33} \end{vmatrix}$$

$$\begin{aligned} &= \mu^6 - 3\lambda\mu^5 + (3\lambda^2 - b_{11} - b_{22} - b_{33})\mu^4 \\ &\quad + (-\lambda^3 + 2b_{11}\lambda + 2b_{22}\lambda + 2b_{33}\lambda - b_{21} - b_{32})\mu^3 \\ &\quad + (-b_{11}\lambda^2 - b_{33}\lambda^2 - b_{33}\lambda^2 + b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} - b_{13}b_{31} + b_{21}\lambda + b_{22}b_{33} \\ &\qquad - b_{23}b_{32} + b_{32}\lambda - b_{31})\mu^2 \\ &\quad + (-b_{11}b_{220}\lambda - b_{11}b_{33}\lambda + b_{12}b_{21}\lambda + b_{13}b_{31}\lambda - b_{22}b_{33}\lambda + b_{23}b_{32}\lambda + b_{11}b_{32} \\ &\qquad - b_{12}b_{31} + b_{21}b_{33} - b_{23}b_{31})\mu \\ &\quad - b_{11}b_{22}b_{33} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33} - b_{12}b_{23}b_{31} - b_{13}b_{21}b_{32} + b_{13}b_{22}b_{31} \\ &= \mu^6 + (-3\lambda)\mu^5 + (3\lambda^2)\mu^4 + (-\lambda^3)\mu^3 \\ &= \mu^3(\mu - \lambda)^3 \end{aligned}$$

and the lemma is valid.

1.7 Proof of Theorem 1 if $i = 7$

The following theorem is proved in [5], Theorem 9.

Theorem 8 *System* (1) *is a weakly delayed system if and only if*

$$b_{11} = 0, \tag{40}$$
$$b_{22} + b_{33} = 0, \tag{41}$$
$$b_{23} - b_{32} = 0, \tag{42}$$
$$b_{22}b_{33} - b_{12}b_{21} - b_{13}b_{31} - b_{23}b_{32} = 0, \tag{43}$$
$$(\lambda - p)(b_{12}b_{21} + b_{13}b_{31}) + q(b_{12}b_{31} - b_{13}b_{21}) = 0, \tag{44}$$
$$b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} = 0. \tag{45}$$

Now we prove that Theorem 1 holds if $i = 7$.

Lemma 7 *Let a matrix A_7 be of type* (9) *and let the entries of a matrix B satisfy* (40)–(45). *Then, the eigenvalues $\mu_i, i = 1, \dots, 6$ of the matrix*

$$\mathcal{A}_7 = \left(\begin{array}{ccc|ccc} \mu_1 & 1 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & \mu_2 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & \mu_3 & b_{31} & b_{32} & b_{33} \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right) = \left(\begin{array}{c|c} A_7 & B \\ \hline I & \Theta \end{array} \right)$$

where I is an identity matrix and Θ is zero matrix are $\mu_1 = \lambda$, $\mu_2 = p + qi$, $\mu_3 = p - qi$, $\mu_4 = \mu_5 = \mu_6 = 0$.

Proof. Computing $\det(\mathcal{A}_7 - \mu I)$, we get

$$\Delta_7 = \det(\mathcal{A}_7 - \mu I) = \begin{vmatrix} \lambda - \mu & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & p - \mu & q & b_{21} & b_{22} & b_{23} \\ 0 & -q & p - \mu & b_{31} & b_{32} & b_{33} \\ 1 & 0 & 0 & -\mu & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 1 & 0 & 0 & -\mu \end{vmatrix}$$

$$= \begin{vmatrix} \lambda - \mu & 0 & 0 & \mu(\lambda - \mu) + b_{11} & b_{12} & b_{13} \\ 0 & p - \mu & q & b_{21} & \mu(p - \mu) + b_{22} & \mu q + b_{23} \\ 0 & -q & p - \mu & b_{31} & -\mu q + b_{32} & \mu(p - \mu) + b_{33} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \cdots = -\begin{vmatrix} \mu(\lambda-\mu)+b_{11} & b_{12} & b_{13} \\ b_{21} & \mu(p-\mu)+b_{22} & \mu q+b_{23} \\ b_{31} & -\mu q+b_{32} & \mu(p-\mu)+b_{33} \end{vmatrix}$$

$$\begin{aligned}
&= \mu^6 + (-\lambda-2p)\mu^5 + (2\lambda p + p^2 + q^2 - b_{11} - b_{22} - b_{33})\mu^4 \\
&\quad + (-\lambda p^2 - \lambda q^2 + 2b_{11}p + b_{22}\lambda + b_{22}p + b_{23}q - b_{32}q + b_{33}\lambda + b_{33}p)\mu^3 \\
&\quad + (-b_{11}p^2 - b_{11}q^2 - b_{22}\lambda p - b_{23}\lambda q + b_{32}\lambda q - b_{33}\lambda p + b_{11}b_{22} + b_{11}b_{33} - b_{12}b_{21} \\
&\qquad - b_{13}b_{31} + b_{22}b_{33} - b_{23}b_{32})\mu^2 \\
&\quad + (-b_{11}b_{22}p - b_{11}b_{23}q + b_{11}b_{32}q - b_{11}b_{33}p + b_{12}b_{21}p - b_{12}b_{31}q + b_{13}b_{21}q \\
&\qquad + b_{13}b_{31}p - b_{22}b_{33}\lambda + b_{23}b_{32}\lambda)\mu \\
&\quad - b_{11}b_{22}b_{33} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33} - b_{12}b_{23}b_{31} - b_{13}b_{21}b_{32} + b_{13}b_{22}b_{31} \\
&= \mu^6 + (-\lambda-2p)\mu^5 + (2\lambda p + p^2 + q^2)\mu^4 + (-\lambda p^2 - \lambda q^2)\mu^3 \\
&= \mu^3(\mu-\lambda)(\mu^2 - 2\mu p + p^2 + q^2) = \mu^3(\mu-\lambda)(\mu-(p+qi))(\mu-(p-qi)).
\end{aligned}$$

Now it is easy to see that the roots of the equation $\det(\mathcal{A}_7 - \mu I) = 0$ are as formulated in the lemma.

2 Utilization of Putzer Algorithm

To compute the powers $\mathcal{D}^k, k \geq 0$ of an s by s matrix $\mathcal{D}$, we recall a Putzer algorithm (see, e.g., [4], p. 118). It calculates the powers using the formula

$$\mathcal{D}^k = \sum_{j=1}^{s} u_j(k) M(j-1), \;\; k \geq 0 \tag{46}$$

where

$$M(0) = I, \tag{47}$$

$$M(1) = (\mathcal{D} - \nu_1 I)M(0), \tag{48}$$

$$M(2) = (\mathcal{D} - \nu_2 I)M(1), \tag{49}$$

$$\cdots$$

$$M(s-1) = (\mathcal{D} - \nu_{s-1})M(s-2) \tag{50}$$

and

$$u_1(k) = \nu_1^k, \;\; k \geq 0, \tag{51}$$

$$u_2(k) = \sum_{i=0}^{k-1} \nu_2^{k-1-i} u_1(i), \quad k \geq 0, \tag{52}$$

$$\dots$$

$$u_s(k) = \sum_{i=0}^{k-1} \nu_s^{k-1-i} u_{s-1}(i), \quad k \geq 0, \tag{53}$$

ν_i, $i = 1, \dots, s$ are eigenvalues of $\mathcal{D}$.

2.1 Powers of the matrix $\mathcal{A}_1$

To compute the powers $\mathcal{A}_1^k, k \geq 0$, we use formulas (46)–(53) and Lemma 1. Then, $k = 6$ and

$$\mathcal{A}_1^k = \sum_{j=1}^{6} u_j(k) M(j-1), \quad k \geq 0$$

where

$$\begin{aligned}
&M(0) = I, \\
&M(1) = (\mathcal{A}_1 - \mu_1 I)M(0) = (\mathcal{A}_1 - \mu_1 I) = (\mathcal{A}_1 - \lambda_1 I), \\
&M(2) = (\mathcal{A}_1 - \mu_2 I)M(1) = (\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I), \\
&M(3) = (\mathcal{A}_1 - \mu_3 I)M(2) = (\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I), \\
&M(4) = (\mathcal{A}_1 - \mu_4 I)M(3) = \mathcal{A}_1 M(3) = \mathcal{A}_1(\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I), \\
&M(5) = (\mathcal{A}_1 - \mu_5 I)M(4) = \mathcal{A}_1^2 M(3) = \mathcal{A}_1^2(\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I)
\end{aligned}$$

and

$$\begin{aligned}
u_1(k) &= \mu_1^k = \lambda_1^k, \quad k \geq 0, \\
u_2(k) &= \sum_{i=0}^{k-1} \mu_2^{k-1-i} u_1(i) = \sum_{i=0}^{k-1} \lambda_2^{k-1-i} \lambda_1^i, \quad k \geq 0, \\
u_3(k) &= \sum_{i=0}^{k-1} \mu_3^{k-1-i} u_2(i) = \sum_{i=0}^{k-1} \lambda_3^{k-1-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j, \quad k \geq 0, \\
u_4(k) &= \sum_{i=0}^{k-1} \mu_4^{k-1-i} u_3(i) = \sum_{i=0}^{k-1} 0^{k-1-i} u_3(i) = u_3(k-1)
\end{aligned}$$

$$= \sum_{i=0}^{k-2} \lambda_3^{k-2-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j, \quad k \geq 0,$$

$$u_5(k) = \sum_{i=0}^{k-1} \mu_5^{k-1-i} u_4(i) = \sum_{i=0}^{k-1} 0^{k-1-i} u_3(i-1) = u_3(k-2)$$

$$= \sum_{i=0}^{k-3} \lambda_3^{k-3-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j, \quad k \geq 0,$$

$$u_6(k) = \sum_{i=0}^{k-1} \mu_6^{k-1-i} u_5(i) = \sum_{i=0}^{k-1} 0^{k-1-i} u_3(i-2) = u_3(k-3)$$

$$= \sum_{i=0}^{k-4} \lambda_3^{k-4-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j, \quad k \geq 0.$$

Finally, we get

$$\begin{aligned}
\mathcal{A}_1^k =& \sum_{j=1}^{6} u_j(k) M(j-1) \\
=& \lambda_1^k I + (\mathcal{A}_1 - \lambda_1 I) \sum_{i=0}^{k-1} \lambda_2^{k-1-i} \lambda_1^i \\
&+ (\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I) \sum_{i=0}^{k-1} \lambda_3^{k-1-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j \\
&+ (\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I) \sum_{i=0}^{k-2} \lambda_3^{k-2-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j \\
&+ \mathcal{A}_1(\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I) \sum_{i=0}^{k-3} \lambda_3^{k-3-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j \\
&+ \mathcal{A}_1^2(\mathcal{A}_1 - \lambda_3 I)(\mathcal{A}_1 - \lambda_2 I)(\mathcal{A}_1 - \lambda_1 I) \sum_{i=0}^{k-4} \lambda_3^{k-4-i} \sum_{j=0}^{i-1} \lambda_2^{i-1-j} \lambda_1^j, \\
& k \geq 0.
\end{aligned}$$

2.2 Solution of an initial problem

Now we find an explicit solution to the initial problem

$$x_i(0) = x_{i,0},\ x_i(-1) = x_{i,-1},\ \ i = 1, 2, 3$$

to system (1), where $A = A_1$.
Define

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} = (E|\Theta).$$

Then, the solution is given by formula

$$x(k) = Q\mathcal{A}_1^k \cdot x^*,\ \ k \geq 0,$$

where $x^* = (x_{1,0}, x_{2,0}, x_{3,0}, x_{1,-1}, x_{2,-1}, x_{3,-1})^T$.

3 Conclusion

For weakly delayed system (1) with matrices A_i (where i is fixed, $i \in \{1, \ldots, 7\}$), B, it is proved that the union of all their eigenvalues is the same as the set of all eigenvalues of relevant matrix $\mathcal{A}_i$ of the non-delayed system (10). The usefulness of this fact is demonstrated for one of the possible cases when the Putzer algorithm is used to solve an initial problem. Results for 3-dimensional discrete systems are new (in [1] – [3] only planar systems are considered).

Acknowledgement

The authors were supported by the Grant FEKT-S-14-2200 of Faculty of Electrical Engineering and Communication, BUT.

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.

[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.

[3] 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.

[4] Elaydi, S. N.: *An Introduction to Difference Equations,* Third Edition, Springer, 2005.

[5] 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.

The Application of Design of Experiments to Analyze Operating Conditions of Technological Process

Alena VAGASKÁ[a] – Miroslav GOMBÁR[b]
[a]Department of Mathematics, Informatics and Cybernetics, Technical University of Košice,
Faculty of Manufacturing Technologies with a seat in Prešov,
Bayerova 1, 080 01 Prešov, Slovakia, alena.vagaska@tuke.sk
[b] Department of Mechanical Engineering, Institute of Technology and
Business in České Budějovice, Okružní 10, 370 01 České Budějovice, Czech Republic,
gombar.mirek@gmail.com

Abstract: *It is the objective of this paper to demonstrate how Design of Experiments (DOE) methodology may be applied in industrial and technical practice. It is often the case that an industrial or technological process is not operated at optimal conditions. Usually a process is affected by many factors which often interact. DOE is a strategy for experimentation, whereby all acting factors are manipulated simultaneously. The usage of DOE to analyze operating conditions of longitudinal turning process is presented in this paper. The influence of input factors (cutting conditions) on the parameters of a surface roughness profile (responses) has been investigated.*

Keywords: Design of Experiments, mathematical-statistical model, longitudinal turning process, significant factors, roughness profile parameters.

Introduction

It is often the case that some processes are very complex and do not exist suitable description or mathematical-physical-chemical models of them, so it is necessary to recognize and identify the relationships between considered variables only experimentally. But experimental work is traditionally done by changing the value of one separate factor at a time until no further improvement is accomplished. This is called the COST approach to experimental work (COST is the acronym for *c*onsider *o*ne *s*eparate factor at a *t*ime) and represents the intuitive way of performing experiments [1]. But this is an inefficient approach. In experimentation for process improvement and discovery it is usually necessary to consider simultaneously the influence of a number of input variables (factors) and output variables (responses). A better approach is to construct a carefully prepared set of experiments, in which all relevant factors are varied simultaneously. This is called statistical experimental design, or, design of experiments (DOE) [2], [3]. The application of DOE to analyse operating conditions of longitudinal turning process is presented in this paper.

1 Design of Experiments – Benefits

It is important to note that the neglect of certain principles in planning and carrying out of experiments may lead to devastation of the whole experimental work. No analyse (no method) of experimental data does not help to correct wrong or ill-prepared experiment. Application of DOE methodology enables us to avoid this risk. DOE methodology provides us to obtain maximum amount of information of high statistical and numerical correctness in an optimal number of individual test runs and the use of statistical principles in the design of experiments ensures that experiments are designed economically, that they are efficient, that individual and joint factor effects can be evaluated and conclusions can be stated with high reliability.

2 Experimental

2.1 Experimental conditions

The influence of cutting conditions – input factors x_1, x_2, x_3 (x_1 – cutting speed v_c, x_2 – feed f, x_3 – depth of cut a_p) on the response (parameters of the resulting surface roughness: $\hat{R}z$ – the maximum height of the roughness profile, $\hat{R}a$ – the mean arithmetic deviation of roughness), i.e. function dependence $\hat{y} = \hat{R}z, \hat{R}a = f(x_1, x_2, x_3)$ during longitudinal turning of steel C45 have been investigated. The actual experiment was carried out under the operating conditions listed in Tab. 1. The parameters of the surface roughness $\hat{R}z, \hat{R}a$ were measured at defined experimental points by usage of roughness meter Mitutoyo Surftest SJ–301. Experimental points were indicated at the intersections of horizontal and vertical lines, the parameters of the surface roughness $\hat{R}z$, $\hat{R}a$ were measured five times at each experimental point. The arithmetic average of five measurements was taken as an individual measurement. Experimentally obtained data represented an input matrix for statistical analysis.

<table>
<tr><td colspan="2"><b>Experiment Code</b></td><td colspan="12"><b>Rz.v_c,f,a_p – 12 050.1</b></td></tr>
<tr><td colspan="2"><b>Used machine-tool:</b></td><td colspan="12">SU 40</td></tr>
<tr><td colspan="2" rowspan="2"><b>Used cutting tool</b></td><td colspan="3"><b>Holder</b></td><td colspan="3"><b>Cutting Blade</b></td><td colspan="3"><b>Cutting Material</b></td><td colspan="3"><b>r_ε [mm]</b></td></tr>
<tr><td colspan="3">MWLNR</td><td colspan="3">KNUX 190 405 EL</td><td colspan="3">P20 according ISO</td><td colspan="3">0.50</td></tr>
<tr><td colspan="2" rowspan="2"><b>Cutting Conditions</b></td><td colspan="4"><b>v_c [m.min^{-1}]</b></td><td colspan="4"><b>a_p [mm]</b></td><td colspan="4"><b>f [mm]</b></td></tr>
<tr><td colspan="4">8.792 – 351.680</td><td colspan="4">0.10 – 3.00</td><td colspan="4">0.100 – 0.500</td></tr>
<tr><td colspan="2"><b>Set the tool to the workpiece axis</b></td><td colspan="4">h = 0 [mm]</td><td colspan="4"><b>Cooling</b></td><td colspan="4">No</td></tr>
<tr><td colspan="2"><b>Machined material</b></td><td colspan="12">12 050.1 C45</td></tr>
<tr><td colspan="2"><b>The measuring instruments</b></td><td colspan="12">Mitutoyo Surftest SJ – 301 to measure parameters of surface roughness</td></tr>
<tr><td><b>Accuracy of calculation</b></td><td>E = 1/1000</td><td colspan="8"><b>The chosen level of significance</b></td><td colspan="4">α = 0.05</td></tr>
<tr><td><b>The number of runs</b></td><td>N = 8</td><td colspan="8"><b>The number of repeated measurements for each experimental unit</b></td><td colspan="4">m = 5</td></tr>
</table>

Tab. 1 Experimental conditions
Source: own

2.2 Construction of full factorial design

In order to identify significant factors affecting the surface roughness of machined material and analyse the relationships between them, the DOE methodology was used. Taking into account the expected non-linear dependencies, the two-level full factorial design in three factors, denoted 2^3, was chosen from a relatively large amount of design types. To perform a two-level full factorial design, a low level and a high level to each factor was assign. These settings were then used to construct an orthogonal array of experiment. Usually, the low level of a factor is denoted by -1, and the high level by +1. Individual test runs were performed on the basis of the design matrix of the experiment created as a combination of individual levels of three investigated factors according to Tab. 2.

Design Matrix				Experimental matrix		
Run No	Factors levels - Coded unit			Factrors – Original unit		
	x_1	x_2	x_3	v_c [m.min^{-1}]	f [mm]	a_p [mm]
1.	-1	-1	-1	8.792	0.1	0.1
2.	+1	-1	-1	351.68	0.1	0.1
3.	-1	+1	-1	8.792	0.5	0.1
4.	+1	+1	-1	351.68	0.5	0.1
5.	-1	-1	+1	8.792	0.1	3.0
6.	+1	-1	+1	351.68	0.1	3.0
7.	-1	+1	+1	8.792	0.5	3.0
8.	+1	+1	+1	351.68	0.5	3.0

Tab. 2 The 2^3 factorial design of experiment
Source: own

By means of DOE, individual runs were performed in random order to eliminate systematic error and to avoid subjective preference of any factor-level. Use scalar products the orthogonality of experiment design was verified, i.e. all columns of design matrix must be perpendicular to each other. Due to the orthogonality of the experimental design we can avoid improper indication of statistical insignificance of factors effects [4].

When least squares analysis is applied to the modelling of effects of several factors it is commonly known as multiple linear regression (MLR). To avoid numerical and statistical incorrectness in computation of regression model (e.g. incorrect indication of statistical insignificance of some factors due to their multicollinearity), it is necessary to perform DOE standardization (coding) of input factors into coded unit before applying regression analyse of experimentally obtained data. This is the proper way of expressing regression coefficients. Then we obtain not only correct statistical significance of regression model, but also correct statistical significance of regression coefficients [4].
DoE coding (standardization) of input factors is based on the coding equation

$$x_d(i) = \frac{x(i) - \frac{x_{\max} + x_{\min}}{2}}{\frac{x_{\max} - x_{\min}}{2}} \tag{1}$$

where $x(i) -$ is an original input variable (factor), $i = 1, ..., n$, $n -$ is the number of input factors, $x_d(i) -$ is a coded variable according to the DOE methodology, $x_{\max}$ – the maximum value of original variable $x(i)$ [physical unit], $x_{\min}$ – the minimum value of original variable $x(i)$ [physical unit]. DoE standardization by formula (1) presents linear transformation of values of origin variable from interval $< \mathrm{x}_{\min}, x_{\max} >$ to interval $< -1, 1 >$ and provides transformation of factors from original physical unit to dimensionless form.

3 Results and discussion

Based on the statistical analysis of experimentally obtained data (exploratory data analysis EDA, screening analysis, analysis of variance ANOVA, DOE analysis) using software such as Matlab, Statistica, QC-Expert, we have indicated important factors affecting the final surface roughness, we analysed how they interact, and obtained computational statistical models predicting the value of roughness parameters at varied levels of factors. Data analysis was performed with a statistically correct approach including analysis of the basic assumptions and subsequent analysis of the classical regression triplet: data, model, residues. Subsequently, as the conditions of normality of repeated measurements and homogeneity of variance of repeated measurements and testing for the presence of outliers had been verified, regression analysis was performed; the results are listed in Tab. 3. During our experimental work various types of regression model was used (linear, quadratic). In this paper is presented the regression model in the form of power function:

$$\hat{R}z, \hat{R}a = 10^{b_0} \cdot v_c^{\,b_1} \cdot f^{\,b_2} \cdot a_p^{\,b_3} \tag{2}$$

Statistical value	$\hat{R}z, \hat{R}a = 10^{b_0} \cdot v_c^{\,b_1} \cdot f^{\,b_2} \cdot a_p^{\,b_3}$		
	Rz	Ra	b_i – the estimation of i–th regression coefficient ± b_i – 95 % confidence interval of regression coefficient estimation t_i – the test statistic for the i-th regression coefficient $t_i = \frac{\lvert b_i \rvert}{s_{b_i}}$ s_{bi} – standard deviation of the i–th regression coefficient $t_{1-\frac{\alpha}{2}}(f)$ – quantile of the t–test distribution $f = N - 1 = 7$, $t_{1-\frac{\alpha}{2}}(f) = t_{0.975}(7) = 2.571$ **H$_0$:** $b_i = 0$, i = 1,2,.., k versus **H$_1$:** $b_i \neq 0$ rejection region of H$_0$: $\lvert t_i \rvert > t_{1-\frac{\alpha}{2}}(f)$, N – the number of test runs
b_0	2.364	1.775	
± b_0	0.424	0.343	
t_0	14.348	13.309	
significance	significant	significant	
b_1	-0.199	-0.162	
± b_1	0.173	0.140	
t_1	-2.956	-2.975	
significance	significant	significant	
b_2	0.787	1.077	
± b_2	0.397	0.321	
t_2	5.101	8.618	
significance	significant	significant	
b_3	0.017	0.040	
± b_3	0.188	0.152	
t_3	0.229	0.678	
significance	insignificant	insignificant	

Tab. 3 Results of regression analysis
Source: own

As we can see, the null hypothesis H_0 was tested against the alternative hypothesis H_1 and it can be concluded that only for one regression coefficient b_3 the hypothesis test with $\alpha = 0.05$ reject the null hypothesis, i.e. the regression coefficient b_3 is statistically insignificant. According to the results presented in Tab. 3 it was possible to develop the mathematical – statistical model, and considering DOE coding of individual factors (1) and natural unit (Tab. 2), the technological model of factors effect on the examined parameters of the resulting roughness profile was obtained in the form of natural scale:

$$\hat{R}z = \frac{10^{2.364(\pm 0.424)} * f^{\,0.787(\pm 0.397)} * a_p^{\,0.017(\pm 0.188)}}{v_c^{\,0.199(\pm 0.173)}} \tag{3}$$

$$\hat{R}a = \frac{10^{1.775(\pm 0.343)} * f^{\,1.077(\pm 0.321)} * a_p^{\,0.040(\pm 0.152)}}{v_c^{\,0.162(\pm 0.140)}} \tag{4}$$

It is convenient to display regression coefficients in a bar chart or plots of factor effects. Based on the results shown in Tab. 3, the Pareto diagrams of individual factors effect on the observed parameters are displayed in Fig. 1. and Fig. 2.

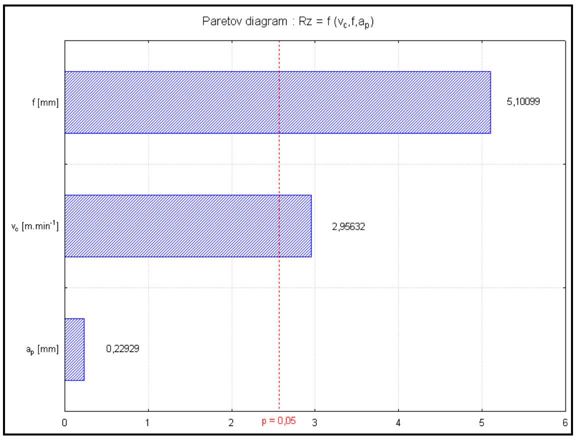

Fig. 1 Cutting conditions effect on the $\hat{R}z$ value

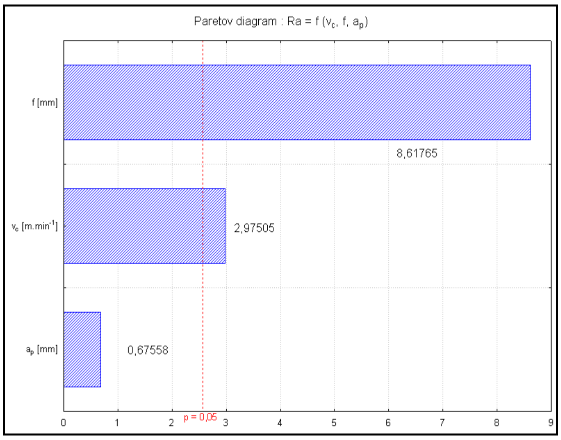

Fig. 2 Cutting conditions effect on the $\hat{R}a$ value

As we can see in the Fig. 1 and Fig. 2, the feed has the main effect on the studied parameters, while the cutting speed is of less importance and statistically insignificant appears to be the depth of cut. In order to express the functional dependence of observed parameters $\hat{R}z$ and $\hat{R}a$ on the cutting conditions, Pearson's correlation coefficients (confidence interval of $\alpha = 0.05$) were determined and their statistical significance was verified by the Student's *t*-Test Criterion. Statistically significant effect of the feed f on the parameter $\hat{R}z$ is confirmed by the high value of the pairwise correlation coefficient $r_{Rz\_f.(v_c,a_p)} = +80.843\%$, a direct dependency between them is obvious. The indirect dependency of parameter $\hat{R}z$ on the cutting speed v_c is evident from the value of pairwise correlation coefficient $r_{Rz\_v_c.(f,a_p)} = -46.854\%$. The relationship between $\hat{R}z$ and the depth of cut a_p is statistically insignificant, $r_{Rz\_a_p.(v_c,f)} = 3.634\%$. The effect of cutting conditions on the parameter $\hat{R}a$ is demonstrated by the value of individual pairwise correlation coefficient: $r_{Ra\_f.(v_c,a_p)} = 91.566\%$, $r_{Ra\_v_c.(f,a_p)} = -31.611\%$, $r_{Ra\_a_p.(v_c,f)} = 7.199\%$.

To verify correctness of regression model (3) and (4), the estimation of multiple correlation coefficient, the coefficient of determination and degree of variability of regression model was performed. For regression model (3), which express the effect of cutting conditions on $\hat{R}z$, the multiple correlation coefficient is $r_{Rz\_v_c,f,a_p} = +93.509\%$, the statistical significance of this estimation was confirmed by *F*-test. The regression model (3) explains 82.418% of the variability of the $\hat{R}z$ values, which is expressed by the adjusted coefficient of determination (AdjR^2) in order to eliminate the influence of multiple regression coefficients on the coefficient of determination R^2. For regression model (4), which express the effect of cutting conditions on $\hat{R}a$, the multiple correlation coefficient is $r_{Ra\_v_c,f,a_p} = +97.137\%$, its 95% confidence interval is (84,528%, 99,498%). The regression model (4) explains 92.097% of the variability of values $\hat{R}a$, i.e. AdjR^2 = 92.097%. The analysis of factor effects on the value of investigated roughness profile parameters confirms the conclusions of previous experimental work, where the effect of cutting conditions on the parameters was examined [], []. Graphical representation of the individual factors effect on $\hat{R}a$ is shown in Fig. 3.

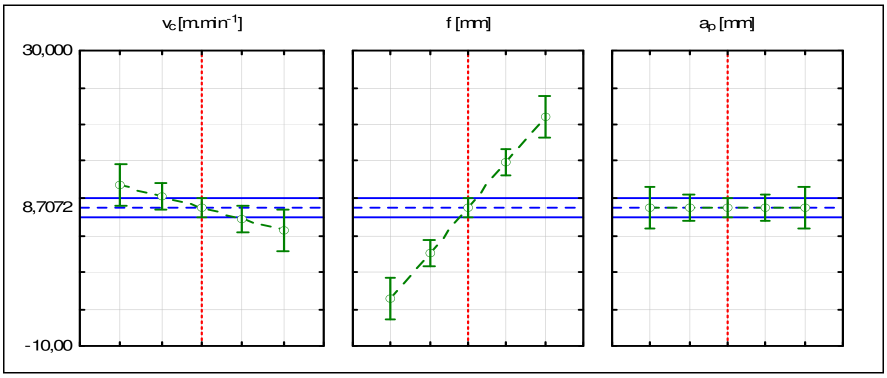

Fig. 3 The factor effects on the resulting value $\hat{Ra}$

Based on Fig. 3, it can be stated that the increase of cutting speed decreases the value of the $\hat{Ra}$ parameter, but along with the increasing of the feed values, the value of the examined $\hat{Ra}$ parameter increases rapidly. Due to the change in the cutting depth, however, no marked change of examined parameters is achieved.

4 Conclusion

Unlike most scientific publications in this field of research (observation of operating conditions of longitudinal turning process), where is considered manipulating of only one factor at a time and its impact on the response, in our work we focused on the influence of all relevant factors and their interactions. The influence of cutting conditions – input factors x_1, x_2, x_3 (x_1– cutting speed v_c, x_2 – feed f, x_3– depth of cut a_p) on the response (parameters of the resulting surface roughness: $\hat{Rz}$– the maximum height of the roughness profile, $\hat{Ra}$– the mean arithmetic deviation of roughness), i.e. function dependence $\hat{y} = \hat{Rz}, \hat{Ra} = f(x_1, x_2, x_3)$ during longitudinal turning of steel C45 have been investigated. In order to identify significant factors affecting the surface roughness of machined material and analyse the relationships between them, the DOE methodology was used. DOE is very useful for this purpose, whereby all such factors are manipulated simultaneously and fewer experiments are required.

This article clarifies some basic principles of DOE application to improve technological process (specifically the longitudinal turning process). Finally it can be stated that a combination of high cutting speeds, small feeds and small depths of cutting appears the most advantageous. The results obtained by our experimental work have important benefits for technical practice, because they were practically verified under conditions of real production. Results conclusion and interpretation were performed without numerical and statistical incorrectness, what was also confirmed by practical experience in the subject matter field of longitudinal turning of steel C45.

LITERATURE

[1] ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., WIKSTRÖM, C. WOLD, S. *Multi-and Megavariate Data Analysis: Part I.* Umetrics Academy, 2006, ISBN -10:91-973730-2-8, 424 p.

[2] BOX, G. E. P., HUNTER, J. S., HUNTER, W. G. *Statistics for Experimenters*. 2008, p. 639, ISBN 978-0-471-71813-0.

[3] ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., WIKSTRÖM, C. WOLD, S. *Design of Experiments*. Umetrics Academy, 2008, ISBN -10:91-973730-4-4, 459 p.

[4] MORÁVKA, J., MAROŠ, B., MICHALEK, K. Vliv neortogonality plánu experimentu na statistickou korektnost modelu. In *Mezinárodní konference Technical Computing Prague 2008*, 2008, str. 73, http://dsp.vscht.cz/konference_matlab/MATLAB08/

Acknowledgement

The work presented in this paper has been supported by grant VEGA 1/0738/14 of Scientific Grant Agency of the Ministry of Education of Slovak Republic and by grant KEGA 026TUKE-4/2016.