Aplimat

volume III

2010

Journal of Applied Mathematics

ISSN: 1337-6365

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

VOLUME 3 (2010), NUMBER 2

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

VOLUME 3 (2010), NUMBER 2

Edited by: Slovak University of Technology in Bratislava

Editorial Office:
Institute of natural sciences, humanities and social sciences
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava
Námestie slobody 17
812 31 Bratislava

Correspodence concerning subscriptions, claims and distribution:
F.X. spol s.r.o
Azalková 21
821 00 Bratislava
journal@aplimat.com

Frequency: One volume per year consisting of three issues at price of 120 EUR, per volume, including surface mail shipment abroad.
Registration number EV 2540/08

Information and instructions for authors are available on the address:
http://www.journal.aplimat.com/

Printed by: FX spol s.r.o, Azalková 21, 821 00 Bratislava

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 3 (2010), NUMBER 2

DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 3 (2010), NUMBER 2

MODELING AND SIMULATION

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 3 (2010), NUMBER 2

ALGEBRA AND ITS APPLICATIONS

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

LIST OF REVIEWERS

Andrade Marina, Professor Auxiliar	ISCTE - Lisbon University Institute, Lisboa, Portugal
Ayud Sani Muhammed	NGO, University of Cape Cost , Accra
Barbagallo Annamaria, Dr.	University of Catania, Italy
Bartošová Jitka, RNDr., PhD.	University of Economics Prague, Jindřichův Hradec, Czech Republic
Bayaraa Nyamish,	Ulaanbaatar, Mongolia
Beránek Jaroslav, Doc. RNDr. CSc.	Masaryk University, Brno, Czech Republic
Brueckler Franka Miriam, DrSc.	University of Zagreb, Zagreb, Poland
Budíková Marie, RNDr., Dr.	Masarykova univerzita, Brno, Czech Republic
Budkina Natalja, Dr. math	Riga, Latvia Technical university, Riga, Latvia
Buy Dmitriy B., PhD.	Kyiv Taras Shevchenko National University, Kyiv
Conversano Elisa	Architettura, Roma, Italy
Csatáryová Mária, RNDr., PhD.	University of Prešov, Prešov, Slovak Republic
Daniele Patrizia, Associate Professor	University of Catania, Italy,
Dobrakovová Jana, Doc., RNDr., CSc.	Slovak University of Technology in Bratislava, Slovak Republic
Doležalová Jarmila, Doc., RNDr., CSc.	VŠB - TU, Ostrava, Czech Republic
Dorociaková Božena, RNDr., PhD.	University of Žilina, Žilina, Slovak Republic
Doupovec Miroslav, doc. RNDr.. CSc.	University of Technology, Brno, Czech Republic
Ferreira Manuel Alberto M., Professor Catedrático	ISCTE- Lisbon University Institute , Lisboa, Portugal
Filipe José António, Prof. Auxiliar	ISCTE - Lisbon University Institute, Lisboa, Portugal

Francaviglia Mauro, Full Prof.	University of Calabria, Calabria, Italy
Franců Jan, prof. RNDr., CSc.	University of Technology, Brno, Czech Republic
Fulier Jozef, prof. RNDr. CSc.	UKF Nitra, Slovak Republic
Habiballa Hashim, RNDr., PaedDr., PhD.	University of Ostrava, Ostrava, Czech Republic
Hennyeyová Klára, doc. Ing. CSc.	Slovak University of Agriculture, Nitra, Slovak Republic
Hinterleitner Irena, PhD.	University of Technology, Brno, Czech Republic
Hošková Šárka, Doc. RNDr., PhD.	University of Defence, Brno, Czech Republic
Hunka Frantisek, Doc. (Assoc. Prof)	University of Ostrava, Ostrava, Czech Republic
Chvosteková Martina, Mgr.	Slovak Academy of Science, Bratislava, Slovak Republic
Jan Chvalina, Prof. DrSc.	University of Technology, Brno, Czech Republic
Janková Mária, Mgr.	Slovak Academy of science, Bratislava, Slovak Republic
Jukic Ljerka	University of Osijek, Osijek, Croatia
Kalina Martin, Doc. RNDr., CSc.	Slovak University of Technology in Bratislava, Slovak Republic
Kapička Vratislav, Prof. RNDr., DrSc.	Masaryk university, Brno, Czech Republic
Klobucar Antoaneta,	University in Osijek, Osijek, Croatia
Klůfa Jindřich, prof. RNDr., CSc.	University of Economics, Prague, Czech Republic
Kováčová Monika, Mgr., PhD.	Slovak University of Technology in Bratislava, Slovak Republic
Kulčár Ladislav, Doc. RNDr. CSc.	M. Bel University Banská Bystrica, Poprad, Slovak Republic
Kureš Miroslav, Assoc. Prof.	Brno, Czech Republic University of Technology, Brno, Czech Republic
Kvasz Ladislav, Doc. PhD. (Assoc. Prof)	University Komenskeho, Bratislava, Slovak Republic
Lopes Ana Paula, PhD.	Polytechnic Institute of Oporto – IPP, Oporto, Portugal
Marcella Giulia Lorenzi, prof.	Università della Calabria, Calabria, Italy

Lungu Nicolae, Prof.	Technical University of Cluj- Napoca, Cluj-Napoca, Romania
Majstorović Snježana	J.J.Strossmayer University, Osijek, Croatia
Mařík Robert, doc. PhD.	Mendel University in Brno, Czech Republic, Brno, Czech Republic
Matvejevs Andrejs, Ing., DrSc.	Riga, Latvia Technical university, Riga, Latvia
Michálek Jiří, RNDr, CSc.	Czech Academy of Sciences, Prague, Czech Republic
Mikeš Josef, Prof. RNDr., DrSc.	Palacky University, Olomouc Czech Republic
Myšková Helena, RNDr., PhD.	Technical University of Košice, Košice
Nemoga Karol, doc. RNDr., PhD.	Slovak Academy of Sciences, Bratislava, Slovak Republic
Neuman František, Prof., RNDr., DrSc.	Mathematical Institute of the Academy, Brno, Czech Republic
Omachelová Milada, Mgr., PhD	Slovak University of Technology in Bratislava, Slovak Republic
Paun Marius, Prof. Dr.	Transilvania University of Brasov, Brasov, Romania
Pavlenko Oksana, Dr. math	Riga, Latvia Technical university, Riga, Latvia
Plavka Ján	Technical University Košice, Košice, Slovak Republic
Plch Roman, RNDr., PhD.	Masaryk University, Brno, Czech Republic
Pokorný Michal, Prof.	University of Zilina, Zilina, Slovak Republic
Potůček Radovan, RNDr., PhD.	University of Defense, Brno, Czech Republic
Půlpán Zdeněk, Prof., RNDr., PhD.r., CSc.	University of Hradec Králové, Hradec Králové, Czech Republic
Račková Pavlína, PhD.r., PhD.	University of Defense, Brno, Czech Republic
Rus Ioan A., Professor	Babes-Bolyai University, Cluj-Napoca, Romania
Růžičková Miroslava, RNDr. Doc. CSc.	University of Žilina, Slovak Republic, Žilina, Slovak Republic
Segeth Karel, Prof., RNDr., CSc.	Academy of Sciences, Praha, Czech Republic
Stankovičová Iveta, Ing., PhD.	Univerzita Komenského, Bratislava, Slovak Republic

Sterba Jan, Ing.	University of Economics in Bratislava, Slovak Republic
Svatošová Libuše, Prof. Ing., CSc.	Czech University of Life Sciences, Praha, Czech Republic
Šooš Ľubomír, Prof. Ing., PhD.	Slovak University of Technology in Bratislava, Slovak Republic
Šuch Ondrej, PhD.	Inštitút Matematiky a Informatiky, Banská Bystrica, Slovak Republic
Tarabella Leonello, Researcher	Research Area pf CNR, Pisa, Italy
Terek Milan, prof., Ing., PhD.	University of Economics, Bratislava, Slovak Republic
Trokanova Katarina, Doc.	Slovak University of Technology in Bratislava, Slovak Republic
Tseytlin Georgiy E., PhD.	Institute of Program System , Kyiv, Ukraine
Urban František, Doc. Ing., CSc.	Slovak University of Technology, Bratislava, Slovak Republic
Vanžurová Alena, Doc. RNDr., CSc.	Palacký University, Olomouc, Czech Republic
Velichová Daniela, Doc. RNDr., CSc.	Slovak University of Technology, Bratislava, Slovak Republic
Wikovsky Viktor, Doc. RNDr., CSc.	Slovak Academy of Sciences, Bratislava, Slovak Republic
Wimmer Gejza, Mgr.	Slovak Academy of Sciences, Bratislava, Slovak Republic

Journal of Applied Mathematics

volume III (2010), number II

MEMORY TERM FOR DYNAMIC OLIGOPOLISTIC MARKET EQUILIBRIUM PROBLEM

BARBAGALLO Annamaria, (I), MAUGERI Antonino, (I)

Abstract. The aim of the paper is to provide a more realistic model for the dynamic oligopolistic market equilibrium problems. In particular, a long-term memory is introduced and the corresponding variational inequality model is discussed in order to study the problem in presence of delay. Moreover, existence and regularity results are proved.

Key words and phrases. Oligopolistic market equilibrium problem, evolutionary variational inequality, existence and continuity, memory term.

Mathematics Subject Classification. Primary 49J40, 49K40, 91B62

1 Introduction

The theory of elastic body was pioneered by Boltzman [9, 10] who gave a first mathematical formulation to hereditary phenomena, where the deformations of a body are studied along with the history of the deformations under which it was subjected in the past. Later Volterra [22, 23, 24] gave his contribution to the theory of elasticity, introducing some hereditary coefficients in form of integral term in the constitutive equations for an elastic body with memory. Starting from the 1960s, see [11, 13], the principle of fading memory was advanced, suggesting that a body is able to recollect only its recent past and thus all the history before can be neglected. As a consequence, the memory term represents only the history in the time interval $(0, t)$ and all the previous events cannot affect the body behavior.

Since then, notable applications in different fields have been studied. In economics we many refer to [8], where the dynamics of market adjustment processes are described via a Volterra integral term. More recently, the integral memory term has been used in order to represent some physical characteristics of the quantities involved in mechanical and engineering problems.

For instance, it may describe the relaxation tensor in viscoelastic contact models as in [1], or the conductivity of an electrolyte in electrochemical machining as in [14, 21].

Inspired by these applied problems, in [18] the authors suggest to introduce the memory term in the framework of network equilibria, thus leading to a refinement of the model. In fact, they explicitly incorporate the contribution of flows from the initial time to the observation time t, which causes the presence of the memory term. Hence we are able to analyze how the current equilibrium solution is affected by past equilibria. The same suggestion can concern the dynamic oligopolistic market equilibrium problem. This problem due to Cournot in [12], recently, has been studied in the dynamic case. The dynamic framework has been developed by [6]. In this paper, the variational formulation has been proved, from which the existence and regularity of the dynamic oligopolistic market equilibrium solution has been derived. The continuity of solution allows to provide a computational procedure to compute the equilibrium solution. In the subsequently paper [7], applying the infinite-dimensional duality results developed in [17], the existence of the Lagrange variables, which allow to describe the behaviour of the market, is provided. Furthermore, some sensitivity results has been proved showing that small changes of the solution happen in correspondence of small changes of the profit function.

2 Dynamic Oligopolistic Market Equilibrium

Here we describe, for the reader's convenience, a dynamic spatial oligopolistic market equilibrium problem, which constitutes an example of imperfect competition and can be viewed as a prototypical game theoretic problem, operating under the Nash equilibrium concept of noncooperative behaviour.

Let us consider m firms P_i, $i = 1, 2, \dots, m$, and n demand markets Q_j, $j = 1, 2, \dots, n$, that are generally spatially separated. Assume that the homogeneous commodity, produced by the m firms and consumed at the n markets, is involved during a period of time $[0, T]$, $T > 0$. Let $p_i(t)$, $t \in [0, T]$, $i = 1, 2, \dots, m$, denote the nonnegative commodity output produced by firm P_i at the time $t \in [0, T]$ and let $q_j(t)$, $t \in [0, T]$, $j = 1, 2, \dots, n$, denote the demand for the commodity at demand market Q_j at the same time $t \in [0, T]$. Let $x_{ij}(t)$, $t \in [0, T]$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, denote the nonnegative commodity shipment between the supply market P_i and the demand market Q_j at the time $t \in [0, T]$. Group the production output into a vector-function $p : [0, T] \to \mathbb{R}^m_+$, the demands into a vector-function $q : [0, T] \to \mathbb{R}^n_+$, and the commodity shipments into a matrix-function $x : [0, T] \to \mathbb{R}^{mn}_+$.

Assuming that we are not in presence of production and demand excesses the following feasibility conditions must hold for every i and j and a.e. in $[0, T]$:

$$p_i(t) = \sum_{j=1}^{n} x_{ij}(t), \tag{1}$$

$$q_j(t) = \sum_{i=1}^{m} x_{ij}(t). \tag{2}$$

Hence, the quantity produced by a firm, at the time $t \in [0, T]$, must be equal to the sum of the commodity from that firm to all the demand markets, at the same $t \in [0, T]$, and the demand at a demand market, at the time $t \in [0,T]$, must be equal to the sum of all the commodity shipments to that demand market, at the same $t \in [0,T]$. Furthermore, assuming that the nonnegative commodity shipment between the supply market P_i and the demand market Q_j has to satisfy time-dependent constrains, namely that:

$$\underline{x}_{ij}(t) \le x_{ij}(t) \le \overline{x}_{ij}(t), \quad \forall i = 1,2,\dots,m, \; j = 1,2,\dots,n, \text{ a.e. in } [0,T], \tag{3}$$

where $\underline{x}, \overline{x} \in L^2([0,T], \mathbb{R}^{mn})$, with $\underline{x}(t) < \overline{x}(t)$ a.e. in $[0,T]$, are considered known with non-negative values. In this paper we consider the functional setting of the Hilbert space $L^2([0,T], \mathbb{R}^{mn})$. Hence, the set of feasible vectors $x \in L^2([0,T], \mathbb{R}^{mn})$ is

$$\mathbb{K} = \Big\{ x \in L^2([0,T], \mathbb{R}^{mn}) : \; 0 \le \underline{x}(t) \le x(t) \le \overline{x}(t), \text{ a.e. in } [0,T] \Big\}. \tag{4}$$

This set is convex, closed and bounded in the Hilbert space $L^2([0,T], \mathbb{R}^{mn})$.

Furthermore, associate with each firm P_i a production cost f_i and let us consider the more general situation where the production cost of the quantity p_i a firm i may depend upon the entire production pattern and the time, namely we assume that

$$f_i = f_i(t, p(t)). \tag{5}$$

Similarly, allow the demand price for the commodity at a demand market to depend, in general, upon the entire consumption pattern and the time, namely we assume that

$$d_j = d_j(t, q(t)), \tag{6}$$

where we have denoted with p and q the vector-functions given by (1) and (2).

Then, we have the following mappings:

$$f : [0,T] \times L^2([0,T], \mathbb{R}^m_+) \to L^2([0,T], \mathbb{R}^m_+),$$

$$d : [0,T] \times L^2([0,T], \mathbb{R}^n_+) \to L^2([0,T], \mathbb{R}^n_+).$$

Let c_{ij}, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$, denote the transaction cost at the same time $t \in [0,T]$, which includes between firm P_i and demand market Q_j. Here we permit the transaction cost to depend, in general, upon the entire shipment pattern and the time, namely,

$$c_{ij} = c_{ij}(t, x(t)), \tag{7}$$

so we have

$$c : [0,T] \times L^2([0,T], \mathbb{R}^{mn}_+) \to L^2([0,T], \mathbb{R}^{nm}_+).$$

The profit $v_i(t, x(t))$, $t \in [0,T]$, $i = 1,2,\dots,m$, of firm P_i at the same time $t \in [0,T]$ is then

$$v_i(t, x(t)) = \sum_{j=1}^{n} d_j(t, q(t)) x_{ij}(t) - f_i(t, p(t)) - \sum_{j=1}^{n} c_{ij}(t, x(t)) x_{ij}(t). \tag{8}$$

Now consider the dynamic oligopolistic market mechanism, in which the m firms supply the commodity in a noncooperative fashion, each one trying to maximize its own profit at the time $t \in [0,T]$. We seek to determine a nonnegative commodity distribution pattern x for which the m firms will be in a state of equilibrium as defined below.

Definition 2.1 *A commodity shipment distribution $x^* \in \mathbb{K}$ is a dynamic oligopolistic market equilibrium if and only if for each $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ and a.e. in $[0, T]$ we have:*

$$v_i(t, x^*(t)) \geq v_i(t, x_i(t), \widehat{x}_i^*(t)), \quad \forall x \in \mathbb{K}, \text{ a.e. in } [0, T], \tag{9}$$

where we denote by $\widehat{x}_i^* = (x_1^*, \ldots, x_{i-1}^*, x_{i+1}^*, \ldots, x_m^*)$, and we set

$$v_i(t, x_i(t), \widehat{x}_i^*(t)) = v_i(t, x_1^*(t), \ldots, x_{i-1}^*(t), x_i(t), x_{i+1}^*(t), \ldots, x_m^*(t)).$$

In the Hilbert space $L^2([0, T], \mathbb{R}^k)$, let us recall that

$$\ll \phi, y \gg := \int_0^T \langle \phi(t), y(t) \rangle dt,$$

is its duality mapping, where $\phi \in (L^2([0, T], \mathbb{R}^k))^* = L^2([0, T], \mathbb{R}^k)$ and $y \in L^2([0, T], \mathbb{R}^k)$.

We recall at this point that a continuously differentiable function v (as for example given by (8)) is called **pseudoconcave** with respect to x_i, $i = 1, 2, \ldots, m$, (see [15]) if the following holds a.e. in $[0, T]$:

$$\langle \frac{\partial v}{\partial x_i}(t, x_1, \ldots, x_i, \ldots x_n), x_i - y_i \rangle \geq 0 \implies v(t, x_1, \ldots, x_i, \ldots x_n) \geq v(t, x_1, \ldots, y_i, \ldots x_n),$$

In the paper [6], the authors have shown that the equilibrium problem defined by (9) can be reformulated as an evolutionary variational inequality. In fact, the following result holds:

Theorem 2.2 *Assume that for each firm P_i the profit function $v_i(t, x(t))$ is pseudoconcave with respect to the variables $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$, $i = 1, 2, \ldots, m$, and continuously differentiable for a.e. $t \in [0, T]$. Assume that ∇v is a Carathèodory function such that*

$$\exists h \in L^2([0, T], \mathbb{R}) : \ \|\nabla v(t, u(t))\| \leq h(t)\|u(t)\|, \quad \forall u \in L^2([0, T], \mathbb{R}^{mn}). \tag{10}$$

Then $x^* \in \mathbb{K}$ is a dynamic Cournot-Nash equilibrium if and only if it satisfies the evolutionary variational inequality

$$\ll -\nabla v(x^*), x - x^* \gg \geq 0, \quad \forall x \in \mathbb{K}. \tag{11}$$

Corollary 2.3 *Assume that for each firm P_i the profit function $v_i(t, x(t))$ is pseudoconcave with respect to the variables $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$, and continuously differentiable, for a.e. $t \in [0, T]$, and ∇v is a Carathèodory function such that*

$$\exists h \in L^2([0, T], \mathbb{R}) : \|\nabla v(t, u(t))\| \leq h(t)\|u(t)\|, \quad \forall u \in L^2([0, T], \mathbb{R}^{mn}).$$

Then the evolutionary variational inequality (11) *is equivalent to*

$$\langle -\nabla v(t, x^*(t)), x(t) - x^*(t) \rangle \geq 0, \quad \forall x(t) \in \mathbb{K}(t), \text{ a.e. in } [0, T], \tag{12}$$

where

$$\mathbb{K}(t) = \Big\{ x(t) \in \mathbb{R}^{mn} : \ 0 \leq \underline{x}(t) \leq x(t) \leq \overline{x}(t) \Big\}.$$

This equivalence is very important in the construction of a discretization procedure to compute numerical solutions for dynamic oligopolistic market equilibrium problems.

3 The memory term in dynamic oligopolistic market equilibria

Let us consider the dynamic oligopolistic market equilibrium problem, that is expressed by means of the evolutionary variational inequality (11):

$$x^* \in \mathbb{K}: \quad \int_0^T \langle -\nabla v(t, x^*(t)), x(t) - x^*(t) \rangle dt \geq 0, \quad \forall x \in \mathbb{K},$$

where $\mathbb{K} = \{x \in L^2([0,T], \mathbb{R}^{nm}): \ \underline{x}(t) \leq x(t) \leq \overline{x}(t), \ \text{ a.e. in } [0,T]\}$.

Therefore, it turns out that the first effect of the presence of the integral term is the adjustment of the operator $v(t, x(t))$ which become

$$v(t, x(t)) + \int_0^t I(t-s)x(s)ds,$$

where $I = [I^r]_{r=1,\dots,m}$ is a vector of nonnegative defined $m \times n$ matrixes with entries $I^r_{ij} \in L^2([0,T], \mathbb{R})$.

This means that the commodity shipments do not only incur in the current-time operator, but are also subject to the impact of all previous equilibrium solutions. As a consequence, the equilibrium conditions are required on the full operator

$$F_r(t, x(t)) = -\nabla\Big(v_r(t, x(t)) - \int_0^t \sum_{i=1}^m \sum_{j=1}^n I^r_{ij}(t-s)x_{ij}(s)ds\Big), \quad r = 1, \dots, m.$$

It is also worth emphasizing the role of the matrixes $I^r(t-s)$, for $r = 1, \dots, m$. In fact, the entries of the matrix I^r_{ij} can be regarded as continuous weights acting on solutions and allow us to represent the history of the past equilibrium patterns and their influence on the current one. The meaning of the integral term is then justified: it expresses, by means of a relaxation over the time interval $(0, t)$, the equilibrium distribution in which commodity shipments incur at time t, and, hence, the effect of the previous framework situation on the present one.

The memory term is also strictly connected with the concept of time shifts and delay patterns. In fact, the integral term represents the displacement, namely the delay, of the equilibrium solution commodity shipments, due to the previous equilibrium state. Therefore, delay effects are not only regarded as perturbation factors for the constraint set, see [20] in connection with traffic network problems, but can also be interpreted as adjustment factors of operators.

In this case the variational inequality problem has the form

$$x^* \in \mathbb{K}: \quad \int_0^T \Big\langle -\nabla\Big[v(t, x^*(t)) + \int_0^t I(t-s)x^*(s)ds\Big], x(t) - x^*(t)\Big\rangle dt \geq 0, \quad \forall x \in \mathbb{K}. \quad (13)$$

The resulting problem explicitly takes account of the contribution of the equilibrium solution from the initial time to the observation time and include it in the operator as an adjustment factor.

Now, we want to investigate on the explicit form of the gradient of the memory-long term. We apply the definition of the Fréchet derivative, and we obtain:

$$\nabla_{x_{hk}}\Big[\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{r}(t-s)x_{ij}(s)\Big] = \sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{r}(t-s)\delta_{ih}\delta_{jk} = I_{hk}^{r}(t-s).$$

Then, we get

$$\int_0^t \nabla_{x_{hk}}\Big[\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{r}(t-s)x_{ij}(s)\Big]ds = \int_0^t I_{hk}^{r}(t-s)ds = \widetilde{I}_{hk}^{r}(t),$$

where we denote by $\widetilde{I}_{hk}^{r}(t)$ the final operator. We observe that the previous operators are independent on the matrix of the commodity shipments. The last remark implies that the operator $F(t,x(t))$ is given by the sum of the gradient of the profit function and a function dependent only on the time, in particular:

$$F_r(t,x(t)) = -\nabla v_r(t,x(t)) - \widetilde{I}^{r}(t), \quad r = 1,\dots,m.$$

4 Existence and regularity results

In this section a theorem for the existence of continuous solutions to the dynamic oligopolistic market problem is provided.

First of all, we recall some definitions (see [16]):

Definition 4.1 *A mapping $A : \mathbb{K} \to X^*$ is pseudomonotone in the sense of Brezis (B-pseudomonotone) iff*

1. *For each sequence u_n weakly converging to u (in short $u_n \rightharpoonup u$) in $\mathbb{K}$ and such that $\limsup_n \langle Au_n, u_n - v\rangle \le 0$ it results that:*

$$\liminf_n \langle Au_n, u_n - v\rangle \ge \langle Au, u-v\rangle, \quad \forall v \in K.$$

2. *For each $v \in \mathbb{K}$ the function $u \to \langle Au, u-v\rangle$ is lower bounded on the bounded subset ok $\mathbb{K}$.*

Definition 4.2 *A mapping $A : \mathbb{K} \to E^*$ is hemicontinuous in the sense of Fan (F-hemicontinuous) iff for all $v \in \mathbb{K}$ the function $u \to \langle Au, u-v\rangle$ is weakly lower semicontinuous on $\mathbb{K}$.*

Moreover we recall the following other kind of continuity, which will be used together with some kind of monotonicity assumptions:

Definition 4.3 *A mapping $A : \mathbb{K} \to E^*$ is hemicontinuous along line segments, iff the function $t \to \langle A(tu + (1-t)v), w\rangle$, $t \in [0,1]$ is continuous for all $u, v \in \mathbb{K}$.*

Definition 4.4 *A mapping $A : \mathbb{K} \to E^*$ is lower hemicontinuous along line segments, iff the function $\xi \to \langle A\xi, u - v\rangle$ is lower semicontinuous for all $u, v \in \mathbb{K}$ on the line segments $[u, v]$.*

Definition 4.5 *The map $A : \mathbb{K} \to E^*$ is said to be pseudomonotone in the sense of Karamardian (K-pseudomonotone) iff for all $u, v \in \mathbb{K}$*

$$\langle Av, u - v\rangle \geq 0 \implies \langle Au, u - v\rangle \geq 0$$

Now, taking into account that our constraint set $\mathbb{K}$ is nonempty, convex and weakly compact, we can present existence results for the dynamic oligopolistic market equilibrium problem.

Theorem 4.6 *If $-\nabla v$ is a B-pseudomonotone mapping and I is a vector of nonnegative defined $m \times n$ matrixes with entries $I^r_{ij} \in L^2([0, T], \mathbb{R})$, then variational inequality problem* (13) *admits solutions.*

To ensure the uniqueness of the solution, we must suppose that the operator $-\nabla v$ is strictly pseudomonotone, namely $\forall x(t) \neq y(t)$, a.e. in $[0, T]$,

$$\langle -\nabla v(t, y(t)), x(t) - y(t)\rangle \geq 0 \implies \langle -\nabla v(t, x(t)), x(t) - y(t)\rangle > 0.$$

Theorem 4.7 *If $-\nabla v$ is an F-hemicontinuous mapping and I is a vector of nonnegative defined $m\times n$ matrixes with entries $I^r_{ij} \in L^2([0, T], \mathbb{R})$, then variational inequality* (13) *admits solutions.*

Theorem 4.8 *If $-\nabla v$ a K-pseudomonotone map which is lower hemicontinuous along line segments and I is a vector of nonnegative defined $m\times n$ matrixes with entries $I^r_{ij} \in L^2([0, T], \mathbb{R})$, then variational inequality* (13) *admits solutions.*

We point out that condition (10) ensures the lower hemicontinuity along line segments of $-\nabla v$. Now, let us recall some conditions under which the dynamic oligopolistic market equilibrium problem has continuous solutions. To do so, we first need to recall the definition of set convergence given by U. Mosco (see [19]).

Definition 4.9 *Let $(X, \|\cdot\|)$ be an Hilbert space and $\mathbb{K}$ a closed, nonempty, convex subset of X. A sequence of nonempty, closed, convex sets $\mathbb{K}_n$ converges to $\mathbb{K}$ in Mosco's sense, as $n \to +\infty$, i.e. $\mathbb{K}_n \to \mathbb{K}$, if and only if*

(M1) for any $x \in \mathbb{K}$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ strongly converging to x in X such that x_n lies in $\mathbb{K}_n$ for all $n \in \mathbb{N}$,

(M2) for any subsequence $\{x_{k_n}\}_{n\in\mathbb{N}}$ weakly converging to x in X, such that x_{k_n} lies in $\mathbb{K}_{k_n}$ for all $n \in \mathbb{N}$, then the weak limit x belongs to $\mathbb{K}$.

The following result holds (see [6]).

Lemma 4.10 *Let $\underline{x}, \overline{x} \in C([0, T], \mathbb{R}^{mn}_+)$, and let $\{t_n\}_{n\in\mathbb{N}}$ be a sequence such that $t_n \to t \in [0, T]$, as $n \to +\infty$. Then, the sequence of sets $\mathbb{K}(t_n) = \big\{x(t_n) \in \mathbb{R}^{mn} : 0 \leq \underline{x}(t_n) \leq x(t_n) \leq \overline{x}(t_n)\big\}$, $\forall n \in \mathbb{N}$, converges to $\mathbb{K}(t) = \big\{x(t) \in \mathbb{R}^{mn} : 0 \leq \underline{x}(t) \leq x(t) \leq \overline{x}(t)\big\}$, as $n \to +\infty$, in Mosco's sense.*

We remark that $\mathbb{K}(t)$, for $t \in [0,T]$, is uniformly bounded provided that $\underline{x}$ and $\overline{x}$ are two continuous matrix-functions.

Now, we able to show the continuity of the solution to the evolutionary variational inequality that models our retarded dynamic oligopolistic market equilibrium problem. The result is a generalized version of Theorem 4.2 in [6].

Theorem 4.11 *Assume that for each firm P_i the profit function $v(t, x(t))$ is strictly pseudo-concave with respect to the variable x for a.e. $t \in [0,T]$, and belonging to $C^1([0,T] \times \mathbb{R}^{mn}_+, \mathbb{R})$ and I is a vector of nonnegative defined $m \times n$ matrixes with entries $I^r_{ij} \in L^2([0,T], \mathbb{R})$. Assume ∇v is a Carathèodory function such that*

$$\exists h \in L^2([0,T], \mathbb{R}) : \|\nabla v(t, u(t))\| \leq h(t)\|u(t)\|, \quad \forall u \in L^2([0,T], \mathbb{R}^{mn}. \tag{14}$$

Then, the unique dynamic Cournot-Nash equilibrium $x^* \in \mathbb{K}$ is continuous in $[0,T]$.

Analogous results have been proved for the parametric variational inequalities (see [5, 18]) and the dynamic traffic equilibrium problem (see [2, 3, 4]).

5 Lipschitz continuity result

This section is devoted to show a Lipschitz continuity result for the dynamic oligopolistic market equilibrium problem. More precisely, we apply a general result proved in [18] to our model. This theorem establishes the Lipschitz continuity of the solution to the following abstract parameterized variational inequality problem:

$$\langle F(t, x^*(t)), x - x^*(t)\rangle \geq 0, \quad \forall x \in K(t), t \in [0,T], \tag{15}$$

where the constraint set $K(t)$, $t \in [0,T]$, is a closed convex and nonempty subset of $\mathbb{R}^n$, $F : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ is a point-to-point mapping, and $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^n$. Specifically, the following result holds (see [18], Theorem 1):

Theorem 5.1 *Let the following assumptions be satisfied:*

(a) F is strongly monotone, i.e., there exists $\alpha > 0$ such that for $t \in [0,T]$, $\langle F(t, x_1) - F(t, x_2), x_1 - x_2\rangle \geq \alpha\|x_1 - x_2\|^2$, $\forall x_1, x_2 \in \mathbb{R}^n$;

(b) F is Lipschitz continuous with respect to t, i.e., there exists $\beta > 0$ such that, for $t \in [0,T]$, $\|F(t, x_1) - F(t, x_2)\| \leq \beta\|x_1 - x_2\|$, $\forall x_1, x_2 \in \mathbb{R}^n$;

(c) F is Lipschitz continuous with respect to t, i.e., there exists $M > 0$ such that, for $t_1, t_2 \in [0,T]$, $\|F(t_2, x) - F(t_1, x)\| \leq M\|x\||t_2 - t_1|$, $\forall x \in \mathbb{R}^n$;

(d) there exists $\kappa \geq 0$ such that, for $t_1, t_2 \in [0,T]$, $\|P_{K(t_1)}(z) - P_{K(t_2)}(z)\| \leq \kappa|t_1 - t_2|$, $\quad \forall z \in \mathbb{R}^n$, where $P_{K(t)}(z) = \arg\ \min_{x \in K(t)} \|z - x\|$, $t \in [0,T]$ denotes the projection onto the set $K(t)$.

Then, the unique solution $x^*(t)$, $t \in [0,T]$, to (15) *is Lipschitz continuous on* $[0,T]$, $t_1 \neq t_2$, *the following estimate holds:*

$$\frac{\|x^*(t_2) - x^*(t_1)\|^2}{|t_2 - t_1|^2} \leq \gamma\Big(\|x^*\|^2_{C^0([0,T],\mathbb{R}^n} + \sup_{\substack{t_1,t_2 \in [0,T] \\ t_1 \neq t_2}} \Big\| \frac{P_{K(t_2)}(z) - P_{K(t_1)}(z)}{t_2 - t_1} \Big\|^2\Big), \tag{16}$$

where $\gamma = \gamma(\alpha, \beta, M, T.L)$.

In order to achieve the Lipschitz continuity of the dynamic market equilibrium solution, we need to estimate the variation rate of projections onto time-dependent constraint set describing the oligopolistic market equilibrium problem. It is noteworthy that $\mathbb{K}(t)$ can be reduced to the case where $0 \leq x_{ij}(t) \leq x^*_{ij}(t)$, $x^*_{ij}(t) = x_{ij}(t) - \underline{x}_{ij}(t)$, $i = 1, \dots, m$ and $j = 1, \dots, n$, with the transformation $x'_{ij}(t) = x_{ij}(t) - \underline{x}_{ij}(t)$, $i = 1, \dots, m$ and $j = 1, \dots, n$. We are able to show that assumption (d) of Theorem 5.1 is fulfilled. In fact we have the following result.

Proposition 5.2 *Let us suppose that x^* is Lipschitz continuous on $[0,T]$ with constant L and let z be an arbitrary point in $\mathbb{R}^n$. Then it results*

$$\|P_{\mathbb{K}(t_2)}(z) - P_{\mathbb{K}(t_1)}(z)\| \leq L|t_1 - t_2|,$$

where $\mathbb{K}(t) = \{x(t) \in \mathbb{R}^{mn} : \ 0 \leq x(t) \leq x^*(t)\}$.

Then, the Lipschitz continuity of the solution to (13) is ensured assuming that the operator $-\nabla v$ satisfies conditions (a), (b), (c) of Theorem 5.1 and I is a vector of nonnegative defined $m \times n$ matrixes with entries $I^r_{ij} \in L^2([0,T], \mathbb{R})$.

References

[1] ARÓS, A., SOFONEA, M., VIAÑO, J.M.: *A class of evolutionary variational inequalities with Volterra-type term.* In Mathematical Models and Methods in Applied Sciences, Vol. 14, No. 4, pp. 557–577, 2004.

[2] BARBAGALLO, A.: *Regularity results for time-dependent variational and quasi-variational inequalities and computational procedures.* In Mathematical Models and Methods in Applied Sciences, Vol. 17, No. 2, pp. 277–304, 2007.

[3] BARBAGALLO, A.: *Regularity results for evolutionary nonlinear variational and quasi-variational inequalities with applications to dynamic equilibrium problems.* In Journal of Global Optimization, Vol. 40, No. 1–3, pp. 29–39, 2008.

[4] BARBAGALLO, A.: *Existence and regularity of solutions to nonlinear degenerate evolutionary variational inequalities with applications to traffic equilibrium problem.* In Applied Mathematics and Computation, Vol. 208, pp. 1–13, 2009.

[5] BARBAGALLO, A., COJOCARU, M.-G.: *Continuity of solutions for parametric variational inequalities in Banach space.* In Journal of Mathematical Analysis and Applications, Vol. 351, pp. 707–720, 2009.

[6] BARBAGALLO, A., COJOCARU, M.-G.: *Dynamic equilibrium formulation of oligopolistic market problem*. In Mathematical and Computer Modelling, Vol. 49, pp. 966–976, 2009.

[7] BARBAGALLO, A., MAUGERI, A.: *Duality theory for the dynamic oligopolistic market equilibrium problem*. To appear in Optimization.

[8] BECKMAN, M.J., WALLACE, J.P.: *Continuous lags and the stability of market equilibrium*. In Economica, New Series, Vol. 36, No. 141, pp. 58–68, 1969.

[9] BOLTZMANN, L.: *Zur Theorie der elastichen Nachwirkung*. In Sitzber. Kaiserl. Akad. Wiss. Wien, Math.-Naturw. Kl., 70, Sect. II (1874), 275–300.

[10] BOLTZMANN, L.: *Zur Theorie der elastichen Nachwirkung*. In Ann. Phys. u. Chem., Vol. 5, pp. 430–432, 1878.

[11] COLEMAN, B.D., NOLL, W.: *Foundations of linear viscoelasticity*. In Rev. of Modern Physics, Vol. 33, No. 2, pp. 239–249, 1961.

[12] COURNOT, A.: *Researches into the mathematical principles of the theory of wealth*. English translation MacMillan, London, 1838.

[13] FICHERA, G.: *Integro-differential problems of hereditary elasticity*. In Atti del Sem. Mat. Fis. Univ. Modena, Vol. XXVI, pp. 363–370, 1977.

[14] GWINNER, J.: *Time-dependent variational inequalities. Some recent trends*. In Daniele P., Giannessi F. and Maugeri A. (Eds.), *Equilibrium problems and variational models*, Academic Publishers, Dordrecht, The Netherlands, pp. 225–264, 2003.

[15] MANGASARIAN, O.: *Pseudo-convex functions*. In J. Soc. Indust. Appl. Math. Ser. A Control, Vol. 3, pp. 281–290, 1965.

[16] MAUGERI, A., RACITI, F.: *On existence theorems for monotone and nonmonotone variational inequalities*. In Journal of Convex Analysis, Vol. 16, pp. 899–911, 2009.

[17] MAUGERI, A., RACITI, F.: *Remarks on Infinite Dimensional Duality*. In Journal of Global Optimization, DOI 10.1007/s108980099442y.

[18] MAUGERI, A., SCRIMALI, L.: *Global Lipschitz Continuity of Solutions to Parameterized Variational Inequalities*. Bollettino dell'Unione Matematica Italiana, Vol. II, No. 9, pp. 45–69, 2009.

[19] RODRIGUES, J.F.: *Obstacle Problems in Mathematical Physics*. North-Holland, Amsterdam, 1987.

[20] SCRIMALI, L.: *Quasi-variational inequalities in transportation networks*.In Mathematical Models and Methods in Applied Sciences, Vol. 14, No. 10, pp. 1541–1560, 2004.

[21] STEINBACH, J.: *On a variational inequality containing a memory term with an application in electro-chemical machining*. In Journal of Convex Analysis, Vol. 5, No. 1, pp. 63–80, 1998.

[22] VOLTERRA, V.: *Sulle equazioni integro-differenziali della teoria della elasticità*. In Rendiconti dell'Accademia Nazionale dei Lincei, Vol. XVIII, No. 2, pp. 295–301, 1909.

[23] VOLTERRA, V.: *Sulle equazioni integro-differenziali della teoria della elasticità nel caso della entropia*. In Rendiconti dell'Accademia Nazionale dei Lincei, Vol. XVIII, No. 2, pp. 577–586, 1909.

[24] VOLTERRA, V.: *Sur les equation intégro-differentielles et leurs applications*. In Acta Math., Vol. XXXV, pp. 295–356, 1912.

Current address

Barbagallo Annamaria, Dr.
Department of Mathematics and Computer Science, University of Catania,
viale Andrea Doria n. 6 - 95125 CATANIA, Italy
tel. number: +390957383094, email: barbagallo@dmi.unict.it

Maugeri Antonino, Prof.
Department of Mathematics and Computer Science, University of Catania,
viale Andrea Doria n. 6 - 95125 CATANIA, Italy
tel. number: +390957383059, email: maugeri@dmi.unict.it

Journal of Applied Mathematics

volume III (2010), number II

STABILITY AND ESTIMATION OF SOLUTIONS OF LINEAR DIFFERENTIAL SYSTEMS WITH CONSTANT COEFFICIENTS OF NEUTRAL TYPE

BAŠTINEC Jaromír, (CZ), DIBLÍK Josef, (CZ), KHUSAINOV Denys, (UA), RYVOLOVÁ Alena, (CZ)

Abstract. In this paper we investigate exponential stability of systems of linear differential equations of neutral type with constant coefficients and a constant delay

$$\dot{x}(t) - D\dot{x}(t-\tau) = Ax(t) + Bx(t-\tau),$$

where $t \geq 0$ is an independent variable, $\tau > 0$ is a constant delay, A, B and D are $n \times n$ constant matrices and $x\colon [-\tau, \infty) \to \mathbb{R}^n$. Estimations of solutions are given as well.

Key words and phrases. System of linear differential equations of neutral type, exponential stability, Lyapunov-Krasovskii functionals.

Mathematics Subject Classification. Primary 34K20, 34K25; Secondary 34K12.

1 Introduction

The purpose of the paper is to analyze systems of linear differential equations of neutral type with constant coefficients and a constant delay

$$\dot{x}(t) - D\dot{x}(t-\tau) = Ax(t) + Bx(t-\tau), \tag{1}$$

where $t \geq 0$ is an independent variable, $\tau > 0$ is a constant delay, A, B and D are $n \times n$ constant matrices and $x\colon [-\tau, \infty) \to \mathbb{R}^n$ is a column vector-solution. The derivative " $\cdot$ " is understood as a left-hand derivative. Let $\varphi\colon [-\tau,0]\to\mathbb{R}^n$ be a continuously differentiable vector-function. The solution $x=x(t)$ of the problem (1), (2) on $[-\tau,\infty)$ where

$$x(t)=\varphi(t),\ \dot{x}(t)=\dot{\varphi}(t),\ t\in[-\tau,0] \tag{2}$$

we define in a classical sense (cf. e.g. [1, 3]) as a function continuous on $[-\tau,\infty)$ and continuously differentiable on $[-\tau,\infty)$ except at points τp, $p=0,1,\dots$, and satisfying the equation (1) everywhere on $[0,\infty)$ except at the above mentioned points.

We will investigate exponential stability of (1) and, moreover, for every positive t we will find estimation of the norm of the difference between the solution $x=x(t)$ of the problem (1), (2) and the steady state in the moment t.

We will use the matrix norm

$$\|A\|:=\sqrt{\lambda_{\max}(A^TA)}\,,$$

where symbols $\lambda_{\max}(\,\cdot\,)$ or $\lambda_{\min}(\,\cdot\,)$ denote maximal and minimal eigenvalues of the corresponding matrix, and the following vector norms

$$\|x(t)\|:=\sqrt{\sum_{i=1}^{n}x_i^2(t)},\quad \|x(t)\|_\tau:=\max_{-\tau\le s\le 0}\{\,\|x(s+t)\|\,\}\,,\quad \|x(t)\|_{\tau,\beta}:=\sqrt{\int\limits_{t-\tau}^{t}e^{-\beta(t-s)}\|x\,(s)\,\|^2\,\mathrm{d}s}$$

where β is a parameter.

The most commonly used method for investigation of stability of functional-differential systems is the method of Lyapunov-Krasovskii functionals [2]. Usually, functionals having a quadratic form generated with terms on the left-hand side of equation (1), and with the integral (over the interval of delay) of a quadratic form are used. The relevant form is

$$V\,[x(t)]=[x(t)-Dx(t-\tau)]^T\,H\,[x(t)-Dx(t-\tau)]+\int\limits_{t-\tau}^{t}x^T(s)Gx(s)\mathrm{d}s \tag{3}$$

where H and G are suitable $n\times n$ positively definite matrices.

Regarding functionals of the form (3) we should remark the following. Using a functional (3), we can obtain affirmations on stability only, and relevant stability statements stating, e.g. that the expression

$$\int\limits_{t-\tau}^{t}x^T(s)Gx(s)\mathrm{d}s$$

is bounded from above, are of an integral type. Due to terms $[x(t)-Dx(t-\tau)]$ in (3) containing differences we are not able to deduce the boundedness of the norm of $x(t)$ itself.

In this paper we will use functionals of Lyapunov-Krasovskii of a quadratic type, which is dependent on running coordinates as well as on their derivatives. This will permit us to derive estimations of solutions of the system (1).

2 Estimations of convergence of solutions of stable systems

At first we give relevant definition of exponential stability:

Definition 2.1 *The zero solution of the system of equations of neutral type* (1) *is called exponentially stable in the metric* C^0, *if there exist constants* $N_i > 0$, $i = 1, 2$ *and* $\gamma > 0$ *such that for arbitrary solution* $x = x(t)$ *of* (1) *the inequality*

$$\|x(t)\| \leq [N_1 \|x(0)\|_\tau + N_2 \|\dot{x}(0)\|_\tau] \exp(-\gamma t/2)$$

holds for $t \geq 0$.

We will give estimation of solutions of linear system (1) on interval $(0, \infty)$ using Lyapunov-Krasovskii functional

$$V[x(t), t] = x^T(t)Hx(t) + \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T(s) G_1 x(s) + \dot{x}^T(s) G_2 \dot{x}(s)\right] \mathrm{d}s \tag{4}$$

with $n \times n$ positively definite matrices G_1, G_2 and H and a constant β. Then it is easy to see that the estimation

$$\lambda_{\min}(H)\|x(t)\|^2 + \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T(s)G_1x(s) + \dot{x}^T(s)G_2x(s)\right] \mathrm{d}s \leq V[x(t), t]$$

$$\leq \lambda_{\max}(H)\|x(t)\|^2 + \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T(s)G_1x(s) + \dot{x}^T(s)G_2x(s)\right] \mathrm{d}s \tag{5}$$

holds. We define an auxiliary $3n \times 3n$-dimensional matrix

$$S(\beta, G_1, G_2, H) := \begin{pmatrix} -A^T H - HA - G_1 - A^T G_2 A & -HB - A^T G_2 B & -HD - A^T G_2 D \\ -B^T H - B^T G_2 A & e^{-\beta\tau} G_1 - B^T G_2 B & -B^T G_2 D \\ -D^T H - D^T G_2 A & -D^T G_2 B & e^{-\beta\tau} G_2 - D^T G_2 D \end{pmatrix}$$

depending on the parameter β and the matrices G_1, G_2, H, and numbers

$$\varphi(H) := \frac{\lambda_{\max}(H)}{\lambda_{\min}(H)}, \varphi_1(G_1, H) := \frac{\lambda_{\max}(G_1)}{\lambda_{\min}(H)}, \varphi_2(G_2, H) := \frac{\lambda_{\max}(G_2)}{\lambda_{\min}(H)}.$$

Now we give a statement on exponential stability of the zero solution of system (1) and estimations of convergence of a solution, which will be proved using Lyapunov-Krasovskii functional (4).

Theorem 2.2 *Let* D *be a nonsingular matrix and let there exist positively definite matrices* G_1, G_2, H *and a parameter* $\beta > 0$ *such that the matrix* $S(\beta, G_1, G_2, H)$ *is also positively definite. Then the zero solution of system* (1) *is exponentially stable in the metric* C^0.

Moreover, for the solution $x = x(t)$ of the problem (1), (2) *the following estimation of convergence holds on* $[0, \infty)$:

$$\|x(t)\| \leq \left[\sqrt{\varphi(H)}\|x(0)\| + \tau\sqrt{\varphi_1(G_1, H)}\,\|x(0)\|_\tau + \tau\sqrt{\varphi_2(G_2, H)}\,\|\dot{x}(0)\|_\tau\right] e^{-\gamma t/2} \tag{6}$$

where

$$\gamma = \min\left\{\beta, \frac{\lambda_{\min}(S(\beta, G_1, G_2, H))}{\lambda_{\max}(H)}\right\}.$$

Proof. Let $t \geq 0$. We rewrite system (1) in the form

$$\dot{x}(t) = D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau), \tag{7}$$

where $t \geq 0$ and calculate the full derivative of functional (4) along solutions of system (7). We obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V[x(t), t] = {} & [D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau)]^T Hx(t) \\ & + x^T(t)H[D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau)] \\ & + \left[x^T(t)G_1x(t) - e^{-\beta\tau}x^T(t-\tau)G_1x(t-\tau)\right] \\ & + \left[\dot{x}^T(t)G_2\dot{x}(t) - e^{-\beta\tau}\dot{x}^T(t-\tau)G_2\dot{x}(t-\tau)\right] \\ & - \beta\int_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s) + \dot{x}^T(s)G\dot{x}(s)\right]\mathrm{d}s.\end{aligned}$$

We substitute value from (7) for $\dot{x}(t)$. We obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V[x(t), t] = {} & [D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau)]^T Hx(t) \\ & + x^T(t)H[D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau)] \\ & + \left[x^T(t)G_1x(t) - e^{-\beta\tau}x^T(t-\tau)G_1x(t-\tau)\right] \\ & + (D\dot{x}(t-\tau) + Ax(t) + Bx(t-\tau))^T G_2(D\dot{x}(t-\tau) + Ax(t)) \\ & + Bx(t-\tau) - e^{-\beta\tau}\dot{x}^T(t-\tau)G_2\dot{x}(t-\tau) \\ & - \beta\int_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s) + \dot{x}^T(s)G\dot{x}(s)\right]\mathrm{d}s.\end{aligned}$$

Now it is easy to verify that

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] = -\left(x^T(t), \quad x^T(t-\tau), \quad \dot{x}^T(t-\tau)\right)$$
$$\times \begin{pmatrix} -A^TH - HA - G_1 - A^TG_2A & -HB - A^TG_2B & -HD - A^TG_2D \\ -B^TH - B^TG_2A & e^{-\beta\tau}G_1 - B^TG_2B & -B^TG_2D \\ -D^TH - D^TG_2A & -D^TG_2B & e^{-\beta\tau}G_2 - D^TG_2D \end{pmatrix}$$
$$\times \begin{pmatrix} x(t) \\ x(t-\tau) \\ \dot{x}(t-\tau) \end{pmatrix} - \beta \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T\left(s\right)G_1x(s) + \dot{x}^T\left(s\right)G_2\dot{x}(s)\right] \mathrm{d}s$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] = -\left(x^T(t), \quad x^T(t-\tau), \quad \dot{x}^T(t-\tau)\right) \times S\left(\beta, H, G_1, G_2\right)$$
$$\times \begin{pmatrix} x(t) \\ x(t-\tau) \\ \dot{x}(t-\tau) \end{pmatrix} - \beta \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T\left(s\right)G_1x(s) + \dot{x}^T\left(s\right)G_2\dot{x}(s)\right] \mathrm{d}s.$$

Since the matrix $S\left(\beta, G_1, G_2, H\right)$ was assumed to be positively definite, for the full derivative of Lyapunov-Krasovskii functional (4) we obtain the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] \leq -\lambda_{\min}\left(S\left(\beta, G_1, G_2, H\right)\right)\left[\|x(t)\|^2 + \|x(t-\tau)\|^2\right.$$
$$\left. + \|\dot{x}\left(t-\tau\right)\|^2\right] - \beta \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T\left(s\right)G_1x(s) + \dot{x}^T\left(s\right)G_2\dot{x}(s)\right] \mathrm{d}s. \tag{8}$$

We will study the two possible (regarding the positive value of β) cases: either

$$\beta > \frac{\lambda_{\min}\left(S\left(\beta, G_1, G_2, H\right)\right)}{\lambda_{\max}(H)} \tag{9}$$

is valid or

$$\beta \leq \frac{\lambda_{\min}\left(S\left(\beta, G_1, G_2, H\right)\right)}{\lambda_{\max}(H)} \tag{10}$$

holds.

1. Let (9) be valid. From inequality (5) there follows that

$$-\|x\left(t\right)\|^2 \leq -\frac{1}{\lambda_{\max}(H)}V\left[x(t),t\right] + \frac{1}{\lambda_{\max}(H)} \int\limits_{t-\tau}^{t} e^{-\beta(t-s)} \left[x^T(s)G_1x(s) + \dot{x}^T(s)G_2\dot{x}(s)\right] \mathrm{d}s.$$

We substitute this expression into inequality (8). Then

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] \le \lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)\left[-\frac{1}{\lambda_{\max}(H)}V\left[x\left(t\right),t\right]\right.$$
$$\left.+\frac{1}{\lambda_{\max}(H)}\int\limits_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s)+\dot{x}^T(s)G_2\dot{x}(s)\right]\mathrm{d}s\right]$$
$$-\beta\int\limits_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s)+\dot{x}^T(s)G_2\dot{x}(s)\right]\mathrm{d}s$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] \le -\frac{\lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)}{\lambda_{\max}(H)}V\left[x\left(t\right),t\right]-\left[\beta-\frac{\lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)}{\lambda_{\max}(H)}\right]$$
$$\times\int\limits_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s)+\dot{x}^T(s)G_2\dot{x}(s)\right]\mathrm{d}s.$$

Due to (9) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x(t),t\right] \le -\frac{\lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)}{\lambda_{\max}(H)}V\left[x\left(t\right),t\right].$$

Integrating this inequality over interval $(0,t)$ we get

$$V\left[x\left(t\right),t\right] \le V\left[x\left(0\right),0\right]\exp\left(-\frac{\lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)}{\lambda_{\max}(H)}t\right) \le V\left[x\left(0\right),0\right]e^{-\gamma t}. \tag{11}$$

2. Let (10) be valid. From (5) we get

$$-\int\limits_{t-\tau}^{t} e^{-\beta(t-s)}\left[x^T(s)G_1x(s)+\dot{x}^T(s)G_2\dot{x}(s)\right]\mathrm{d}s \le -V\left[x\left(t\right),t\right]+\lambda_{\max}\left(H\right)\|x(t)\|^2.$$

We substitute this expression into inequality (8). We obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x\left(t\right),t\right] \le -\lambda_{\min}\left(S\left(\beta,G_1,G_2,H\right)\right)\|x\left(t\right)\|^2+\beta\left[-V\left[x\left(t\right),t\right]+\lambda_{\max}(H)\|x\left(t\right)\|^2\right]$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x\left(t\right),t\right] \le -\beta V\left[x\left(t\right),t\right]-\left\{\lambda_{\min}\left(S\left[\beta,G_1,G_2,H\right]\right)-\beta\lambda_{\max}(H)\right\}\|x\left(t\right)\|^2.$$

Since (10) holds then

$$\frac{\mathrm{d}}{\mathrm{d}t}V\left[x\left(t\right),t\right] \le -\beta V\left[x\left(t\right),t\right].$$

Integrating this inequality over interval $(0, t)$ we get

$$V\left[x\left(t\right), t\right] \le V\left[x\left(0\right), 0\right] e^{-\beta t} \le V\left[x\left(0\right), 0\right] e^{-\gamma t}. \tag{12}$$

Connecting both inequalities (11), (12) we conclude that in both cases (9), (10) we have

$$V\left[x\left(t\right), t\right] \le V\left[x\left(0\right), 0\right] e^{-\gamma t}. \tag{13}$$

Now using inequality (13) we obtain an estimation of convergence of solutions of system (1). From (5) follows that

$$\|x(t)\|^2 \le \frac{1}{\lambda_{\min}(H)} \left[\lambda_{\max}(H)\|x(0)\|^2 + \lambda_{\max}(G_1)\left\|x(0)\right\|^2_{\tau,\beta} + \lambda_{\max}\left(G_2\right)\left\|\dot{x}(0)\right\|^2_{\tau,\beta}\right] e^{-\gamma t},$$

or

$$\left\|x\left(t\right)\right\| \le \left[\sqrt{\varphi\left(H\right)}\left\|x\left(0\right)\right\| + \sqrt{\varphi_1\left(G_1, H\right)}\left\|x\left(0\right)\right\|_{\tau,\beta} + \sqrt{\varphi_2\left(G_2, H\right)}\left\|\dot{x}\left(0\right)\right\|_{\tau,\beta}\right] e^{-\gamma t/2}.$$

The last inequality implies

$$\left\|x\left(t\right)\right\| \le \left[\sqrt{\varphi\left(H\right)}\left\|x\left(0\right)\right\| + \tau\sqrt{\varphi_1\left(G_1, H\right)}\left\|x\left(0\right)\right\|_{\tau} + \tau\sqrt{\varphi_2\left(G_2, H\right)}\left\|\dot{x}(0)\right\|_{\tau}\right] e^{-\gamma t/2}.$$

So the inequality (6) is proved and, consequently, the zero solution of system (1) is exponentially stable in the metric C^0.

3 Example

We will investigate system (1) where $n = 2$, $\tau = 1$,

$$D = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \qquad A = \begin{pmatrix} -1 & 0.1 \\ 0.1 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix},$$

i.e., the system

$$\dot{x}_1(t) = 0.5\dot{x}_1(t-1) - x_1(t) + 0.1x_2(t) + 0.1x_1(t-1), \tag{14}$$
$$\dot{x}_2(t) = 0.5\dot{x}_2(t-1) + 0.1x_1(t) - x_2(t) + 0.1x_2(t-1), \tag{15}$$

with initial conditions (2). Set $\beta = 0.1$ and

$$G_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad G_2 = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}, \qquad H = \begin{pmatrix} 2 & 0.1 \\ 0.1 & 5 \end{pmatrix}.$$

The following computations were performed by using MATLAB & SIMULINK R2009a.

The eigenvalues of matrix G_1 are $\lambda_{\min}(G_1) = \lambda_{\max}(G_1) = 1$, the eigenvalues of matrix G_2 are $\lambda_{\min}(G_2) \doteq 0.5858$ and $\lambda_{\max}(G_2) \doteq 3.4142$, and the eigenvalues of matrix H are $\lambda_{\min}(H) \doteq 1.9967$ and $\lambda_{\max}(H) \doteq 5.0033$.

The matrix $S(\beta, G_1, G_2, H)$ takes the form

$$S(\beta, G_1, G_2, H) \doteq \begin{pmatrix} 2.15 & -1.11 & -0.11 & 0.06 & -0.55 & 0.3 \\ -1.11 & 6.17 & 0.08 & -0.21 & 0.4 & -1.05 \\ -0.11 & 0.08 & 0.8948 & -0.01 & -0.05 & -0.05 \\ 0.06 & -0.21 & -0.01 & 0.8748 & -0.05 & -0.15 \\ -0.55 & 0.4 & -0.05 & -0.05 & 0.6548 & 0.6548 \\ 0.3 & -1.05 & -0.05 & -0.15 & 0.6548 & 1.9645 \end{pmatrix}.$$

Eigenvalues of $S(\beta, G_1, G_2, H)$ are

$$\begin{aligned} \lambda_1(S) &\doteq 6.7377, \\ \lambda_2(S) &\doteq 2.2297, \\ \lambda_3(S) &\doteq 1.8651, \\ \lambda_4(S) &\doteq 0.8967, \\ \lambda_5(S) &\doteq 0.8352, \\ \lambda_6(S) &\doteq 0.1445. \end{aligned}$$

Because all eigenvalues are positive, matrix $S(\beta, G_1, G_2, H)$ is positively definite.

All conditions of Theorem 2.2 are satisfied so the zero solution of system (14), (15) is asymptotically stable in the metric C^0. Further we have

$$\varphi(H) \doteq \frac{5.003}{1.9967} \doteq 2.5056, \quad \varphi_1(G_1.H) \doteq \frac{1}{1.9967} \doteq 0.5008,$$

$$\varphi_2(G_2, H) \doteq \frac{3.4142}{1.9967} \doteq 1.7099, \quad \gamma \doteq \min\left\{0.1, \frac{0.1445}{5.0033}\right\} \doteq 0.0289.$$

Finally, from (6) there follows, that the inequality

$$\|x(t)\| \leq \left[\sqrt{2.5056}\|x(0)\| + \sqrt{0.5008}\,\|x(0)\|_1 + \sqrt{1.7099}\,\|\dot{x}(0)\|_1\right] e^{-0.0289t/2}$$

holds on $[0, \infty)$.

Acknowledgement

This research was supported by the Grant 201/08/9469 of Czech Grant Agency and by the Council of Czech Government MSM 0021630503, MSM 0021630519, MSM 0021630529.

References

[1] J. HALE, S.V. LUNEL, *Introduction to Functional Differential Equations.* Springer-Verlag, 1993.

[2] N. N. KRASOVSKII, *Stability of Motion.* Standford University Press, 1963.

[3] V. KOLMANOVSKII, A. MYSHKIS, *Introduction to the Theory and Applications of Functional Differential Equations.* Kluwer Academic Publishers, 1999.

Current address

Baštinec Jaromír, doc. RNDr., CSc.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 8, 616 00 Brno, Czech Republic
tel.: +420-541143222
email: bastinec@feec.vutbr.cz

Diblík Josef, Prof. RNDr., DrSc.
Brno University of Technology, Czech Republic
tel.: +420-514147601, +420-541143155
email: diblik.j@fce.vutbr.cz, diblik@feec.vutbr.cz

Khusainov Denys Ya., Prof. DrSc.
Faculty of Cybernetics,
Kiev National University,
Vladimirskaya Str. 64, 01601 Kiev, Ukraine.
email: Khusainov@unicyb.kiev.ua

Ryvolová Alena, Mgr.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 8, 616 00 Brno, Czech Republic
email: ryvolova.a@seznam.cz

Journal of Applied Mathematics

volume III (2010), number II

AVERAGING, MERGER AND STABILITY OF LINEAR DYNAMICAL SYSTEMS WITH SMALL MARKOV JUMPS

CARKOVS Jevgenijs, (LV), PAVLENKO Oksana, (LV)

Abstract. In this paper it is shown that the asymptotical methods may be succesfully used for exponential mean square stability analysis of the linear dynamical system in $\mathbb{R}^n$ of impulse type which dynamical characteristics are dependent on the step Markov process $\{y(t),\, t \geq 0\}$.

Key words and phrases. linear differential equations, small random perturbations, impulse equations, mean square stability, Lyapunov methods, quadratic Lyapunov functionals, averaging procedure.

Mathematics Subject Classification. 60H10, 60H30.

1 Introduction

Let $\{y_\varepsilon(t),\, t \geq 0\}$ be series of right continuous homogeneous Markov processes [4] on the countable space $\mathbf{Y} \subset \mathbb{R}$ depending on parameter $\varepsilon \in (0,\, 1)$ with weak infinitesimal operators Q_ε defined on any element of the space $\mathbf{V}$ of bounded mappings $v : \mathbf{Y} \to \mathbb{R}$ by the equality

$$Q_\varepsilon v(y) := a(y,\, \varepsilon) \sum_{z \in \mathbf{Y}} [v(z) - v(y)] p(y,\, z,\, \varepsilon), \tag{1}$$

and let us suppose that $0 < a_1 := \inf_{\substack{y \in \mathbf{Y} \\ \varepsilon \in (0,\,1)}} a(y,\, \varepsilon) \leq \sup_{\substack{y \in \mathbf{Y} \\ \varepsilon \in (0,\,1)}} a(y,\, \varepsilon) := a_2 < \infty$. For any fixed $\varepsilon \in (0,\, 1)$ the Markov process with infinitesimal operator (1) is piecewise constant process [4, 7] with switching moments $\{\tau_j^\varepsilon,\ j \in \mathbb{N}\}$. These random variables can be recurrently defined by equalities

$$\tau_0^\varepsilon = 0, \quad \mathbf{P}_y(\tau_j^\varepsilon - \tau_{j-1}^\varepsilon > t) = e^{a(y,\,\varepsilon)t}, \quad j \in \mathbb{N},\, y \in \mathbf{Y},\, t \geq 0.$$

Now we will describe the series of Impulse Dynamical Systems (IDS) in $\mathbb{R}^n$ with parameter $\varepsilon \in (0, 1)$ this paper deals with. The phase coordinates $x_\varepsilon(t)$ of this systems satisfy:

1) the initial condition

$$x_\varepsilon(0) = x \tag{2}$$

2) the differential equation

$$\frac{dx_\varepsilon}{dt} = A(y(t),\, \varepsilon))\, x_\varepsilon \tag{3}$$

for all $t \in (\tau^\varepsilon_{j-1},\, \tau^\varepsilon_j),\ \ j \in \mathbb{N}$;

3) the conditions of jumps

$$x_\varepsilon(t) = x_\varepsilon(t-0) + B(y(t),\, y(t-0)),\, \varepsilon)x_\varepsilon(t-0) \tag{4}$$

for all $t \in \{\tau^\varepsilon_j,\ \ j \in \mathbb{N}\}$, where the matrices $A(y,\, \varepsilon),\ B(z,\, y,\, \varepsilon)$ are defined as the series

$$A(y,\, \varepsilon) = \sum_{k=1}^{\infty} A_k(y)\, \varepsilon^k,\ B(z,\, y,\, \varepsilon) = \sum_{k=1}^{\infty} B_k(z,\, y)\, \varepsilon^k.$$

with matrix coefficients satisfying the inequalities

$$\sup_{y\in\mathbf{Y}} ||A_k(y)|| := \alpha_k < \infty,\ \sup_{z,y\in\mathbf{Y}} ||B_k(z,\, y)|| := \beta_k < \infty,\ k \in \mathbb{N} \tag{5}$$

and also the series composed of $\alpha_k,\ \beta_k$ are convergent. It is easily to make sure of existence and uniqueness of the above defined process $x_\varepsilon(t)$ for all $t \geq 0$.

The IDS (3)-(4) we will named as *exponentially mean square stable* if there exist such positive numbers M and ρ that $\mathbf{E}_y\, |x_\varepsilon(t+s,\, s,\, x)|^2 \leq M\, e^{-\rho t}|x|^2$ for any $x \in \mathbb{R}^n,\ y \in \mathbf{Y}$ and $t \geq s \geq 0$. Due to homogeneity of the Markov process $\{x_\varepsilon(t),\, y(t)\}$ the above inequality is equivalent to the inequality $\mathbf{E}_y\, |x_\varepsilon(t,\, 0,\, x)|^2 \leq M\, e^{-\rho t}|x|^2$.

Let us denote by $\mathbf{Q}$ the space of the symmetric $n \times n$ matrix-valued continuous functions $\{q(y),\, y \in \mathbf{Y}\}$ with the subset $\mathbf{K} := \{q \in \mathbf{Q} :\ (q(y)x,\, x) \geq 0, \forall x \in \mathbb{R}^n, \forall y \in \mathbf{Y}\}$ of nonnegative-definite matrices.

The set of inner points of $\mathbf{K}$ can be defined as $\dot{\mathbf{K}} := \{q \in \mathbf{K} : \exists c > 0,\ q \gg cI\}$.

The following theorem was proved in [6].

Theorem. *Equation (1) is exponentially mean square stable if and only if there exist $q \in \dot{\mathbf{K}}$ and $r \in \dot{\mathbf{K}}$ such that*

$$\mathbf{A}_\varepsilon\, q = -r, \tag{6}$$

where

$$\begin{aligned}(\mathbf{A}_\varepsilon)\, q(y) &= A^T(y,\, \varepsilon)q(y) + q(y)\, A(y,\, \varepsilon)\\ &+a(y,\, \varepsilon) \sum_{z\in\mathbf{Y}} [(I + B^T(z,\, y,\, \varepsilon))\, q(z)\, (I + B(z,\, y,\, \varepsilon)) - q(z)]\, p(y,\, z,\, \varepsilon) + Q_\varepsilon\, q(y).\end{aligned} \tag{7}$$

Corollary. *IDS (3)-(4) is exponentially mean square stable if and only if there exists solution $q \in \dot{\mathbf{K}}$ of equation (6) with $r = I$.*

Equation (6) will be named *the Lyapunov equation for mean square stability investigation of IDS (3)-(4).*

We will suppose that the infinitesimal generator (1) also can be represented as the uniformly on $y \in \mathbf{Y}$, $\varepsilon \in (0,\, 1)$ convergent series

$$Q_\varepsilon\, v(y) = Q\, v(y) + \sum_{k=1}^{\infty} Q_k\, v(y)\, \varepsilon^k$$

where

$$\begin{aligned} Q\, v(y) &= a(y) \sum_{y \in \mathbf{Y}} [v(z) - v(y)]\, p(y,\, z), \qquad (8) \\ Q_k\, v(y) &= \sum_{y \in \mathbf{Y}} [v(z) - v(y)] p_k(y,\, z), \;\; k \in \mathbb{N}, \end{aligned}$$

$p(y,\, z)$ is transition probability of some embedded Markov chains and $p_k(y,\, z),\; k \in \mathbb{N}$ are some positive measures on $\mathbf{Y}$. The operator (8) can be considered [4, 7] as the infinitesimal generator of some homogeneous piece wise constant Markov process $y_0(t),\, t \geq 0$. Let us assume that this operator has 0 as an isolated simple eigenvalue of multiplicity h, h eigenfunctions with nonintersecting supports $\mathbf{Y}_j,\;\; j = \overline{1, h}$ defined by equalities $f_j\,(y) = \begin{cases} 1, & \text{for} y \in \mathbf{Y}_j \\ 0, & \text{for} y \in \mathbf{Y}_k,\;\; k \neq j. \end{cases}$ and the remaining part of its spectrum is situated in the half-plane $\mathbf{C}_{-\rho}$ for some positive ρ. The conjugate operator Q^* also [4] has 0 as an isolated eigenvalue of multiplicity h and h invariant measures $\mu_k(y)$ with the same supports $\mathbf{Y}_k$, $k = \overline{1, h}$. If $h = 1$ then [4, 7]the Markov process $y_0(t)$ is uniformly exponentially ergodic. It has unique invariant measure $\mu(y)$.

The operator (7) can be decomposed in terms of powers of ε

$$\mathbf{A}_\varepsilon = \sum_{k=0}^{\infty} \varepsilon^k\, \mathbf{G}_k \tag{9}$$

where the operators $\mathbf{G}_m \in \mathbf{L}(\mathbf{Q})$, $m \geq 0$ are defined by the formulae

$$\begin{aligned} (\mathbf{G}_0\, q)(y) &:= A_0^T\, q(y) + q(y)\, A_0 + (Q\, q)(y), & (10) \\ (\mathbf{G}_1\, q)(y) &:= A_1^T(y)\, q(y) + q(y)\, A_1(y) + (Q_1 q)(y) \\ &\quad + a(y) \sum_{z \in \mathbf{Y}} [B_1^T(z,\, y)\, q(z) + q(z)\, B_1(z,\, y)]\, p(y,\, z), & (11) \\ (\mathbf{G}_2\, q)(y) &:= A_2^T(y)\, q(y) + q(y)\, A_2(y) + (Q_2 q)(y) \\ &\quad + a(y) \sum_{z \in \mathbf{Y}} [B_2^T(z,\, y)\, q(z) + q(z)\, B_2(z,\, y) + B_1^T(z,\, y)\, q(z)\, B_1(z,\, y)]\, p(y,\, z) \\ &\quad + \sum_{z \in \mathbf{Y}} [B_1^T(z,\, y)\, q(z) + q(z)\, B_1(z,\, y)]\, p_1(y,\, z), & (12) \end{aligned}$$

$$
\begin{aligned}
(\mathbf{G}_m\, q)(y) \;\; := \;\; & A_m^T(y)\, q(y) + q(y)\, A_m(y) + (Q_m q)(y) + a(y) \sum_{z\in\mathbf{Y}} [B_m^T(z,\, y)\, q(z) + q(z)\, B_m(z,\, y) \\
& + \sum_{k=1}^{m-1} B_{m-k}^T(z,\, y)\, q(z)\, B_k(z,\, y)]\, p(y,\, z) \\
& + [B_1^T(z,\, y)\, q(z) + q(z)\, B_1(z,\, y)]\, p_{m-1}(y,\, z), \\
& + \sum_{z\in\mathbf{Y}} \sum_{t=2}^{m-1} [B_t^T(z,\, y)\, q(z) + q(z)\, B_t(z,\, y) \\
& + \sum_{k=1}^{t-1} B_{t-k}^T(z,\, y)\, q(z)\, B_k(z,\, y)]\, p_{m-t}(y,\, z), \qquad m \geq 3. \qquad (13)
\end{aligned}
$$

Hence like in [7] one can prove the next result

Lemma. *Let $\lambda_0 \notin \sigma(\mathbf{A}_\varepsilon)$ for all sufficiently small $\varepsilon > 0$. Under the above assumptions there exists positive ε_0 such that the solution q^ε of the equation $\mathbf{A}_\varepsilon q^\varepsilon - \lambda_0 q^\varepsilon = f$ for any $f \in \mathbf{Q}$ and $\varepsilon \in (0,\, \varepsilon_0)$ has the form $q^\varepsilon = \sum\limits_{k=-d}^{\infty} \varepsilon^k q_k$ with some $d \in \mathbf{N}$.*

2 Averaging, merger and stability

Let us suppose that the differential equation (3) has the form

$$\frac{dx_\varepsilon}{dt} = \varepsilon\, A_1(y_\varepsilon(t))\, x_\varepsilon + \varepsilon^2\, A_2(y_\varepsilon(t))\, x_\varepsilon \qquad (14)$$

for all $t \in (\tau_{j-1}^\varepsilon,\, \tau_j^\varepsilon)$, $j \in \mathbb{N}$, and the conditions of jumps (4) have the form

$$x_\varepsilon(t) = x_\varepsilon(t-0) + \varepsilon\, B_1(y_\varepsilon(t),\, y_\varepsilon(t-0))\, x_\varepsilon(t-0) + \varepsilon^2\, B_2(y_\varepsilon(t),\, y_\varepsilon(t-0))\, x_\varepsilon(t-0) \qquad (15)$$

At the beginning we will assume that $h = 1$, that is, the Markov process $y_0(t)$ has a unique invariant measure $\mu(dy)$ and will denote by $\hat{y_0}(t)$ the stationary Markov function, corresponding to this measure. This means that for any $t \in \mathbb{R}$ and $A \subset \mathbf{Y}$ one can write $\mathbf{P}(\hat{y_0}(t) \in A) = \mu(A)$. We shall also denote

$$C_j(y) := A_j(y) + a(y) \sum_{z\in\mathbf{Y}} B_j(z,\, y)\, p(y,\, z),$$

$$\overline{A}_j = \sum_{y\in\mathbf{Y}} A_j(y)\, \mu(dy), \;\; \overline{C}_j = \sum_{y\in\mathbf{Y}} C_j(y)\, \mu(dy), \qquad j \;=\; 1, 2,$$

If $\overline{C}_1 = 0$ one can define [4] the matrix

$$
\begin{aligned}
F \;\; := \;\; & \overline{C}_2 + \sum_{y\in\mathbf{Y}} \sum_{z\in\mathbf{Y}} (\Pi\, C_1)(z)\, p_1(y,\, z) \mu(y) \\
& + \sum_{y\in\mathbf{Y}} \{A_1(y)\, (\Pi\, C_1)(y) + a(y) \sum_{z\in\mathbf{Y}} B_1(z,\, y)\, (\Pi\, C_1)(z)\, p(y,\, z)\}\, \mu(y)
\end{aligned}
$$

and the operator $H \in \mathbf{L}(\mathbf{M}_n(\mathbb{R}))$

$$\begin{aligned} H\,q \quad := & \sum_{y\in\mathbf{Y}}\{A_1^T(y)\,q\,(\Pi\,C_1)(y) + (\Pi\,C_1)(y)\,q\,A_1(y) \\ & + a(y)\sum_{z\in\mathbf{Y}}[B_1^T(z,\,y)\,q\,(\Pi\,C_1)(z) + (\Pi\,C_1^T)(z)\,q\,B_1(z,\,y)]\,p(y,\,z)\,\mu(y) \end{aligned}$$

for $q \in \mathbf{M}_n(\mathbb{R})$, where the operator-potential Π.

For the first example let us consider the algorithm described in [6] with $d = 1$. In this case in order to investigate stability of the IDS (14)-(15) on the first step we must deal with the equation

$$Qq(y) = 0. \tag{16}$$

According to our previous assumptions about the operator Q one can conclude that the solution of (16) is an arbitrary symmetric matrix $q(y) \equiv q$. The conjugate equation to (16) has the form

$$Q^*\,p(y) = 0 \tag{17}$$

and its solution can be represented as $p(y) = p\,\mu(y)$ with an arbitrary symmetric matrix p and the invariant measure μ described above.

In the second step we must analyze the possibility of solving the equation

$$Q\,q_0(y) = -I - (\mathbf{G}_1\,q)(y)), \tag{18}$$

that is, due to Fredholm alternative it should be

$$Tr\,\{\sum_{y\in\mathbf{Y}}\{A_1^T(y)\,q + q\,A_1(y) + a(y)\sum_{z\in\mathbf{Y}}[B_1^T(z,\,y)\,q + q\,B_1(z,\,y)]\,p(y,\,z)\}\,\mu(y) + I\}\,p = 0$$

for an arbitrary matrix p. This is equivalent to existence of a symmetric matrix $\overline{q}$ as solution of the equation

$$\overline{C}_1^T\,q + q\,\overline{C}_1 = -I. \tag{19}$$

This equation has positive definite solution if and only if

$$\sigma(\overline{C}_1) \subset \{\mathbf{C} : \Re\lambda < 0\} \tag{20}$$

which is equivalent to asymptotic stability of the ordinary differential equation

$$\frac{d\overline{x}}{dt} = \overline{A}_1\,\overline{x}. \tag{21}$$

If

$$\sigma(\overline{A}_1) \cap \{\mathbf{C} : \Re\lambda > 0\} \neq \emptyset \tag{22}$$

the equation (21) is not asymptotically stable and then quation (19) has a nonpositive defined matrix $\overline{q}$ as its solution. Thus, in both cases (20) and (22) equation (19) has solution $\overline{q}$ and we can find the solution $q_0(y)$ of equation (18). Then the matrix

$$\frac{1}{\varepsilon}\,\overline{q} + q_0(y) \tag{23}$$

allows us to infer on the stability of the IDS (14)-(15). It is clear that if the matrix $\overline{q}$ is positive definite or nonpositive definite then the matrix (23) also has this property for sufficiently small positive ε. Therefore

- *if the averaged equation (21) is asymptotically stable then the IDS (14)-(15) is exponentially mean square stable for sufficiently small positive ε;*
- *if the averaged equation (21) has exponentially growing solutions then equation (14)-(15) also has exponentially mean square growing solutions for sufficiently small positive ε.*

If

$$\sigma(\overline{C}_1) \subset \{\mathbf{C} : \Re\lambda \leq 0\}, \quad \sigma(\overline{C}_1) \cap \{\mathbf{C} : \Re\lambda = 0\} \neq \emptyset \tag{24}$$

then equation (19) has no solutions and therefore $d > 1$ and we must analyze the equation

$$Q\,q(y) = -A_1^T(y)\,q - q\,A_1(y). \tag{25}$$

Due to the assumption (24) the equation

$$\overline{A}_1^T\,\tilde{q} + \tilde{q}\,\overline{A}_1 = 0 \tag{26}$$

has as solution a nonnegative definite matrix $\tilde{q}$ [2]. Then equation (25) must have solution because

$$Tr\,\{(\overline{A}_1^T\,\tilde{q} + \tilde{q}\,\overline{A}_1)\,p\} = 0$$

for any matrix p. Let us assume that the Markov process $\{y(t)\}$ is uniformly exponentially ergodic, that is, its transition probability satisfies the inequality

$$|P(t,y,A) - \mu(A)| \leq e^{-\rho t}$$

uniformly on $A \in \mathfrak{G}, y \in \mathbf{Y}$ for some $\rho > 0$ and all $t \geq 0$. In this case one can define [3] *the potential* Π of Markov process by the equality

$$(\Pi g)(y) := \int_0^\infty \int_{\mathbf{Y}} g(z)P(t,y,dz)\,dt = \int_0^\infty \mathbf{E}_y g(y(t))\,dt \tag{27}$$

for all g satisfying the condition $\int_{\mathbf{Y}} g(z)\mu(dz) = 0$. Next we extend the potential (27) on all $v \in \mathbf{C}(\mathbf{Y})$ by the equality

$$(\Pi v)(y) := \int_0^\infty \{\int_{\mathbf{Y}} v(z)P(t,y,dz) - \int_{\mathbf{Y}} v(z)\mu(dz)\}\,dt. \tag{28}$$

It is clear that $\Pi v \in D(Q)$ and

$$Q\Pi v = -v + \overline{v},$$

where

$$\overline{v} = \int_{\mathbf{Y}} v(z)\mu(dz).$$

By using the definition (28) of the extension potential Π the solution of (25) can be written in the form

$$q_{-1}(y) := (\Pi A_1^T)(y)\,\tilde{q} + \tilde{q}\,(\Pi A_1)(y).$$

This function can also be rewritten in the form [3]

$$q_{-1}(y) = \int_0^\infty \mathbf{E}_y\,\{(A_1^T(y(t)) - \overline{A}_1^T)\,\tilde{q} + \tilde{q}\,(A_1(y(t)) - \overline{A}_1)\,\}\,dt.$$

The right part of this formula is a linear continuous operator acting on $\tilde{q}$.

Next we must analyze the equation

$$Q\,q(y) = -I - A_2^T(y)\,\tilde{q} - \tilde{q}\,A_2(y) - A_1^T(y)\,q_{-1}(y) - q_{-1}(y)\,A_1(y). \tag{29}$$

By using the Fredholm alternative one has to verify the orthogonality of the right part of (29) to the matrix measure $p\,\mu(dy)$ for arbitrary matrix p. This equation has a solution if and only if there exists a matrix $\tilde{q}$ which satisfies equation (26) and equation

$$\overline{A}_2^T\,\tilde{q} + \tilde{q}\,\overline{A}_2 + \overline{(\Pi A_1)A_1}^T\,\tilde{q} + \tilde{q}\,\overline{(\Pi A_1)A_1} + \overline{\Pi A_1^T\,\tilde{q}\,A_1} + \overline{A_1^T\,\tilde{q}\,\Pi A_1} = -I, \tag{30}$$

where overline denotes an averaging according to measure $\mu\,(dy)$. Therefore, the following conclusions about stability of (21) can be drawn under the conditions (24):

-if the system (26),(30) has positive definite solution $\tilde{q}$ then equation (21) is exponentially mean square stable for sufficiently small positive ε;

-if the system (26),(30) has nonpositive definite solution $\tilde{q}$ then equation (21) has exponentially mean square growing solutions for sufficiently small positive ε.

In the papers [1, 3] it is proven that under condition $\overline{A}_1 = 0$ the solutions of the system (26),(30) when represented in the form $x(t/\varepsilon^2)$ converge weakly as $\varepsilon \to 0$ to the corresponding solutions of the stochastic equation

$$d\hat{x}(t) = (F + \bar{A}_2)\,\hat{x}(t)\,dt + \sum_{j=1}^{n} D_j\,\hat{x}(t)\,dw_j(t) \tag{31}$$

and if the latter equation is exponentially mean square stable then this property also holds for the system (26),(30). The same result can be obtained using the above analysis of (26),(30) with $\overline{A}_1 = 0$ since then (23) is automatically satisfied and equation (26) is fulfilled for any matrix $\tilde{q}$. Equation (30) is the Lyapunov equation for analysis of mean square stability of equation (31) [5]. It can be easily seen that both equations (31) and (26),(30) have the same asymptotic behaviour as $t \to \infty$. Hence, *under the condition $\overline{A}_1 = 0$:*

- if the stochastic approximation of the system (26),(30), given by (31), is asymptotically mean square stable then the system (26),(30) is exponentially mean square stable for sufficiently small positive ε;

- if the stochastic approximation of the system (26),(30), given by (31), has exponentially mean square growing solutions then the system (26),(30) also has exponentially mean square growing solutions for sufficiently small positive ε.

3 Example

Let us analyse the stability of the system of type (14)-(15), described below.
Let us consider the Markov process with two states space $\mathbf{Y} = \{0; 1\}$, defined by the infinitesimal matrix $Q = \begin{pmatrix} -\alpha & \alpha \\ 1-\alpha & \alpha - 1 \end{pmatrix}$ where $\alpha \in (0; 1)$

Then the transition probabilities are $p(0,0) = 0$, $p(1,0) = 1$, $p(0,1) = 1$, $p(1,1) = 0$ and intensities of switching are: $a(0) = \alpha$, $a(0) = -\alpha$. Therefore the invariant measure of this Markov process is given by the equalities $\mu(0) = 1 - \alpha$, $\mu(1) = \alpha$.

Let the above Markov process be switching process for two dimensional MIDS of type (14)-(15), given by the equations:

$$\frac{dx}{dt} = \varepsilon A(y(t))x$$

for all $t \in (\tau_{j-1},\ \tau_j),\ \ j \in \mathbb{N}$;

$$x(t) = (I + B(y(t),\, y(t-0)))x(t-0)$$

for all $t \in \{\tau_j,\ \ j \in \mathbb{N}\}$,
where $A(y) = Ay = \begin{pmatrix} 0 & 0 \\ 0 & -2\delta \end{pmatrix}$, $B(y) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_2 \end{pmatrix} zy$

1) In the first step we must deal with the equation (16). The solution is an arbitrary matrix with equal elements $Q = \begin{pmatrix} q & q \\ q & q \end{pmatrix}$. Then we solve the conjugate equation (17). $p(y) = \begin{pmatrix} \frac{(1-\alpha)^2}{\alpha^2} & \frac{1-\alpha}{\alpha} \\ \frac{1-\alpha}{\alpha} & 1 \end{pmatrix} p$

2) In the second step we must find the solution of (19). $C_1(y) = \begin{pmatrix} 0 & 0 \\ 0 & -2\delta \end{pmatrix} y$,
$\bar{C}_1 = \begin{pmatrix} 0 & 0 \\ 0 & -2\delta\alpha \end{pmatrix}$.

$$\begin{pmatrix} 0 & 0 \\ 0 & -2\delta \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_2 \end{pmatrix} + \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -2\delta \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

We obtain, that the solution of this equation do not exists. So we need to continue the algorithm and find the solution of (26)-(30). But from the equation (26):

$$2 \begin{pmatrix} 0 & 0 \\ 0 & -2\delta \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_2 \end{pmatrix} = 0$$

follows, that $q_{21} = q_{22} = 0$ of this solution. So the matrix q could not be positive defined and MIDS can not has any stable solution.

References

[1] ANISIMOV, V.V.: *Switching processes: averaging principle, diffusion approximation and applications*, Kluver, The Netherlands, Acta Aplicandae Mathematica (1995),no. 40, 95–141.

[2] BLANKENSHIP G., PAPANICOLAOU, G. C.: *Stability and control of stochastic systems with wide-band noise disturbances I.*, SIAM J. Appl. Math., **34** (178), no.3, 437–476.

[3] GIKHMAN I.I., SKOROKHOD A.V.: *Stochastic Differential Equations*, Springer Verlag, New York, 1972.

[4] DYNKIN E. B.: *Markov Processes*, Academic Press, New York, 1965.

[5] KATAFYGIOTIS L., TSARKOV, Ye.: *Mean square stability of linear dynamical systems with small Markov perturbations.I. Bounded coefficients,* Random Oper. and Stoch. Equ. **4**, (1996), no.2, 149–170.

[6] PAVLENKO O., TSARKOV, Ye.: *Positive covariation semigroup and Lyapunov equation for Markov impulsive dynamical systems,* Acta Universitatis Latviensis. Mathematics, **633** (2001), 51–65.

[7] SKOROKHOD, A. V.: *Asymptotic Methods of Theory of Stochastic Differential Equations*, AMS, Providance, 1994.

Current address

CARKOVS Jevgenijs
Probability and Statistics Dept, The Riga Technical University, 1/4 Meža iela, Rīga LV-1048 Latvia

PAVLENKO Oksana
Probability and Statistics Dept, The Riga Technical University, 1/4 Meža iela, Rīga LV-1048 Latvia
email Oksana.Pavlenkortu.lv

Journal of Applied Mathematics

volume III (2010), number II

REPRESENTATION OF SOLUTIONS OF LINEAR DISCRETE SYSTEMS WITH CONSTANT COEFFICIENTS, A SINGLE DELAY AND WITH IMPULSES

DIBLÍK Josef, (CZ), MORÁVKOVÁ Blanka, (CZ)

Abstract. The purpose of this paper is to develop a method for the construction of solutions of linear discrete systems with constant coefficients, with pure delay and with impulses. Solutions are expressed with the aid of a special function called a discrete matrix delayed exponential.

Key words and phrases. Impulses, linear discrete system with constant coefficients.

Mathematics Subject Classification. Primary 39A10; Secondary 39A11

1 Introduction

We use the following notation throughout this paper: For integers s, q, $s \leq q$ we define a set $\mathbb{Z}_s^q := \{s, s+1, \dots, q-1, q\}$. Similarly we define sets $\mathbb{Z}_{-\infty}^q := \{\dots, q-1, q\}$ and $\mathbb{Z}_s^\infty := \{s, s+1, \dots\}$. The function $[\,\cdot\,]$ is the greatest integer function.

Consider the initial Cauchy problem

$$\Delta x(k) = Bx(k-m), \quad k \in \mathbb{Z}_0^\infty, \tag{1}$$

$$x(k) = \varphi(k), \quad k \in \mathbb{Z}_{-m}^0 \tag{2}$$

where $m \geq 1$ is a fixed integer, $B = (b_{ij})$ is a constant $n \times n$ matrix, $x\colon \mathbb{Z}_{-m}^\infty \to \mathbb{R}^n$, $\varphi\colon \mathbb{Z}_{-m}^0 \to \mathbb{R}^n$ and $\Delta x(k) = x(k+1) - x(k)$.

We add impulses $J_i \in \mathbb{R}^n$ to x at points having a form $i(m+1)+1$ where the index $i \geq 0$ is defined as $i = \left[\frac{k-1}{m+1}\right]$ for every $k \in \mathbb{Z}_0^{\infty}$, i.e., we set

$$x(i(m+1)+1) = x(i(m+1)+1-0) + J_i \tag{3}$$

and investigate the solution of the problem (1) – (3).

Before we deal with the solution of the problem (1) – (3), we will give the definitions and a theorem needed to solve our problem. We will also show an example to get a better understanding of the problem.

Definition 1.1 *For arbitrary integers n and k, we define the binomial coefficient $\binom{n}{k}$:*

$$\binom{n}{k} := \begin{cases} \dfrac{n!}{k!(n-k)!} & \textit{if} \quad n \geq k \geq 0, \\ 0 & \textit{otherwise}. \end{cases} \tag{4}$$

In this paper, we use a special matrix function called a discrete function exponential. Such a discrete matrix function was first defined in [1], [2].

Definition 1.2 *For an $n \times n$ constant matrix B, $k \in \mathbb{Z}$ and fixed $m \in \mathbb{N}$, we define the discrete matrix delayed exponential e_m^{Bk} as follows:*

$$\mathrm{e}_m^{Bk} := \begin{cases} \Theta \quad \textit{if} \quad k \in \mathbb{Z}_{-\infty}^{-m-1}, \\ I \quad \textit{if} \quad k \in \mathbb{Z}_{-m}^{0}, \\ I + B \cdot \binom{k}{1} \quad \textit{if} \quad k \in \mathbb{Z}_1^{m+1}, \\ I + B \cdot \binom{k}{1} + B^2 \cdot \binom{k-m}{2} \quad \textit{if} \quad k \in \mathbb{Z}_{(m+1)+1}^{2(m+1)}, \\ I + B \cdot \binom{k}{1} + B^2 \cdot \binom{k-m}{2} + B^3 \cdot \binom{k-2m}{3} \quad \textit{if} \quad k \in \mathbb{Z}_{2(m+1)+1}^{3(m+1)}, \\ \dots \\ I + B \cdot \binom{k}{1} + B^2 \cdot \binom{k-m}{2} + \dots + B^\ell \cdot \binom{k-(\ell-1)m}{\ell} \\ \qquad \textit{if} \quad k \in \mathbb{Z}_{(\ell-1)(m+1)+1}^{\ell(m+1)}, \ \ell = 0, 1, 2, \dots \end{cases} \tag{5}$$

where Θ is $n \times n$ null matrix and I is $n \times n$ unit matrix.

The Definition 1.2 of the discrete matrix delayed exponential can be shortened as

$$\mathrm{e}_m^{Bk} := \begin{cases} \Theta & \text{if} \quad k \in \mathbb{Z}_{-\infty}^{-m-1}, \\ I + \sum_{j=1}^{\ell} B^j \cdot \binom{k-(j-1)m}{j} & \text{if} \quad k \in \mathbb{Z}_{(\ell-1)(m+1)+1}^{\ell(m+1)}, \ \ell = 0, 1, 2, \dots . \end{cases}$$

Next, Theorem 1.3 is proved in [1].

Theorem 1.3 *Let B be a constant $n \times n$ matrix. Then, for $k \in \mathbb{Z}_{-m}^{\infty}$,*

$$\Delta\, \mathrm{e}_m^{Bk} = B\mathrm{e}_m^{B(k-m)}. \tag{6}$$

The following example illustrates the influence of impulses on the solution and serves as a motivation for the formulation of a general case.

Example 1.4 *We consider a particular case of (1) if $n = 1$, $B = b$, $m = 3$ together with an initial problem (2) for $\varphi(k) = 1$, $k \in \mathbb{Z}_{-3}^{0}$ and with impulses $J_i \in \mathbb{R}$ at points $i(m+1)+1 = 4i+1$ where $i \geq 0$, $i = \left[\frac{k-1}{m+1}\right] = \left[\frac{k-1}{4}\right]$:*

$$\Delta x(k) = bx(k-3), \tag{7}$$

$$x(-3) = x(-2) = x(-1) = x(0) = 1, \tag{8}$$

$$x(4i+1) = x(4i+1-0) + J_i, \tag{9}$$

where $b \in \mathbb{R}$, $b \neq 0$. Rewriting the equation (7) as

$$x(k+1) = x(k) + bx(k-3)$$

and solving it by the method of steps, we conclude that the solution of the problem, can be written in the form:

$$x(k) = b^0\binom{k+3}{0} \qquad \text{if} \quad k \in \mathbb{Z}_{-3}^{0},$$

$$x(k) = b^0\binom{k+3}{0} + b^1\binom{k}{1} + J_0\, b^0\binom{k-1}{0} \qquad \text{if} \quad k \in \mathbb{Z}_{1}^{4},$$

$$x(k) = b^0\binom{k+3}{0} + b^1\binom{k}{1} + b^2\binom{k-3}{2} + J_0\left[b^0\binom{k-1}{0} + b^1\binom{k-4}{1}\right] + J_1\, b^0\binom{k-5}{0}$$
$$\text{if} \quad k \in \mathbb{Z}_{5}^{8},$$

$$x(k) = b^0\binom{k+3}{0} + b^1\binom{k}{1} + b^2\binom{k-3}{2} + b^3\binom{k-6}{3} + J_0\left[b^0\binom{k-1}{0} + b^1\binom{k-4}{1}\right.$$
$$\left. + b^2\binom{k-7}{2}\right] + J_1\left[b^0\binom{k-5}{0} + b^1\binom{k-8}{1}\right] + J_2\, b^0\binom{k-9}{0} \qquad \text{if} \quad k \in \mathbb{Z}_{9}^{12},$$

$$\vdots$$

$$
\begin{aligned}
x(k) = {} & b^0\binom{k+3}{0} + b^1\binom{k}{1} + b^2\binom{k-3}{2} + \cdots + b^\ell\binom{k-3(\ell-1)}{\ell} \\
& + J_0\left[b^0\binom{k-1}{0} + b^1\binom{k-4}{1} + b^2\binom{k-7}{2} + \cdots + b^{\ell-1}\binom{k-4-3(\ell-2)}{\ell-1}\right] \\
& + J_1\left[b^0\binom{k-5}{0} + b^1\binom{k-8}{1} + b^2\binom{k-11}{2} + \cdots + b^{\ell-2}\binom{k-8-3(\ell-3)}{\ell-2}\right] \\
& + J_2\left[b^0\binom{k-9}{0} + b^1\binom{k-12}{1} + b^2\binom{k-15}{2} + \cdots + b^{\ell-3}\binom{k-12-3(\ell-4)}{\ell-3}\right] \\
& + J_3\left[b^0\binom{k-13}{0} + b^1\binom{k-16}{1} + b^2\binom{k-19}{2} + \cdots + b^{\ell-4}\binom{k-16-3(\ell-5)}{\ell-4}\right] \\
& + \cdots \\
& + J_i\left[b^0\binom{k-4(i+1)+3}{0} + b^1\binom{k-4(i+1)}{1} + b^2\binom{k-4(i+1)-3}{2}\right. \\
& \qquad \left. + b^3\binom{k-4(i+1)-6}{3} + \cdots + b^{\ell-(i+1)}\binom{k-4(i+1)-3(\ell-(i+2))}{\ell-(i+1)}\right] \\
& \qquad\qquad \textit{if} \quad k \in \mathbb{Z}_{4(\ell-1)+1}^{4(\ell-1)+4},\ \ell = 0,1,2,\ldots,\ i = \left[\tfrac{k-1}{4}\right],\ i \geq 0.
\end{aligned}
$$

The solution of the problem (7) – (9) *can be shortened to*

$$
x(k) = \sum_{j=0}^{\ell} b^j\binom{k-3(j-1)}{j} + \sum_{q=0}^{i} J_q \sum_{j=0}^{\ell-(q+1)} b^j\binom{k-4(q+1)-3(j-1)}{j}, \tag{10}
$$

for $k \in \mathbb{Z}_{4(\ell-1)+1}^{4(\ell-1)+4},\ \ell = 0,1,2,\ldots,\ i = \left[\frac{k-1}{4}\right],\ i \geq 0.$

2 Representation of a Solution of a Homogeneous Initial Problem

Theorem 2.1 *Let B be a constant $n \times n$ matrix, m be a fixed integer, $J_i \in \mathbb{R}^n$. Then the solution of the initial Cauchy problem with impulses*

$$
\Delta x(k) = Bx(k-m), \quad k \in \mathbb{Z}_0^{\infty}, \tag{11}
$$

$$
x(k) = \varphi(k), \quad k \in \mathbb{Z}_{-m}^{0}, \tag{12}
$$

$$
x(i(m+1)+1) = x(i(m+1)+1-0) + J_i, \quad J_i \in \mathbb{R}^n, \quad i \geq 0, \quad i = \left[\tfrac{k-1}{m}\right] \tag{13}
$$

can be expressed in the form:

$$
x(k) = \mathrm{e}_m^{Bk}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(k-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_m^{B(k-(q+1)(m+1))} \tag{14}
$$

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

Proof. We substitute (14) into the left-hand side $\mathcal{L}$ of the equation (11):

$$\begin{aligned}
\mathcal{L} &= \Delta x(k) \\
&= \Delta \left[\mathrm{e}_m^{Bk}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(k-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_m^{B(k-(q+1)(m+1))} \right] \\
&= [\text{according to the Theorem 1.3}] \\
&= \Delta \mathrm{e}_m^{Bk}\varphi(-m) + \sum_{j=-m+1}^{0} \Delta\mathrm{e}_m^{B(k-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \Delta\mathrm{e}_m^{B(k-(q+1)(m+1))} \\
&= B\mathrm{e}_m^{B(k-m)}\varphi(-m) + \sum_{j=-m+1}^{0} B\mathrm{e}_m^{B(k-m-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q B\mathrm{e}_m^{B(k-m-(q+1)(m+1))} \\
&= B \left[\mathrm{e}_m^{B(k-m)}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(k-m-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_m^{B(k-m-(q+1)(m+1))} \right].
\end{aligned}$$

Now we substitute (14) into the right-hand side $\mathcal{R}$ of the equation (11):

$$\begin{aligned}
\mathcal{R} &= Bx(k-m) \\
&= B \left[\mathrm{e}_m^{B(k-m)}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(k-m-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_m^{B(k-m-(q+1)(m+1))} \right].
\end{aligned}$$

Since $\mathcal{L} = \mathcal{R}$, (14) is a solution of (11), (12).

Now we have to prove that (13) holds, too. We substitute (14) into the left-hand side $\mathcal{L}^*$ and right-hand side $\mathcal{R}^*$ of (13):

$$\begin{aligned}
\mathcal{L}^* &= x(i(m+1)+1) \\
&= \mathrm{e}_m^{B(i(m+1)+1)}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(i(m+1)+1-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))},
\end{aligned}$$

$$\begin{aligned}
\mathcal{R}^* &= x(i(m+1)+1-0) + J_i \\
&= \mathrm{e}_m^{B(i(m+1)+1)}\varphi(-m) + \sum_{j=-m+1}^{0} \mathrm{e}_m^{B(i(m+1)+1-m-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i-1} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))} \\
&\quad + J_i.
\end{aligned}$$

Since

$$\begin{aligned}\sum_{q=0}^{i} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))} &= \sum_{q=0}^{i-1} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))} + J_i \mathrm{e}_m^{B(i(m+1)+1-(i+1)(m+1))} \\ &= \sum_{q=0}^{i-1} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))} + J_i \mathrm{e}_m^{B(-m)} \\ &= \left[\text{according to the Definition 1.2 },\ \mathrm{e}_m^{B(-m)} = I\right] \\ &= \sum_{q=0}^{i-1} J_q \mathrm{e}_m^{B(i(m+1)+1-(q+1)(m+1))} + J_i\end{aligned}$$

it is obvious that $\mathcal{L}^* = \mathcal{R}^*$ and (13) holds.

Example 2.2 *We consider the problem* (11) – (13) *where* $n = 1$*,* $B = b$*,* $m = 3$*,* $\varphi(k) = 1$ *for* $k \in \mathbb{Z}_{-3}^{0}$*. Then* (14) *takes a form:*

$$x(k) = \mathrm{e}_3^{bk}\varphi(-3) + \sum_{j=-3+1}^{0} \mathrm{e}_3^{b(k-3-j)}\Delta\varphi(j-1) + \sum_{q=0}^{i} J_q \mathrm{e}_3^{b(k-(q+1)(3+1))}. \tag{15}$$

This problem was also solved in Exaple 1.4. We will show that the representations (15) *and* (10) *are equivalent.*

We write out all the addition terms of (15)*:*

$$\begin{aligned}\mathrm{e}_3^{bk}\varphi(-3) &= [\textit{according to the Definition 1.2}\,] \\ &= 1 + b\binom{k}{1} + b^2\binom{k-3}{2} + b^3\binom{k-6}{3} + \cdots + b^\ell\binom{k-3(\ell-1)}{\ell} \\ &= b^0\binom{k+3}{0} + b^1\binom{k}{1} + b^2\binom{k-3}{2} + b^3\binom{k-6}{3} + \cdots + b^\ell\binom{k-3(\ell-1)}{\ell} \\ &= \sum_{j=0}^{\ell} b^j\binom{k-3(j-1)}{j},\end{aligned}$$

$$\begin{aligned}\sum_{j=-3+1}^{0} \mathrm{e}_3^{b(k-3-j)}\Delta\varphi(j-1) &= \sum_{j=-2}^{0} \mathrm{e}_3^{b(k-3-j)}\Delta\varphi(j-1) \\ &= \mathrm{e}_3^{b(k-1)}\Delta\varphi(-3) + \mathrm{e}_3^{b(k-2)}\Delta\varphi(-2) + \mathrm{e}_3^{b(k-3)}\Delta\varphi(-1) \\ &= \mathrm{e}_3^{b(k-1)}(\varphi(-2)-\varphi(-3)) + \mathrm{e}_3^{b(k-2)}(\varphi(-1)-\varphi(-2)) \\ &\qquad + \mathrm{e}_3^{b(k-3)}(\varphi(0)-\varphi(-1)) \\ &= 0,\end{aligned}$$

$$
\begin{aligned}
\sum_{q=0}^{i} J_q \mathrm{e}_3^{b(k-(q+1)(3+1))} &= \sum_{q=0}^{i} J_q \mathrm{e}_3^{b(k-4(q+1))} \\
&= [\textit{according to the Definition 1.2 }] \\
&= \sum_{q=0}^{i} J_q \sum_{j=0}^{\ell} b^j \binom{k-4(q+1)-3(j-1)}{j} \\
&= \sum_{q=0}^{i} J_q \Bigg[b^0 \binom{k-4(q+1)+3}{0} + b^1 \binom{k-4(q+1)}{1} + b^2 \binom{k-4(q+1)-3}{2} \\
&\qquad + \cdots + b^{\ell-(q+1)} \binom{k-4(q+1)-3(\ell-(q+1)-1)}{\ell-(q+1)} \\
&\qquad + b^{\ell-(q+1)+1} \binom{k-4(q+1)-3(\ell-(q+1))}{\ell-(q+1)+1} \\
&\qquad + \cdots + b^{\ell} \binom{k-4(q+1)-3(\ell-1)}{\ell} \Bigg].
\end{aligned}
$$

We prove that the binomical coeficients

$$\binom{k-4(q+1)-3(\ell-(q+1))}{\ell-(q+1)+1}, \ldots, \binom{k-4(q+1)-3(\ell-1)}{\ell}$$

are equal to zero. These coeficients can be writen as

$$\binom{k-4(q+1)-3(\ell-(q+1)+p-1)}{\ell-(q+1)+p}, \quad \textit{where } p = 1, 2, \ldots, q+1.$$

Since $k \in \mathbb{Z}_{4(\ell-1)+1}^{4(\ell-1)+4}$*, we can write* $k = 4(l-1)+h$ *where* $h = 1, 2, 3, 4$*. Thus*

$$
\begin{aligned}
\binom{k-4(q+1)-3(\ell-(q+1)+p-1)}{\ell-(q+1)+p} &= \binom{4(l-1)+h-4(q+1)-3(\ell-(q+1)+p-1)}{\ell-(q+1)+p} \\
&= \binom{\ell-q-3p-2+h}{l-q+p-1} \\
&= [\textit{because } -3p-2+h < p-1] \\
&= 0.
\end{aligned}
$$

Hence

$$\sum_{q=0}^{i} J_q \mathrm{e}_3^{b(k-(q+1)(3+1))} = \sum_{q=0}^{i} J_q \sum_{j=0}^{\ell-(q+1)} b^j \binom{k-4(q+1)-3(j-1)}{j}.$$

Then the solution (14) *is in the form*

$$x(k) = \sum_{j=0}^{\ell} b^j \binom{k-3(j-1)}{j} + \sum_{q=0}^{i} J_q \sum_{j=0}^{\ell-(q+1)} b^j \binom{k-4(q+1)-3(j-1)}{j},$$

which is the solution (10) *of the problem* (7) – (9) .

Acknowledgement

Research was supported by the Grant 201/10/1032 of Czech Grant Agency (Prague) and by the Council of Czech Government MSM 00216 30503 and MSM 00216 30519.

References

[1] J. DIBLÍK, D. YA. KHUSAINOV, *Representation of solutions of linear discrete systems with constant coefficients and pure delay*. Adv. Difference Equ. 2006, Art. ID 80825, 13 pp.

[2] J. DIBLÍK, D. YA. KHUSAINOV, *Representation of solutions of discrete delayed system* $x(k+1) = Ax(k) + Bx(k-m) + f(k)$ *with commutative matrices*. J. Math. Anal. Appl. 318 (2006), no. 1, 63–76.

[3] J. DIBLÍK, D. YA. KHUSAINOV, Z. ŠMARDA, *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.

Current address

Diblík Josef, Prof. RNDr., DrSc.
Brno University of Technology, Czech Republic
tel.: +420-514147601, +420-541143155
email: diblik.j@fce.vutbr.cz, diblik@feec.vutbr.cz

Morávková Blanka, Mgr.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 8, 616 00 Brno, Czech Republic
email: xmorav17@stud.feec.vutbr.cz

Department of Mathematics and Descriptive Geometry
Faculty of Civil Engineering
Brno University of Technology
Žižkova 17, 662 37 Brno, Czech Republic
email: moravkova.b@fce.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

SOME GENERALIZATIONS OF BANACH FIXED POINT THEOREM

FAJMON Břetislav, (CZ), ŠMARDA Zdeněk, (CZ)

Abstract. In the paper the fixed point theorems are presented, which assert that every complete metric space is a fixed point space for the class of contractive mappings. The obtained results outgoing from the classical Banach fixed point theorem generalize contractive conditions which do not imply the continuity of an operator. The illustrative example is given, as well.

Key words and phrases. Fixed point theorems, Banach contraction principle.

Mathematics Subject Classification. 47H10.

1 Introduction

The Banach contraction principle is the simplest and one of the most versatile elementary results in fixed point theory. Being based on an iteration process, it can be implemented on a computer to find the fixed point of a contractive map. It produces approximations of any required accuracy, and moreover, even the number of iterations needed to get a specified accuracy can be determined.

First we recal some basic notions. Consider the operator equation

$$u = Tu, \quad u \in X, \tag{1}$$

where X is a complete metric space. Solve (1) by means of the following iteration method:

$$u_{n+1} = Tu_n, \quad n = 0, 1, \dots, \tag{2}$$

where $u_0 \in X$. Each solution of (1) is called a *fixed point* of the operator T.

Theorem 1.1 (Banach fixed point theorem). *Let (X, d) be a complete metric space $M \subseteq X$ and $T : M \to M$ be a map satisfying*

$$d(Tx, Ty) \leq kd(x, y), \quad for\ all\ x, y \in M, \tag{3}$$

where $0 \leq k < 1$ is a constant. Then, the following hold true:

(i) Existence and uniqueness. *Equation (1) has exactly one fixed point $u \in M$.*

(ii) Convergence of the iteration method. *For each given $u_0 \in M$ the sequence (u_n) constructed by (2) converges to the unique solution u of equation (1).*

(iii) Error estimates. *For all $n = 0, 1, \ldots$ we have so-called a priori error estimate*

$$d(u_n, u) \leq k^n(1 - k)^{-1}d(u_1, u_0), \tag{4}$$

and the so-called a posteriori error estimate

$$d(u_{n+1}, u) \leq k(1 - k)^{-1}d(u_{n+1}, u_n). \tag{5}$$

(iv) Rate of convergence. *For all $n = 0, 1, \ldots$ we have*

$$d(u_{n+1}, u) \leq kd(u_n, u).$$

This theorem was proved by Banach in 1920. The Banach fixed point theorem is also called the *contraction principle*.

The a apriori estimate (4) makes it possible to use the knowledge of initial value u_0 along $u_1 = Tu_0$ to determine the maximal number of steps of iteration required to attain a desidered level of precision.

In contrast to this, the a posteriori estimate (5) allows us to use computed values u_n and u_{n+1} to determine the accuracy of approximation u_{n+1}.

Theorem 1.1 suffers from one drawback the contractive condition (3) forces T to be continuous on M. It was then natural to ask if there exist or not weaker contractive conditions which do not imply the continuity of T. This was answered by R.Kannan [5] in 1968 , who proved a fixed point theorem which extends Theorem 1.1 to mappings that need not be continuous on M (but are continuous in their fixed point) by considering instead of (3) the next contractive condition: there exists a constant $b \in [0, 1/2)$ such that

$$d(Tx, Ty) \leq b[d(x, Tx) + d(y, Ty)], \quad \text{for all } x, y \in M. \tag{6}$$

A lot of papers were devoted to obtaining fixed point theorems for various classes of contractive type conditions that do not require the continuity of T (see [2],[9],[10]). One of them , actually a sort of dual of Kannan fixed point theorem, due to Chatterjea [3], is based on a condition similar to (6): there exists a constant $c \in [0, 1/2)$ such that

$$d(Tx, Ty) \leq c[d(x, Ty) + d(y, Tx)], \quad \text{for all } x, y \in M. \tag{7}$$

For a presentation and comparison of such kind of fixed point theorems, see [2],[6],[7],[8]. Ciric [4] determined the contractive condition: there exists $0 \leq h < 1$ such that for all $x, y \in M$

$$d(Tx, Ty) \leq h \max\{d(x,y), d(x,Tx), d(y,Ty), d(x,Ty), d(y,Tx)\}. \tag{8}$$

A mapping satisfying (8) is commonly called *quasi contraction.* It is obvious that each of the conditions (3),(6) and (7) does imply (8). One of the most general contraction conditions has been obtained by Berinde [1].

Theorem 1.2 *Let (X,d) be a complete metric space and let $T : X \to X$ be a Ciric almost contraction, that is, a mapping for which there exists a constant $\alpha \in (0,1]$ and $L \geq 0$ such that*

$$d(Tx, Ty) \leq \alpha M(x,y) + Ld(y,Tx) \quad \text{for all } x, y \in X,$$

where

$$M(x,y) = \max\{d(x,y), d(x,Tx), d(y,Ty), d(x,Ty), d(y,Tx)\}.$$

Then

1) *Fix(T)* $= \{x \in X : Tx = x\} \neq \emptyset$.
2) *For any $x_0 \in X$, the Picard iteration (x_n) given by (2) converges to some $x^* \in Fix(T)$.*
3) *The following estimate holds:*

$$d(x_n, x^*) \leq \frac{\alpha^n}{(1-\alpha)^2} d(x, Tx), \quad n = 1, 2, \ldots$$

2 Illustrative example

Consider the following singular initial value problem

$$y'(t) = \mathcal{F}\left(t, y(t), \int_{0^+}^{t} K(t,s,y(t),y(s))ds, \right), \quad y(0^+) = 0, \tag{9}$$

where

(I) $\mathcal{F} : \Omega \to R^n$, $\mathcal{F} \in C^0(\Omega)$,
$\Omega = \{(t, u_1, u_2) \in J \times R^n \times R^n : |u_1| \leq \phi(t),\ |u_2| \leq \psi(t)\}$, $J = (0, t_0]$, $t_0 > 0$, $0 < \phi(t) \in C^0(J)$, $\phi(0^+) = 0$, $0 < \psi(t) \in C^0(J)$, $|\cdot|$ denotes the usual norm in R^n, $|\mathcal{F}(t, \overline{u}_1, \overline{u}_2) - \mathcal{F}(t, \overline{\overline{u}}_1, \overline{\overline{u}}_2| \leq M_1|\overline{u}_1 - \overline{\overline{u}}_1| + M_2|\overline{u}_2 - \overline{\overline{u}}_2|$ for all $(t, \overline{u}_1, \overline{u}_2,)$, $(t, \overline{\overline{u}}_1, \overline{\overline{u}}_2) \in \Omega$, $M_i \geq 0$, $i = 1, 2$.

(II) $K : \Omega_1 \to R^n$, $K \in C^0(\Omega_1)$,
$\Omega_1 = \{(t, s, v_1, v_2) \in J \times J \times R^n \times R^n : |v_1| \leq \phi(t)\ |v_2| \leq \phi(t).\}$, $|K(t, s, \overline{v}_1, \overline{v}_2) - K(t, s, \overline{\overline{v}}_1, \overline{\overline{v}}_1| \leq N_1|\overline{v}_1 - \overline{\overline{v}}_1| + N_2|\overline{v}_2 - \overline{\overline{v}}_2|$ for all $(t, s, \overline{v}_1, \overline{v}_2)$, $(t, s, \overline{\overline{v}}_1, \overline{\overline{v}}_2) \in \Omega_1$, $N_i \geq 0$, $i = 1, 2$.

The initial value problem (9) is equivalent to the system of integral equations

$$y(t) = \int_{0^+}^{t} \mathcal{F}\left(s, y(s), \int_{0^+}^{s} K(s, w, y(s), y(w))dw, \right) ds. \tag{10}$$

We want to solve singular initial problem (9) along with the iteration method

$$y_{n+1}(t) = \int_{0^+}^{t} \mathcal{F}\left(s, y_n(s), \int_{0^+}^{s} K(s, w, y_n(s), y_n(w))dw\right) ds, \quad n = 0, 1, \ldots, \tag{11}$$

where $y_0(x) \equiv 0$.

Theorem 2.1 *Let the functions* $\mathcal{F}(t, u_1, u_2)$, $K(t, s, v_1, v_2)$ *satisfy conditions* (I), (II) *and, moreover*

(i)

$$|\mathcal{F}| \leq g_1(t)|u_1| + g_2(t)|u_2|, \quad 0 < g_i(t) \in C^0(J),\ i = 1, 2, \quad \int_{0^+}^{t} g_1(s)\phi(s)ds \leq \alpha\phi(t),$$

$$\int_{0^+}^{t} g_2(s)\psi(s)ds \leq \beta\phi(t), \quad \alpha + \beta \leq 1.$$

(ii) Let H *be the Banach space of continuous vector-valued functions*

$$h : J_0 \to R^n,\ J_0 = [0, t_0],\ |h(t)| \leq \phi(t)$$

on J *with the norm*

$$||h||_\lambda = \max_{t \in J_0}\{e^{-\lambda t}|h(t)|\},$$

where $\lambda > 0$ *is an arbitrary parameter.*

(iii) There is a real number k *such that* $k = \left(\frac{M_1 + t_0 M_2 N_1}{\lambda} + \frac{M_2 N_2 + M_2 N_1}{\lambda^2}\right) < 1$.

Then the following hold true:

Singular initial value problem (9) *has a unique solution* $y(t)$, $t \in J$.

The sequence (y_n) *constructed by (11) converges to* $y(x)$.

For all $n = 0, 1, \ldots$ *we get the following error estimates:*

$$||y_n - y|| \leq k^n(1 - k)^{-1}||y_1||,$$

$$||y_{n+1} - y|| \leq k(1 - k)^{-1}||y_{n+1} - y_n||.$$

Proof. Define the operator T by the right-hand side of (10)

$$T(h)=\int_{0^+}^{t}\mathcal{F}\left(s,h(s),\int_{0^+}^{s}K(s,w,h(s),h(w))dw,\right)ds,$$

where $h\in H$. The transformation T maps H continuously into itself because

$$\begin{aligned}|T(h)| &\le \int_{0^+}^{t}\left|\mathcal{F}\left(s,h(s),\int_{0^+}^{s}K(s,w,h(s),h(w))dw,\right)\right|ds\\ &\le \int_{0^+}^{t}\left[g_1(s)|h(s)|+g_2(s)\int_{0^+}^{s}|K(s,w,h(s),h(w))|dw\right]ds\\ &\le \int_{0^+}^{t}(g_1(s)\phi(s)+g_2(s)\psi(s))\,ds\le(\alpha+\beta)\phi(t)\le\phi(t)\end{aligned}$$

for every $h\in H$. We shall prove that

$$||T(h_2)-T(h_1)||_\lambda\le||h_2-h_1||_\lambda\left(\frac{M_1+t_0M_2N_1}{\lambda}+\frac{M_2N_2+M_2N_1}{\lambda^2}\right)$$

for all $h_1,h_2\in H$. Using (I), (II) and the definition $||.||_\lambda$ we have

$$\begin{aligned}|T(h_2)-T(h_1)| &\le \int_{0^+}^{t}\left|\mathcal{F}\left(s,h_2(s),\int_{0^+}^{s}K(s,w,h_2(s),h_2(w))dw,\right)\right.\\ &\qquad\left.-\mathcal{F}\left(s,h_1(s),\int_{0^+}^{s}K(s,w,h_1(s),h_1(w))dw,\right)\right|ds\\ &\le \int_{0^+}^{t}\left(M_1|h_2(s)-h_1(s)|+M_2\int_{0^+}^{s}|K(s,w,h_2(s),h_2(w))-K(s,w,h_1(s),h_1(w))|dw\right)ds\\ &\le \int_{0^+}^{t}\left(M_1|h_2(s)-h_1(s)|+M_2\int_{0^+}^{s}(N_1|h_2(s)-h_1(s)|+N_2|h_2(w)-h_1(w)|)dw\right)ds\\ &\le M_1||h_2-h_1||_\lambda\int_{0^+}^{t}e^{\lambda s}ds+M_2N_1||h_2-h_1||_\lambda\int_{0^+}^{t}se^{\lambda s}ds+M_2N_2||h_2-h_1||_\lambda\int_{0^+}^{t}\int_{0^+}^{s}e^{(\lambda w)}dwds\\ &=||h_2-h_1||_\lambda\left(M_1(\frac{e^{\lambda t}}{\lambda}-\frac{1}{\lambda})+M_2N_1(\frac{te^{\lambda t}}{\lambda}-\frac{e^{\lambda t}}{\lambda^2}+\frac{1}{\lambda^2})+M_2N_2(\frac{e^{\lambda t}}{\lambda^2}-\frac{1}{\lambda^2}-\frac{t}{\lambda})\right)\\ &<||h_2-h_1||_\lambda e^{\lambda t}\left(\frac{M_1+t_0M_2N_1}{\lambda}+\frac{M_2N_2+M_2N_1}{\lambda^2}\right).\end{aligned}$$

Thus

$$||T(h_2)-T(h_1)||_\lambda=\max_{t\in J_0}\{e^{-\lambda t}|T(h_2)-T(h_1)|\}\le||h_2-h_1||_\lambda\left(\frac{M_1+t_0M_2N_1}{\lambda}+\frac{M_2N_2+M_2N_1}{\lambda^2}\right).$$

Now we choose $\lambda>0$ so that

$$\left(\frac{M_1+t_0M_2N_1}{\lambda}+\frac{M_2N_2+M_2N_1}{\lambda^2}\right)<1$$

and define

$$k := \left(\frac{M_1 + t_0 M_2 N_1}{\lambda} + \frac{M_2 N_2 + M_2 N_1}{\lambda^2} \right).$$

The assertions of Theorem 2.1. follow now from Theorem 1.1.

Acknowledgement

This research has been supported by the Czech Ministry of Education in the frames of MSM002160503 Research Intention MIKROSYN New Trends in Microelectronic Systems and Nanotechnologies and MSM0021630529 Research Intention Inteligent Systems in Automatization.

References

[1] BERINDE, V., *General constructive fixed point theorems for Ciric-type almost contractions in metric spaces*,Carpathian J.Math. **24**(2008), No.2, 10-19.

[2] BERINDE, V., *Iterative Approximation of Fixed Points,* 2nd Ed. Springer Verlag, Berlin Heilderberg New York, 2007.

[3] CHATTERJEA, S.K., *Fixed-point theorems,* C.R.Acad. Bulgare Sci. **25** (1972), 727-730.

[4] CIRIC, Lj.B., *A generalization of of Banach's contraction principle,* Proc. Am. Math. Soc. **45** (1974), 267-273.

[5] KANNAN, R., *Some results on fixed points,* Bull. Calcutta Math. Soc. **10** (1968), 71-76.

[6] MESZAROS, J., *A comparison of various defintions of contactive type mappings,* C.R.Acad. Bulgare Sci. **84** (1992), 167-194.

[7] RHOADES, B.E., *Contractive definitions reviseted,* Contemporary Mathematics **21** (1983)189-205.

[8] RHOADES, B.E., *A comparison of various definitions of contractive mappings,* Trans. Amer. math. Soc. **226** (1997), 257-290.

[9] RUS. I.A., *Principles and Applications of the Fixed Point Theory,* (in Romanian), Editura dacia, Cluj-Napoca, 1979.

[10] RUS, I.A., *Generalized Contractions and Applications,* Cluj University press, Cluj-Napoca, 2001.

Current address

RNDr. Břetislav Fajmon, Ph.D.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology, Technická 8, 616 00 BRNO
e-mail: fajmon@feec.vutbr.cz

Doc. RNDr. Zdeněk Šmarda, CSc.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology, Technická 8, 616 00 BRNO
e-mail: smarda@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

NON-NEGATIVITY PRESERVATION OF THE DISCRETE NONSTATIONARY HEAT EQUATION IN 1D AND 2D

FARAGÓ István, (HU), KOROTOV Sergey, (FI), SZABÓ Tamás, (HU)

Abstract. In this paper we analyse the preservation of the non-negativity property for the semidiscrete and fully discretized (by finite differences and finite elements) numerical solutions of the linear parabolic problem in one and two space dimensions. In particular, we derive the exact (necessary and sufficient) conditions for the 1D problem, and also give sufficient conditions for the 2D case, under which the non-negativity preservation property of the fully discretized problem is valid.

Key words and phrases. parabolic problems, semidiscretization, fully discretized problem, linear and bilinear finite elements, finite differences, non-negativity preservation, tridiagonal matrix, block-tridiagonal matrix.

Mathematics Subject Classification. 15A06, 65M06, 65M60

1 Introduction

Besides the covergence, another natural requirement in the process of numerical solution of partial differential equations, is a presevation of basic qualitative properties of the original (physical) solution, assuming that they are inherent to the continuous mathematical model. As an example, consider the advection-diffusion-reaction equation arising, e.g., in the large air-pollution modeling [22]:

$$\frac{\partial c}{\partial t} = -\mathrm{div}\,(vc) + \mathrm{div}\,(D\,\mathrm{grad}\,c) + R(c), \quad t \in (0,T], \quad c(0,x) = c_0(x), \tag{1}$$

where the vector-valued function $c(t,x)$ denotes the concentration of compounds, $v = v(t,x)$ presents the current velocity of the medium, D is the so-called diffusion coefficient matrix, and the function R describes the chemical reactions between the compounds and includes the parametrized deposition and emission. (As a rule, we use equation (1) in the componentwise sense.) We solve the above problem by suitably chosen numerical method. Since c denotes the concentration, which is always non-negative, it is natural to require the non-negativity from the numerical approximations of c as well.

The technique, often used to solve problems like (1) (based on several physical processes) is the so-called *operator splitting method*, see e. g. [4]: we choose a suitable time-step $\Delta t > 0$, and for $j = 1, 2, ...$ solve the sequence of subproblems

$$\frac{\partial u_1^{(j)}}{\partial t} = -\mathrm{div}\,(v u_1^{(j)}), \quad u_1^{(j)}((j-1)\Delta t, x) = u_3^{(j-1)}((j-1)\Delta t, x), \tag{2}$$

$$\frac{\partial u_2^{(j)}}{\partial t} = \mathrm{div}\,(D\ \mathrm{grad}\ u_2^{(j)}), \quad u_2^{(j)}((j-1)\Delta t, x) = u_1^{(j)}(j\Delta t, x), \tag{3}$$

$$\frac{\partial u_3^{(j)}}{\partial t} = R(u_3^{(j)}), \quad u_3^{(j)}((j-1)\Delta t, x) = u_2^{(j)}(j\Delta t, x), \tag{4}$$

on the intervals $[(j-1)\Delta t, j\Delta t]$, where $u_3^{(0)}(0, x) = c_0(x)$. In this procedure, $u_3^{(j)}$ yields an approximation to $c(j\Delta t, x)$.

For numerical solution of each of the above subproblems, we can choose a certain suitable numerical method (naturally, all three may be completely different each from other). Obviously, if all numerical techniques used for solving (2)–(4) are non-negativity preserving, then the whole computational scheme, used for solving (1), is non-negativity preserving.

In the above sequence of subproblems (2)–(4), the central role belongs to subproblem (3), which, in the simplest setting, has the following form

$$\frac{\partial u}{\partial t} = \triangle u \quad \text{in } \Omega_T = (0, T) \times \Omega, \tag{5}$$

$$u = 0 \quad \text{on } \Gamma_T = (0, T) \times \partial\Omega, \tag{6}$$

$$u|_{t=0} = u_0 \quad \text{on } \Omega, \tag{7}$$

where $\Omega \subset \mathbb{R}^d$ is a segment (for $d = 1$, and the problem is referred to as the one-dimensional), or a rectangle (for $d = 2$, and the problem is called two-dimensional), the symbol $\triangle$ denotes the Laplace operator, and u_0 is the initial function defined from the splitting procedure.

This equation has an importance of its own. For instance, it describes the heat conduction process, therefore hereafter we refer to it as the heat conduction equation. For this equation, the non-negativity preservation principle reads as follows: for any non-negative initial function u_0, the solution u has to be non-negative in Ω_T as well, see, e.g. [19].

A typical numerical technique for solving (5)–(7) presents a combination of separate discretizations in space and time. For the first one, we can employ the finite element method (based e.g. on the linear elements in the one-dimensional case, and on bilinear elements - for the two-dimensional case), or the standard finite difference method with the mesh-size h. As a result, we get the following Cauchy problem for the *semidiscrete solution* u_h

$$\frac{du_h}{dt}(t) = \triangle_h u_h(t), \qquad t \in (0, T), \tag{8}$$

where the initial value $u_h(0)$ is given and $\triangle_h$ denotes the corresponding discrete Laplace operator (represented by a matrix). Applying some suitable time discretization method with the time-step τ to problem (8), we finally arrive at the *fully discretized problem*, presenting the following algebraic iterative procedure

$$X_1 y^{j+1} = X_2 y^j, \tag{9}$$

where X_1 and X_2 are given matrices, and the vector y^j represents the approximation to $u_h(j\tau)$.

Our aim is to formulate, for a fixed standard parameter $\theta \in [0,1]$, such conditions on the discretization parameters h and τ, under which the corresponding semidiscrete and fully discretized solutions preserve the non-negativity property.

In this paper we study these problems. We show that in the one-dimensional case, the exact (necessary and sufficient) conditions (with respect to τ and h, for any fixed θ) can be obtained. The results are based on finding the *exact represenation* of the matrix $X = X_1^{-1} X_2$, where the crucial point consists of computing the inverse X_1^{-1}.

In the two-dimensional case, a similar problem is more difficult, due to more complicated structures of the corresponding matrices. A certain sufficient condition is given in [8] and it is based on the following requirements: $X_2 \geq 0$ and X_1 is a monotone matrix. However, finding the necessary and sufficient condition is still an open problem.

In our paper, we analyse first the non-negativity preservation for the semidiscrete solutions of (8), and, additionally, we establish direct connection between the non-negativity preservation of the semidiscrete solutions (8) for the one and two-dimensional cases. In Section 3 we give the exact condition for the non-negativity preservation in 1D and we give "the bounds" for the finite difference and linear finite element methods. In Section 4 we formulate the conditions under which the discrete problems are non-negativity preserving in 2D. In the final section we illustrate numerically the theoretically derived results.

We note that the non-negativity preservation property has a close relation to the validity of the discrete maximum principle. This topic is addressed, e.g., in [10], [5], [6] for the parabolic problem, and in [12], [13] - for the elliptic problems.

2 Non-negativity preservation for the semidiscrete solutions

In this section, we consider the non-negativity preservation for the Cauchy problem (8), where $\triangle_h$ arises from the space uniform discretization of the Laplace operator on rectangular mesh. The solution has a form

$$u_h(t) = \exp(\triangle_h t) u_0, \quad t \in (0,T), \tag{10}$$

where $\exp(\triangle_h t)$ denotes the exponent of the matrix $\triangle_h t$. Therefore, the problem of the non-negativity preservation is equivalent to finding conditions on dicretization methods for which the matrix exponential $\exp(\triangle_h t)$ is non-negative.

The following useful lemma holds (cf. [1, p. 172]).

Lemma 2.1 *Let A be an arbitrary square matrix with the entries a_{ij}. Then* $\exp(At)$ *is non-negative for any $t \geq 0$ if and only if the condition*

$$a_{ij} \geq 0 \quad \text{for all} \quad i \neq j \tag{11}$$

holds.

Proof: By the definition of the exponential, we have

$$\exp(At) = I + At + ...,$$

where I denotes the identity matrix. This series immediately shows the necessity of condition (11). Let now s be a scalar such that $A + sI$ is a non-negative matrix. Then, obviously, $\exp((A + sI)t)$ is non-negative if $t \geq 0$. Moreover, $\exp(-sIt)$ is also non-negative, and the matrices $(A + sI)t$ and $-sIt$ commute. Therefore, due to the identity

$$\exp(At) = \exp((A + sI)t - sIt) = \exp((A + sI)t) \cdot \exp(-sIt),$$

the sufficiency of condition (11) is also proven.

■

In the following we consider the application of this lemma to the one-dimensional case.

2.1 One-dimensional case

If problem (5)–(7) is considered in the one-dimensional case, the structure of the matrix $\triangle_h$ is well-known for the both finite difference and finite element (space) discretizations. Namely, using the standard denotations

$$Q = tridiag(1, -2, 1), \quad M = \frac{1}{6} tridiag(1, 4, 1), \tag{12}$$

we have:

for the finite difference method

$$\triangle_h = \frac{1}{h^2} Q, \tag{13}$$

for the linear finite element method

$$\triangle_h = \frac{1}{h^2} M^{-1} Q. \tag{14}$$

The matrix $\triangle_h$ from (13) obviously satisfies the condition (11), at the same time, the matrix $\triangle_h$ from (14) (which can be computed explicitly) is known to have its entries changing the sign chessboard-likely [7], i.e., it does not satisfy the condition (11).

Thus, using Lemma 2.1, we obtain the following result.

Theorem 2.1 *For the one-dimensional problem (8), the semidiscrete numerical solutions, obtained by the finite difference discretization, preserve the non-negativity property. However, this property is not preserved, in general, for the numerical solutions resulting from the linear finite element discretization.*

Remark 2.1 The lumped mass method for the linear finite element method results in the semidiscrete problem which coincides with the finite difference semidiscrete problem. Therefore, this approach makes possible to improve the qualitative property of the finite element discretization.

Remark 2.2 In [2] there is given the condition of the positivity of the time derivative of the semidiscrete solutions. Clearly, this condition can be regarded as a sufficient condition of the non-negativity preservation of the semidiscrete solutions.

Remark 2.3 Let us introduce the following denotation

$$T_1(p) = tridiag(1, p, 1), \tag{15}$$

where $p \in \mathbb{R}$, i.e., $Q = T_1(-2)$. Then, the semidiscretization in the form

$$\frac{du_h}{dt}(t) \quad = \quad \frac{1}{h^2}\, T_1(p) u_h(t), \qquad t \in (0, L), \tag{16}$$

is non-negativity preserving for any value of the parameter p. Therefore, instead of (5), we can consider a more general equation

$$\frac{\partial u}{\partial t} \quad = \quad \triangle u \; + ku \quad \text{in} \;\; \Omega_L = (0, L) \times \Omega, \tag{17}$$

where k is *any* constant, and prove that the finite difference semi-discretization for such an equation is non-negativity preserving, because the approximation of the new term effects only the diagonal elements of the matrix in (16).

Remark 2.4 However, the non-negativity property for the above problem is known to hold only for some $k \geq k_0$ for the original continuous problem (see, e.g., [19]), and, of course, the preservation of this property in the numerical realization is somewhat meaningless for certain values of k.

2.2 Two-dimensional case

We now consider the discretization on the uniform mesh (of the step-size h) of problem (5)–(7) in the two-dimensional case. We introduce the following denotation

$$T_2(p) = tridiag(I, T_1(p), I) \tag{18}$$

for a block tridiagonal matrix (from $\mathbb{R}^{n^2 \times n^2}$), where $p \in \mathbb{R}$.

We consider the both semidiscretization methods. Clearly, if we apply the usual finite difference method, then in the Cauchy-problem (8) the matrix $\triangle_h$ has the form

$$\triangle_h = \frac{1}{h^2}\, T_2(-4),$$

that is, all its off-diagonal elements are non-negative. If we use the linear finite element method, then we get $\triangle_h = (1/h^2)M^{-1}T_2(-4)$.

In what follows, we establish the relation between the exponentials of the matrices $T_1(p)$ from $\mathbb{R}^{n\times n}$ and $T_2(p)$ from $\mathbb{R}^{n^2\times n^2}$, i.e., the matrix exponential between the 1D and 2D discrete Laplacians. Using the denotation

$$A(p) = tridiag(1, p+2, 1) = T_1(p) + 2I, \tag{19}$$

we have

$$\begin{aligned} T_2(p) = tridiag(I, A(p) - 2I, I) &= \\ = tridiag(0, A(p), 0) + tridiag(I, -2I, I) &= I \otimes A(p) + Q \otimes I, \end{aligned} \tag{20}$$

where $\otimes$ denotes the Kronecker product of matrices, see e.g. [9, 17]. In order to attribute the matrix exponential of the matrix $T_2(p) \in \mathbb{R}^{n^2\times n^2}$ to the matrix exponentials of the matrices $A(p)$ and Q from $\mathbb{R}^{n\times n}$, i.e., the two-dimensional problem to the one-dimensional one, we prove the folowing lemma.

Lemma 2.2 *For the matrices $T_2(p), A(p)$, and Q, the relation*

$$\exp(T_2(p)) = \exp(Q) \otimes \exp(A(p)) \tag{21}$$

holds.

Proof: For any matrices A, B, C, D of the same size, we have [9, p. 228],

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

Therefore

$$\begin{aligned} (I \otimes A(p))(Q \otimes I) &= Q \otimes A(p), \\ (Q \otimes I)(I \otimes A(p)) &= Q \otimes A(p). \end{aligned} \tag{22}$$

Consequently, the corresponding exponential can be written by use of the binomial rule as follows

$$\begin{aligned} \exp(T_2(p)) &= \exp((I \otimes A(p)) + (Q \otimes I)) = \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(I \otimes A(p) + Q \otimes I)^n = \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{1}{k!(n-k)!}(I \otimes A(p))^k (Q \otimes I)^{n-k} = \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{1}{k!(n-k)!} Q^k \otimes A(p)^{n-k}. \end{aligned} \tag{23}$$

On the other hand, by the definition of the tensor product, we have

$$\exp(Q) \otimes \exp(A(p)) = \sum_{i=0}^{\infty} \frac{1}{i!} Q^i \otimes \sum_{j=0}^{\infty} \frac{1}{j!} A(p)^j = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{i! j!} Q^k \otimes A(p)^j. \tag{24}$$

Since the right-hand sides in (23) and (24) are equal, we obtain the relation (21). ■

Obviously, the tensor product of two matrices is non-negative if and only if both involved matrices are non-negative. Moreover, $\exp(Q)$ is non-negative and $\exp(A(p))$ is non-negative if and only if $\exp(T_1(p))$ is non-negative. Based on Lemma 2.1, we obtained the following statement.

Theorem 2.2 *For the two-dimensional problem on rectangular mesh, the semidiscrete numerical solution, obtained by the regular finite difference discretization, preserves the non-negativity property. However, this property is not preserved, in general, for the linear finite element discretization.*

Remark 2.5 Clearly, the statement of Theorem 2.2 is also valid for the more general equation (17).

3 The non-negativity preservation in 1D case

We consider the non-negativity preservation property of the discretization of the one-dimensional heat conduction problem with first homogeneous boundary conditions. (For the simplicity, the constant coefficient is assumed to be equal one.) Then we get problem (9) with the matrices from $\mathbb{R}^{N \times N}$ in the form:

$$X_1 = \frac{1}{\Delta t} M - \theta Q, \quad X_2 = \frac{1}{\Delta t} M + (1-\theta) Q, \tag{25}$$

where $M = I$ for the finite difference method and it has the form from (12) for the linear finite element method. Hence, these matrices have the following entries:

- for the finite difference method

$$\begin{aligned} X_1 &= \text{tridiag}\left[-\frac{\theta}{h^2}, \frac{1}{\Delta t} + 2\frac{\theta}{h^2}, -\frac{\theta}{h^2}\right], \\ X_2 &= \text{tridiag}\left[\frac{1-\theta}{h^2}, \frac{1}{\Delta t} - 2\frac{1-\theta}{h^2}, \frac{1-\theta}{h^2}\right], \end{aligned} \tag{26}$$

- for the linear finite element method the corresponding matrices are

$$\begin{aligned} X_1 &= \text{tridiag}\left[\frac{1}{6\Delta t} - \frac{\theta}{h^2}, \frac{2}{3\Delta t} + 2\frac{\theta}{h^2}, \frac{1}{6\Delta t} - \frac{\theta}{h^2}\right], \\ X_2 &= \text{tridiag}\left[\frac{1}{6\Delta t} + \frac{1-\theta}{h^2}, \frac{2}{3\Delta t} - 2\frac{1-\theta}{h^2}, \frac{1}{6\Delta t} + \frac{1-\theta}{h^2}\right]. \end{aligned} \tag{27}$$

For the non-negativity preservation property we require the condition

$$X = X_1^{-1} X_2 \geq 0. \tag{28}$$

Let us notice that the matrices in (26) and (27) have special structure: only the entries of the main-, super- and sub-diagonals differ from zero and the elements standing on the same diagonal are equal. Moreover, these matrices are symmetric, too. Such kind of matrix is called *uniformly continuant, symmetrical tridiagonal matrix* and they have some special qualitative properties, which will be considered in the sequel.

3.1 Non-negativity of the iteration matrix in general form

We consider the real, uniformly continuant, symmetrical tridiagonal matrices

$$X_1 = z \cdot \mathrm{tridiag}[-1, 2w, -1]; \quad X_2 = s \cdot \mathrm{tridiag}[1, p, 1] \tag{29}$$

with the assumptions

$$z > 0, \quad s > 0; \quad w > 1. \tag{30}$$

Our aim is to define for this case those conditions under which the iteration matrix X is non-negative.

We introduce the following one-pair matrix $G = (G_{ij})$, depending on the parameter w:

$$G_{i,j} = \begin{cases} \gamma_{i,j}, & \text{if } i \leq j \\ \gamma_{j,i}, & \text{if } j \leq i \end{cases} \tag{31}$$

$(i, j = 1, 2, \ldots, N)$, where

$$\gamma_{i,j} = \frac{\mathrm{sh}(i\vartheta)\mathrm{sh}(N+1-j)\vartheta}{\mathrm{sh}\vartheta\,\mathrm{sh}(N+1)\vartheta}, \quad \vartheta = \mathrm{arch}(w), \quad \text{with } w > 1. \tag{32}$$

We have the relation $X_1^{-1} = (1/z)G$ (see [17]), thus a direct computation verifies the validity of the following

Lemma 3.1 *For the matrices X_1 and X_2 of the form* (29) *the iteration matrix $X = X_1^{-1}X_2$ can be expressed as*

$$X = \frac{s}{z}\left[(2w + p)G - I\right]. \tag{33}$$

Hence, taking into the account the conditions (30), we get the following statement.

Lemma 3.2 *Under the condition* (30) *the iteration matrix $X \in \mathbb{R}^{N\times N}$ for arbitrary fixed N is non-negative if and only if the conditions*

$$2w + p > 0 \tag{34}$$

and

$$\gamma_{i,i} \geq \frac{1}{2w + p}, \quad i = 1, 2, \ldots, N \tag{35}$$

are fulfilled.

Now we analyze the expression on left hand side in condition (35).

Lemma 3.3 *For the diagonal elements of the matrix X the relation*

$$\min\{\gamma_{i,i},\ \ i=1,2,\dots,N\}=\gamma_{1,1}=\gamma_{N,N} \tag{36}$$

holds.

Proof. Introducing the functions $h_1(y) = K_1\mathrm{sh}(Cy)\ \mathrm{sh}(C(N+1-y))$ and $h_2(y) = K_2y(N+1-y)$ on the interval $[1,N]$, (where K_1 , K_2 and C are some positive constants), one can check that both functions take their maxima at the same point y=(N+1)/2. Moreover, on the interval [1,(N+1)/2) they are monotonically increasing, while on the interval ((N+1)/2,N] they are monotonically decreasing. Using this fact and the expressions for $\gamma_{i,i}$, we get the statement. ■

Combining Lemma 3.2 and Lemma 3.3, we obtain

Theorem 3.1 *Under the conditions* (30)*, for arbitrary fixed N the iteration matrix $X\in\mathbb{R}^{N\times N}$ is non-negative if and only if the conditions* (34) *and*

$$a(N):=\frac{\mathrm{sh}(N\vartheta)}{\mathrm{sh}((N+1)\vartheta)}\geq\frac{1}{2w+p} \tag{37}$$

are satisfied.

Obviously, (34) and (37) are necessary and sufficient conditions of the non-negativity for some fixed dimension N. Let us turn to the examination of the varying N. Due the relations

$$\frac{\mathrm{sh}(N\vartheta)}{\mathrm{sh}((N+1)\vartheta)}=\mathrm{ch}(\vartheta)-\coth((N+1)\vartheta)\mathrm{sh}(\vartheta), \tag{38}$$

we have

$$\sup\{\frac{\mathrm{sh}(N\vartheta)}{\mathrm{sh}((N+1)\vartheta)};\ N\in\mathbb{N}\}=\mathrm{ch}(\vartheta)-\mathrm{sh}(\vartheta)=\exp(-\vartheta). \tag{39}$$

Since the sequence $a(N)$ is monotonically increasing, it converges to its limit (which is its superior) monotonically. Thus, the conditions (34) and (37), that is, the necessary and sufficient conditions for some fixed N, serve as sufficient condition of the non-negativity of the matrices $X\in\mathbb{R}^{N_1\times N_1}$ for all $N_1\geq N$.

Let us observe that

$$\begin{aligned}\exp(-\vartheta)=\exp(-\mathrm{arch}(w))&=\exp\left(\ln\left[w+\sqrt{w^2-1}\right]^{-1}\right)\\&=\left[w+\sqrt{w^2-1}\right]^{-1}.\end{aligned} \tag{40}$$

Therefore, from some sufficiently large $N_0\in\mathbb{N}$ the relation $X\geq 0$ may be true only if the condition

$$\left[w+\sqrt{w^2-1}\right]^{-1}>\frac{1}{2w+p}, \tag{41}$$

i.e., the condition

$$p > -w + \sqrt{w^2 - 1} \tag{42}$$

is fulfilled. This proves the following

Theorem 3.2 *Assume that the conditions in* (30) *are satisfied. If, for some number $N_0 \in \mathbb{N}$, the conditions* (34) *and* (37) *are satisfied, then, all matrices $X \in \mathbb{R}^{N\times N}$ with $N \geq N_0$, are non-negative. Moreover, there exists such a number N_0, if and only if the condition* (42) *holds.*

Remark 3.1 *Since*

$$a(1) = \frac{\mathrm{sh}\vartheta}{\mathrm{sh}(2\vartheta)} = \frac{1}{2\mathrm{ch}\vartheta} = \frac{1}{2w},$$

therefore, (37) *results in the condition*

$$p \geq 0. \tag{43}$$

Remark 3.2 *Due to the relation*

$$a(2) = \frac{\mathrm{sh}(2\vartheta)}{\mathrm{sh}(3\vartheta)} = \frac{2\mathrm{ch}(\vartheta)}{4\mathrm{ch}^2(\vartheta) - 1} = \frac{2w}{4w^2 - 1},$$

condition (37) *results in the assumption*

$$p \geq -\frac{1}{2w}. \tag{44}$$

That is, $X \in \mathbb{R}^{N\times N}$ is non-negative for all $N = 2, 3, \dots,$ if and only if X_1 is an M-matrix and (44) *is valid.*

Remark 3.3 *The conditions* (43) *and* (44) *(corresponding to the cases $N = 1$ and $N = 2$, respectively) are sufficient conditions for the non-negativity of the matrix X in any larger dimension. For increasing N, the new conditions, which we obtain, are approaching to the necessary condition of non-negativity. Using* (38) *and* (39) *we can characterize the rate of the convergence: it is equal to the rate of convergence of the sequence $\{\coth(N\vartheta),\ n = 1, 2, \dots\}$ to one. Clearly,*

$$\coth(N\vartheta) = 1 + \frac{2}{[\exp(\vartheta)]^{2N} - 1}.$$

Using (40),

$$\exp(\vartheta) = w + \sqrt{w^2 - 1} =: \beta. \tag{45}$$

Hence, the sequence of the bounds of the sufficient conditions converges linearly with the ratio $1/\beta^2$ to the bound of the necessary condition.

3.2 Non-negativity of difference schemes in 1D

The results of the previous part can be used in the qualitative analysis of the finite difference and linear finite element mesh operators in 1D, given by the formula (26) and (27), respectively.

First we investigate the finite difference method. According to (26), the corresponding matrices are uniformly continuant and, using the notation $q = \Delta t/h^2$, they can be written in the form (29) with the choice

$$z = \frac{\theta q}{\Delta t}, \quad s = \frac{(1-\theta)q}{\Delta t}, \quad w = \frac{1+2\theta q}{2\theta q}, \quad p = \frac{1-2(1-\theta)q}{(1-\theta)q}. \tag{46}$$

First we consider two special choices for the parameter θ. For the case $\theta = 0$, according to (26), we have $X_1 = \frac{1}{\Delta t} I$ and $X_2 = \frac{1}{\Delta t} I - (1-\theta)Q$. Hence, X is non-negative if and only if the condition

$$q \leq \frac{1}{2} \tag{47}$$

is satisfied. For the case $\theta = 1$ we get $X_1 = \frac{1}{\Delta t} I + Q$ and $X_2 = \frac{1}{\Delta t} I$. Because such X_1 is monotone matrix, therefore we do not have any condition for the choice of the parameters h and Δt.

In what follows we pass to the analysis of the case $\theta \in (0,1)$. For this case, the conditions of (30) clearly are satisfied. Moreover, let us notice, that, under the choice (46) we have $2w + p = 1/\theta(1-\theta)q$, hence the condition (34) is always satisfied.

Using (43), we directly get that the condition of the non-negativity preservation for all $N = 1, 2, \ldots$ is the condition

$$q \leq \frac{1}{2(1-\theta)}. \tag{48}$$

However, the non-negativity preservationfor all $N = 2, 3, \ldots$ should be guaranteed by the weaker condition (44), which, in our case yields the inequality

$$\frac{1-2(1-\theta)q}{(1-\theta)q} \geq -\frac{\theta q}{1+2\theta q}. \tag{49}$$

Solving this problem, we get the upper bound

$$q \leq \frac{-1+2\theta+\sqrt{1-\theta(1-\theta)}}{3\theta(1-\theta)}, \tag{50}$$

which is larger than the bound in (47).

Our aim is to get the largest value for q under which the non-negativity preservation for sufficiently large values N still holds. Therefore we put the values w and p from (46) into the necessary condition (42). Then we should solve the inequality

$$\frac{1-2(1-\theta)q}{(1-\theta)q} \geq -\frac{1+2\theta q}{2\theta q} + \frac{\sqrt{1+4\theta q}}{2\theta q}. \tag{51}$$

θ	$N=1$	$N=2$	$N=\infty$
0	0.5	0.5	0.5
$0.5-(12q)^{-1}$	0.8333	0.9574	0.9661
0.5	1	$2\sqrt{3}/3$	$2(2-\sqrt{2})$
1	no bound	no bound	no bound

Table 1: Upper bounds for q in several finite difference methods providing the non-negativity.

The solution of (51) gives the bound

$$q \leq \frac{1-\sqrt{1-\theta}}{\theta(1-\theta)}. \tag{52}$$

We can summarize our results in the following

Theorem 3.3 *The finite difference method is non-negativity preserving for each $N \geq 1$ if and only if the condition* (48) *holds. It is non-negativity preserving for each $N \geq 2$ only under the condition* (50)*. There exists a number $N_0 \in \mathbb{N}$ such that the method is non-negativity preserving for each $N \geq N_0$, if and only if the condition* (52) *is satisfied.*

We demonstrate our results on some special choice of θ. Namely, we define the upper bounds for

- explicit Euler method ($\theta = 0$);
- fourth order method $\theta = 1/2 - 1/(12q)$, $q > 1/6$;
- Crank-Nicolson second order method ($\theta = 0.5$);
- implicit Euler method ($\theta = 1$).

The results are shown in Table 1.

We pass to the investigation of the linear finite element mesh operator. According to (27), the corresponding matrices are also symmetric, uniformly continuant, tridiagonal.

First we consider the special choices $\theta = 0$ and $\theta = 1$.

For $\theta = 0$ we get $X_1 = (1/6\Delta t)\text{tridiag}[1,4,1]$, i.e., we are not able to guarantee the monotonicity of X_1, which is required in the Remark 3.2. When $\theta = 1$, then $X_2 = (1/6\Delta t)\text{tridiag}[1,4,1]$, hence, the monotonicity of X_1 is the necessary and sufficient condition of the non-negativity preservation of the of the method. Hence, we get the condition $q \geq 1/6$.

Now we assume that $\theta \in (0,1)$.
When $q = 1/(6\theta)$ then $X_1 = (1/\Delta t)I$, hence the only condition of the non-negativity preservation is $X_2 \geq 0$. This can be guaranteed only by the condition $q \leq (3(1-\theta))^{-1}$. When $q = (3(1-\theta))^{-1}$ then $X_2 = (1/6\Delta t)\text{tridiag}[1,4,1]$, hence the only condition is the monotonicity of X_1. As we can see, for this case this matrix is M-matrix, therefore, there is no additional condition for the non-negativity preservation.

In what follows we may assume that $\theta \in (1/3, 1)$ and

$$\frac{1}{6\theta} < q < \frac{1}{3(1-\theta)}. \tag{53}$$

Then we can use the form (29) with the choice

$$\begin{aligned} z &= \frac{1}{6\Delta t} - \frac{\theta}{h^2}, \quad s = \frac{1}{6\Delta t} + \frac{1-\theta}{h^2}, \\ w &= \frac{\frac{1}{3} + \theta q}{\theta q - \frac{1}{6}}, \quad p = \frac{\frac{2}{3} - 2(1-\theta)q}{(1-\theta)q + \frac{1}{6}}. \end{aligned} \tag{54}$$

For this choice the assumption (30) is valid and $2w + p = [(\theta q - 1/6)((1-\theta)q + 1/6)]^{-1} > 0$. Therefore (34) is always satisfied. Let us notice that under the condition (53) the condition $z > 0$ is also satisfied.

Using (43), we get that the condition of the non-negativity preservation for all $N = 1, 2, \ldots$ is (43), which results in the upper bound

$$q \le \frac{1}{3(1-\theta)}. \tag{55}$$

The non-negativity preservation for all $N = 2, 3, \ldots$ should be guaranteed by the weaker condition (44), which, in our case yields the upper bound

$$q \le \frac{3(-1+2\theta) + \sqrt{9 - 16\theta(1-\theta)}}{12\theta(1-\theta)}, \tag{56}$$

which is larger than the bound in (55).

Our aim is to get the largest value for q under which the non-negativity preservation for sufficiently large values N is still valid. Therefore we put the values w and p from (54) into the necessary condition (42). Hence, we obtain that for any fixed $\theta \in (0, 1))$ the suitable q are the solution of the inequality

$$\begin{aligned} &\theta(1-\theta)q^2 - 1/6(\theta + 4)q + A \le 0; \\ &A = \sqrt{q\theta + 1/12}[1/6 + (1-\theta)q]. \end{aligned} \tag{57}$$

We can summarize our results in the following

Theorem 3.4 *The linear finite element method is non-negativity preserving for any $\theta \in [1/3, 1]$,*

- *for each $N \ge 1$ if and only if the condition*

$$\frac{1}{6\theta} \le q \le \frac{1}{3(1-\theta)}; \tag{58}$$

θ	$N=1$	$N=2$	$N=\infty$
0	not allowed	not allowed	not allowed
0.5	$1/3 \leq q \leq 2/3$	$1/3 \leq q \leq \sqrt{5}/3$	$1/3 \leq q \leq 0.748$
1	$1/6 \leq q$	$1/6 \leq q$	$1/6 \leq q$

Table 2: Upper and lower bounds for q in several linear finite element methods providing the non-negativity.

- *for each $N \geq 2$ if and only if the condition*

$$\frac{1}{6\theta} \leq q \leq \frac{3(-1+2\theta)+\sqrt{9-16\theta(1-\theta)}}{12\theta(1-\theta)} \tag{59}$$

holds. There exists a number $N_0 \in \mathbb{N}$ such that the method is non-negativity preserving for each $N \geq N_0$ if and only if the condition (57) *is satisfied.*

We demonstrate our results again on some special choice of θ. The results are shown in Table 2.

4 The non-negativity preservation in 2D FEM case

We consider the non-negativity preservation property of the discretization of the two-dimensional heat conduction problem with pure homogenous Dirichlet boundary conditions. For the simplicity, the constant coefficient is assumed to be equal one, however, a more general analysis was done in a previous work [20].

The general form of the two-dimensional heat conduction equation on $\Omega \times (0,T)$, where $\Omega := (0,L_x) \times (0,L_y)$, is

$$\begin{aligned} \frac{\partial u}{\partial t} = \triangle u, \quad (x,y) \in \Omega,\ t \in (0,T), \\ u|_{\Gamma_\Omega} = 0,\ t \in [0,T) \\ u(x,y,0) = u_0(x,y),\ (x,y) \in \Omega, \end{aligned} \tag{60}$$

where u is the temperature of the analyzed domain, t and x,y denote the time and space variables, respectively.

In the course of the analysis of the problem the space was divided into $2 \cdot (n_x+1) \cdot (n_y+1)$ triangle elements. Then we get the problem (9) with the matrices from $\mathbb{R}^{n_x^2 \times n_y^2}$ in the form for the linear finite element method the corresponding matrices are

$$X_1 = \frac{1}{\Delta t} M - \theta Q, \tag{61}$$

$$X_2 = \frac{1}{\Delta t} M + (1-\theta) Q, \tag{62}$$

where, for bilinear shape functions [15]

$$Q = tridiag(Q_I, Q_A, Q_I) \tag{63}$$

and

$$M = h_x h_y tridiag(M_D^T, M_A, M_D) \tag{64}$$

respectively, where

$$Q_A = \frac{h_y}{h_x} tridiag\left(1, -2\left[1 + \frac{h_x^2}{h_y^2}\right], 1\right), \tag{65}$$

$$Q_I = \frac{h_x}{h_y} tridiag(0, -1, 0), \tag{66}$$

$$M_A = \frac{1}{12} tridiag(1, 6, 1), \ M_D = \frac{1}{12} tridiag(0, 1, 1), \tag{67}$$

moreover, h_x and h_y are the lengths of the spatial approximations. For one dimensional linear spline functions see [11].
It is clear that for non-negativity preservation property we require the condition

$$X = X_1^{-1} X_2 \geq 0. \tag{68}$$

The sufficient conditions of the non-negativity of X are the following:

$$X_1^{-1} \geq 0 \ \text{ and } \ X_2 \geq 0. \tag{69}$$

Remark 4.1 *The decomposition of $X_1 - X_2 = \Delta t Q$ with the property* (69)*, is called a regular matrix splitting [16].*

For X_2 it is easy to give a condition that guarantees its non-negativity by analyzing the elements of the matrix. By a direct computation we get the condition

$$\frac{h_y h_x}{2\Delta t} - 2\left(\frac{hx}{hy} + \frac{hy}{hx}\right)(1-\theta) \geq 0, \tag{70}$$

which yields the upper bound

$$\frac{h_x h_y}{4\left(\frac{h_x}{h_y} + \frac{h_y}{h_x}\right)(1-\theta)} \geq \Delta t. \tag{71}$$

It is not possible to obtain a sufficient condition for the non-negativity of the matrix X_1^{-1} by the so-called M-matrix method [21]. This also follows from the fact that X_1 contains positive elements in its off-diagonal. Therefore, a sufficient condition for the inverse-positivity of matrix X_1 will be obtained by some other criteria.

Lemma 4.1 *[14] Let A be an n-by-n matrix, denote A_d and A^- the diagonal and the negative off-diagonal part of the matrix A, respectively.*
Let $A^- = A^z + A^s = (a^z_{ij}) + (a^s_{ij})$. If

$$a_{ij} \leq \sum_{k=1}^{n} a^z_{ik} a^{-1}_{kk} a^s_{kj}, \ \text{for all } a_{ij}, i \neq j, \tag{72}$$

then A is a product of two M-matrices, i.e., A is monotone.

We will analyze the monotonicity of X_1 with the help of this lemma. We can do it because it is a square matrix and it can be decomposed into the diagonal part, the positive off-diagonal part, the upper triangular and lower triangular negative parts. All the conditions of the lemma are satisfied if

$$\frac{1}{12} \leq \frac{\left(\frac{1}{12} - \frac{\Delta t\theta}{h_x^2}\right)\left(\frac{1}{12} - \frac{\Delta t\theta}{h_y^2}\right)}{\frac{1}{12} + 2\left(\frac{h_x}{h_y} + \frac{h_y}{h_x}\right)\frac{\Delta t\theta}{h_x h_y}}, \tag{73}$$

which implies the lower bound

$$\frac{h_y^2}{12\theta}\left(\frac{3}{2}\left(\frac{h_x^2}{h_y^2}+1\right) + \sqrt{\frac{9}{4}\left(\frac{h_x^4}{h_y^4}+1\right) + \frac{19}{2}\left(\frac{h_x^2}{h_y^2}\right)}\right) \leq \Delta t. \tag{74}$$

Hence, the next statement is proven.

Theorem 4.3 *Let us assume that the conditions* (71) *and* (74) *hold. Then for the problem* (60) *on a rectangular domain with an arbitrary non-negative initial condition the linear finite element method results in a non-negative solution on any time level.*

Remark 4.2 *If $\theta = 1$, there is no upper bound for the time-step size, nor any condition for the ratio of the lengths of the spatial approximations.*

Remark 4.3 *If the conditions of the theorem hold, then the following complementary conditions are also satisfied:*

- *For the ratio of the lengths of the spatial approximations*

$$\sqrt{\omega - \sqrt{\omega^2 - 1}} \leq \frac{h_x}{h_y} \leq \sqrt{\omega + \sqrt{\omega^2 - 1}}, \tag{75}$$

where

$$\omega = \frac{\frac{10}{9}T^2 + 2T - 1}{-\frac{10}{9}T^2 - 2T} \tag{76}$$

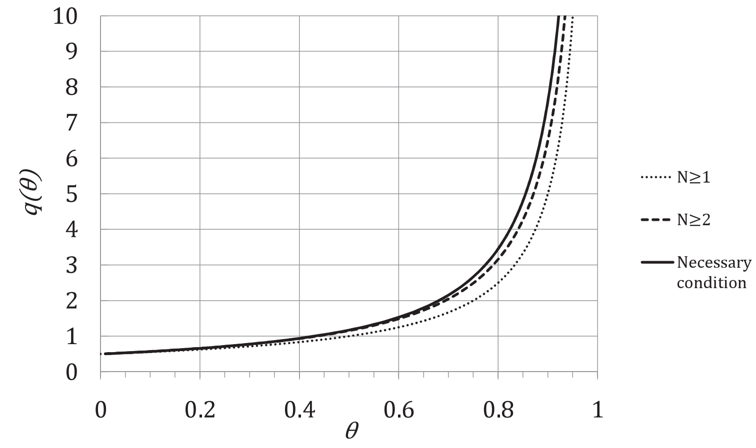


Figure 1: Condition for the choice of the discretization step-sizes for FDM in 1D.

and

$$T = \frac{1-\theta}{\theta}. \tag{77}$$

This yields a geometrical restriction for the shape of the partition of the space domain in the linear FEM.

- *For θ, which is the parameter of the applied numerical method, we have the bound*

$$\theta \geq \frac{1}{\sqrt{\frac{63}{50}} + \frac{1}{10}}. \tag{78}$$

Since the right-hand side is greater than 0.818, this implies that for the Crank-Nicolson method ($\theta = 0.5$) we cannot guarantee the non-negativity by this principle [3].

5 Numerical experiments

In the following figures we illustrate the possible choice of the discretization step-sizes $q = \Delta t/h^2$ in the one-dimensional case for finite difference and for linear finite element methods, for different values of θ.

In the course of the numerical experiments in 2D ($n_x = 20, n_y = 25, h_x = 0.1, h_y = 0.04$) for the homogenous initial condition $u_0(x, y) = 300K$ was considered. For the numerical experiments, the tridiagonal matrix algorithm (TDMA) was used for the inversion of the sparse tridiagonal matrices [18]. The following figures are in three dimensions, in Fig. 3 the first two dimensions are the spatial ones (x, y) and the third is the temperature at the nodes. First, we apply the Crank-Nicolson method and a relatively long time step ($\theta = 0.5, \Delta t = 10^{-2}$, Timesteps $= 1$), which results in a negative X_2.

For the sake of completeness, in Fig. 4 we applied the time-step size from the interval (71) and (74) ($\theta = 1, \Delta t = 0.02$, Timesteps $= 10$) , and it can be seen we have got a more realistic solution.

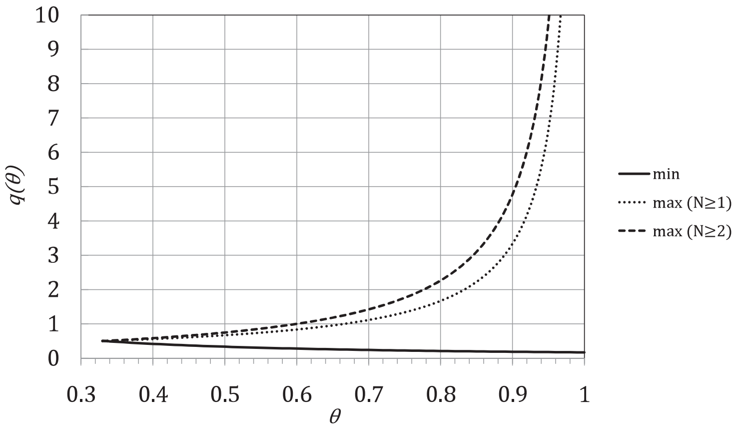


Figure 2: Condition for the choice of the discretization step-sizes for linear FEM in 1D.

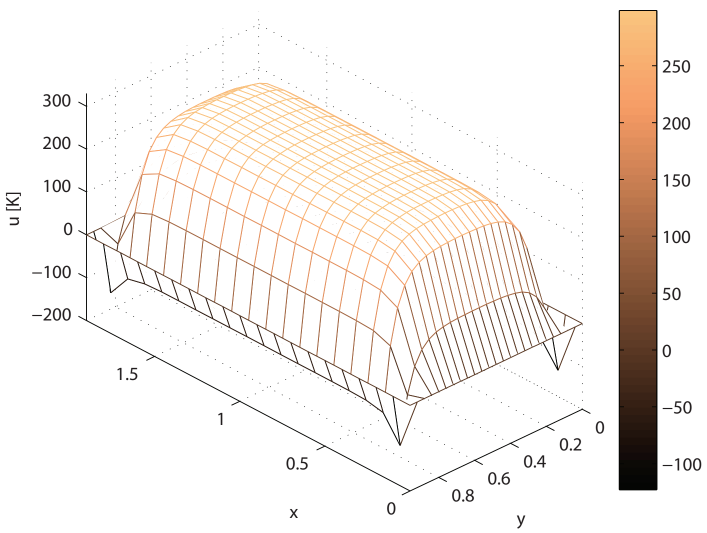


Figure 3: The solution obtained by the Crank-Nicolson method and relatively long time step.

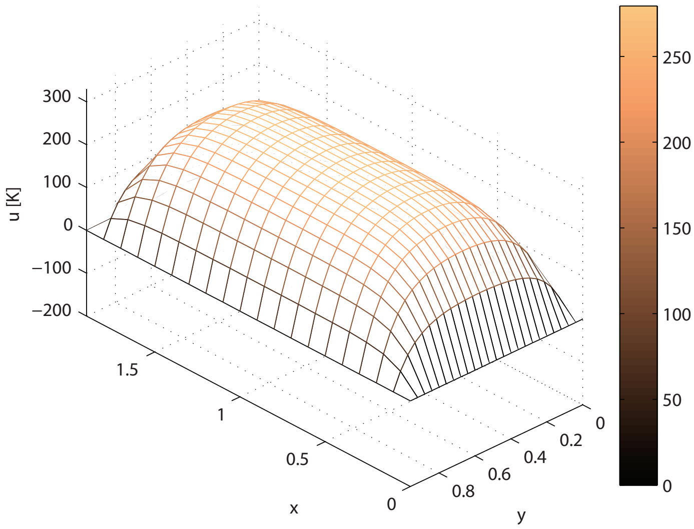


Figure 4: The solution obtained by applying a time step from the interval (71) and (74).

6 Conclusions

In this paper, we have considered the non-negativity preservation property for the linear parabolic PDE's. We established the direct connection of the non-negativity preservation for the semidiscrete solutions between the one and two-dimensional cases. Namely, we proved that the two-dimensional problem has this property if the corresponding one-dimensional problem is non-negativity preserving. The conditions posed are satisfied for all arbitrary (linear FEM and FDM) methods. Moreover, we gave the explicit formula for the inversion of the block-tridiagonal matrices with scalar tridianal matrices placed along their diagonals.

Acknowledgement

The first author was supported by Hungarian National Research Fund OTKA No. K67819.
The second author was supported by project no. 124619 from the Academy of Finland.
The first and the third authors were supported by Jedlik project "ReCoMend" 2008–2011.

References

[1] BELLMAN, R.: *Introduction to Matrix Analysis*, Mcgraw-Hill, 1960.

[2] BERZINS, M.: Modified mass matrices and positivity preservation for hyperbolic and parabolic PDE's, *Commun. Numer. Meth. Engng.*, 17 (2001), 659–666.

[3] FARAGÓ, I.: Qualitative analysis of the Crank-Nicolson method for the heat conduction equation. In: Lect. Notes Comp. Sci., 5434, Springer Verlag, Berlin (2009), 44–55.

[4] FARAGÓ, I., HAVASI, Á.: *Operator Splittings and Their Applications*, Nova Science Publisher Inc., New York, 2009, 112 pp.

[5] FARAGÓ, I., HORVÁTH, R.: Discrete maximum principle and adequate discretizations of linear parabolic problems, *SIAM Sci. Comput.*, 28 (2006), 2313–2336.

[6] FARAGÓ, I., HORVÁTH, R.: Continuous and discrete parabolic operators and their qualitative properties, *IMA Numer. Anal.*, 29 (2009), 606–631.

[7] FARAGÓ, I., KOMÁROMI, N.: Non-negativity of the numerical solution of the parabolic problems. In: Numerical Methods (eds. D. Greenspan, P. Rózsa), North-Holland, 59 (1991), 173–179.

[8] FARAGÓ, I., HORVÁTH, R.: On the non-negativity conservation of finite element solutions of parabolic problems. In: Finite Element Methods: Three-Dimentional Problems (eds. P. Neittaanmäki, M. Křížek), GAKUTO Sci. Appl., Tokyo (2001), 78–86.

[9] FIEDLER, M.: *Special Matrices and Their Application in the Numerical Mathematics*, Martinus Nijhoff Publishers, 1986.

[10] FUJII, H.: Some remarks on finite element analysis of time-dependent field problems. In: Theory and Practice in Finite element Structural Analysis, Univ. Tokyo Press, Tokyo (1973), 91–106.

[11] HARITON, H.: *Some Qualitative Properties of The Numerical Solution to the Heat Conduction Equation*, PhD thesis, Budapest Eötvös Loránd University, 1995.

[12] KOROTOV, S., KRIZEK, M., NEITTAANMÄKI, P.: Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle, *Math. Comp.*, 70 (2001), 107–119.

[13] KRIZEK, M., QUN LIN: On diagonal dominance of stiffness matrices in 3D, *East-West J. Numer. Math.*, 3 (1995), 59–69.

[14] LORENZ, J.: Zur Inversmonotonie diskreter Probleme, *Numer. Math.*, 27 (1977), 227–238 (in German).

[15] MARCHUK, G.: *Numerical Methods*, Muszaki Konyvkiado, Budapest, 1976 (in Hungarian).

[16] VARGA, R.: *Matrix Iterative Analysis*, Prentice Hall, 1962.

[17] RÓZSA, P.: *Linear Algebra and its Applications*, Muszaki Konyvkiado, 1976 (in Hungarian).

[18] SAMARSKIY, A. A.: *Theory of Different Schemes*, Moscow, Nauka, 1977 (in Russian).

[19] SMOLLER, J.: *Shock Waves and Reaction-Diffusion Equations*, Springer Verlag, 1981.

[20] SZABÓ, T.: Some sufficient conditions for the preservation of the non-negativity of the numerical solution of the heat conduction equations, *J. Comput. Appl. Math.* (submitted in 2009).

[21] SZABÓ, T.: On the discretization time-step in the finite element theta-method of the discrete heat equation. In: Lect. Notes Comp. Sci. 5434, Springer Verlag, Berlin, 2009, 564–571.

[22] ZLATEV, Z.: *Computer Treatment of Large Air Pollution Models*, Kluwer Academic Publishers, 1995.

Current address

István Faragó
Department of Applied Analysis and Computational Mathematics
Eötvös Loránd University
H-1117, Budapest, Pázmány P. s. 1/c. Hungary
e-mail: faragois@cs.elte.hu

Sergey Korotov
Department of Applied Analysis and Computational Mathematics
Eötvös Loránd University
H-1117, Budapest, Pázmány P. s. 1/c. Hungary
e-mail: smkorotov@gmail.com

Department of Mathematics
Tampere University of Technology
P.O. Box 553, FIN–33101 Tampere, Finland
e-mail: sergey.korotov@tut.fi

Tamás Szabó
Department of Applied Analysis and Computational Mathematics
Eötvös Loránd University
H-1117, Budapest, Pázmány P. s. 1/c. Hungary
e-mail: szabot@cs.elte.hu

Journal of Applied Mathematics

volume III (2010), number II

CHAOS, ANTI-CHAOS AND RESOURCES: DEALING WITH COMPLEXITY

FILIPE José António (P), FERREIRA Manuel Alberto M. (P), COELHO M., (P), PEDRO Maria Isabel (P)

Abstract: Chaos is ordinarily disorder or confusion; scientifically it represents a disarray connection, but basically it involves much more than that. Change and time are closely linked and they are essential when considered together to understand the foundations of chaos theory. The theories of dynamic systems have been applied to numerous areas of knowledge. In the 80's, several exact sciences (physics, chemistry or biology, for example) and some social sciences (economics or management or even the sociology) still had their own objects of study and their own methods of analysis and each one of them was different from the others. The Science has been branched and specialized, so that each one uses to have its own world. Recently new forms of analysis, looking for an integrated study have emerged (Filipe, 1). It is the case of chaos theory, which is applied here to natural resources, in order to understand the complexity of natural phenomena following this perspective. Anti-chaos theory is also introduced to show how systems in nature often tend to self-organization.

Keywords: Chaos theory, anti-chaos theory, complexity, dynamical systems, complex adaptive systems

1 Introduction

Chaos theory and complexity theory itself reflect the phenomena that in many activities (such as fisheries) are translated into dynamic forms of analysis and reflect a very complex and widespread reality, specific of complex systems. That reality falls within a range of situations integrated in a broader context, which is expressed in the theory itself but also in terms of their own realities (fisheries, for example), dynamic, complex and often chaotic features in its essence.

The chaos theory stresses that the world does not necessarily work as a linear relationship with perfectly defined or with direct relations in terms of expected proportions between causes and effects. The chaos occurs when a system is very sensitive to the initial conditions. These initial conditions are the measured values for a given initial time. The presence of chaotic systems in nature seems to place a limit on our ability to apply physical deterministic laws to predict movements with any degree of certainty. Indeed, one of the most interesting subjects in the study of chaotic systems is the question of whether the presence of chaos may or may not produce ordered structures and patterns on a wider scale. In past times, the dynamic systems showed up completely unpredictable and the only ones who could aspire to be understood were those that were represented by linear relationships, which are not the rule.

Complexity science has the potential to strengthen many areas of several sciences. The use of complexity theory concepts is changing the focus of research in all scientific disciplines and is leading to important practical outcomes.

This century theoretical physics is coming out from the chaos revolution. The study of complexity is the way and the computer the main tool. Thermodynamics, as a vital part of theoretical physics, will be involved in the transformation. In this complexity analysis, Anti-chaos theory conducts to the understanding of how systems perform a self organization and a structured system.

2 Chaos, Complexity and Dynamical Systems

The chaos theory allows realizing the endless alternative ways leading to a new form or new ways that will be disclosed and that eventually emerge from the chaos as a new structure. The reality is a process in which structure and chaos rotate between form and deformation in an eternal cycle of death and renewal. Conditions of instability seem to be the rule and, in fact, a small inaccuracy in the departure conditions tends to grow to a huge scale. Basically, two insignificant changes in the initial conditions for the same system tend to end in two situations completely different. This situation is known as the "butterfly wing effect". A small movement of the wings of a butterfly can have huge consequences.

It is the microscopic turbulence having effects in a macroscopic scale - an effect called by Grabinski (3) as “hydrodynamics”. Mathematically, the "butterfly wing effect" corresponds to the effect of chaos, which can be expressed as follows.

Given the initial conditions

$$x_1, x_2, x_3, \ldots, x_N$$

it is possible to calculate the final condition given by

$$\textit{final result} = f\,(x_1, x_2, x_3, \ldots x_N)$$

If the initial conditions x_i have a margin of error (variation), the final result will be influenced by the existence of this margin. If these margins in x_i are as small as the margin of error in the final result, we have a non-chaotic situation. Otherwise if the margins of error in x_i are small but the final result has a big variation, there is a chaotic situation. Therefore, small variations in initial conditions can lead to a major effect in the final outcome. Sometimes small changes in x_i have exponential effects on the final result due to the passage of time.

This effect can be demonstrated mathematically[1] using the Lyapunov Exponent[2] (see Grabinski, 4). Given the initial value x_0 and being ε its arbitrarily small variation, we are conducted to an initial value between x_0 and $x_0+\varepsilon$. The general form of Lyapunov indicator is presented by

[1] Several statistics may indicate chaos and can express how chaotic a system is. One of the most important statistics to measure the magnitude of chaos is at present Lyapunov exponents. Other statistics could be presented such as the Kolmogorov-Sinai entropy or the mutual information or redundancy.

$$x_{n+1} = f(x_n)$$

that after N iterations leads to a value for x_N between

$$f^N(x_0) \qquad \text{and} \qquad f^N(x_{0+\varepsilon})$$

being the difference between these two values

$$f^N(x_{0+\varepsilon}) - f^N(x_0) \ \Xi \ \varepsilon . e^{N\lambda(x_0)}$$

where λ is a parameter depending on x_0.

Dividing both sides by the variation ε and assuming the limit $\varepsilon \to 0$, we have a differential quotient. Making its logarithm and assuming the limit $N \to \infty$, we get the final definition of the Lyapunov Exponent

$$\lambda(x_0) = \lim_{N\to\infty} \frac{1}{N} \log \left| \frac{d\, f^N(x_0)}{dx_0} \right|$$

and there is chaos when $\lambda > 0$.

Through this function, the chaos exists when arbitrary small variations in initial conditions grow exponentially with a positive exponent.

Grabinski also points that the nonlinearity is the main characteristic of a chaotic situation. Mathematically, the nonlinear functions to be considered chaotic should be based on variables with some resistance. The author also argues that it is not enough to describe the chaotic situations, such as turbulence, but it is necessary to find ways to better cope with the nonlinearity. A smooth flow of a river (non-chaotic) that can be described in quantities like the flow velocity can reach a chaotic behavior with variations of many situations. The best example is a waterfall where the speed of the flow reaches a certain point. In a smoothly flowing river it is easy to calculate or predict the flow velocity of the river at any point. However, to calculate it in a river with a waterfall, it is necessary to introduce chaos. In an attempt to make this calculation, man has focused on the construction of super computers that have shown to be useless due the infinity of factors that may cause turbulence in the flow of the river. Thus, the analysis of frequency on the change of flow's velocity is much more promising than the analysis of velocities themselves.

Moreover, Grabinski shows the situation in which there is chaos on a microscopic scale but not on a macroscopic scale - the hydrodynamics. An example is a glass of water resting on a table, a non-chaotic event. A slight disturbance on the table causes a small flow on a macroscopic level in the water. However, a microscopic observation reveals a great agitation of millions of molecules, a chaotic event. This is a situation where there is chaos on the microscopic scale but a smooth flow on the macroscopic scale.

Mathematically Grabinski presents hydrodynamics equations which combine the chaos theory with business situations. For that, he presents the value function of a company (*v*) depending on two variables, the revenue (*r*) and the number of employees (*n*). Its general form is

[2] A Lyapunov Exponent is a number that reflects the rate of divergence or convergence, averaged over the entire attractor, of two neighbouring phase space trajectories. Trajectories divergence or convergence have to follow an exponential law, for the exponent to be definable.

$$v(r,n) = v_0 + a_{10}.r + a_{01}.n + a_{11}.r.n + a_{20}.r^2 + a_{02}.n^2 + ...$$

being a_{ij} general parameters. For $n = 0$ (no employees) or $r = 0$ (no revenue) the company doesn't exist because the function value is equal to 0. So some terms of the function must be removed ($v_0 = a_{10} = a_{01} = a_{20} = a_{02} = ... = 0$). The general form comes as

$$v(r,n) = a_{11}.r.n + a_{21}.r^2.n + a_{12}.r.n^2 + a_{22}.r^2.n^2 + ...$$

Now, because of symmetry, r and n may be negative. A negative employee means that the employee is paying to work and negative revenue means that the company is paying the customer to consume. So the previous formula can lead to negative results if r and n change signs simultaneously. Only these terms are allowed for which the sums of the powers of r and v are even numbers. Thus the general expression of the equation is

$$v(r,n) = a_{11}.r.n + a_{22}.r^2.n^2 + ...$$

3 Complexity and Ecological Systems

It is usual to consider that the really complex systems could be the biological ones, particularly the systems in which people are present: human body, human groups, society itself or people culture.

Many scientists see today, with particular interest, the chaos theory as a way to explain environment. Therefore, the chaos theory stresses the fundamental laws of nature and natural processes and requires a course for a constant evolution and recreation of nature.

In biological area and in order to frame some methodological developments, it must be mentioned, first of all, that some characteristics associated with some species support strategic survival features that are exploited by the present theory. Its aim is to find the reasons and the way in which these strategies are developed and the resulting consequences. The species use their biological characteristics resulting from evolutionary ancient processes to establish defense strategies. It is particularly interesting to see the behavior of schooling species and the way they delineate a consistent strategy for the group and specie as a whole, which is self-organizing, an anti-chaos feature, and which can be understood according to a focus based on the systems properties.

The ecology where many things are random and uncertain, in which everything interacts with everything at the same time is, itself, a fertile area for a cross search to the world explanations (Filipe *et al*, 2).

Lansing (5) states that the initial phase of the research of nonlinear systems was based on the deterministic chaos, and it was later redirected to new outbreaks of research focusing on the systems properties, which are self-organizing, the so called anti-chaos. It also says that the study of complex adaptive systems, discussed in the context of non-linear dynamic systems, has become a major focus of interest resulting from the interdisciplinary research in the social sciences and the natural sciences.

The theory of systems in general represents the natural world as a series of reservoirs and streams governed by various feedback processes. However, the mathematical representations were ignoring the role of these adjustment processes.

The theory of complex adaptive systems, part of the theory of systems, has in specific account the diversity and heterogeneity of systems rather than representing them only by reservoirs. It explicitly considers the role of adaptation on the control of the dynamics and of the responses of these heterogeneous reservoirs. This theory allows ecologists to analyze the reasons inherent to the process at the lower levels of the organization that lead to patterns at higher levels of organization and ecosystems. The adaptive systems represent one of the means to understand how the organization is produced to a large scale and how it is controlled by processes that operate at lower levels of organization. According to Lansing (5), came to be a general idea involving physical and mathematical complexity that is hidden behind very simple systems.

Considering a system composed by many interactive parts, if it is sufficiently complex, it may not be practical or even not be possible to know the details of each interaction place. Moreover, the interactions can generate local non-linear effects that, often, it becomes impossible to find a solution even for simple systems. However, diverting us from causal forces that move the individual elements, if we focus on the system behavior as a whole we can highlight certain global behavior standards. However, these behavior standards may hide an associated cost: it can not be expected to understand the causes at the level of individual behavior.

Indeed, the systems do not match the simple decomposition of the whole into parts. Therefore do not correspond to the mere sum of the parts, as living systems are not the juxtaposition of molecules and atoms. Since the molecule to the biosphere, the whole is organized and each level of integration leads to properties that can not be analyzed only from mechanisms that have explanatory value in the lower levels of integration. This corresponds to the appearance of new features to the level of the set that does not exist at the level of the constituent elements. Lansing (5) believes that the adoption in the social sciences of the idea that complex global patterns can emerge with new properties from local interactions had a huge impact here.

The ecological systems are comparable to systems self-organized as they are open systems which arise far from thermodynamic equilibrium. On self-organized and self-regulated systems, the reciprocal interactions within the system between the structures and the processes contribute to the regulation of its dynamics and the maintenance of its organization; partly due to the phenomena of feedback (see Lévêque, 6). These systems seem to develop themselves in accordance with the properties referred to the anti-chaotic systems. Indeed, we have auto-regulated systems that channel different initial conditions for the same stage, instead of what is happening with chaotic systems, which are very sensitive to initial conditions (see Kauffman, 7). These systems would be relatively robust for a particular type of disturbance, to which the components of the system fit, creating a meta-stability that depends not only on the internal interactions within the system but also on external forces that can regulate and strengthen the internal factors of cohesion (see Lévêque, 6).

4 Some Notes about Fisheries

Some people see nature as not casual and unpredictable. The natural processes are complex and dynamic, and the causal relations and sequential patterns may extend so much in time that may seem to be non-periodical. The data appear as selected random works, disorderly, not causal in their connections and chaotic. In nature, for example, for live resources in the sea, the vision provided by nature leads to consider the fish stocks, time, the market and the various processes of fisheries management as likely to be continuously in imbalance rather than behave in a linear fashion and in a constant search for internal balance. It is this perspective that opens the way for the adoption of the chaos theory in fisheries. However, the models of chaos do not deny, for themselves, some of the linearity resulting from the application of usual bionomic models. What is considered is that there are no conditions to implement all significant variables in a predictive model. Moreover, in finding that a slight change in initial conditions caused by a component of the system may cause major changes and deep consequences in the system itself.

As it has been explained previously, the butterfly effect represents the more sensitive dependence on initial conditions in chaos theory. Small variations on initial conditions in a dynamical system may produce large variations in the system long term behavior.

So, the application of chaos theory to fisheries is considered essential, by many researchers. The chaos theory depends on a multitude of factors, all major (and in the prospect of this theory all very important at the outset) on the basis of the wide range of unpredictable effects that they can cause.

Given the emergence of new forms of predation, species got weaker and weaker because they are not prepared with mechanisms for effective protection for such situations. In fisheries there is a predator, man, with new fishing technologies who can completely destabilize the ecosystem. By using certain fisheries technologies, such as networks of siege, allowing the capture of all population individuals which are in a particular area of fishing, the fishers cause the breakdown of certain species, particularly the pelagic, the ones normally designated by schooling species.

To that extent, with small changes in ecosystems, this may cause the complete deterioration of stocks and the final collapse of ecosystems, which in extreme cases can lead to extinction. These species are concentrated in high density areas in small space. These are species that tend to live in large schools.

Usually, large schools allow the protection against large predators. The mathematical theory, which examines the relationship between schools and predators, due to Brock and Riffenburgh (see Clark, 8), indicates that the effectiveness of predators is a reverse function of the size of the school. Since the amount of fish that a predator can consume has a maximum average value, overcoming this limit, the growth of school means a reduction in the rate of consumption by the predator. Other aspects defensive for the school such as intimidation or confusing predators are also an evidence of greater effectiveness of schools.

However this type of behavior has allowed the development of very effective fishing techniques. With modern equipment for detecting schools (sonar, satellites, etc.) and with modern artificial fibers' networks (strong, easy to handle and quick placement), fishing can keep up advantageous for small stocks (Bjorndal, 9; Mangel and Clark, 10). As soon as schools become scarce, stocks become less protected. Moreover, the existence of these modern techniques prevents an effect of stock in the costs of businesses, as opposed to the so-called search fisheries, for which a fishery involves an action of demand and slow detection. Therefore, the existence of larger populations is essential for fishermen because it reduces the cost of their detection (Neher, 11). However, the easy detection by new technologies means that the costs are not anymore sensitive to the stock size (Bjorndal and Conrad, 12).

This can be extremely dangerous due to poor biotic potential of the species subject to this kind of pressure. The reproductive capacity requires a minimum value below which the extinction is inevitable (Filipe *et al*, 13). Since the efficiency of the school is proportional to its size, the losses due to the effects of predation are relatively high for low levels of stocks. This implies non-feedback in the relation stock-recruitment, which causes a break in the curves of income-effort, so that an infinitesimal increase on fishing effort leads to an unstable condition that can lead to its extinction (Filipe *et al*, 14).

Considering the fished value function of a company (v) depending on two variables, the fishing effort (r) and the fish stock (n), a simple model for fisheries, analogous to the presented in section 2, can be built

$$v(r,n) = v_0 + a_{10}.r + a_{01}.n + a_{11}.r.n + a_{20}.r^2 + a_{02}.n^2 + ...$$

being a_{ij} general parameters. Now it makes no sense to consider negative values for the variables. For $n = 0$ (no fish stock) or $r = 0$ (no fishing effort) the company fished value doesn't exist because the function value is equal to 0.

Consequently, $v_0 = a_{10} = a_{01} = a_{20} = a_{02} = ... = 0$ and now

$$v(r,n) = a_{11}.r.n + a_{21}.r^2.n + a_{12}.r.n^2 + a_{22}.r^2.n^2 + ...$$

5 Conclusion

Chaos theory got its own space among sciences and has become itself an outstanding science. However there is much left to be discovered. Anyway, many scientists consider that chaos theory is one of the most important developed sciences on the twentieth century.
Aspects of chaos are shown up everywhere around the world and chaos theory has changed the direction of science, studying chaotic systems and the way they work.
It is not possible to say yet if chaos theory may give solutions to problems that are posed by complex systems. Nevertheless, understanding the way chaos discusses the characteristics of complexity and analyzes open and closed systems and structures is an important matter of present discussion.
Finally, some words to say that anti-chaos theory is directed to research focusing on the systems properties, which are self-organizing. In nature, systems seem to look for a durable organization and stability. This may be seen, for example, in ecological systems (see schooling fish species, for example).
This work shows in fact how natural resources are complex and how complexity theory deals with ecological phenomena.

References

[1] FILIPE, J. A. (2006), *O Drama dos Recursos Comuns. Um caso de aplicação da Teoria dos Jogos aos comuns da pesca. PhD thesis*. Lisboa: ISCTE.

[2] FILIPE, J. A., COELHO, M. AND FERREIRA, M. A. M. (2005), *Sistemas Dinâmicos, Caos e os Comuns da Pesca*. Revista de Economia Global e Gestão. N.º 2/2005. Lisboa: ISCTE.

[3] GRABINSKI, M. (2004), *Is There Chaos in Management or Just Chaotic Management?*. Complex Systems, intelligence and Modern Technology Applications. Paris.

[4] GRABINSKI, M. (2008), *Chaos – Limitation or Even End of Supply Chain Management*. High Spees Flow of Material, Information and Capital. Istanbul.

[5] LANSING, J. S. (2003), *Complex adaptive systems*, Annual Review Anthropology. http://www.ic.arizona.edu/lansing/GompAdSys.pdf.
[6] LÉVÊQUE, G. (2002), *Ecologia: do ecossistema à biosfera*. Lisboa: Instituto Piaget.
[7] KAUFFMAN, S. (1993), *The origins of order: self-organization and selection in evolution*. New York: Oxford Univ. Press.
[8] CLARK, C. W. (1974), *Possible effects of schooling on the dynamics of exploited fish populations*, Journal du Conseil Internatinal pour L'Exploration de la Mer, 36(1).
[9] BJORNDAL, T. (1987), *Production economics and optimal stock size in a North Atlantic fishery*, Scandinavian Journal of Economics, 89(2).
[10] MANGEL, M. and CLARK, G. (1983), *Uncertainty, search and information in fisheries*, Journal du Conseil International pour L'Exploration de la Mer, 41.
[11] NEHER, P. (1990), *Natural Resource Economics: Conservation and Exploitation*. Cambridge: Cambridge University Press.
[12] BJORNDAL, T. and CONRAD, J. (1987), *The dynamics of an open access fishery*, Canadian Journal of Economics, 20(1).
[13] FILIPE, J. A., FERREIRA, M. A. M. and COELHO, M. (2008), *The Relevance of Chaos Theory to Explain Problems of Overexploitation in Fisheries***.** Working Paper, WP/24/2008/DE/SOCIUS. Lisboa: ISEG.
[14] FILIPE, J. A., FERREIRA, M. A. M., COELHO, M. and PEDRO, M. I. C. (2009), *Complexity, Theory of Chaos and Fishing*". In Porath, D. and Bayer, A., „International Suplement" of the „update" of the scentific bi-annual magazine of FH Mainz, University of Applied Sciences. FH Mainz, University of Applied Sciences. Mainz.

Current address

José António Filipe, Professor Auxiliar
ISCTE - IUL – Lisbon University Institute
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa, Portugal
Tel.+351 217 903 000, e-mail: jose.filipe@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático
ISCTE – IUL – Lisbon University Institute
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa, Portugal,
Tel. +351 217 903 000, e-mail: manuel.ferreira@iscte.pt

Manuel Coelho, Professor Auxiliar
SOCIUS/ISEG – Portugal, Rua do Quelhas, 6, 1200-781, LISBOA, Portugal
Phone: +(351) 213925800, Fax: +(351) 213922808, e-mail: coelho@iseg.utl.pt

Maria Isabel Pedro, Professor Auxiliar
CEGIST/IST - Portugal
Av. Rovisco Pais, LISBOA, Portugal
Phone: +(351) 214233267, Fax: +(351) 218417979, e-mail: ipedro@ist.utl.pt

Journal of Applied Mathematics

volume III (2010), number II

ADOMIAN DECOMPOSITION METHOD FOR CERTAIN SINGULAR INITIAL VALUE PROBLEMS

FILIPPOVA Olga (CZ), ŠMARDA Zdeněk, (CZ)

Abstract. In this paper an efficient modification of Adomian decomposition method is introduced for singular initial value problems. Solutions are constructed in the form of a convergent series. The approach is illustrated with few examples.

Key words and phrases. Adomian decomposition method, singular initial problem .

Mathematics Subject Classification. 65L05.

1 Introduction

In the recent years the studies of singular initial value problems in the second-order ordinary differential equations have attracted the attention of many mathematicians and physicists. A large amount of literature developed concerning Adomian decomposition method [1,2,3] and the related modifications [9,14,15] to investigate various scientific models. One of singular equations is the Lane-Emden type equations formulated as

$$y'' + \frac{2}{x}y' + f(y) = 0, \quad y(0) = A, \ y'(0) = B, \quad 0 < x \leq 1, \tag{1}$$

where A, B are constants, $f(y)$ is a continuous function. Equation (1) with specializing $f(y)$ is a model several phenomena in the theory of stellar structure, isothermal gas spheres and theory of therminionic currents [5,11]. Most algorithms currently in use for handling solving of the Lane-Emden equations are based on either series solutions or perturbation techniques. Wazwaz [15] has given a general study to construct exact a series solutions to (1) by employing the Adomian decomposition method.

On the other hand, studies have been carried out on another class of singular initial value problems of the form

$$y'' + \frac{2}{x}y' + f(x,y) = g(x), \quad y(0) = A,\ y'(0) = B, \quad 0 < x \le 1, \tag{2}$$

where $f(x,y)$ is a continuous function, $g(x) \in C[0,1]$. It is important to note that (2) with boundary conditions has attracted many mathematicians and has been studied from various points of view. Russel and Shampine [12] have investigated (2) for the linear function $f(x,y) = ky+h(x)$. Three-point difference methods of second order have been used by Chawla and Katti [6], Iyengar and Jain [10], El-Sayed [4].

2 Adomian decomposition method

Now we recall basic principles of the Adomian decomposition method [1] for solving differential equations. Consider the general equation $Tu = g$, where T represents a general nonlinear differential operator involving both linear and nonlinear terms. The linear term is decomposed into $L + R$ where L is easily invertible and R is the reminder of the linear operator. For convenience, L may be taken as the highest order derivation. Thus the equation may be written as

$$Lu + Ru + Nu = g, \tag{3}$$

where Nu represents the nonlinear terms. From (3) we have

$$Lu = g - Ru - Nu. \tag{4}$$

Since L is invertible the equivalent expression is

$$u = L^{-1}g - L^{-1}Ru - L^{-1}Nu. \tag{5}$$

A solution u can be expressed as following series

$$u = \sum_{n=0}^{\infty} u_n, \tag{6}$$

with reasonable u_0 which may be identified with respect to the definition of L^{-1} , g and $u_n,\ n > 0$ is to be determined. The nonlinear term Nu will be decompsed by the infinite series of Adomian polynomials

$$Nu = \sum_{n=0}^{\infty} A_n, \tag{7}$$

where A_n's are obtained by writing

$$v(\lambda) = \sum_{n=0}^{\infty} \lambda^n u_n \tag{8}$$

$$N(v(\lambda)) = \sum_{n=0}^{\infty} \lambda^n A_n. \tag{9}$$

Here λ is a parameter introcuded for convenience. From (8) and (9) we have

$$A_n = \frac{1}{n!}\frac{d^n}{d\lambda^n}\left[N\sum_{k=0}^{\infty}\lambda^k u_k\right]_{\lambda=0}, \quad n \geq 0. \tag{10}$$

The A_n's are given as

$$\begin{aligned} A_0 &= F(u_0) \\ A_1 &= u_1F'(u_0) \\ A_2 &= u_2F'(u_0) + \frac{u_1^2}{2!}F''(u_0) \\ A_3 &= u_3F'(u_0) + u_1u_2F''(u_o) + \frac{u_1^3}{3!}F'''(u_0) \\ &\vdots \end{aligned} \tag{11}$$

Now substituting (6) and (7) into (5) we get

$$\sum_{n=0}^{\infty} u_n = u_0 + L^{-1}R\left(\sum_{n=0}^{\infty} u_n\right) - L^{-1}\sum_{n=0}^{\infty} A_n. \tag{12}$$

Consequently, with a suitable u_0 we can write

$u_1 = -L^{-1}Ru_0 - L^{-1}A_0$

$\vdots$

$u_{n+1} = -L^{-1}Ru_n - L^{-1}A_n$

$\vdots$

3 Modification ADM for singular differential equations of the second order

Consider equation(2) then according to the above mentioned method we can rewrite equationn (2) as

$$Ly = -f(x,y) + g(x), \tag{13}$$

where the differential operator L is defined by

$$L = x^{-2}\frac{d}{dx}\left(x^2\frac{d}{dx}\right). \tag{14}$$

The inverse operator L^{-1} is therefore considered a two-fold integral operator defined by

$$L^{-1}(.) = \int_0^x x^{-2}\int_0^x x^2(.)dxdx. \tag{15}$$

Operating with L^{-1} on (13) it follows

$$y(x) = A + Bx + L^{-1}g(x) - L^{-1}f(x, y). \tag{16}$$

The Adomian decomposition method introduces the solution $y(x)$ by an infinite series of components

$$y(x) = \sum_{n=0}^{\infty} y_n(x), \quad f(x, y) = \sum_{n=0}^{\infty} A_n, \tag{17}$$

where the components $y_n(x)$ of the solution $y(x)$ will be determined recurrently, and A_n are Adomian polynomials that can be constructed for various classes of nonlinearity according to specific algorithms set by Adomian [1] and recently calculated by Wazwaz [15].

Substituting (17) into (16) we obtain

$$\sum_{n=0}^{\infty} y_n(x) = A + Bx + L^{-1}g(x) - L^{-1}\sum_{n=0}^{\infty} A_n. \tag{18}$$

To determine the components $y_n(x)$, we use Adomian decomposition method that suggests the use of the recursive relation

$$y_0(x) = A + Bx + L^{-1}g(x),$$

$$y_{k+1}(x) = -L^{-1}(A_k), \quad k \geq 0,$$

which gives

$$\begin{aligned} y_0(x) &= A + Bx + L^{-1}g(x), \\ y_1(x) &= -L^1(A_0), \\ y_2(x) &= -L^1(A_1), \\ y_3(x) &= -L^1(A_2), \\ &\vdots \end{aligned} \tag{19}$$

Combining (19) with (11) will enable us to determine the components $y_n(x)$ recursively. For numerical purposes the n-th approximant

$$\phi_n = \sum_{k=0}^{n-1} y_k,$$

can be used to approximate the solution.

In the above discussion it was shown that with the proper choice of the differential operator L , it is possible to overcome the singularity question and to attain practically a series solution by computing components of $y(x)$ as far as we like. The convergence concept has been discussed by Cherruault [7,8] among others.

4 Numerical illustration

Example 1. Consider the following singular initial value problem:

$$y'' + \frac{2}{x}y' - 10y = 12 - 10x^4, \quad y(0) = y'(0) = 0. \tag{20}$$

Then we obtain the operator form of (20)

$$Ly = 12 - 10x^4 + 10y. \tag{21}$$

Applying L^{-1} on both sides of (21) we get

$$Ly = 2x^2 - \frac{5}{21}x^6 + 10L^{-1}(y). \tag{22}$$

With respect to the above mentioned method we obtain recursive relationship

$$\begin{aligned} y_0(x) &= 2x^2 - \frac{5}{21}x^6 \\ y_{k+1}(x) &= 10L^{-1}(y_k), \quad k \geq 0. \end{aligned} \tag{23}$$

Consequently, the first few components are as follows:

$$\begin{aligned} y_0 &= 2x^2 - \frac{5}{21}x^6 \\ y_1 &= x^4 - \frac{25x^8}{756} \\ y_2 &= \frac{5x^6}{21} - \frac{25x^{10}}{8316} \\ y_3 &= \frac{25x^8}{756} - \frac{125x^{12}}{648648} \\ y_4 &= \frac{25x^{10}}{8316} - \frac{625x^{14}}{3243240} \end{aligned} \tag{24}$$

Other components can be evaluated in a similar manner. It is obvious that the noise terms $-\frac{5}{21}x^6$, $-\frac{25x^8}{756}$ and $-\frac{25x^{10}}{8316}$ appear in y_2, y_3 and y_4 with the opposite signs. Canceling these therm we get the exact solution

$$y(x) = 2x^2 + x^4.$$

Example 2. Consider the nonlinear singular initial value problem:

$$y'' + \frac{2}{x}y' + 4(2e^y + e^{y/2}) = 0, \quad y(0) = y'(0) = 0.$$

We obtain the operator form

$$Ly = -4(2e^y + e^{y/2}) \tag{25}$$

Applying L^{-1} on both sides of (25) we get

$$y = -4L^{-1}(2e^y + e^{y/2})$$

Using the decomposition series for the linear function $y(x)$ and the polynomial series for the nonlinear term , we obtain the recursive relationship

$$\begin{aligned} y_0(x) &= 0 \\ y_{k+1}(x) &= -4L^{-1}(A_k), \quad k \geq 0. \end{aligned} \tag{26}$$

The Adomian polynomials for the nonlinear term $2e^y + e^{y/2}$ are computed as follows:

$$\begin{aligned} A_0 &= 2e^{y_0} + e^{y_0/2} \\ A_1 &= y_1\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) \\ A_2 &= y_2\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) + \frac{y_1^2}{2!}\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) \\ A_3 &= y_3\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) + y_1y_2\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) + \frac{y_1^3}{3!}\left(2e^{y_0} + \frac{1}{2}e^{y_0/2}\right) \\ &\vdots \end{aligned} \tag{27}$$

Substituting (27) into (26) we get

$$\begin{aligned} y_0 &= 0 \\ y_1 &= -4L^{-1}(A_0) = -2x^2 \\ y_2 &= -4L^{-1}(A_1) = x^4 \\ y_3 &= -4L^{-1}(A_2) = -\frac{2}{3}x^6 \\ y_4 &= -4L^{-1}(A_3) = \frac{1}{2}x^8 \\ y_5 &= -4L^{-1}(A_4) = -\frac{2}{5}x^{10} \\ y_6 &= -4L^{-1}(A_5) = \frac{1}{3}x^{12} \\ &\vdots \end{aligned}$$

Then the solution in a series form is given by

$$y(x) = -2\left(x^2 - \frac{1}{2}x^4 + \frac{1}{3}x^6 - \frac{1}{4}x^8 + \frac{1}{5}x^{10} - \frac{1}{6}x^{12} + \dots\right).$$

Hence the exact solution has the form

$$y(x) = -2\ln(1 + x^2).$$

Acknowledgement

This research has been supported by the Czech Ministry of Education in the frames of MSM002160503 Research Intention MIKROSYN New Trends in Microelectronic Systems and Nanotechnologies and MSM0021630529 Research Intention Inteligent Systems in Automatization.

References

[1] ADOMIAN,G., *A review of the decomposition method and some recent results for nonlinear equation*, Math. Comput. Modelling **13**(7) , 1990, 17-43.

[2] ADOMIAN,G., RACH, R., *Noise terms in decomposition series solution,* Comput. Math. Appl. **24**(11), 1992, 61-64.

[3] ADOMIAN,G., RACH, R., SHAWAGFEH, N.T., *On the analytic solution of the Lane-Emden equation* , Found. Phys. Lett. **8**(2), 1995, 161-185.

[4] EL-SAYED, S.M., *Multi-integral methods for nonlinear boundary-value problems, A fourth order method for a singular two points boundary value problem,* J. Comput. Math. 71, 1999, 259-269.

[5] CHANDRASEKHAR, S., *Introduction to the Study of Stellar Structure,* Dover, New York , 1967.

[6] CHAWLA,M.M., KATTI, C.P., *A finite-difference method for a class of singular boundary-value problem,* IMA, J. Numer. Anal. 4, 1984, 457-467.

[7] CHERRUAULT, Y., *Convergence of Adomian's method,* Math. Comput. Modelling, 14, 1990 , 83-93.

[8] CHERRUAULT, Y., SACCOMANDI, G., *Some new results for convergence of Adomian's method applied to integral equations,* Math. Comput, Modelling **16**(2), 1992, 85-92.

[9] HOSSEINI, M.M., *Adomian decomposition method with Chebyshev polynomials,* Appl. Math. Comput. **175**, 2006, 1685-1693.

[10] IENGAR, S.R.K., JAIN, P., *Spline difference methods for singular two point boundary value problems,* Numer. Math. 500, 1987, 363-372.

[11] RICHARDSON, O.U., *The Emission of Electricity from Hot Bodies,* London, 1921.

[12] RUSSEL, R.D., SHAMPINE, L.F., *Numerical methods for singular boundary problems,* SIAM J. Numer. Anal.12, 1975, 13-21.

[13] WAZWAZ, A.M., *A new algorithm for solving differential equations of Lane-Emden type,* Appl. Math. Comput. 111, 2000, 53-61.

[14] WAZWAZ, A.M., *Analytical approximations and Pade approximations for Volterra's population model,* Appl. Math. Comput. **100** 1999, 13-25.

[15] WAZWAZ, A.M., *A new method for solving singular value problems in the second-order ordinary differential equations,* Appl. Math. Comput. **128** 2002, 45-57.

Current address

Ing. Olga Filippova
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology, Technická 8, 616 00 BRNO
e-mail: xfilip03@stud.feec.vutbr.cz

Doc. RNDr. Zdeněk Šmarda, CSc.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology, Technická 8, 616 00 BRNO
e-mail: smarda@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

BISECTION METHOD FOR FINDING INITIAL DATA GENERATING BOUNDED SOLUTIONS OF DISCRETE EQUATIONS

HLAVIČKOVÁ Irena, (CZ)

Abstract. Numerous papers are devoted to the asymptotic behavior of solutions of discrete equations. In many papers, one can find sufficient conditions guaranteeing the existence of at least one solution the graph of which stays in a prescribed domain. Not so much attention has been paid to the problem of determining corresponding initial data generating such solutions. Here we will try to find such data with help of a numerical method which resembles the well-known bisection method used for solving nonlinear equations.

Key words and phrases. Discrete equation, bounded solutions, initial data, bisection method.

Mathematics Subject Classification. Primary 39A10, 39A11.

1 Introduction

Throughout this paper, we use the following notation: for integers s, q, $s \le q$ we define $\mathbb{Z}_s^q := \{s, s+1, \ldots, q\}$, where the possibility $q = \infty$ is admitted, too.
We investigate the asymptotic behavior for $k \to \infty$ of the solutions of the equation

$$\Delta u(k) = f(k, u(k)) \tag{1}$$

where $k \in \mathbb{Z}_a^\infty$, $a \in \mathbb{N}$ is fixed, $\Delta u(k) = u(k+1) - u(k)$, and $f\colon \mathbb{Z}_a^\infty \times \mathbb{R} \to \mathbb{R}$.
The solution of Eq. (1) is defined as an infinite sequence of numbers

$$\{u(a), u(a+1), u(a+2), \ldots\}$$

such that for any $k \in \mathbb{Z}_a^\infty$, equality (1) holds.
The existence and uniqueness of the solution of Eq. (1) with a prescribed initial condition

$$u(a) = u^a \in \mathbb{R} \tag{2}$$

on $\mathbb{Z}_a^\infty$ is obvious. If f is continuous with respect to its second argument then the initial problem (1), (2) depends continuously on initial data.
Let $b(k), c(k)$ be real functions defined on $\mathbb{Z}_a^\infty$ such that $b(k) < c(k)$ for every $k \in \mathbb{Z}_a^\infty$. We define

$$\omega := \{(k,u) : k \in \mathbb{Z}_a^\infty,\, u \in \omega(k)\} \quad \text{with} \quad \omega(k) := \{u : b(k) < u < c(k)\}$$

and the closure

$$\overline{\omega} := \{(k,u) : k \in \mathbb{Z}_a^\infty,\, u \in \overline{\omega}(k)\} \quad \text{with} \quad \overline{\omega}(k) := \{u : b(k) \le u \le c(k)\}.$$

The aim is to find a solution $u = u(k)$ of Eq. (1) such that

$$u(k) \in \omega(k) \quad \text{for every } k \in \mathbb{Z}_a^\infty.$$

Similar problems are studied in many papers, e.g. [1], [3] or [6].

2 The Existence Theorem

The following theorem concerning asymptotic behavior of solutions of Eq. (1) is a particular case of more general results in [4, Theorem 2] and [5].

Theorem 2.1 *Let us suppose that $f : \overline{\omega} \to \mathbb{R}$ is continuous with respect to its second argument. If, moreover,*

$$f(k, b(k)) - b(k+1) + b(k) < 0 \tag{3}$$

and

$$f(k, c(k)) - c(k+1) + c(k) > 0 \tag{4}$$

for every $k \in \mathbb{Z}_a^\infty$, then there exists an initial condition

$$u(a) = u^* \in \omega(a) \tag{5}$$

such that the corresponding solution $u = u^*(k)$ satisfies the relation

$$u^*(k) \in \omega(k) \quad \textit{for every } k \in \mathbb{Z}_a^\infty. \tag{6}$$

Let us explain the geometrical meaning of conditions (3) and (4):
Consider a solution $u = u(k)$ of Eq. (1) such that $u(s) = b(s)$ for some $s \in \mathbb{Z}_a^\infty$. Then, due to Eq. (1), the next member of this solution, $u(s+1)$, is

$$u(s+1) = u(s) + f(s, u(s)) = b(s) + f(s, b(s)).$$

According to (3), $b(s) + f(s, b(s)) < b(s+1)$, i.e. $u(s+1) < b(s+1)$. This means that $u(s+1) \notin \omega(s+1)$.
Analogously we can show that if we have a solution $u = u(k)$ such that $u(s) = c(s)$ for some $s \in \mathbb{Z}_a^\infty$ then for this solution, $u(s+1) > c(s+1)$, i.e. $u(s+1) \notin \omega(s+1)$.
The proof of Theorem 2.1 is done by contradiction. It is supposed that no solution stays in ω and under this supposition a continuous mapping of the interval $[b(a), c(a)]$ onto the set $\{b(a), c(a)\}$ is found which is impossible.
Unfortunately, Theorem 2.1 just states that there exists a solution staying in the domain ω but it does not give us any recipe how to find the appropriate initial condition (5). This gap is particularly filled e.g. in [2] where the case of linear equation is studied. Here we present another approach which is more general. Our method will be applicable to any equation satisfying the conditions of Theorem 2.1.

3 Bisection Method for Finding the Initial Data

We will describe an algorithm how to find u^* so that the solution generated by the initial condition (5) stays in the domain ω. The value u^* will be found as a limit of an infinite sequence $\{u_i^a\}_{i=1}^\infty$ (although sometimes the process can be finite).
The method of finding u^* will be similar to the well-known bisection method for solving non-linear equations of the form $f(x) = 0$. Let us start with an interval that certainly contains the sought "root" u^*. According to Theorem 2.1, it is the interval $[b(a), c(a)]$. Denote

$$u_{L,1}^a := b(a) \quad \text{and} \quad u_{U,1}^a := c(a).$$

(L as "lower", U as "upper" bound). So we have the interval $[u_{L,1}^a, u_{U,1}^a]$ and, similarly as in the bisection method, we will construct a sequence of intervals $[u_{L,i}^a, u_{U,i}^a]$, $i = 1, 2, \dots$, containing the "root" u^*. The next interval will be obtained by bisecting the previous one and choosing the correct half of it.
Denote the solutions of Eq. (1) given by the initial conditions $u(a) = u_{L,i}^a$ and $u(a) = u_{U,i}^a$ as $u = u_{L,i}(k)$ and $u = u_{U,i}(k)$, respectively.
Due to conditions (3) and (4), we have

$$u_{L,1}(a+1) < b(a+1) \quad \text{and} \quad u_{U,1}(a+1) > c(a+1).$$

Now we will bisect the interval $[u_{L,1}^a, u_{U,1}^a]$. Denote its center as

$$u_1^a := \frac{u_{L,1}^a + u_{U,1}^a}{2}.$$

Consider the solution $u = u_1(k)$ of Eq. (1) given by the initial condition $u(a) = u_1^a$. There are three possibilities:
I) $u_1(k) \in \omega(k)$ for every $k \in \mathbb{Z}_a^\infty$. In this case $u^* = u_1^a$, we have a solution with the desired property (6) and we can stop the process.

II) There exists an $r \in \mathbb{Z}_a^\infty$ such that $u_1(k) \in \omega(k)$ for $k = a, \ldots, r-1$, but $u_1(r) \leq b(r)$, i.e. $u_1(r) \notin \omega(r)$. In this case we set

$$u^a_{L,2} := u^a_1, \quad u^a_{U,2} := u^a_{U,1}.$$

III) There exists an $s \in \mathbb{Z}_a^\infty$ such that $u_1(k) \in \omega(k)$ for $k = a, \ldots, s-1$, but $u_1(s) \geq c(s)$. This time we change the upper bound of the interval:

$$u^a_{U,2} := u^a_1, \quad u^a_{L,2} := u^a_{L,1}.$$

Now, either we have the desired u^*, or we have a new interval $[u^a_{L,2}, u^a_{U,2}]$ with the property that the solution $u = u_{L,2}(k)$ exceeds the lower bound $b(r)$ of the domain ω for some $r \in \mathbb{Z}_a^\infty$, meanwhile the solution $u = u_{U,2}(k)$ exceeds the upper bound $c(s)$ for some $s \in \mathbb{Z}_a^\infty$. Such interval has to contain a point u^* for which the corresponding solution $u = u^*(k)$ stays in ω.
Further, we will proceed inductively. Having the interval $[u^a_{L,i}, u^a_{U,i}]$, we bisect it and denote its center as

$$u^a_i := \frac{u^a_{L,i} + u^a_{U,i}}{2}.$$

For the solution $u = u_i(k)$ given by the initial condition $u(a) = u^a_i$, we have three possibilites: either it stays in ω, or it exceeds its lower bound, or it exceeds its upper bound. According to this, either we have found $u^* = u^a_i$, or we set $u^a_{L,i+1} := u^a_i, u^a_{U,i+1} := u^a_{U,i}$, or we set $u^a_{U,i+1} := u^a_i, u^a_{L,i+1} := u^a_{L,i}$, respectively.
Continuing this process, either we get the sought initial point u^* in a finite number of steps, or we get infinite sequences $\{u^a_{L,i}\}_{i=1}^\infty$, $\{u^a_{U,i}\}_{i=1}^\infty$ and $\{u^a_i\}_{i=1}^\infty$. These sequences are obviously convergent as $\{u^a_{L,i}\}_{i=1}^\infty$ is a nondecreasing sequence bounded from above by $c(a)$, $\{u^a_{U,i}\}_{i=1}^\infty$ is a nonincreasing sequence bounded from below by $b(a)$ and $u^a_{L,i} < u^a_i < u^a_{U,i}$ for every $i \in \mathbb{N}$. In this case, $u^* = \lim_{i\to\infty} u^a_i$.

4 Practical Implementation of the Algorithm

Programming the above described method, we are limited by the possibilities of computers. In the ideal case, we would bisect the intervals until either we find a solution with property (6), or the length of the interval $[u^a_{L,i}, u^a_{U,i}]$ is less than some chosen $\varepsilon > 0$. But, practically, for a given initial condition (2), we can compute the values of the corresponding solution of Eq. (1) for $k = a, a+1, \ldots$, but it is clear that it is impossible to compute to infinity. We have to stop sometimes. Thus, given a fixed $n \in \mathbb{Z}_a^\infty$, we are able to find a point $\tilde{u}^*$ such that the solution $u = \tilde{u}^*(k)$ satisfies the condition

$$\tilde{u}^*(k) \in \omega(k), \quad k \in \mathbb{Z}_a^n. \tag{7}$$

The algorithm (in C++ similar pseudocode) is as follows:

```
inside = false;    //indicates whether the solution stays inside ω
u_L = b(a);
u_U = c(a);
while ( !inside && (u_U − u_L) > ε) {
   ũ* = (u_L + u_U)/2;
   ũ*(a) = ũ*;
   for (k = a; k < n; k++) {
      ũ*(k + 1) = ũ*(k) + f(k, ũ*(k));
      if (ũ*(k + 1) < b(k + 1)) {
         u_L = ũ*;
         break;
      }
      else if (ũ*(k + 1) > c(k + 1)) {
         u_U = ũ*;
         break;
      }
   }
   if (n == k)    // the whole for-cycle passed without breaking
      inside = true;
}
if ( !inside)
   ũ* = (u_L + u_U)/2;
```

If the while-cycle ends due to the condition $u_U - u_L < \varepsilon$ then we are sure that $|\tilde{u}^* - u^*| < \varepsilon/2$. But if the cycle ends because of the "inside" condition, then we just have a solution for which (7) holds. But the validity for every $k \in \mathbb{Z}_a^\infty$ cannot be guaranteed and neither is guaranteed that the numerically obtained point $\tilde{u}^*$ is close to the precise value of u^*.

5 Numerical Experiment

Example 5.1 *Consider the equation*

$$\Delta u(k) = -\frac{3}{2}\,k + \sqrt{u^2(k) + k^2}, \quad k \in \mathbb{Z}_3^\infty. \tag{8}$$

First we will prove that there exists a solution $u = u^*(k)$ such that for every $k \in \mathbb{Z}_3^\infty$

$$k < u^*(k) < 2k, \tag{9}$$

and then we will find the approximate value of the initial condition u^* generating this solution.

We will show that the assumptions of Theorem 2.1 are satisfied. Put

$$a := 3, \quad b(k) := k, \quad c(k) := 2k, \quad f(k,u) := -\frac{3}{2}\,k + \sqrt{u^2 + k^2}$$

and

$$\omega(k) := \{u : k < u < 2k\}$$

Function f is obviously continuous. Let us verify conditions (3) *and* (4).
Substituting into inequality (3)*, we get*

$$f(k, b(k)) - b(k+1) + b(k) = -\frac{3}{2}k + \sqrt{k^2 + k^2} - (k+1) + k = k(\sqrt{2} - \frac{3}{2}) - 1 < 0$$

which holds for every $k \in \mathbb{N}$.
Substituting into (4)*, we get*

$$f(k, c(k)) - c(k+1) + c(k) = -\frac{3}{2}k + \sqrt{4k^2 + k^2} - (2k+2) + 2k = k(\sqrt{5} - \frac{3}{2}) - 2 > 0.$$

This is fulfilled for $k \in \mathbb{Z}_3^\infty$.
Hence, all the assumptions of Theorem 2.1 are satisfied and thus there exists an initial condition $u(3) = u^*$ such that the corresponding solution satisfies (9).

Now let us find the approximate value of u^* with help of the method described in Section 3. Put $\varepsilon = 0.01$.

Starting values: $u_L = 3$, $u_U = 6$, $\tilde{u}^* = 4.5$				*New values:* $u_L = 4.5$, $u_U = 6$, $\tilde{u}^* = 5.25$			
k	$b(k)$	$\tilde{u}^*(k)$	$c(k)$	k	$b(k)$	$\tilde{u}^*(k)$	$c(k)$
3	*3*	*4.5*	*6*	*3*	*3*	*5.25*	*6*
4	*4*	*5.40833*	*8*	*4*	*4*	*6.79669*	*8*
5	*5*	*6.13514*	*10*	*5*	*5*	*8.68308*	*10*
6	*6*	*6.54968*	*12*	*6*	*6*	*11.2028*	*12*
7	*7*	*6.43215*	*14*	*7*	*7*	*14.9113*	*14*
Stop, $\tilde{u}^*(7) < b(7)$				*Stop,* $\tilde{u}^*(7) > c(7)$			

Analogously we continue, until we get the interval $[u_L, u_U] \approx [4.763; 4.769]$ *the length of which is less than ε. Thus we get* $u^* \approx 4.766$.

We can also slightly change the algorithm by leaving out the stop criterion $u_U - u_L < \varepsilon$. Then we stop only in the case that we find a solution satisfying (7) *for some chosen $n \in \mathbb{Z}_a^\infty$. For such changed algorithm with $n = 50$, we get the approximate value* $u^* \approx 4.76748736115405$.

References

[1] BAŠTINEC, J., DIBLÍK, J.: *Asymptotic formulae for a particular solution of linear non-homogeneous discrete equations.* In Advances in difference equations, IV., Comput. Math. Appl. Vol. 45, No. 6-9, pp. 1163-1169, 2003.

[2] BAŠTINEC, J., DIBLÍK, J., RŮŽIČKOVÁ, M.: *Initial data generating bounded solutions of linear discrete equations.* In Opuscula Math. Vol. 26, No. 3, pp. 395-406, 2006.

[3] DIBLÍK, J.: *Discrete retract principle for systems of discrete equations.* In Advances in difference equations, III., Comput. Math. Appl. Vol. 42, No. 3-5, pp. 515-528, 2001.

[4] DIBLÍK, J.: *Asymptotic behaviour of solutions of discrete equations.* In Funct. Differ. Equ., No. 11, pp. 37-48, 2004.

[5] DIBLÍK, J., HLAVIČKOVÁ, I.: *Combination of Liapunov and retract methods in the investigation of the asymptotic behavior of solutions of systems of discrete equations*. In Dyn. Syst. Appl., Vol. 18, No. 3-4, pp. 507-537, 2009.
[6] DIBLÍK, J., RŮŽIČKOVÁ, M., ŠMARDA, Z.: *Ważewski's method for systems of dynamic equations on time scales*. In Nonlinear Anal. (2009),doi:10.1016/j.na.2009.01.098 (in press).

Current address

Irena Hlavičková, Mgr., PhD.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 8
616 00 Brno
Czech Republic
tel.: +420-541143152
email: hlavicka@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

OSCILLATION OF NONLINEAR NEUTRAL DIFFERENTIAL SYSTEMS

ILAVSKÁ Iveta, (SK), NAJMANOVÁ Anna, (SK), OLACH Rudolf, (SK)

Abstract. The article deals with the nonlinear neutral differential delay systems. The oscillatory properties of such systems are investigated. Some sufficient conditions which guarantee the oscillatory behaviour are established. Some examples illustrate the main results.

Key words and phrases: differential system; nonlinear neutral system; oscillation; sufficient conditions.

Mathematics Subject Classification: Primary 34K15, 34K10.

1 Introduction

In the present paper we consider the nonlinear neutral differential delay systems of the form

$$\begin{aligned} [x_1(t) - a(t)x_1(t-\tau)]' &= p_1(t)|x_2(t)|^{\alpha-1}x_2(t), \\ x_2'(t) &= -p_2(t)|x_1(t-\sigma)|^{\beta-1}x_1(t-\sigma),\ t \geq t_0, \end{aligned} \tag{1}$$

where $a, p_i \in C([t_0,\infty),[0,\infty))$, $p_i(t) \not\equiv 0$, $i = 1,2$, $\alpha > 0$, $\beta > 0$, $\sigma > \tau > 0$.
By a solution of the system (1) we mean a function $x = (x_1, x_2) \in C([t_1 - \sigma, \infty), R^2)$ for some $t_1 \geq t_0 + \sigma$ such that $x_1(t) - a(t)x_1(t-\tau)$ and $x_2(t)$ are continuously differentiable on $[t_1,\infty)$ and such that the system (1) is satisfied for $t \geq t_1$.
By W we denote the set of all solutions $x = (x_1, x_2)$ of (1). A component $x_1(t)$ or $x_2(t)$ of x is said to be oscillatory if it has arbitrarily large zeros; otherwise it is called nonoscillatory.
In the recent years many autors are studying the differential systems similar to (1). However they are mostly interesting about the asymptotic behaviour of such systems, e.g. in [3,5,7] and in the papers cited therein. It seems that only in a few articles the oscillatory properties are

treated, e.g. in [2,4]. Thus our aim is to study the oscillation of components of the solutions of the systems (1).
In addition the nonlinear neutral differential systems can be interpreted as the models for the population biology [6]. It is recognised that time delays are natural components of the processes of biology, ecology, physiology, economics, mechanics, etc. This is due to influence of the past history of the processes on their evolution.
We shall need the following lemma in the next section.

Lemma 1.1 $[2,4]$ *Suppose that* $P \in C([t_0,\infty),[0,\infty))$, $\delta > 0$ *and*

$$\liminf_{t\to\infty} \int_{t-\delta}^{t} P(s)\,ds > \frac{1}{e}.$$

Then the following inequality

$$u'(t) + P(t)u(t-\delta) \le 0,\ t \ge t_0,$$

has no eventually positive solutions.

2 The oscillatory properties

In this section we establish two oscillatory theorems.

Theorem 2.1 *Suppose that* $x = (x_1, x_2) \in W$ *and the component* $x_2(t)$ *is oscillatory. Then* $x_1(t)$ *is also oscillatory.*

Proof. Assume that $x_1(t)$ is a nonoscillatory component of x. Let $x_1(t) > 0$ for $t \ge t_0$. Then by the second equation of (1) we have

$$x_2'(t) \le 0,\ t \ge t_1 \ge t_0,$$

where t_1 is sufficiently large. This implies that $x_2(t) > 0$ or $x_2(t) < 0$ for $t \ge t_2 \ge t_1$, i.e. $x_2(t)$ is nonoscillatory. Let now $x_1(t) \le 0$, $t \ge t_0$. Then using the same argument as previously we get $x_2'(t) \ge 0$, $t \ge t_1$, which implies a nonoscillatory character of $x_2(t)$. Thus the theorem is proved.

Theorem 2.2 *Suppose that* $0 < a(t) \le 1$, $\alpha\beta = 1$, *and*

$$\int_{t_0}^{\infty} p_i(t)\,dt = \infty, \quad i = 1,2,$$

$$\liminf_{t\to\infty} \int_{t+\tau+\gamma-\sigma}^{t} p_1(s) \left(\int_s^{s+\gamma} p_2(\xi)\,d\xi \right)^{\alpha} ds > \frac{1}{e}, \tag{2}$$

where $0 < \gamma < \sigma - \tau$. *Then* $x_1(t)$ *is oscillatory.*

Proof. Without loss of generality we may assume that $x_1(t) > 0$ for $t \geq t_0$, i.e. $x_1(t)$ is nonoscillatory. Then the second equation of (1) implies

$$x_2^{'}(t) \leq 0,\ t \geq t_1 \geq t_0,$$

where t_1 is sufficiently large. So the next two cases are possible:

1. $x_2(t) < 0,\ t \geq t_2 \geq t_1,$
2. $x_2(t) > 0,\ t \geq t_2.$

Next for the simplicity we will use the notation

$$y(t) = x_1(t) - a(t)x_1(t-\tau),\ t \geq t_2.$$

Case 1. We have $x_2(t) < 0$ for $t \geq t_2$. Then

$$y^{'}(t) = p_1(t)|x_2(t)|^{\alpha-1}x_2(t) \leq 0,\ t \geq t_2.$$

It follows that for $y(t)$ there are two possible cases:

(i) $y(t) < 0,\ t \geq t_3 \geq t_2,$

(ii) $y(t) > 0,\ t \geq t_3.$

If the case (i) holds, then there are constants $c > 0$ and $0 < a_1 \leq 1$ such that

$$\begin{aligned} x_1(t) - a(t)x_1(t-\tau) &\leq -c,\ t \geq t_3, \\ x_1(t) &\leq -c + a_1 x_1(t-\tau),\ t \geq t_3. \end{aligned}$$

By induction we obtain

$$x_1(t_2 + n\tau) \leq -c\sum_{i=0}^{n-1} a_1^i + a_1^n x_1(t_2).$$

We conclude that $x_1(t_2 + n\tau) < 0$ for large n, which contradics the fact that $x_1(t) > 0$ for $t \geq t_0$. Hence the case (ii) holds, i.e. $y(t) > 0,\ t \geq t_3$ and the function $y(t)$ is nonincreasing on $[t_3, \infty)$. Integrating the first equation of (1) from t_3 to ∞ we get

$$\begin{aligned} -y(t_3) &\leq \int_{t_3}^{\infty} p_1(s)|x_2(s)|^{\alpha-1}x_2(s)\,ds, \\ y(t_3) &\geq -\int_{t_3}^{\infty} p_1(s)|x_2(s)|^{\alpha-1}x_2(s)\,ds \geq |x_2(t_3)|^{\alpha}\int_{t_3}^{\infty} p_1(s)\,ds. \end{aligned}$$

It implies that

$$\int_{t_3}^{\infty} p_1(s)\,ds < \infty,$$

which contradicts the condition of theorem. The Case 1 cannot occur.

Case 2. Let $x_2(t) > 0$ for $t \geq t_2$. Then $y^{'}(t) \geq 0$ and $y(t)$ is nondecreasing on $[t_2, \infty)$. Thus for function $y(t)$ two cases are possible:

(j) $y(t) > 0$, $t \geq t_3 \geq t_2$,

(jj) $y(t) < 0$, $t \geq t_3$.

Consider the case (j). From relation $x_1(t) - a(t)x_1(t-\tau) = y(t)$ we obtain that $x_1(t) \geq y(t)$, $t \geq t_3$. For sufficiently large $t_3 \geq t_2$ the second equation of (1) implies that

$$\begin{aligned} x_2^{'}(t) &= -p_2(t)x_1^{\beta}(t-\sigma) \leq -p_2(t)y^{\beta}(t-\sigma), \\ -x_2^{'}(t) &\geq p_2(t)y^{\beta}(t-\sigma),\ t \geq t_3. \end{aligned}$$

Integrating the last inequality form t_3 to ∞ we get

$$x_2(t_3) \ \geq \ \int_{t_3}^{\infty} p_2(s)y^{\beta}(s-\sigma)\,ds \geq y^{\beta}(t_3-\sigma)\int_{t_3}^{\infty} p_2(s)\,ds,$$

which is a contradiction.
Finally consider the case (jj). Then for sufficiently large $t_3 \geq t_2$ we obtain

$$-a(t)x_1(t-\tau) < y(t),\ \ x_1(t-\tau) > -y(t)$$

and

$$x_1(t-\sigma) > -y(t+\tau-\sigma).$$

So we have

$$\begin{aligned} x_2^{'}(t) &= -p_2(t)x_1^{\beta}(t-\sigma) \leq -p_2(t)|y(t+\tau-\sigma)|^{\beta}, \\ -x_2^{'}(t) &\geq p_2(t)|y(t+\tau-\sigma)|^{\beta},\ t \geq t_3. \end{aligned}$$

Integrating the last inequality from t to ∞ we obtain

$$\begin{aligned} x_2(t) &\geq \int_t^{\infty} p_2(s)|y(s+\tau-\sigma)|^{\beta}\,ds, \\ p_1(t)x_2^{\alpha}(t) &\geq p_1(t)\left(\int_t^{\infty} p_2(s)|y(s+\tau-\sigma)|^{\beta}\,ds\right)^{\alpha}, \\ y^{'}(t) &\geq p_1(t)\left(\int_t^{\infty} p_2(s)|y(s+\tau-\sigma)|^{\beta}\,ds\right)^{\alpha},\ t \geq t_3. \end{aligned}$$

Then it follows that

$$\begin{aligned} y^{'}(t) &\geq p_1(t)\left(\int_t^{t+\gamma} p_2(s)|y(s+\tau-\sigma)|^{\beta}\,ds\right)^{\alpha} \geq p_1(t)\left(|y(t+\tau+\gamma-\sigma)|^{\beta}\int_t^{t+\gamma} p_2(s)\,ds\right)^{\alpha}, \\ -y^{'}(t) &+ p_1(t)\left(\int_t^{t+\gamma} p_2(s)\,ds\right)^{\alpha} |y(t+\tau+\gamma-\sigma)|^{\alpha\beta} \leq 0, \\ -y^{'}(t) &- p_1(t)\left(\int_t^{t+\gamma} p_2(s)\,ds\right)^{\alpha} y(t+\tau+\gamma-\sigma) \leq 0,\ t \geq t_3. \end{aligned} \tag{3}$$

Set

$$v(t) = -y(t),\ \delta = \sigma-\tau-\gamma,\ p(t) = p_1(t)\left(\int_t^{t+\gamma} p_2(s)\,ds\right)^{\alpha}.$$

Then the inequality (3) can be written as follows

$$v'(t) + p(t)v(t-\delta) \le 0,\ t \ge t_3. \tag{4}$$

With regard to condition (2), the Lemma 1.1 implies that the inequality (4) cannot have a positive solution $v(t)$. Thus the case (jj) is also impossible. The proof of theorem is complete.

Example 1. Consider the nonlinear neutral differential system

$$\begin{aligned} [x_1(t) - \frac{1}{2}x_1(t-1)]' &= |x_2(t)|^{\alpha-1}x_2(t), \\ x_2'(t) &= -|x_1(t-3)|^{\beta-1}x_1(t-3),\ t \ge 0, \end{aligned}$$

where $\alpha = 2$, $\beta = \frac{1}{2}$. The conditon (2) has a form

$$\lim_{t\to\infty} \inf \int_{t+\gamma-2}^{t} \gamma^2\, ds\ = (2-\gamma)\gamma^2 > \frac{1}{e},$$

and e.g. for $\gamma = 1$ is satisfied. Other conditions of Theorem 2.2 are also satisfied. Thus the component $x_1(t)$ is oscillatory.

Example 2. Consider the nonlinear neutral differential system

$$\begin{aligned} [x_1(t) - \frac{1}{2t}x_1(t-1)]' &= 4|x_2(t)|^{\alpha-1}x_2(t), \\ x_2'(t) &= -|x_1(t-2)|^{\beta-1}x_1(t-2),\ t \ge 1, \end{aligned}$$

where $\alpha = \frac{1}{3}$, $\beta = 3$. The conditon (2) has a form

$$\lim_{t\to\infty} \inf \int_{t+\gamma-1}^{t} 4\gamma^{\frac{1}{3}}\, ds\ = 4(1-\gamma)\gamma^{\frac{1}{3}} > \frac{1}{e},$$

and e.g. for $\gamma = \frac{1}{2}$ is satisfied. Other conditions of Theorem 2.2 are also satisfied. Thus the component $x_1(t)$ is oscillatory.

Acknowledgements

The research was supported by the grant 1/0090/09 of the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and project APVV-0700-07 of the Slovak Research and Development Agency.

References

[1] DIBLÍK, J., SVOBODA, Z.: *Positive solutions of p-type retarded functional differential equations.* Nonlinear Analysis **64**(2006), pp. 1831–1848.

[2] ERBE, L. H., KONG, Q. K., ZHANG, B. G.: *Oscillation Theory for Functional Differential Equations.* Marcel Dekker, New York, 1995.
[3] HANUŠTIAKOVÁ, Ľ., OLACH, R.: *Asymptotic and oscillatory properties of neutral differential systems.* Studies of the University of Žilina, Mathematical Series, Vol. **19**(2005), pp. 9–18.
[4] MARUŠIAK, P., OLACH R.: *Functional differential equations* (Slovak lang.). EDIS -ŽU, 2000.
[5] MIHALÍKOVÁ B.: *A little about neutral differential system.* Studies of the University of Žilina, Mathematical Series, Vol. **18**(2004), pp. 27–36.
[6] MURRAY, J. D.: Mathematical Biology. New York, Springer-Verlag, 2002.
[7] ŠPANÍKOVÁ E., ŠAMAJOVÁ, H.: *Asymptotic behaviour of nonoscillatory solutions of neutral differential systems.* Studies of the University of Žilina, Mathematical Series, Vol. **15**(2002), pp. 1–7.

Current address

RNDr. Iveta Ilavská
Faculty of Science
Department of Mathematics
Univerzitná 8215/1
010 26 Žilina
e-mail: iveta.ilavska@fpv.uniza.sk

Mgr. Anna Najmanová
Faculty of Science
Department of Mathematics
Univerzitná 8215/1
010 26 Žilina
e-mail: anna.najmanova@fpv.uniza.sk

doc. RNDr. Rudolf Olach, CSc.
Faculty of Science
Department of Mathematics
Univerzitná 8215/1
010 26 Žilina
e-mail: rudolf.olach@fpv.uniza.sk

Journal of Applied Mathematics

volume III (2010), number II

ON EXISTENCE OF SPECIAL SOLUTIONS OF CERTAIN COUPLED SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS

KECSKÉS Norbert, (SK)

Abstract. In the paper we investigate the asymptotic behavior of special solutions of certain systems of linear differential equations with constant coefficients in two dimensional space. We show that in, so called, non-critical cases the solution with a given property always exists, as well as in critical cases, except one.

Key words: differential system, differential equation, geodetics, eigenvalues and vectors, solution,

Mathematics Subject Classification: 97M60.

1 Introduction

Let's consider two systems of linear differential equations of the form

$$\boldsymbol{w}'(t) = \boldsymbol{A} \cdot \boldsymbol{w}(t) + \boldsymbol{v}(t) \tag{1}$$
$$\boldsymbol{v}'(t) = \boldsymbol{B} \cdot \boldsymbol{v}(t) \tag{2}$$
$$\boldsymbol{w}(0) = \boldsymbol{0} \tag{3}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ are square 2x2 matrices with constant coefficients. Moreover we assume $\boldsymbol{A}+\boldsymbol{B}$ to be a skew-symmetric matrix, i.e. the condition

$$\boldsymbol{A}+\boldsymbol{B} = -(\boldsymbol{A}+\boldsymbol{B})^T \tag{4}$$

holds. We call the solutions of (1) satisfying the initial condition (3) special solutions. Our aim is to prove the existence of at least one special solution of (1) with the property $\liminf\limits_{t\to\infty}\left\|e^{-\omega t}\boldsymbol{w}(t)\right\| > 0$, where $\| \ \|$ denotes the Euclidean norm and ω is the maximum of real parts of all eigenvalues of the matrices $\boldsymbol{A}, \boldsymbol{B}$. Since we assume (4) to hold, there are four possible cases:

$$0 = \omega = \max \operatorname{Re} \lambda_i = \max \operatorname{Re} \mu_i \tag{5}$$

$$0 < \omega = \max \operatorname{Re} \lambda_i = \max \operatorname{Re} \mu_i \tag{6}$$

$$0 < \omega = \max \operatorname{Re} \mu_i > \max \operatorname{Re} \lambda_i \tag{7}$$

$$0 < \omega = \max \operatorname{Re} \lambda_i > \max \operatorname{Re} \mu_i \tag{8}$$

where μ_i, λ_i denote the eigenvalues of $\boldsymbol{A}, \boldsymbol{B}$, respectively. We call the cases (5) and (6) critical and the cases (7) and (8) noncritical. The above given systems (1) - (3) originate from differential geometry and under certain conditions they represent a linearized Jacobi equation of a vector field along a geodesic curve [2].

2 Special solution

We derive the explicit solution of (1) under the assumption that no eigenvalues are multiple and the spectra of the matrices are disjoint.

Lemma 1. *Let μ_1, μ_2 be the eigenvalues of $\boldsymbol{A}$ and let λ_1, λ_2 be the eigenvalues of $\boldsymbol{B}$. Let $\boldsymbol{a}_1, \boldsymbol{a}_2$ be the eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{b}_1, \boldsymbol{b}_2$ be the eigenvectors of $\boldsymbol{B}$, then the special solution of* (1) *is of the form*

$$\boldsymbol{w}(t) = \sum_{k=1}^{2} \left(\sum_{i=1}^{2} \frac{e^{\lambda_i t} - e^{\mu_k t}}{\lambda_i - \mu_k} d_{ik} c_i \right) \boldsymbol{a}_k \tag{9}$$

Proof. The general solution of (1) is of the form $\boldsymbol{w}(t) = \boldsymbol{w}_c(t) + \boldsymbol{w}_p(t)$, where $\boldsymbol{w}_c(t)$ denotes the solution of $\boldsymbol{w}'(t) = \boldsymbol{A} \cdot \boldsymbol{w}(t)$ and $\boldsymbol{w}_p(t)$ denotes any particular solution of (1). The general solution of (2) has the form $\boldsymbol{v}(t) = \sum_{i=1}^{2} c_i \boldsymbol{b}_i e^{\lambda_i t}$.

Let's set the eigenvectors $\boldsymbol{a}_1, \boldsymbol{a}_2$ as the basis of E_2. Now $\boldsymbol{b}_i = \sum_{k=1}^{2} d_{ik} \boldsymbol{a}_k$, for $i = 1,2$, where d_{ik} are real or complex numbers. We find the required particular solution by the method of undetermined coefficients. Let

$$\boldsymbol{w}_p(t) = \sum_{i=1}^{2} \boldsymbol{u}_i e^{\lambda_i t} = \sum_{i=1}^{2} \left(\sum_{k=1}^{2} \alpha_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t}, \text{ where } \alpha_{ik} \text{ are unknown numbers.}$$

Upon differentiation and plugging into (1) we get

$$\sum_{i=1}^{2} \lambda_i \left(\sum_{k=1}^{2} \alpha_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t} = \boldsymbol{A} \cdot \sum_{i=1}^{2} \left(\sum_{k=1}^{2} \alpha_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t} + \sum_{i=1}^{2} c_i \left(\sum_{k=1}^{2} d_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t} .$$

Since $\boldsymbol{a}_k$ are eigenvectors of $\boldsymbol{A}$, we have $\boldsymbol{A}\boldsymbol{a}_k = \mu_k \boldsymbol{a}_k$, for $k = 1,2$. Hence

$$\sum_{i=1}^{2} \lambda_i \left(\sum_{k=1}^{2} \alpha_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t} = \sum_{i=1}^{2} \left(\sum_{k=1}^{2} \alpha_{ik} \mu_k \boldsymbol{a}_k \right) e^{\lambda_i t} + \sum_{i=1}^{2} c_i \left(\sum_{k=1}^{2} d_{ik} \boldsymbol{a}_k \right) e^{\lambda_i t},$$

from which $\lambda_i \alpha_{ik} = \mu_k \alpha_{ik} + d_{ik} c_i$, and $\alpha_{ik} = \dfrac{d_{ik} c_i}{\lambda_i - \mu_k}$, for $k, i = 1,2$.

The general solution of (1) then assumes the form

$$\boldsymbol{w}(t) = \boldsymbol{w}_c(t) + \boldsymbol{w}_p(t) = \sum_{k=1}^{2} s_k \boldsymbol{a}_k e^{\mu_k t} + \sum_{i=1}^{2} \left(\sum_{k=1}^{2} \frac{d_{ik} c_i}{\lambda_i - \mu_k} \boldsymbol{a}_k \right) e^{\lambda_i t},$$

where s_k are real or complex numbers. Now we impose the condition $\boldsymbol{w}(0) = \boldsymbol{0}$, then

$$0 = \sum_{k=1}^{2} s_k \mathbf{a}_k + \sum_{i=1}^{2} \left(\sum_{k=1}^{2} \frac{d_{ik} c_i}{\lambda_i - \mu_k} \mathbf{a}_k \right),$$

from which $s_k = -\sum_{i=1}^{n} \dfrac{d_{ik} c_i}{\lambda_i - \mu_k}$, for $k = 1,2$.

The system (1) with the initial condition $\boldsymbol{w}(0) = \boldsymbol{0}$ has the solution

$$\boldsymbol{w}(t) = \sum_{k=1}^{n} \left(-\sum_{i=1}^{n} \frac{d_{ik} c_i}{\lambda_i - \mu_k} \right) \boldsymbol{a}_k e^{\mu_k t} + \sum_{i=1}^{n} \left(\sum_{k=1}^{n} \frac{d_{ik} c_i}{\lambda_i - \mu_k} \boldsymbol{a}_k \right) e^{\lambda_i t},$$

which is equivalent to (9). ■

3 Results in Euclidean space E_2

3.1. The case of real eigenvalues

Let $\mu_1 > \mu_2 \in \mathbf{R} \wedge \lambda_1 > \lambda_2 \in \mathbf{R}$. According to the abovementioned restrictions (see Lemma 1), the case of real eigenvalues is always noncritical. We distinguish two sub cases, namely (7) and (8). In case of (7) we have $0 < \omega = \mu_1 > \max \operatorname{Re} \lambda_i$. From (9) for $c_1 \neq 0, c_2 \neq 0$ we get

$$\overline{\boldsymbol{w}}(t) = e^{-\mu_1 t} \boldsymbol{w}(t) = \sum_{k=1}^{2} \left(\sum_{i=1}^{2} \frac{e^{(\lambda_i - \mu_1)t} - e^{(\mu_k - \mu_1)t}}{\lambda_i - \mu_k} d_{ik} c_i \right) \boldsymbol{a}_k$$

and upon neglecting the exponential terms of the form $e^{-k\,t}$ for $t \to \infty$ we are left with

$$\widetilde{\boldsymbol{w}}(t)=\left(\frac{-1}{\lambda_1-\mu_1}d_{11}c_1+\frac{-1}{\lambda_2-\mu_1}d_{21}c_2\right)\boldsymbol{a}_1 \tag{10}$$

In order to have $\liminf\limits_{t\to\infty}\left\|e^{-\omega t}\boldsymbol{w}(t)\right\|=\liminf\limits_{t\to\infty}\left\|\widetilde{\boldsymbol{w}}(t)\right\|>0$, at least one of the coefficients d_{11} or d_{21} must be nonzero, but this is assured by the skew – symmetry of the matrices. The condition $\liminf\limits_{t\to\infty}\left\|\widetilde{\boldsymbol{w}}(t)\right\|>0$ is equivalent to the statement that the curve $\widetilde{\boldsymbol{w}}(t)$ doesn't pass through the origin for any $t>t_0$, where t_0 is arbitrary preset real number.

From the geometric viewpoint (10) represents a point off the origin on a straight line generated by the vector $\boldsymbol{a}_1$. So in this case the condition $\liminf\limits_{t\to\infty}\left\|e^{-\omega t}\boldsymbol{w}(t)\right\|>0$ is fulfilled.

Now let's consider $0<\omega=\lambda_1>\max\operatorname{Re}\mu_i$ (8). If in (9) $c_2=0$ we get

$$\overline{\boldsymbol{w}}(t)=e^{-\lambda_1 t}\boldsymbol{w}(t)=\sum_{k=1}^{2}\frac{1-e^{(\mu_k-\lambda_1)t}}{\lambda_1-\mu_k}d_{1k}c_1\,\boldsymbol{a}_k \text{ and}$$

$$\widetilde{\boldsymbol{w}}(t)=c_1\left(\frac{1}{\lambda_1-\mu_1}d_{11}\boldsymbol{a}_1+\frac{1}{\lambda_1-\mu_2}d_{12}\boldsymbol{a}_2\right) \tag{11}$$

Obviously, we again have a point lying off the origin and automatically $d_{11}\neq 0\vee d_{12}\neq 0$, otherwise the vector $\boldsymbol{b}_1$ would be zero. Note that in this case it is necessary to set $c_1\neq 0$. If $c_1=0\wedge c_2\neq 0$, we would have a solution with $\liminf\limits_{t\to\infty}\left\|e^{-\alpha t}\boldsymbol{w}(t)\right\|=0$. But for $c_1\neq 0$ the condition $\liminf\limits_{t\to\infty}\left\|e^{-\alpha t}\boldsymbol{w}(t)\right\|>0$ always holds.

3.2. The case of complex eigenvalues

Let $\mu_{1,2}=\alpha\pm\mathrm{i}\beta\ \wedge\ \lambda_{1,2}=\gamma\pm\mathrm{i}\delta$. In this case $\gamma=-\alpha$ due to the skew-symmetry of the matrices. The case of complex eigenvalues with nonzero real parts is always non-critical.
We consider the case $0<\omega=\alpha>\gamma$. From (9) for $c_2=0$ we get

$$\overline{\boldsymbol{w}}(t)=e^{-\alpha t}\boldsymbol{w}(t)=\sum_{k=1}^{2}\frac{e^{(\lambda_1-\alpha)t}-e^{(\mu_k-\alpha)t}}{\lambda_1-\mu_k}d_{1k}c_1\,\boldsymbol{a}_k \text{ and}$$

$$\widetilde{\boldsymbol{w}}(t)=-c_1\left(\frac{e^{\mathrm{i}\beta t}}{\lambda_1-\mu_1}d_{11}\boldsymbol{a}_1+\frac{e^{-\mathrm{i}\beta t}}{\lambda_1-\mu_2}d_{12}\boldsymbol{a}_2\right) \tag{12}$$

where $d_{11}\neq 0\vee d_{12}\neq 0$, otherwise the vector $\boldsymbol{b}_1$ would be zero. In (12) we are dealing with a complex function. Upon decomposition into real and imaginary parts, necessary simplifications and taking $\operatorname{Re}\{\widetilde{\boldsymbol{w}}(t)\}$ we get

$$\mathrm{Re}\{\widetilde{\boldsymbol{w}}(t)\} = -c_1(\boldsymbol{u}_1 \cos\beta\, t + \boldsymbol{u}_2 \sin\beta\, t) \tag{13}$$

Equation in (13) represents parametric equations of an ellipse centered at the origin, hence $\liminf_{t\to\infty} \| e^{-\alpha t} \boldsymbol{w}(t) \| > 0$ holds for any $c_1 \neq 0$. It can be shown that the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2$ in (13) are linearly independent provided that $\boldsymbol{A} + \boldsymbol{B}$ is skew-symmetric. The case $0 < \omega = \gamma > \alpha$ leads to similar results; the real solution is again an ellipse centered at the origin.

3.2.1. The case of pure imaginary eigenvalues

In this case $\mu_{1,2} = \pm\,\mathrm{i}\beta$ and $\lambda_{1,2} = \pm\,\mathrm{i}\delta$. We show that in this case no solution has the property $\liminf_{t\to\infty} \| e^{-\omega t} \boldsymbol{w}(t) \| > 0$. We have

$$\overline{\boldsymbol{w}}(t) = \boldsymbol{w}(t) = \sum_{k=1}^{2} \left(\sum_{i=1}^{2} \frac{e^{\lambda_i t} - e^{\mu_k t}}{\lambda_i - \mu_k} d_{ik} c_i \right) \boldsymbol{a}_k =$$

$$= \left(\frac{e^{\mathrm{i}\delta t} - e^{\mathrm{i}\beta t}}{\mathrm{i}(\delta - \beta)} d_{11} c_1 + \frac{e^{-\mathrm{i}\delta t} - e^{\mathrm{i}\beta t}}{\mathrm{i}(-\delta - \beta)} d_{21} c_2 \right) \boldsymbol{a}_1 + \left(\frac{e^{\mathrm{i}\delta t} - e^{-\mathrm{i}\beta t}}{\mathrm{i}(\delta + \beta)} d_{12} c_1 + \frac{e^{-\mathrm{i}\delta t} - e^{-\mathrm{i}\beta t}}{\mathrm{i}(-\delta + \beta)} d_{22} c_2 \right) \boldsymbol{a}_2 \tag{14}$$

Now for $e^{\mathrm{i}\delta t} = e^{\mathrm{i}\beta t} \wedge e^{\mathrm{i}\delta t} = e^{-\mathrm{i}\beta t}$ we get $\overline{\boldsymbol{w}}(t) = 0$. This pair of equalities is equivalent to $\cos\delta\, t - \cos\beta\, t = 0 \quad \wedge \quad \sin\delta\, t \pm \sin\beta\, t = 0$. If $\dfrac{\beta}{\delta}$ is rational, it is possible to find such $t_0 > 0$, for which $\overline{\boldsymbol{w}}(t_0) = 0$, so the curve periodically returns to the origin. From (14) we see that for $t = 0$ the curve starts from the origin and for $t_0 = \dfrac{2k_0\pi}{\beta + \delta} \in \mathbf{Z}$ it returns back. So $\overline{\boldsymbol{w}}(t)$ is a closed curve passing through the origin and hence $\liminf_{t\to\infty} \| e^{-\omega t} \boldsymbol{w}(t) \| = 0$. In case of $\dfrac{\beta}{\delta}$ is irrational, the curve $\overline{\boldsymbol{w}}(t)$ doesn't have a finite period but it is bounded around the origin. It can be proved that for any $\varepsilon > 0$ there exists such t_ε that $\|\overline{\boldsymbol{w}}(t_\varepsilon)\| < \varepsilon$, which implies $\liminf_{t\to\infty} \| e^{-\omega t} \boldsymbol{w}(t) \| = 0$. So in case of pure imaginary eigenvalues no solution has the property $\liminf_{t\to\infty} \| e^{-\omega t} \boldsymbol{w}(t) \| > 0$.

3.3. The case of real and complex eigenvalues

When the eigenvalues of the matrices are combined, the critical cases (5) and (6) are interesting. Let $\mu_1, \mu_2 \in \mathbf{R} \wedge \lambda_{1,2} = \gamma \pm \mathrm{i}\delta$ and $\mu_1 = \gamma$. Let $c_1 \neq 0, c_2 = 0$, then

$$\overline{\boldsymbol{w}}(t) = e^{-\mu_1 t} \boldsymbol{w}(t) = \sum_{k=1}^{2} \frac{e^{(\lambda_1 - \mu_1)t} - e^{\mu_k}}{\lambda_1 - \mu_k} d_{1k} c_1\, \boldsymbol{a}_k = \frac{e^{(\lambda_1 - \mu_1)t} - 1}{\lambda_1 - \mu_1} d_{11} c_1\, \boldsymbol{a}_1 + \frac{e^{(\lambda_1 - \mu_1)t} - e^{(\mu_2 - \mu_1)}}{\lambda_1 + 3\mu_1} d_{12} c_1\, \boldsymbol{a}_2 \text{ and}$$

$$\tilde{\mathbf{w}}(t) = \frac{e^{i\delta t} - 1}{\lambda_1 - \mu_1} d_{11} c_1 \mathbf{a}_1 + \frac{e^{i\delta t}}{\lambda_1 + 3\mu_1} d_{12} c_1 \mathbf{a}_2 ,$$

where d_{11}, d_{12} are non-zero complex conjugates, otherwise the vector $\boldsymbol{b}_1$ would be zero. Since $e^{i\delta t} \neq 0$, it implies that $\tilde{\boldsymbol{w}}(t) \neq 0$ and $\liminf_{t\to\infty} \left\| e^{-\omega t} \boldsymbol{w}(t) \right\| > 0$. In this case $\mathrm{Re}\{\tilde{\boldsymbol{w}}(t)\}$ represents parametric equations of an ellipse shifted in direction of the vector $\boldsymbol{a}_1$. The case $\lambda_1, \lambda_2 \in \mathbf{R} \wedge \mu_{1,2} = \alpha \pm i\beta$ and $\lambda_1 = \alpha$ is for $c_1 = 0, c_2 \neq 0$ trivial and yields the same result as in (13).

3 Conclusion

The abovementioned analysis in two-dimensional space can be interpreted as follows: Let's have a Lie group with left invariant metric. Each geodetic curve is associated with two matrices with a skew-symmetric sum. If the eigenvalues of the matrices are not pure imaginary and $\liminf_{t\to\infty} \left\| e^{-\omega t} \boldsymbol{w}(t) \right\| > 0$ is true, it is possible to find a system of geodesic curves which have a common origin with a given geodesic curve and are exponentially unstable.

4 References

[1] GROMOLL, D. – KLINGBERG, W. – MEYER, W.: *Rimanova geometria v celom*, Moskva, Mir, 1971 (ruský preklad), 344 s.

[2] MALIČKÝ, P.: *Jacobi fields on Lie groups with a left invariant Riemannian metrics,* (zaslané do Geometrie dedicada)

[3] CHAVEL. I.: *Riemannian Geometry*: *A Modern Introduction,* Cambridge University Press,Cambridge, 366 s.

[4] GOODE, S. : *Differential Equations and Linear Algebra* Prentice Hall : New Jersey, 2000, 703 s., ISBN 0-13-263757-X

[5] DEMLOVÁ, M. – NAGY, J. : *Algebra,* SNTL : Praha, 1982, 192 s.

[6] GREGUŠ, M. – ŠVEC, M. – ŠEDA, V. : *Obyčajné diferenciálne rovnice* Alfa : Bratislava, 1985, 376 s.

Current address

Norbert Kecskés, Mgr.
Department of Mathematics,
Faculty of Economics and Management, Slovak University of Agriculture in Nitra,
Tr. Andreja Hlinku 2, 949 76 Nitra, Slovak Republic
e-mail: Norbert.Kecskes@fem.uniag.sk

Journal of Applied Mathematics

volume III (2010), number II

COMPUTER SIMULATIONS OF LINEAR STOCHASTIC DIFFERENTIAL EQUATIONS

KOLÁŘOVÁ Edita, (CZ)

Abstract. This paper deals with linear scalar stochastic differential equations. We present three examples, find their analytic solutions using the Itô formula and present the numerical simulations of the solutions. We also compute the confidence intervals for the solutions of these equations.

Key words and phrases. Stochastic differential equation, Itô formula, numerical simulation, confidence interval.

Mathematics Subject Classification. 60H10.

1 Introduction

Stochastic differential equations (SDEs) describe physical systems by taking into account some randomness of the system. A general scalar SDE has the form

$$\mathrm{d}X(t) = f(t, X(t))\,\mathrm{d}t + g(t, X(t))\,\mathrm{d}W(t),\ \ X(0) = X_0,$$

where $f : \langle 0, T\rangle \times \mathbb{R} \to \mathbb{R}$ is the drift coefficient and $g : \langle 0, T\rangle \times \mathbb{R} \to \mathbb{R}$ is the diffusion coefficient. $W(t)$ is the so called Wiener process, a stochastic process representing the noise. (A stochastic process $W(t)$ is called the Wiener process if it has independent increments, $W(0) = 0$ and $W(t) - W(s)$ distributed $N(0, t - s),\ 0 \le s < t$). We can represent the SDE in the integral form

$$X(t) = X(t_0) + \int_{t_0}^{t} f(s, X(s))\,\mathrm{d}s + \int_{t_0}^{t} g(s, X(s))\,\mathrm{d}W(s), \tag{1}$$

where the first integral is an ordinary Riemann integral. Since the sample paths of a Wiener process do not have bounded variation on any time interval, the second integral cannot be a

Riemann-Stieljtes integral. K. Itô proposed a way to overcome this difficulty with the definition of a new type of integral, a stochastic integral which is now called the Itô integral (see [6]). The solution of a stochastic differential equation is a stochastic process.
Although the Itô integral has some very convenient properties, the usual chain rule of classical calculus doesn't hold. Instead, the appropriate stochastic chain rule, known as Itô formula, contains an additional term, which, roughly speaking, is due to the fact that the square of the stochastic differential $(\mathrm{d}W(t))^2$ is equal to $\mathrm{d}t$.

The 1-dimensional Itô formula. Let the stochastic process $X(t)$ be a solution of the stochastic differential equation $\mathrm{d}X(t) = f(t, X(t))\,\mathrm{d}t + g(t, X(t))\,\mathrm{d}W(t)$ for some suitable functions f, g (see [4], p.44). Let $h(t,x) : (0,\infty) \times \mathbf{R} \to \mathbf{R}$ be a twice continuously differentiable function. Then

$$Y(t) = h(t, X(t))$$

is a stochastic process, for which

$$dY(t) = \frac{\partial h}{\partial t}(t, X(t))\,\mathrm{d}t + \frac{\partial h}{\partial x}(t, X(t))\,\mathrm{d}X(t) + \frac{1}{2}\,\frac{\partial^2 h}{\partial x^2}(t, X(t))(\,\mathrm{d}X(t))^2,$$

where $(\,\mathrm{d}X(t))^2 = (\,\mathrm{d}X(t)) \cdot (\,\mathrm{d}X(t))$ is computed according to the rules

$$\mathrm{d}t \cdot\ \mathrm{d}t =\ \mathrm{d}t \cdot\ \mathrm{d}W(t) =\ \mathrm{d}W(t) \cdot\ \mathrm{d}t = 0, \quad \mathrm{d}W(t) \cdot\ \mathrm{d}W(t) =\ \mathrm{d}t.$$

2 Linear SDEs

The general form of a scalar linear Itô stochastic differential equation is

$$\mathrm{d}X(t) = \big(a_1(t)X(t) + a_2(t)\big)\,\mathrm{d}t + \big(b_1(t)X(t) + b_2(t)\big)\,\mathrm{d}W(t),\ X(0) = X_0 \tag{2}$$

where the coefficients $a_1(t), a_2(t), b_1(t), b_2(t)$ are functions of time or constants. If $a_2(t) \equiv 0$ and $b_2(t) \equiv 0$, the equation (2) reduces to the homogeneous bilinear Itô SDE. A general solution of a linear stochastic differential equation, like in the case of a deterministic linear differential equation, can be determined explicitly with the help of an integrating factor or a fundamental solution of an associated homogeneous differential equation.

2.1 Moment equations

Let us suppose, that the initial condition $X(0) = X_0$ has a finite second moment, then we can compute the first and second moment of $X(t)$, the solution of (2). First we take the expectation of the integral form of the equation (2) and use the zero expectation property of the Itô integral. We obtain an ordinary differential equation for the expected value $m(t) = E[X(t)]$

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = a_1(t)m(t) + a_2(t),\ m(0) = E[X_0].$$

Similarly we can compute the second moment of the solution $X(t)$.
$p(t) = E[X^2(t)]$ is the unique solution of the ordinary linear differential equation

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = \Big(2a_1(t) + b_1^2(t)\Big)p(t) + 2m(t)\Big(a_2(t) + b_1(t)b_2(t)\Big) + b_2^2(t),\ p(0) = E[X_0^2],$$

where $m(t) = E[X(t)]$.

2.2 Confidence intervals

If $b_1(t) \equiv 0$, we say that (2) is a linear stochastic differential equation with additive noise. If moreover the initial condition $X(0) = X_0$ is normally distributed or constant, the solution of (2) is a Gaussian stochastic process. That means, that $X(t)$ is distributed $N(m(t), \sigma^2(t))$ for every $t \in \langle 0, T\rangle$, where $m(t) = E[X(t)]$ and $\sigma^2(t) = E[X(t)^2] - m^2(t)$. Based on the properties of the normal distribution, we can compute for any t

$$P(|X(t) - m(t)| < 1.96\ \sigma(t)) = 2\ \Phi(1.96) - 1 = 0.95,$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}s^2}\ \mathrm{d}s.$$

As we are able to compute $p(t) = E[X(t)^2]$ and $m(t) = E[X(t)]$, we can predict the interval $(m(t) - \varepsilon, m(t) + \varepsilon)$, where the trajectories of the solution lie with a probability 95 %.

3 Simulations of the stochastic solution

To simulate the solution of a stochastic differential equation numerical techniques have to be used (see [4]). The simplest numerical scheme, the stochastic Euler scheme, is based on numerical methods for ordinary differential equations.
Let us consider an equidistant discretisation of the time interval $\langle 0, T\rangle$ as

$$t_n = nh,\ \text{where } h = \frac{T}{n} = t_{n+1} - t_n = \int_{t_n}^{t_{n+1}} \mathrm{d}t$$

and the corresponding discretisation of the Wiener process as

$$\Delta W_n = W(t_{n+1}) - W(t_n) = \int_{t_n}^{t_{n+1}} \mathrm{d}W(s).$$

To be able to apply any stochastic numerical scheme, first we have to generate the random increments of W as independent Gauss random variables with mean $E[\Delta W_n] = 0$ and $E[(\Delta W_n)^2] = h$.

The Euler scheme for the scalar stochastic differential equation (1) has the form

$$X_{n+1} = X_n + f(t_n, \mathbf{X}_n)h + g(t_n, \mathbf{X}_n)\Delta W_n.$$

For measuring the accuracy of a numerical solution to an SDE we use the strong order of convergence. We say that a stochastic numerical scheme converges with strong order γ, ($\gamma > 0$) if there exist real constants $K > 0$ and $\delta > 0$, so that

$$E[|X_T - X_T^h|] \leq K h^\gamma, \quad h \in (0, \delta)$$

where the numerical solution is denoted by X_T^h.
The Euler scheme converges with strong order $\gamma = \frac{1}{2}$.

4 Some examples

Example 1. We want to solve the following linear Itô stochastic differential equation with constant coefficients ($a_1(t) = A$, $a_2(t) = C$, $b_2(t) = B$; $A < 0$, $B, C > 0$) and with additive noise ($b_1(t) \equiv 0$):

$$\mathrm{d}X(t) = \big(AX(t) + C\big)\,\mathrm{d}t + B\,\mathrm{d}W(t),\ X(0) = 0. \tag{3}$$

We find the solution of (3) using the Itô calculus. Denote $h(t, x) = e^{-At}x$, and compute its derivative at point $(t, X(t))$ using the Itô formula.

$$\mathrm{d}h(t, X(t)) = \mathrm{d}\left(e^{-At}X(t)\right) = e^{-At}(-A)X(t)\,\mathrm{d}t + e^{-At}\,\mathrm{d}X(t) + 0\,(\mathrm{d}X(t))^2 =$$

$$-AX(t)e^{-At}\,\mathrm{d}t + e^{-At}\left(\big(AX(t) + C\big)\,\mathrm{d}t + B\,\mathrm{d}W(t)\right) = Ce^{-At}\,\mathrm{d}t + Be^{-At}\,\mathrm{d}W(t).$$

This in integral form gives us the solution

$$e^{-At}X(t) - X(0) = C\int_0^t e^{-As}\,\mathrm{d}s + B\int_0^t e^{-As}\,\mathrm{d}W(s),$$

and after some trivial computations we get

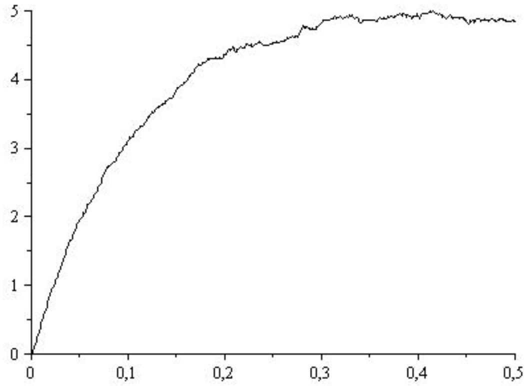
$$X(t) = \frac{C}{A}\Big(e^{At} - 1\Big) + B\int_0^t e^{A(t-s)}\,\mathrm{d}W(s). \tag{4}$$

We can compute the expectation directly from the stochastic solution (4) and we get

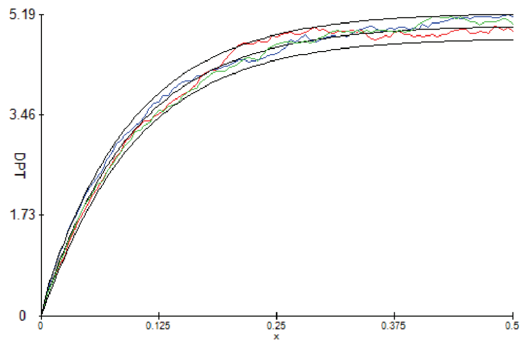
$$m(t) = \frac{C}{A}\Big(e^{At} - 1\Big).$$

The second moment of the solution is the solution of the ordinary first order linear differential equation

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = 2\,A\,p(t) + 2\,C\,m(t) + B^2,\ p(0) = 0,$$



One sample path of the solution



Some trajectories of the solution and the 95 % prediction interval

where $m(t) = \frac{C}{A}\left(e^{At} - 1\right)$.

Example 2. Let us solve the linear Itô SDE:

$$\mathrm{d}X(t) = -X(t)\,\mathrm{d}t + e^{-t}\,\mathrm{d}W(t),\ X(0) = 0. \tag{5}$$

Denote $h(t, x) = e^t x$, and compute the derivative of this function at point $(t, X(t))$.

$$\mathrm{d}h(t, X(t)) = \mathrm{d}\left(e^t X(t)\right) = e^t X(t)\,\mathrm{d}t + e^t\,\mathrm{d}X(t) = e^t X(t)\,\mathrm{d}t + e^t\big(-X(t)\big)\,\mathrm{d}t + e^t\,\mathrm{d}W(t) = e^t\,\mathrm{d}W(t).$$

The integral form of this equation is

$$e^t X(t) - X(0) = \int_0^t \mathrm{d}W(s) \quad \Rightarrow \quad X(t) = e^{-t}\,W(t).$$

Since $X(0) = 0$, we have the expectation $E[X(t)] = 0$ and the second moment $E[X^2(t)] = te^{-2t}$.

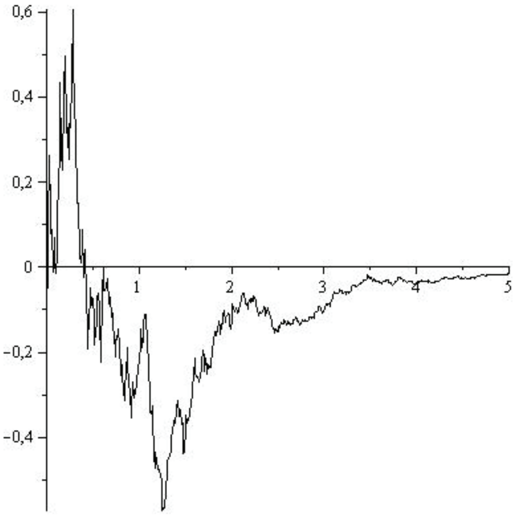
Example 3. Let us consider the linear Itô SDE:

$$\mathrm{d}X(t) = \mathrm{d}t - \frac{1}{4}X(t)\,\mathrm{d}W(t),\ X(0) = 0. \tag{6}$$

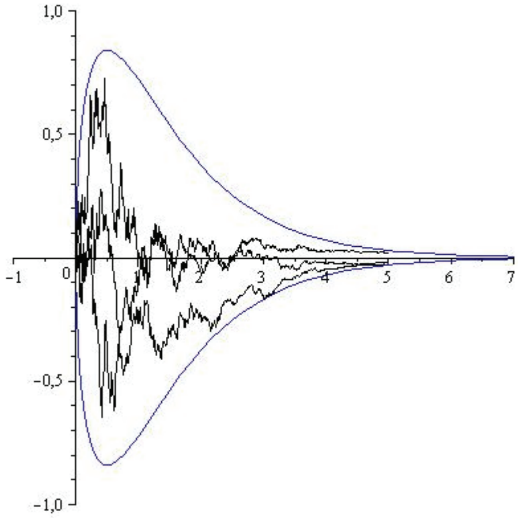
The solution of this equation is more complicated, then in the examples 1 or 2 with additive noises. We use the method, described in [6], page 77., exercise 5.16. We define a function $F(t) = e^{\frac{1}{32}t + \frac{1}{4}W(t)}$, called integrating factor, and compute $\mathrm{d}\big(F(t)X(t)\big) = F(t)\,\mathrm{d}t$. This way for the function $Y(t) = F(t)X(t)$ we got an ordinary differential equation. We solve it and then compute $X(t) = F^{-1}(t)Y(t)$. So we have

$$X(t) = \int_0^t e^{\frac{1}{32}(s-t) + \frac{1}{4}(W(s) - W(t)}\,\mathrm{d}s. \tag{7}$$

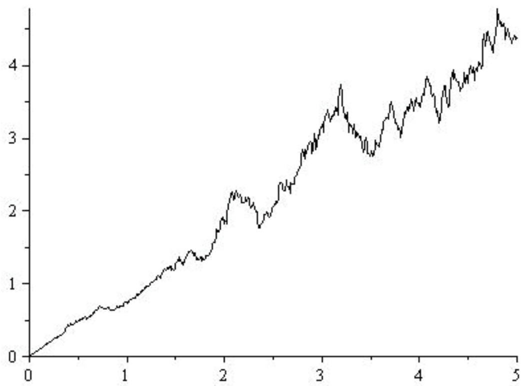
Since $X(0) = 0$, we have $E[X(t)] = t$. The solution isn't a Gaussian process and we are not able to fined the 95 % prediction interval the way, as in the examples 1 or 2. Here the Chebyshev's



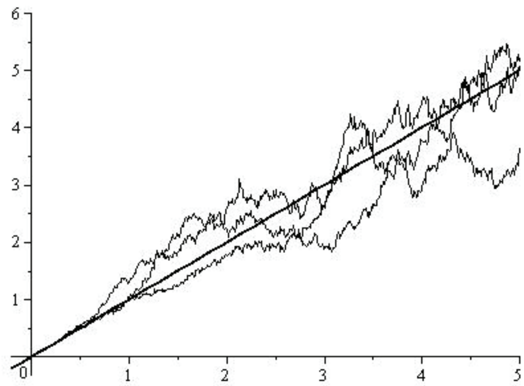
One sample path of the solution



Three trajectories of the solution and the 95 % prediction interval



One sample path of the stochastic solution



Three trajectories of the stochastic solution and its expectation

inequality can be used to predict such an interval (see. [5]).

Acknowledgement

The paper was supported by the research plan MSM 0021630516.

References

[1] ARNOLD, L.: *Stochastic Differential Equations: Theory and Applications*, John Wiley & Sons, 1974

[2] CYGANOWSKI, S., KLOEDEN, P., OMBACH, J.: *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Springer-Verlag, 2002.

[3] HALLIDAY, D., RESNICK, R., WALKER, J.: *Fundamentals of Physics*, John Wiley & Sons, 1997.

[4] KLOEDEN, P., PLATEN, E., SCHURZ, H.: *Numerical Solution Of SDE Through Computer Experiments,* Springer-Verlag, 1997.

[5] KOLAROVA, E.: *Statistical Estimates of Stochastic Solutions of RL Electrical Circuits,*the procedings od the IEEE konference ICIT 2006, Bombay, Indie, 2006.

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

Current address

Edita Kolářová, RNDr., PhD.
Brno University of Technology
Faculty of Electrical Engineering and Communication
Department of Mathematics
Technická 8, 616 00 Brno, Czech Republic
e-mail: kolara@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

AUTOREGRESSIVE MODELS OF RISK PREDICTION AND ESTIMATION USING MARKOV CHAIN APPROACH

MATVEJEVS Andrejs, (LV), SADURSKIS Karlis, (LV)

Abstract. The possibility of identifying nonlinear time series using nonparametric estimates of the conditional mean and conditional variance is studied. Most nonlinear models satisfy the assumptions needed to apply nonparametric asymptotic theory. Sampling variations of the conditional quantities are studied by simulation and explained by asymptotic arguments for the first-order nonlinear autoregressive processes. The paper deals with the identification and prediction problems of the autoregressive models of nonlinear time series using nonparametric estimates of the conditional mean and conditional variance.

Key words. Time series, Markov chain, transition probability, regression model, statistical estimation

Mathematics Subject Classification: 60J10, 62F12, 11K45

1 Introduction

Predictive modelling has attracted significant attention from the most if not all of risk management researchers. Especially the methods and algorithms of time series analysis remain an important tool and is widely applicable in financial econometrics for assessment and prediction of risk. As it well known the time series analysis is limited by the choice of regressive model and even for the simplest Markov model requires identifying multivariate distribution function. The proposal project is devoted to analysis of semiparametric stationary Markov dynamical systems used in the contemporary applications of mathematics for risk control and prediction with the help of auto regressive models. Our identification strategy uses a nonlinear auto regressive model with no classical heteroskedastic errors. Whereas the classical approach to regression analysis assumes that the form of the relationship between collections of variables is known apart from a few unknown parameters that must be estimated from the data, our approach uses more modern techniques that employ copula based nonparametric curve fitting to produce estimators as well as to assess the validity of parametric models. Owing to coupling different marginal distributions with different copula functions, copula-based auto regressive processes help us to model not only a wide variety of marginal behaviours but also such dependence properties as clusters, positive or negative tail dependence and so on. These tools of stochastic dynamical systems analysis are characterized by nonparametric marginal distributions and parametric copula functions, while the copulas capture all the scale-free temporal dependence of the processes. Given the estimators of the marginal distribution and the copula dependence parameter, one can easily construct an estimator of the transition distribution of the time series and hence estimators of nonlinear conditional moment and conditional quintile functions. Besides we intend to combine nonparametric auto regressive process estimation with Monte Carlo simulation and an empirical analysis of stochastic dynamics. Because of this flexibility, our approach may be very relevant in the finance and insurance community, where modelling and estimating the dependence structure between several univariate time series are of great interest.

2 Preliminaries

One of main problems of modern econometrics is development of time series $\{x_t, t \in \mathrm{Z}\}$ methods of analysis through regression models without a priori information about the form of dependence of the conditional expected value from its past values. Therefore it is necessary to deal with the estimation of unknown function in nonlinear difference equation of the first order with usual kind of information about the distribution law. In many applied problems of regression analysis for time series $\{x_t, t \in N\}$ already in simplest case [1,3,4,5]

$$x_t = f(x_{t-1}) + h_t, \tag{1}$$

where h_t are the uncorrelated tailings, on the average equal to the zero.

The problem of analysis of time series described higher got the name of nonparametric estimation of autoregression. The needed function can be defined through the conditional expected value $f(x_{t-1}) := E\{x_t \mid \mathcal{F}^{t-1}\}$. To use the sequence of sigma-algebra $\{\mathcal{F}^t, t \in \mathbb{Z}\}$ and conditional dispersion $\sigma_t^2 := E\{h_t^2 \mid \mathcal{F}^{t-1}\}$, tailings h_t can be present [7] in form work of "white noise" $\{\xi_t, t \in \mathbb{Z}\}$ in equation (1), (i.e. sequences of the independent identically distributed (i.i.d.) random values with zero mean and by single dispersion) and with conditional standard deviation: $h_t := \sigma_t\ \xi_t$.
This property of tailing's dispersion is called [7] as conditional heteroskedasticity and can be modulate through linear difference equations with coefficients, linearly depending on white noise (GARCH (p,q) processes).

3 Description of model

We will suppose that is observed random process of type

$$x_{n+1} = f(x_n) + \sigma_n \xi_{n+1} \quad , \tag{2}$$

ξ_n is a random error of observations, (i.i.d.) . $\mathrm{E}\{\xi_n\}=0$, $f(x_n)$ is a nonlinear function of the elements of chain .
Equation (2) can be interpreted so, that a random sequence depends on the «history». Also we can write that the conditional expected value of random variable looks like

$$\mathrm{E}\left\{x_{n+1} \mid F^n\right\} = \mathrm{E}\left\{x_{n+1} \mid x_n\right\} = \sum_y p(x_n, y) \cdot y = f(x_n) \tag{3}$$

that determines non-linearity of functional dependence x_{n+1} from x_n. The purpose of our researches is to describe the dynamics of chain $\{x_n\}$. This means to find the functional dependence $f(x_n)$, due to equation (3).
For searching for of function $f(x_n)$ we need to create separate discrete intervals of values and then on every interval we can use either least-squares or minimize specially built functional as an integral with the kernels of different form. We will consider the model of phase space discretization and presentation of him in form eventual number of no splitting areas $\{S_k, k=1,...,r\}$ which can be examined as the states of some Markov chain.

The probabilistic behaviour of a Markov chain is determined by the transition probability matrix ***P*** and a probability distribution over the initial state X_0 , if we are given X_0 and ***P***, we may want to determine the probability distribution for each random variable X_n or possibly we may be interested in the limiting distribution of X_n as $n \to \infty$,, if such a distribution exists. Within this context, if a chain is irreducible and aperiodic and thus ergodic, then there exists a unique row vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_r)$, such as

$$\lim_{m\to\infty} p_{ij}^{(m)} = \pi_j, \quad i, j = 1,2,...,r,$$

where $p_{ij}^{(m)}$ is the *(i,j)*th element of $\boldsymbol{P}^t$, $p_{ij}^{(m)} = P(X_m = j \mid X_0 = i)$
and

$$0 \le \pi_j \le 1; \qquad \sum_j \pi_j = 1, \quad j = 1,2,...,r$$

and

$$\boldsymbol{\pi} = \boldsymbol{\pi} P .$$

When these probabilities $p_{ij}^{(m)}$ are not depending on „m”, they are called as stationary probabilities and the Markov chain is homogeneous. She is fully determined by the matrix of transition probabilities.

4 Unbiased estimations of the transitions probabilities

Most of users wish to use the maximum likelihood estimations for the stationary transitions probabilities ([6], [8]). But maximum likelihood estimations are consistent, but not unbiased estimations. In this article we consider the possibility of constructing the consistent unbiased estimations of the transition probabilities of the Markov chain.

Let us consider the homogeneous Markov chain with a number of states $\{E_i, i = 1,2,...,s+1\}$ and with the positive matrix of transition probabilities $P = \|p_{ij}\|\ (i,j = 1,2,...,s+1)$, $p_{ij}>0$. Then this Markov chain will be ergodic, and there is a unique set of positive final probabilities $\{p_{ij}\}$, not depending on initial vector of probabilities $p_i(1)$.

Let us denote $\boldsymbol{m_i}$ *is* a number of appearances *of state* $\boldsymbol{E_i}$ after n tests, and m_{ij} *is* a number of transitions from the state *of* $\boldsymbol{E_i}$ to the state *of* $\boldsymbol{E_j}$ after n tests. We will count up a number *of* different chainlets of length $\boldsymbol{n}$, made from the $\boldsymbol{s+1}$ states, having the set number of transitions *of* $\boldsymbol{m_{ij}}$ and beginnings in the state $\boldsymbol{E_i}$ and endings in the state $\boldsymbol{E_j}$.

The numbers *of* $\boldsymbol{m_i}$ and $\boldsymbol{m_{ij}}$ meet the following conditions:

$$\sum_{j=1}^{s+1} m_{ij} = m_i, \quad \text{for } i \neq j_0; \qquad \sum_{j=1}^{s+1} m_{j_0 j} = m_{j_0} - 1$$

$$\sum_{i=1}^{s+1} m_{ij} = m_j, \quad \text{for } j \neq i_0; \qquad \sum_{i=1}^{s+1} m_{ii_0} = m_{i_0} - 1$$

Count of number K different chainlets of length $\boldsymbol{n}$ can be calculated on induction on the number of the states of Markov chain.

For $\boldsymbol{n=2}$:

$$K^{(2)}_{i_0 j_0} = c^{m_{i0}-m_{i0i0}-1}_{m_{i0}-1} . C^{m_{i0}-m_{i0i0}-1}_{m_{j0}-1}$$

For $\boldsymbol{n>2}$ ($i_0 \neq j_0$):

$$K = \frac{(m_{j_0}-1)\prod\limits_{i=1}^{s+1} m_j!}{m_{j_0} m_{s+1} \prod\limits_{i,j=1}^{s+1} m_{ij}!} \begin{vmatrix} 1-\frac{m_{11}}{m_1} & \dots & -\frac{m_{1i_0}}{m_1} & \dots & -\frac{m_{1j_0}}{m_1} & \dots & -\frac{m_{1s}}{m_1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{m_{i_01}}{m_{i_0}} & \dots & 1-\frac{m_{i_0i_0}}{m_{i_0}} & \dots & -\frac{m_{i_0j_0}}{m_{i_0}} & \dots & -\frac{m_{i_0s}}{m_{i_0}} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{m_{j_01}}{m_{j_0}-1} & & -\frac{m_{j_0i_0}+1}{m_{j_0}-1} & & 1-\frac{m_{j_0j_0}-1}{m_{j_0}-1} & & -\frac{m_{j_0s}}{m_{j_0}-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{m_{s1}}{m_s} & & -\frac{m_{si_0}}{m_s} & & -\frac{m_{sj_0}}{m_s} & & 1-\frac{m_{ss}}{m_s} \end{vmatrix}$$

For $n>2$ at $i_0 = j_0$ the line of determinant with this number will assume the following:

$$-\frac{m_{j_01}}{m_{j_0}-1} \quad \dots \quad 1-\frac{m_{j_01}}{m_{j_0}-1} \quad \dots \quad -\frac{m_{j_0s}}{m_{j_0}-1},$$

and other line will remain back.

So, we are continued to consider the homogeneous Markov chain with a number of states $\{E_i, i = 1,2,...,s+1\}$ and with the matrix of transition probabilities $P = \|p_{ij}\| \ (i, j = 1,2,...,s+1)$. Indeed, we assume that Markov chain is situated in state E_i at initial moment of time. Looking after the homogeneous Markov chain during ***n*** steps, will get some sequence of events $E_{i_0} \to E_{i_1} \to \cdots \to E_{j_0}$, so called as transitions trajectories. In the case of Markov chain with the fixed number of steps *of* ***n,*** the results of observations can be written down as a matrix of observations $M = \|m_{ij}\| \left(i, j = \overline{1, s+1} \right)$, where $\boldsymbol{m_{ij}}$ is a number of transitions of Markov process from the state $\boldsymbol{E_i}$ in the state $\boldsymbol{E_j}$. Thus it should be noted that the matrix of observations and initial state of Markov chain fully determines the final state of the observed process. On results of observations it is required to estimate an unknown matrix $P = \|p_{ij}\|$.

Transition probabilities $P_{i_0 j_0}^{(n)}(M)$ from the state E_{i_0} in the state E_{j_0} after ***n*** steps with the matrix of observations *of M* can be calculated by the formula.

$$P_{i_0 j_0}^{(n)}(M) = K_{i_0 j_0}(M) \prod_{k,j=1}^{s+1} p_{kj}^{m_{kj}}, \tag{4}$$

where $K_{i_0 j_0}(M)$ is number of trajectories, which come from the state E_{i_0} in the state E_{j_0} The formula (4) is true due the following probabilities..

$$P(x_0, x_1, ..., x_n) = P(x_0) \prod_{t=1}^{n} P(x_t \mid x_{t-1}) \tag{5}$$

$$P(x_0, x_1, ..., x_n \mid M) = P(x_0) \prod_{i,j} p_{ij}^{m_{ij}} \tag{6}$$

The next formula allow to calculate the number of trajectories by the following kind,

$$K_{kj}(N) = A_{jk}(S) \cdot \frac{\prod_{k=1}^{m} \omega_k !}{\prod_{k,j=1}^{m} n_{kj} !}, \tag{7}$$

where $A_{jk}(S)$ is algebraic addition of element with indexes *(j,k)* in a matrix

$S = \|s_{ij}\| \ (\boldsymbol{i, j = 1,...,m})$, where

$$s_{ij} = \begin{cases} \delta_{ij}, & if \ \omega_i = 0, \\ \delta_{ij} - \dfrac{n_{ij}}{\omega_i}, & if \ \omega_i > 0. \end{cases} \tag{8}$$

In the formula (8) δ_{ij} is the "kroneker" character, and $\omega_i = \sum_{j=1}^{n} n_{ij} -$ number of exits from the state of $\boldsymbol{E_i}$.

Another kind of formulas for estimation of matrix $P = \|p_{ij}\|$ can be concluded by the following. We will denote $R(l)$ for the set of matrix of observation $L = \|l_{kj}\| \ (k, j = 1,...,m)$, with the property $\sum_{k,j=1}^{m} l_{kj} = l$, which present realization of Markov chain with the initial state $\boldsymbol{E_i}$.

The unbiased estimation of transition probability takes place from the state *of* $\boldsymbol{E_k}$ in the state *of* $\boldsymbol{E_j}$ after $\boldsymbol{l}$ steps for the Markov chain:

$$\hat{P}_{kj}^{(l)} = \frac{\sum\limits_{L \in R(l)} P_{kj}^{(l)}(L) \cdot P_{kj_n}^{(n-l)}(N-L)}{P_{kj_n}^{(n)}}, \ N \in R(n),$$

where $\boldsymbol{j_n}$ - is the index of final state of the Markov chain.

5 Conclusions

And so, we can show two following conclusions.

Conclusion 1. The unbiased estimation of transition probability $\hat{P}_{kj}^{(l)}$ can be also calculated by the following formula:

$$\widehat{P}_{kj}^{(l)} = \frac{\sum\limits_{L \in R(l)} K_{kj}(L).K_{kj_n}(N-L)}{K_{kj}(N)}, \qquad N \in R(n).$$

Conclusion 2. The unbiased estimation of the element $\hat{p}_{kj}$ of the transition probability matrix is the following.

$$\hat{\Theta}_{kj} = \frac{K_{kj_n}(N-L)}{K_{kj_n}(N)}, \ N \in R(n),$$

Where the matrix $\boldsymbol{L}$ has the dimension m , and the element $l_{kj} = 1$, and all other elements are equal to zero.

6 References

[1] AUESTAD, B. and TJØSTHEIM, D. (1990). *Identification of nonlinear time series: First order characterization and order estimation, Biometrika* **77**: 669-687.

[2] BIERENS, H. J. (1987). *Kernel estimators of regression functions,* Advances in Econometrics, Cambridge University Press.

[3] GRANGER, C. and TEROASVIRTA, T. (1992). *Modeling Nonlinear Dynamic Relationships,* Oxford University Press, Oxford.

[4] PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*, Academic Press, NY. 1988

[5] TJØSTHEIM, D. (1990). *Nonlinear time series and Markov chains,* Advanced Applied Probability **22**: pp.587-611.

[6] BENEDETTI, J. K. (1977). *On the nonparametric estimation of regression functions,* Journal of the Royal Statistical Society, Series B **39**: 248-253.

[7] BOLLERSLEV, T. (1986). *Generalized autoregressive conditional heteroscedasticity,* Journal of Econometrics **31**: 307-327.

[8] ANDRESON, T., Goodman, L. (1967). *Statistical Inference About Markov Chain.* The Annals of Mathematical Statistics **38:** 89-110.

[9] CARKOVA, V., SWERDAN, M. O*n mean square stability of linear stochastic difference equations.* Theory of Stochastic Processes, **11**(**27**), 2005, 6 – 11.

[10] ЛУМЕЛЬСКИЙ, Я.П., ЧИЧАГОВ В.В. *Статистическое оценивание в схеме марковских случайных блужданий*.// Статистические методы / Перм. Ун-т. Пермь, 1986, 36 – 53.

Current address

Matvejevs Andrejs, Dr.sc.ing., professor;
Probability and Mathematical Statistics Department, DITF, Riga Technical University,
1/4 Meža iela, Riga,LV-1048, Latvia; Ph.+37126015121
e-mail: matv@tvnet.lv

Sadurskis Karlis, Dr. math., professor
Probability and Mathematical Statistics Department, DITF, Riga Technical University,
1/4 Meža iela, Riga,LV-1048, Latvia; Ph.+37126015121
e-mail: skarlis@latnet.lv

Journal of Applied Mathematics

volume III (2010), number II

THE STEP METHOD FOR A FUNCTIONAL-DIFFERENTIAL EQUATION FROM PRICE THEORY

MUREŞAN Anton S., (RO)

Abstract. In this paper we consider the following functional-differential equation which appears in the price theory for a single commodity market and in the dynamics of some economical systems

$$x'(t) = f(t, x(t), x(t-h), x(t+h)), t \in [a, b] \subset \mathbb{R}_+^*, \tag{1}$$

with the conditions

$$x(t) = \varphi(t), t \in [a-h, a], \quad x(t) = \psi(t), t \in [b, b+h], \tag{2}$$

where f, φ *and* ψ *are given continuous functions,* $f \in C([a, b] \times \mathbb{R}_+^3, \mathbb{R})$, $\varphi \in C[a-h, a]$, $\psi \in C[b, b+h]$ and $h > 0$.
By using the step methods *we give some new results about existence and uniqueness of the solution* x *of this Wheeler-Feynman problem,* $x \in C([a-h, a], \mathbb{R}_+) \cap C^1([a, b], \mathbb{R}_+) \cap C([b, b+h], \mathbb{R}_+)$.

Key words and phrases. functional-differential equations, Picard operator, step method, price fluctuation.

Mathematics Subject Classification. 34K05, 34K15, 90A12, 90B24, 30B62.

1 Introduction

The fluctuations of the price and supply of various commodities have been studied by many authors and have attracted the attention of economists. Some authors often attributed these fluctuations of price and supply to random factors (for example, weather for agricultural commodities). Other authors considered that economic cycling or fluctuations might be a dynamical behavior characteristic of unstable economic systems.

The possibility that economic may reflect underlying periodic or chaotic dynamic in nonlinear economic systems has been explored in various context.

The interest about the potential role of production delays in generating fluctuations in economic indicators in the techniques from dynamical systems theory has almost exclusively ignored.

But it is known that many problems from economics and biology lead to mathematical models described by functional-differential equations. For example we have the Malthus' model ($x'(t) = ax(t), x(t_0) = x_0, a > 0$), the Verhulst's model ($x'(t) = [a - bx(t)]x(t), x(t_0) = x_0, a > 0, b > 0$), or with delay model ($x'(t) = [a - bx(t-h)]x(t), x(t_0) = x_0, a > 0, b > 0, h > 0$), and so on.

The developments in nonlinear dynamics and in applied mathematics have played an important role in obtaining of many results.

In [2], (1989), J.B. Bélair and M.C. Mackey, considered a model in order to study the dynamics of price, production and consumption for a particular commodity governed by the equation

$$\frac{1}{P}\frac{dP}{dt} = f(D(P_D), S(P_S)), \tag{3}$$

where P is the function which means the price of commodity and D and S, respectively denote the *demand* and *supply functions* for this commodity. They assumed that, for market price, $P(t)$, the relative variations ($\frac{1}{P}\frac{dP(t)}{dt}$) are go-verned by the equation (3), where f is a given function (*price change function).* The arguments of functions D and S are given by P_D (*demand price)* and P_S (*supply price*), respectively, rather than simply the current market price P.

They compared this model with other models, that have been studied in economic literature, and arrived at the conclusion that the following Haldane's model

$$\frac{dp}{dt} = -Ap - B\int_0^\infty g(x)p(t-x)dx, \tag{4}$$

where p is the deviation of commodity price of equilibrium value, is a special case of the model (3).

In the paper ([10], 1989) M.C. Mackey developed a price adjustment model for a single commodity market with state dependent production and storage delays. Conditions for the equilibrium price to be stable are derived in terms of a variety of economic parameters.

A special case of the general model was considered by A.M. Farahani and E.A. Grove in [3], (1992):

$$\frac{P'(t)}{P(t)} = \frac{a}{b + P^n(t)} - \frac{cP^m(t-\tau)}{d + P^m(t-\tau)}, \quad t \geq 0, \tag{5}$$

where $a, b, c, d, \sigma, m \in \mathbb{R}_+$ and $n \geq 1$. The authors gave sufficient and necessary and sufficient conditions in which there exist positive solutions of the equation (5) that oscillate to the unique positive equilibrium solution of the problem (5) with the condition $P(t) = \varphi(t), t \in [-\tau, 0]$.

In [12] we studied a special case of fluctuation model for the price with retard of the form

$$p'(t) = p(t)\left(\frac{a}{b+p^q(t)} - \frac{cp^r(g(t))}{d+p^r(g(t))}\right) \tag{6}$$

and proved that there exists a positive, bounded, unique solution.
I. A. Rus and C. Iancu in [23] studied a more general model of the form

$$x'(t) = F(x(t), x(t-\tau))x(t), t \in \mathbb{R}, \tag{7}$$

$$x(t) = \varphi(t), t \in [-\tau, 0]. \tag{8}$$

They proved the existence and uniqueness for the solution x^* of the problem (7)+(8) and established some relations between the equilibrium solution and the coincidence points.

Some mathematical models that appear in price theory have been considered in the papers [14], [22], [26]. For other results in this field we quote here the papers of A.C. Fowler and M.C. Mackey [4], (2002) and D. Moreno [11], (2002).

The aim of this paper is to study and to present some existence and uniqueness results for the solution of following problem, which appears in dynamics of both economical and biological systems:

$$x'(t) = x(t)[D(x(t)) - S(x(t-h), x(t+h))], t \in [a,b] \subset \mathbb{R}_+^*, \tag{9}$$

$$x(t) = \varphi(t), t \in [a-h, a], \quad x(t) = \psi(t), t \in [b, b+h], \tag{10}$$

where $h > 0$, $D \in C(\mathbb{R}_+, \mathbb{R}_+)$, $S \in C(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R}_+)$, linebreak $\varphi \in C([a-h, a], \mathbb{R}_+)$, $\psi \in C([b, b+h], \mathbb{R}_+)$.

These results are obtained by using the Picard operators' technique (see I.A. Rus [18]-[20]) and an abstract model for the step methods (see I.A.Rus [21]).

2 Basic needed results from Picard and weakly Picard operators' theory

Let (X, d) be a metric space and $A : X \to X$ an operator. We denote by F_A the fixed point set of A.

Definition 2.1 *(I.A. Rus [19]) The operator A is a* ***Picard operator*** *if there exists $x^* \in X$ such that*

1) $F_A = \{x^*\}$; 2) the successive approximation sequence $(A^n(x_0))_{n\in\mathbb{N}}$ converges to x^*, for all $x_0 \in X$. □

Definition 2.2 *(I.A. Rus, [18]) The operator A is a* ***weakly Picard operator*** *if the sequence $(A^n(x_0))_{n\in\mathbb{N}}$ converges for all $x_0 \in X$ and its limit (which may depend on x_0) is a fixed point of A.* □

Theorem 2.3 *(Contraction principle) Let (X, d) be a complete metric space and $A : X \to X$ a contraction. Then A is a Picard operator.* □

Theorem 2.4 *(Fibre contraction theorem) (M.W. Hirsch and C.C. Pugh [7], I.A. Rus [20]) Let (X, d) be a metric space, (Y, ρ) be a complete metric space and $T : X \times Y \to X \times Y$. We suppose that:*

(i) $T(x, y) = (T_1(x), T_2(x, y))$;
(ii) $T_1 : X \to X$ is a weakly Picard operator;
(iii) there exists $c \in]0, 1[$ such that

$$\rho(T_2(x, y), T_2(x, z)) \leq c\rho(y, z), \tag{11}$$

for all $x \in X$ and all $y, z \in Y$.

Then the operator T is a weakly Picard operator. Moreover, if T_1 is a Picard operator, then T is a Picard operator. $\square$

3 Abstract models of step method

In the paper [21] Ioan A. Rus deals with some abstract models of step method which imply the convergence of succesive approximations sequences.

We use this results for the problem (9)+(10) to obtain existence and uniqueness of the solution with forward and backward step methods.

For these reasons we decompose our problem into two problems of the following form:

$$x'(t) = x(t)[D(x(t)) - S(x(t - h), \psi(b))], t \in [a, b], \tag{12}$$

$$\begin{aligned} x(t) &= \varphi(t), t \in [a - h, a], \\ x &\in C([a - h, b], \mathbb{R}_+) \cap C^1([a, b], \mathbb{R}_+), \end{aligned} \tag{13}$$

and

$$x'(t) = x(t)[D(x(t)) - S(\varphi(a), x(t + h))], t \in [a, b], \tag{14}$$

$$\begin{aligned} x(t) &= \psi(t), t \in [b, b + h], \\ x &\in C^1([a, b], \mathbb{R}_+) \cap C([b, b + h], \mathbb{R}_+). \end{aligned} \tag{15}$$

We suppose that the following conditions are satisfied:
(C_1) $D \in C(\mathbb{R}_+, \mathbb{R}_+)$, $S \in C(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R}_+)$, $\varphi \in C([a-h, a], \mathbb{R}_+)$, $\psi \in C([b, b+h], \mathbb{R}_+)$;
(C_2) $(\exists)$ $L_S > 0 : |S(v_1, .) - S(v_2, .)| \leq L_S|v_1 - v_2|$, $(\forall)$ $v_1, v_2 \in \mathbb{R}_+$;
(C_3) $(\exists)$ $L_D > 0 : |D(u_1) - D(u_2)| \leq L_D|u_1 - u_2|$, $(\forall)$ $u_1, u_2 \in \mathbb{R}_+$.

Let $m \in \mathbb{N}^*$ be such that:

$$a + (m - 1)h < b \quad \text{and} \quad a + mh \geq b.$$

We denote $t_{-1} := a - h$, $t_0 := a$, $t_i := a + ih$, $i = \overline{1, m - 1}$, $t_m := b$.

On $C([t_{-1}, t_m], \mathbb{R}_+)$ we consider the Bielecki norm

$$||x||_B := \max_{t_{-1} \le t \le t_m} (|x(t)| e^{-\tau |t - t_0|}),$$

and on $C([t_{i-1}, t_i], \mathbb{R}_+)$ the norm

$$||x||_{iB} := \max_{t_{i-1} \le t \le t_i} (|x(t)| e^{-\tau |t - t_{i-1}|}).$$

The equation (12) is equivalent with the fixed point equation

$$x = E_{D,S}(x), \quad x \in C([t_{-1}, t_m], \mathbb{R}_+),$$

and the problem (12)+(13) is equivalent with the fixed point equation

$$x = B_{D,S}(x), \quad x \in C([t_{-1}, t_m], \mathbb{R}_+),$$

where

$$E_{D,S}(x)(t) := \begin{cases} x(t), & t \in [t_{-1}, t_0] \\ x(t_0) + \int_{t_0}^{t} x(s)[D(x(s)) - S(x(s-h), \psi(b))]ds, \\ \qquad t \in [t_0, t_m], \end{cases}$$

and

$$B_{D,S}(x)(t) := \begin{cases} \varphi(t), & t \in [t_{-1}, t_0] \\ \varphi(t_0) + \int_{t_0}^{t} x(s)[D(x(s)) - S(x(s-h), \psi(b))]ds, \\ \qquad t \in [t_0, t_m]. \end{cases}$$

The following results are well known:

Theorem 3.1 *In the conditions* $(C_1) - (C_3)$ *we have:*
(i) the problem (12)+(13) has in $C([t_{-1}, t_m], \mathbb{R}_+)$ *a unique solution* x^* *and*

$$x^* \in C([t_{-1}, t_m], \mathbb{R}_+) \cap C^1([t_0, t_m], \mathbb{R}_+);$$

(ii) the succesive approximations sequence

$$x^{n+1}(t) := \begin{cases} \varphi(t), & t \in [t_{-1}, t_0] \\ \varphi(t_0) + \int_{t_0}^{t} x^n(s)[D(x^n(s)) - S(x^n(s-h), \psi(b))]ds, \\ \qquad t \in [t_0, t_m], \end{cases}$$

converges to x^*, *for all* $x^0 \in C([t_{-1}, t_m], \mathbb{R}_+)$;
(iii) the operator $E_{D,S}$ *is a weakly Picard operator and* $B_{D,S}$ *is a Picard operator.* □

In what follows we suppose that we are in the conditions $(C_1) - (C_3)$.
The forward steps method for the problem (12)+(13) consists in:

$$(e_0) \qquad x_0(t) = \varphi(t), \quad t \in [t_{-1}, t_0]$$

$$(e_1) \quad x_1(t) = \varphi(t_0) + \int_{t_0}^{t} x_1(s)[D(x_1(s)) - S(\varphi(s-h), \psi(b))]ds, t \in [t_0, t_1],$$

$$(e_2) \quad x_2(t) = x_1^*(t_1) + \int_{t_1}^{t} x_2(s)[D(x_2(s)) - S(x_1^*(s-h), \psi(b))]ds, t \in [t_1, t_2],$$

....

$$\begin{aligned}(e_m)\ x_m(t) &= x_{m-1}^*(t_{m-1}) + \int_{t_{m-1}}^{t} x_m(s)[D(x_m(s)) - S(x_{m-1}^*(s-h), \psi(b))]ds,\\ t &\in [t_{m-1}, t_m],\end{aligned}$$

where $x_i^* \in C([t_{i-1}, t_i], \mathbb{R}_+)$ is the unique solution of the equation $(e_i), i = \overline{1, m}$.

Remark 3.2 *We can put* x_{i-1}^n *instead of* $x_{i-1}^*, i = \overline{2, m}$, *in the conclusion (ii) of the previous theorem (see [21]).* □

So we have the following result:

Theorem 3.3 *In the conditions* $(C_1) - (C_3)$ *the problem (12)+(13) has in* $C([t_{-1}, t_m], \mathbb{R}_+)$ *a unique solution* x^*, *given by*

$$x^*(t) := \begin{cases} \varphi(t), t \in [t_{-1}, t_0], \\ x_1^*(t), t \in [t_0, t_1], \\ \dots \\ x_m^*(t), t \in [t_{m-1}, t_m], \end{cases}$$

and the functions $x_i^*, i = \overline{1, m}$, *are the limit of the successive approximations sequences*

$$x_i^{n+1}(t) := x_{i-1}^n(t_{i-1}) + \int_{t_{i-1}}^{t} x_i^n(s)[D(x_i^n(s)) - S(x_{i-1}^n(s-h), \psi(b))]ds, t \in [t_{i-1}, t_i],$$

in $(C([t_{i-1}, t_i], \mathbb{R}_+), ||.||_B), i = \overline{1, m}$. □

Now we consider the Cauchy problem (14)+(15).
Let $m \in \mathbb{N}^*$ be such that

$$b - (m-1)h > a \quad \text{and} \quad b - mh \le a.$$

We denote

$$t_0 := a, t_1 := b - (m-1)h, ..., t_m := b, t_{m+1} := b + h,$$

and let be the Banach spaces $C([t_{i-1}, t_i], \mathbb{R}_+), i = \overline{1, m+1}$.
The equation (14) is equivalent with the fixed point equation

$$x = F_{D,S}(x), x \in C([t_0, t_{m+1}], \mathbb{R}_+)$$

and the problem (14)+(15) is equivalent with the fixed point equation

$$x = G_{D,S}(x), x \in C([t_0, t_{m+1}], \mathbb{R}_+),$$

where

$$F_{D,S}(x)(t) := \begin{cases} x(t), & t \in [t_m, t_{m+1}] \\ x(t_m) + \int_{t_m}^{t} x(s)[D(x(s)) - S(\varphi(a), x(s+h))]ds, \\ & t \in [t_0, t_m] \end{cases}$$

and

$$G_{D,S}(x)(t) := \begin{cases} \psi(t), & t \in [t_m, t_{m+1}] \\ \psi(t_m) + \int_{t_m}^{t} x(s)[D(x(s)) - S(\varphi(a), x(s+h))]ds, \\ & t \in [t_0, t_m]. \end{cases}$$

The backward steps method for the problem (14)+(15) consists in the following:

$$x_{m+1}(t) = \psi(t), t \in [t_m, t_{m+1}],$$

$$\begin{aligned} x_m(t) &= \psi(t_m) + \int_{t_m}^{t} x_m(s)[D(x_m(s)) - S(\varphi(a), \psi(s+h))]ds, \\ t &\in [t_{m-1}, t_m] \end{aligned}$$

$$\begin{aligned} x_{m-1}(t) &= x_m^*(t_{m-1}) + \int_{t_{m-1}}^{t} x_{m-1}(s)[D(x_{m-1}(s)) - S(\varphi(a), x_m^*(s+h))]ds, \\ t &\in [t_{m-2}, t_{m-1}], \end{aligned}$$

....

$$\begin{aligned} x_1(t) &= x_2^*(t_1) + \int_{t_1}^{t} x_1(s)[D(x_1(s)) - S(\varphi(a), x_2^*(s+h))]ds, \\ t &\in [t_0, t_1], \end{aligned}$$

where x_{m-i}^* is the unique solution of the integral equation in the i-step.

The following results hold:

Theorem 3.4 *In the conditions* $(C_1) - (C_3)$ *we have that:*
(i) the problem (14)+(15) has in $C([t_0, t_{m+1}], \mathbb{R}_+)$ *a unique solution* x^*, *where*

$$x^*(t) := \begin{cases} \psi(t), t \in [t_m, t_{m+1}] \\ x_m^*(t), t \in [t_{m-1}, t_m] \\ \ldots \\ x_1^*(t), t \in [t_0, t_1], \end{cases}$$

(ii) the functions x_i^* *are the limits of succesive approximations sequences*

$$x_{m+1}^{n+1}(t) = \psi(t), t \in [t_m, t_{m+1}],$$

$$\begin{aligned} x_m^{n+1}(t) &= \psi(t_m) + \int_{t_m}^{t} x_m^n(s)[D(x_m^n(s)) - S(\varphi(a), \psi(s+h))]ds, \\ t &\in [t_{m-1}, t_m], \end{aligned}$$

$$\begin{aligned} x_{m-1}^{n+1}(t) &= x_m^*(t_{m-1}) + \int_{t_{m-1}}^{t} x_{m-1}^n(s)[D(x_{m-1}^n(s)) - S(\varphi(a), x_m^*(s+h))]ds, \\ t &\in [t_{m-2}, t_{m-1}], \end{aligned}$$

....

$$\begin{aligned} x_1^{n+1}(t) &= x_2^*(t_1) + \int_{t_1}^{t} x_1^n(s)[D(x_1^n(s)) - S(\varphi(a), x_2^*(s+h))]ds, \\ t &\in [t_0, t_1]. \end{aligned}$$

□

Remark 3.5 *We can put* x_{i+1}^n *instead of* $x_{i+1}^*, i = \overline{1,m}$ *in the conclusion (ii) of the previous theorem (see [21]).* □

Theorem 3.6 *In the conditions* $(C_1) - (C_3)$ *the problem (14)+(15) has in* $C([t_0, t_{m+1}], \mathbb{R}_+)$ *a unique solution* x^*,

$$x^*(t) := \begin{cases} \psi(t), t \in [t_m, t_{m+1}] \\ x_m^*(t), t \in [t_{m-1}, t_m] \\ \dots \\ x_1^*(t), t \in [t_0, t_1], \end{cases}$$

and the functions $x_i^*, i = \overline{1,m}$, *are the limits of the succesive approximations sequences*

$$x_{m+1}^{n+1}(t) = \psi(t), t \in [t_m, t_{m+1}],$$

$$\begin{aligned} x_m^{n+1}(t) &= \psi(t_m) + \int_{t_m}^{t} x_m^n(s)[D(x_m^n(s)) - S(\varphi(a), \psi(s+h))]ds, \\ t &\in [t_{m-1}, t_m], \end{aligned}$$

$$\begin{aligned} x_{m-1}^{n+1}(t) &= x_m^n(t_{m-1}) + \int_{t_{m-1}}^{t} x_{m-1}^n(s)[D(x_{m-1}^n(s)) - S(\varphi(a), x_m^n(s+h))]ds, \\ t &\in [t_{m-2}, t_{m-1}], \end{aligned}$$

....

$$\begin{aligned} x_1^{n+1}(t) &= x_2^n(t_1) + \int_{t_1}^{t} x_1^n(s)[D(x_1^n(s)) - S(\varphi(a), x_2^n(s+h))]ds, \\ t &\in [t_0, t_1]. \end{aligned}$$

□

4 Existence and uniqueness: case of naive consumer

In the paper [3] (1992), A.M. Farahani and E.A. Grove studied the following special case of a general model of the dynamics of price, productions and consumption commodity

$$x'(t) = \left[\frac{a}{b + x^{\alpha}(t)} - \frac{cx^{\beta}(t-h)}{d + x^{\beta}(t-h)}\right] x(t), t \in [0, T] \tag{16}$$

$$x(t) = \varphi(t), \quad t \in [-h, 0], \tag{17}$$

where $a, b, c, d, h, \beta \in \mathbb{R}_+$, $\alpha \in [1, \infty)$.

Using the results of Section 3 we can state the following theorems:

Theorem 4.1 *The problem $(16) + (17)$ has in $C[t_{-1}, t_m]$ a unique solution x^* and $x^* \in C[t_{-1}, t_m] \cap C^1[t_0, t_m]$. Moreover, the successive approximation sequence*

$$x^{n+1}(t) := \begin{cases} \varphi(t), \ t \in [t_{-1}, t_0] \\ \varphi(t_0) + \displaystyle\int_{t_0}^{t} \left[\frac{a}{b + (x^n(s))^{\alpha}} - \frac{c(x^n(s-h))^{\beta}}{d + (x^n(s-h))^{\beta}}\right] x^n(s)ds, \end{cases}$$

converges to x^*, for all $x_0 \in C[t_{-1}, t_m]$.

Proof. The conditions of Theorem 5 are satisfied.
The forward steps method for the problem $(16) + (17)$ consists in:

(e_0) $x_0(t) = \varphi(t)$, $t \in [t_{-1}, t_0]$

(e_1) $x_1(t) = \varphi(t_0) + \displaystyle\int_{t_0}^{t} \left[\frac{a}{b + x_1^{\alpha}(s)} - \frac{c\varphi^{\beta}(s-h)}{d + \varphi^{\beta}(s-h)}\right] x_1(s)ds$, $t \in [t_0, t_1]$

(e_2) $x_2(t) = x_1^*(t) + \displaystyle\int_{t_1}^{t} \left[\frac{a}{b + x_2^{\alpha}(s)} - \frac{cx_1^{\beta}(s-h)}{d + x_1^{*\beta}(s-h)}\right] x_2(s)ds$, $t \in [t_1, t_2]$,

....

(e_m) $x_m(t) = x_{m-1}^*(t) + \displaystyle\int_{t_{m-1}}^{t} \left[\frac{a}{b + x_m^{\alpha}(s)} - \frac{cx_{m-1}^{*\beta}(s-h)}{d + x_{m-1}^{*\beta}(s-h)}\right] x_m(s)ds$,

$t \in [t_{m-1}, t_m]$, where $x_i^* \in C[t_{i-1}, t_i]$ is the unique solution of the equation (e_i), $i = \overline{1, m}$.□

Theorem 4.2 *We have that:*

(i) the problem $(16) + (17)$ has in $C[t_{-1}, t_m]$ a unique solution x^* where

$$x^*(t) = \begin{cases} \varphi(t), & t \in [t_{-1}, t_0], \\ x_1^*(t), & t \in [t_0, t_1], \\ \dots & \\ x_m^*(t), & t \in [t_{m-1}, t_m] \end{cases}$$

(ii) the function x_i^* is the limit of the successive approximations sequence

$$x_i^{n+1}(t) := x_{i-1}^*(t_{i-1}) + \int_{t_{i-1}}^{t} \left[\frac{a}{b + (x_i^n(s))^\alpha} - \frac{c(x_{i-1}^*(s-h))^\beta}{d + (x_{i-1}^*(s-h))^\beta} \right] x_i^n(s) ds,$$

in $C([t_{i-1}, t_i], \|\cdot\|_B)$, $i = \overline{1, m}$. $\square$

Consider the problem

$$x'(t) = x(t)[D(x(t) - S(x(t-h), x(t+h))], t \in [a, b] \subset \mathbb{R}_+^*, \tag{9}$$

$$x(t) = \varphi(t), t \in [a-h, a], \quad x(t) = \psi(t), t \in [b, b+h], \tag{10}$$

where $h > 0$, $D \in C(\mathbb{R}_+, \mathbb{R}_+)$, $S \in C(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R}_+)$, $\varphi \in C([a-h, a], \mathbb{R}_+)$, $\psi \in C([b, b+h], \mathbb{R}_+)$.
Here

$$D(x(t)) = \frac{\delta}{\zeta + x^\alpha(t)},$$

$$S(x(t-h), x(t+h)) = S_1(x(t-h)) - S_2(x(t+h)),$$

$$S_1(x(t-h)) = \frac{\eta x^\beta(t-h)}{\theta + x^\beta(t-h)},$$

$$S_2(x(t+h)) = \frac{\upsilon x^\gamma(t+h)}{\omega + x^\gamma(t+h)},$$

$t \in [a, b] \subset \mathbb{R}_+^*, h > 0, \delta, \zeta, \eta, \theta, \upsilon, \omega, \beta, \gamma > 0, \alpha \geq 1.$

In the relationship of S_1 we take the relative speed of variation for supply function before the reference moment t, with the lenght $h > 0$. In the relationship of S_2 we take the relative speed of variation for supply function that can be after the same lenght $h > 0$ after the reference moment t.

In this model $E(t, x(t)) = x(t)D(x(t))$ represents the elasticity function of demand, in respect to price, when demand has a linear dependence in respect to the price. Consequently, $D(x(t))$ is the relative speed of variation for the demand function and has the economic semnification as the elasticity of monetary unity for price $(E(t, x(t))/x(t))$.

Our model (9)+(10) is more general as that studied in [12], in which appear only the retard argument, by considering a retard and an advanced argument. Thus one obtains the model which have been studied in [24]

$$x'(t) = f(t, x(t), x(t-h), x(t+h)), t \in [a, b] \subset \mathbb{R}_+^*, \tag{18}$$

with the conditions

$$x(t) = \varphi(t), t \in [a-h, a], \quad x(t) = \psi(t), t \in [b, b+h], \tag{19}$$

where f, φ and ψ are given continuous functions, $f \in C(\mathbb{R} \times \mathbb{R}_+^3, \mathbb{R})$, $\varphi \in C([a-h, a], \mathbb{R}_+)$, and $\psi \in C([b, b+h], \mathbb{R}_+), h \in \mathbb{R}_+^*$.

Using the results of previous sections we conclude that:

Theorem 4.3 *In the conditions* $(C_1) - (C_3)$ *the problem (9)+(10) has a unique solution which can be obtained by the succesive approximation method starting from any element* $x^0 \in C([a-h, b+h], \mathbb{R}_+)$. $\square$

Remark 4.4 *The new problem of naive consumer (9)+(10) has a unique solution.* $\square$

Remark 4.5 *We can consider some questions about the smoothness of the solution and some aspects relative to the stability and oscillarity of the equilibrium solution for the extended naive consumer model of the form (9)+(10).* $\square$

References

[1] BELAIR, J., MACKEY, M.C., *A model for the regulation of mammalian platelet production*, Ann. N.J. Acad. Sci. 504(1987), 280-282

[2] BÉLAIR, J., MACKEY, M. C., *Consumer memory and price fluctuations in commodity markets: An integro - differential model*, J. Dynamics Diff. Equations, 3 (1989), 299-325

[3] FARAHANI, A. M., GROVE, E. A., *A simple model for price fluctuation in a single commodity market*, Contemporary Math. 129 (1992), 97-103

[4] FOWLER, A.C., MACKEY, M.C., *Relaxation oscillations in a class of delay differential equations*, SIAM J. Appl. Math., 63, 1(2002), 299-323

[5] GOPALSAMY, K., *Stability and oscillations in delay differential equations of population dynamics*, Kluwer Acad. Publishers, London, 1992

[6] GOPALSAMY, K., *On an equation modelling haematopoiesis*, Automedica, 15(1993), 251-257

[7] HIRSCH, M. W., PUGH, C. C., *Stable manifolds and hyperbolic sets*, Proc. Symp. in Pure Math., 14 (1970), 133-163

[8] KIMMEL, M., WAZEWSKA-CZYZEWSKA, M., *Stochastic approach to the process of red cell destruction*, Zastos. Mat. 17, 2(1982), 217-225

[9] LASOTA, A., MACKEY, M.C., WAZEWSKA-CZYZEWSKA, M., *Minimizing therapeutically induced anemia*, J. Math. Biol., 13, 2(1981/82), 149-158

[10] MACKEY, M. C., *Commodity price fluctuations: price dependent delays and nonlinearities as explanatory factors*, Journal of Economic Theory, 48 (1989), no. 2, 497-509

[11] MORENO, D., *Prices, delay and the dynamics of trade*, Journal of Economic Theory 104(2002), 304-339

[12] MURESAN, A.S., *On some models of price fluctuations in a market economy*, Studia Univ. Babeş-Bolyai, Mathematica, 2(1993), 15-20

[13] MURESAN, A.S., *Mathematical models in economics governed by differential equations with deviating argument*, Bull. Appl & Comp. Math., BAM, 2000, XCIII, 221-234

[14] MURESAN, A.S., *Commodity price fluctuations with delays and nonlinearities*, Proceedings of APLIMAT Conference, 6-9 February 2007, Bratislava, PART I, 195-200

[15] MURESAN, A.S., *On some integral equations from price theory, I* Proceedings of APLIMAT Conference, 5-8 February 2008, Bratislava

[16] MURESAN, A.S., *Step method for a functional-differential equation from mathematical economics*, Carpathian J. Math. 24(2008), 356-362
[17] MURESAN, A.S., IANCU, C., *A new model of price fluctuation for a single commodity market*, Sem. Fixed Point Theory, 3(2002), 277-280
[18] RUS, I. A., *Weakly Picard mappings*, Comment Math. Univ. Carolinae, 34 (1993), no. 4, 769-773
[19] RUS, I. A., *Picard operators and applications*, Babeş - Bolyai Univ. of Cluj-Napoca, Preprint 3 (1996)
[20] RUS, I. A., *A fibre generalized contraction theorem and applications*, Mathematica Tome 41 (64), no. 1 (1999), 85-90
[21] RUS, I.A., *Abstract models of step method which imply the convergence of successive approximations*, Fixed Point Theory, 9(2008), 293-307
[22] RUS, I. A., MURESAN, A.S., *Technique of differential equations with deviating argument in economics*, I, Studia Univ. Babeş-Bolyai, Oeconomica, XXXIV, 1(1989), 65-70
[23] RUS, I.A., IANCU, C., *A functional-differential model for price fluctuations in a single commodity market*, Studia Univ. Babeş-Bolyai, Mathematica, XXXVIII, 2(1993), 9-14
[24] RUS, I.A., IANCU, C., *Wheeler-Feynman problem for mixed first order functional-differential equations,* T. Popoviciu Itinerant Seminar on Functional Equations, Approximation and Convexity, Cluj-Napoca, May 23-29 (2000), 197-200
[25] SURGIE, J. , *Oscillating solutions of scalar delay-differential equations with state dependence*, Appl. Anal. 27(1988) 217-227
[26] TARTA, A. A., *Functional - differential equations that appear in price theory*, Sem. Fixed Point Theory, Cluj-Napoca, 3 (2002), 375-380
[27] WALTHER, H.O., *Bifurcation from a heteroclimic solution in differential delay equations*, Trans. Amer. Math. Soc. 290(1985), 213-233
[28] WAZEWSKA-CZYZEWSKA, M., LASOTA, A., *Mathematical problems of the dynamics of a system of red blood cells*, (Polish), Mat. Stos., 6, 3(1976), 23-40

Current address

Anton S. Mureşan
Babeş-Bolyai University of Cluj-Napoca
Faculty of Economic Sciences and Business Administration
Department of Statistics, Forecasting and Mathematics
e-mail: anton.muresan@econ.ubbcluj.ro

Journal of Applied Mathematics

volume III (2010), number II

A FREDHOLM-VOLTERRA INTEGRO-DIFFERENTIAL EQUATION WITH LINEAR MODIFICATION OF THE ARGUMENT

MUREŞAN Viorica, (RO)

Abstract. In this paper we study existence, uniqueness and data dependence for the solution of the following problem:

$$x'(t) = f(t, x(t), x(\lambda t), \int_0^t K_1(t, s, x(s), x(\lambda s))ds, \int_0^1 K_2(t, s, x(s), x(\lambda s))ds),$$

$$t \in [0, 1],\ 0 < \lambda < 1\ ; \qquad x(0) = 0.$$

The continuity and the differentiability of solution with respect to parameter are also studied.

Key words and phrases. integro-differential equations, fixed points, Picard operators.

Mathematics Subject Classification. 34K05, 34K15, 47H10

1 Introduction

In the past fifty years several papers have been devoted to the study of initial value problems for differential and integro-differential equations with or without modifications of the arguments (see [2]-[8], [16]). Integro-differential equations and integro-differential equations of mixed type have been studied in [5], [6], [10], [12], [15]. Some singular integro-differential equations were presented in [22] and some singular Fredholm-Volterra integro-differential equations were studied in [9]. For an application of integro-differential equations to problems arising from physics see [1] and [7]. The fundamental tools used in the existence and in the existence and uniqueness proofs are essentially fixed point theorems.

In this paper we deal with the following problem:

$$\begin{aligned}
x'(t) &= f\left(t, x(t), x(\lambda t), \int_0^t K_1(t,s,x(s),x(\lambda s))ds,\right. \\
&\quad \left.\int_0^1 K_2(t,s,x(s),x(\lambda s))ds\right), \\
t &\in [0,1],\ 0<\lambda<1; \qquad x(0)=0.
\end{aligned}$$

By using the Picard operators' technique (see I.A.Rus [17],[18]), we obtain existence, uniqueness and data dependence results for the solution of the above problem. The differentiability of solution with respect to parameter is also studied.

Let (X,d) be a metric space and $A : X \to X$ an operator. We denote
$F_A := \{x \in X | A(x) = x\}$ - the fixed point set of A;
$A^0 := 1_X, A^1 := A, A^{n+1} := A \circ A^n,\ n \in \mathbb{N},\ \ P(X) = \{Y \subseteq X\ /Y \neq \varnothing\}$, $I(A) = \{Y \in P(X)/$ $A(Y) \subseteq Y\}$, $O_A(x) = \{x, A(x), A^2(x), ..., A^n(x), ...\}$ - the A-orbit of $x \in X$.

Definition 1.1 *(I.A.Rus [18]) The operator A is a* ***Picard operator*** *if there exists $x^* \in X$ such that:*

(i) $F_A = \{x^*\}$;

(ii) the sequence $(A^n(x_0))_{n\in\mathbb{N}}$ converges to x^*, for all $x_0 \in X$.

Definition 1.2 *(I.A.Rus [19]) The operator A is a* ***weakly Picard*** *operator if the sequence $(A^n(x_0))_{n\in\mathbb{N}}$ converges for all $x_0 \in X$ and its limit (which may depend on x_0) is a fixed point of A.*

The following results are useful in what follows:

Theorem 1.1. (I.A.Rus [17]) (data dependence theorem) *Let (Y,d) be a complete metric space and A, $B : Y \to Y$ two operators. We suppose that:*
(i) A is a contraction with the constant a and $F_A = \{x_A^*\}$;
(ii) B has fixed points and $x_B^* \in F_B$;
(iii) there exists $\eta > 0$ such that $d(A(x), B(x)) \leq \eta$, for all $x \in Y$.

Then $d(x_A^*, x_B^*) \leq \frac{\eta}{1-a}$.

Theorem 1.2. (I.A.Rus [20]) (the continuity with respect to parameter) *Let (X,d) be a metric space, (Λ,τ) a topological space and $A : X \times \Lambda \to X$ an operator. We suppose that:*
(i) there exists $\alpha \in (0,1)$ such that
$d(A(x,\mu), A(z,\mu)) \leq \alpha\, d(x,z)$, for all $x, z \in X$ and all $\mu \in \Lambda$;
(ii) $A(x,.) : \Lambda \to X$ is continuous, for all $x \in X$.
Then:
(a) $F_{A(\ .\ ,\ \mu\ )} = \{\ x_\mu^*\ \}$,for all $\mu \in \Lambda$;

(b) the operator $P : \Lambda \rightarrow X$, $P(\mu) = x^*_\mu$ *is continuous.*

Theorem 1.3. (Hirsch and Pugh [11], I.A.Rus [21]) (fibre contraction theorem) *Let* (X, d) *be a metric space,* (Y, ρ) *be a complete metric space and* $T : X \times Y \rightarrow X \times Y$*. We suppose that:*

(i) $T(x, y) = (T_1(x), T_2(x, y))$;

(ii) $T_1 : X \rightarrow X$ *is a weakly Picard operator*;

(iii) there exists $c \in (0, 1)$ *such that*

$\rho\,(T_2(x, y), T_2(x, z)) \leq c\,\rho(y, z)$, *for all* $x \in X$ *and all* $y, z \in Y$.

Then the operator T *is a weakly Picard operator. Moreover, if* T_1 *is a Picard operator, then* T *is a Picard operator.*

2 Existence and uniqueness

We consider a Banach space $(X, ||.||)$. Let $||.||_B$ and $||.||_C$ be, the Bielecki and the Tchebyshev norms on $C([0, 1], X)$, defined by

$$||x||_B = \max_{t\in[0,1]} ||x(t)||e^{-\tau t}, \tau \in \mathbb{R}_+,$$

respectively

$$||x||_C = \max_{t\in[0,1]} ||x(t)||.$$

We denote d_B and d_C their corresponding metrics.

We consider the following set

$C_L([0, 1], X) := \{x \in C([0, 1], X) \mid \ ||x(t_1) - x(t_2)|| \leq L|t_1 - t_2|$, for all $t_1, t_2 \in [0, 1]\ \}$, where $L > 0$.

If $d \in \{ d_B , d_C \}$, then $(C([0, 1], X), d)$ and $(C_L([0, 1], X), d)$ are complete metric spaces.

We denote $I = [0, 1]$, $D_2 = I \times I$ and $D_1 = \{(t, s) \in D_2 |\ 0 \leq s \leq t \leq 1\}$.

It is well known that $x \in C^1(I, X)$ is a solution of the problem

$$\begin{aligned} x'(t) &= f\left(t, x(t), x(\lambda t), \int_0^t K_1(t, s, x(s), x(\lambda s))ds,\right. \\ &\qquad \left.\int_0^1 K_2(t, s, x(s), x(\lambda s))ds\right), \\ t &\in [0, 1], 0 < \lambda < 1; \end{aligned} \tag{2.1}$$

$$x(0) = 0,$$

if and only if $x \in C(I, X)$ is a solution of the following integro-differential equation:

$$\begin{aligned} x(t) &= \int_0^t f\left(\xi, x(\xi), x(\lambda\xi), \int_0^\xi K_1(\xi, s, x(s), x(\lambda s))ds,\right. \\ &\left.\int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds\right) d\xi, \\ t &\in [0,1], 0 < \lambda < 1. \end{aligned} \tag{2.2}$$

In what follows we will study the equation (2.2). We suppose that:

(C$_1$) $f \in C(I \times X^4, X)$, $K_i \in C(D_i \times X \times X, \mathbb{R})$, $i = 1, 2$;

(C$_2$) there exists $L_0 > 0$ such that

$$|| f(t, u_1, u_2, u_3, u_4) - f(t, v_1, v_2, v_3, v_4)|| \leq L_0 \sum_{i=1}^{4} ||u_i - v_i||,$$

for all u_i, $v_i \in X$, $i = \overline{1,4}$ and all $t \in I$;

(C$_3$) there exists $M > 0$ such that $||f(t, u_1, u_2, u_3, u_4)|| \leq M$, for all $u_i \in X$, $i = \overline{1,4}$ and all $t \in I$;

(C$_4$) there exists $L_i > 0$, $i = 1, 2$ such that

$|| K_i(t, s, v, w) - K_i(t, s, \overline{v}, \overline{w})|| \leq L_i (|| v - \overline{v}|| + || w - \overline{w}||)$, for all $t, s \in I$ and all $v, \overline{v}, w, \overline{w} \in I$, $i = 1, 2$.

Consider the continuous operator

$$A : (C_L(I, X), || \, . \, ||_C) \rightarrow (C_L(I, X), || \, . \, ||_C)$$

defined by

$$\begin{aligned} A(x)(t) \ : &= \int_0^t f\left(\xi, x(\xi), x(\lambda\xi), \int_0^\xi K_1(\xi, s, x(s), x(\lambda s))ds,\right. \\ &\left.\int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds\right) d\xi. \end{aligned}$$

So, we obtain the following fixed point problem:

$$x = A(x).$$

We have

Theorem 2.1. *If the conditions (C_1)-(C_4) are satisfied with $M \leq L$ and*

(C_5) $2L_0(1 + L_1 + L_2) < 1$,

then the problem (2.1) *has a unique solution* x^* *in* $C_L(I,X)$ *and this solution can be obtained by the successive approximations method starting from any element of* $C_L(I,X)$.

Proof We have

$$||A(x)(t_1) - A(x)(t_2)\,|| \;\leq\; M|t_1-t_2| \;\leq L|t_1-t_2|,$$

for all $t_1,t_2 \in I$, that ensures the invariance of $C_L(I,X)$ by the operator A. By using (C_2) and (C_4) we obtain:

$$||A(x)(t) \;-\; A(z)(t)\,|| \;\leq$$

$$\leq \int_0^t ||\; f(\xi,x(\xi),x(\lambda\xi),\int_0^\xi K_1(\xi,s,x(s),x(\lambda s))ds,\int_0^1 K_2(\xi,s,x(s),x(\lambda s))ds) -$$

$$-f(\xi,z(\xi),z(\lambda\xi),\int_0^\xi K_1(\xi,s,z(s),z(\lambda s))ds,\int_0^1 K_2(\xi,s,z(s),z(\lambda s))ds)||\;d\xi \;\leq$$

$$\leq L_0\int_0^t [||x(\xi)-z(\xi)|| \;+\; ||x(\lambda\xi)-z(\lambda\xi)|| \;+$$

$$+L_1\int_0^\xi (||x(s)-z(s)|| \;+\; ||x(\lambda s)-z(\lambda s)||\;)\;ds+$$

$$+L_2\int_0^1 (||x(s)-z(s)|| \;+\; ||x(\lambda s)-z(\lambda s)||\;)\;ds]\,d\xi \leq$$

$$\leq\; 2\;L_0\;(1+L_1+L_2)\;||\;x-z\;||_C\;,$$

for all $x,z\in C_L(I,X)$ and all $t\in I$.

It follows that

$$||A(x) - \;A(z)\,||\,_C \;\;\leq\; 2\;L_0\;(1+L_1+L_2)\;||\;x-z\;||_C\;,$$

for all $x,z\in C_L(I,X)$.

By applying Contraction principle, we have that A is a Picard operator.□

Now we consider

$$A:(C_L(I,X),||\,.\,||_B) \;\;\rightarrow\;\; (C_L(I,X),||\,.\,||_B)$$

Then the condition (C_5) can be replaced by

(C_5') there exists $\tau>0$ such that

$$\frac{L_0}{\tau}\;[\;1+\frac{1}{\lambda}+\frac{L_1}{\tau}\;(1+\frac{1}{\lambda^2})+L_2\;e^\tau\;(1+\frac{1}{\lambda})\;]\;<\;1.$$

3 Data dependence

Now, we consider both (2.2) and the following equation:

$$\begin{aligned} x(t) \;&=\; \int_0^t g\left(\xi, x(\xi), x(\lambda\xi), \int_0^{\xi} K_1(\xi, s, x(s), x(\lambda s))ds,\right. \\ &\qquad \left.\int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds\right) d\xi, \\ t \;&\in\; [0,1], 0 < \lambda < 1 \end{aligned} \tag{3.1}$$

where $g \in C\ (I \times X^4, X)$ and $K_i \in C(D_i \times X^2, \mathbb{R})$, $i = 1, 2,\ 0 < \lambda < 1$ are the same as in (2.2).

We have

Theorem 3.1. *We suppose that:*
(i) all the conditions in Theorem 2.1 *are satisfied and* $x^* \in C_L(I, X)$ *is the unique solution of* (2.1);
(ii) there exists $M_1 > 0$ *such that* $||\ g(s, u_1, u_2, u_3, u_4)\ || \ \leq M_1$, *for all* $u_i \in\ X$, $i = \overline{1,4}$ *and all* $s \in I$;
(iii) $M_1 \leq L$;
(iv)

$$||\ g(s, u_1, u_2, u_3, u_4) - g(s, v_1, v_2, v_3, v_4)|| \ \leq\ L_0 \sum_{i=1}^{4} ||u_i - v_i||,$$

for all $u_i, v_i \in\ X$, $i = \overline{1,4}$ *and all* $s \in I$, *with* $L_0 > 0$ *as in* (C_2);
(v) there exists $\eta > 0$ *such that:*

$$||\ f(s, u_1, u_2, u_3, u_4) -\ g(s, u_1, u_2, u_3, u_4)|| \ \leq\ \eta,$$

for all $u_i \in\ X$, $i = \overline{1,4}$ *and all* $s \in I$.
If y^* *is the solution of the equation* (3.1), *then*

$$||\ x^* -\ y^*||_C \leq \frac{\eta}{1 - 2\ L_0\ (1 + L_1 + L_2)}.$$

Proof Consider the operators
A, $B : (C_L(I, X), ||\ .\ ||_C) \ \to (C_L(I, X), ||\ .\ ||_C)$, defined by

$$\begin{aligned} A(x)(t) \ :\ &= \int_0^t f\left(\xi, x(\xi), x(\lambda\xi), \int_0^{\xi} K_1(\xi, s, x(s), x(\lambda s))ds,\right. \\ &\left.\int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds\right) d\xi \end{aligned}$$

$$B(x)(t) \ : \ = \int_0^t g\left(\xi, x(\xi), x(\lambda\xi), \int_0^\xi K_1(\xi, s, x(s), x(\lambda s))ds, \right.$$
$$\left. \int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds\right) d\xi$$

We have
$|| \ A(x)(t) - B(x)(t) \ || \ \le \ \eta,$ for all $t \in I$ and consequently
$|| \ A(x) - B(x) \ ||_C \le \eta$. So, we apply Theorem 1.1.□

4 Continuity and differentiability with respect to parameter

We denote $J = [\alpha, \beta]$ and consider the following integro-differential equation with parameter:

$$\begin{aligned} x(t,\mu) \ = \ & \int_0^t f\left(\xi, x(\xi), x(\lambda\xi), \int_0^\xi K_1(\xi, s, x(s), x(\lambda s))ds, \right. \\ & \left. \int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds, \mu\right) d\xi, \\ & (t,\mu) \in I \times J, \ 0 < \lambda < 1 \end{aligned} \tag{4.1}$$

We suppose that:
(D$_1$) $f \in C(I \times X^4 \times J, X), K_i \in C(D_i \times X \times X, \mathbb{R}),\ i = 1, 2;$
(D$_2$) there exists $S_0 > 0$ such that

$$|| \ f(t, u_1, u_2, u_3, u_4, \mu) - f(t, v_1, v_2, v_3, v_4, \mu)|| \ \le \ S_0 \sum_{i=1}^{4} ||u_i - v_i||,$$

for all $u_i,\ v_i \in X,\ i = \overline{1,4}$ and all $t \in I, \mu \in J;$
(D$_3$) there exists $\Gamma > 0$ such that $||f(t, u_1, u_2, u_3, u_4, \mu)|| \ \le \ \Gamma,$ for all $u_i \in X, i = \overline{1,4}$ and all $t \in I, \mu \in J;$
(D$_4$) there exists $S_i > 0,\ i = 1, 2$ such that
$||K_i(t, s, v, w) - K_i(t, s, \overline{v}, \overline{w})|| \ \le \ S_i(|| \ v - \overline{v}|| \ + \ || \ w - \overline{w}||),$ for all $t, s \in I$ and all $v, \overline{v}, w, \overline{w} \in X$, $i = 1, 2.$

Here $C_L(I \times J, X) : = \ \{x \in C(I \times J, X) \mid ||x(t_1, \mu) - x(t_2, \mu)|| \ \le L \mid t_1 - t_2|$, for all t_1, $t_2 \in I$ and $\mu \in J\}$, where $L > 0$.

Consider the continuous operator

$$A : C_L(I \times J, X) \ \to \ C_L(I \times J, X)$$

defined by

$$A(x)(t,\mu) := \int_0^t f\left(\xi, x(\xi), x(\lambda\xi), \int_0^\xi K_1(\xi, s, x(s), x(\lambda s))ds, \right.$$

$$\int_0^1 K_2(\xi, s, x(s), x(\lambda s))ds, \mu\Big) d\xi.$$

We have

Theorem 4.1. *If the conditions* (D_1)*-*(D_4) *are satisfied with* $\Gamma \le L$, *and*
(D_5) $2S_0(1 + S_1 + S_2) < 1$,
then the equation (4.1) *has a unique solution* x^* *in* $C_L(I \times J, X)$ *and this solution can be obtained by the successive approximations method starting from any element of* $C_L(I \times J, X)$.

By applying Theorem 1.2 we obtain:

Theorem 4.2. *If all the conditions in Theorem 4.1 are satisfied and moreover* $f(t, u_1, u_2, u_3, u_4, .) : J \to X$ *is continuous for all* $t \in I$ *and all* $u_i \in X$, $i = \overline{1,4}$, *then the operator* $P : J \to C_L(I \times J, X)$, $P(\mu) = x^*_\mu$ *is continuous.*

Remark 4.1 *In the paper [9] has been investigated the continuous dependence of solutions with respect to a parameter for a Fredholm-Volterra integro-differential equation without modification of the argument. The method used there can be applied to (4.1) too.*

To prove the differentiability with respect to parameter we use the fibre contraction theorem (Theorem 1.3). So, we have:

Theorem 4.3. *Suppose that all the conditions in Theorem 4.1 are satisfied and moreover the following conditions hold:*
(D_6) $f(t, ., ., ., ., .) \in C^1(X^4 \times J, X)$, *for all* $t \in I$ *and* $|\frac{\partial f}{\partial u_i}(t, u_1, u_2, u_3, u_4, \mu)| \le Q$, $i = \overline{1,4}$, *for all* $t \in I, u_i \in X$, $i = \overline{1,4}$ *and all* $\mu \in J$;
(D_7) $K_i(\xi,s, ., .) \in C^1(X \times X)$, *for all* $(\xi,s) \in D_i$, $i = 1, 2$ *and* $|\frac{\partial K_i}{\partial v}(\xi, s, v, w)| \le V$, $i = 1, 2$ *and* $|\frac{\partial K_i}{\partial w}(\xi, s, v, w)| \le V$, $i = 1, 2$;
(D_8) $2Q(1 + 2V) < 1$.
Then $x^*(t, .) \in C^1(J)$, *for all* $t \in I$.

Proof. If we suppose that there exists $\frac{\partial x^*}{\partial \mu}(t, .)$ then from (4.1) we obtain

$$\frac{\partial x^*}{\partial \mu}(t, \mu) = \int_0^t \left\{ \frac{\partial f}{\partial u_1}(\omega) \frac{\partial x^*}{\partial \mu}(\xi, \mu) + \frac{\partial f}{\partial u_2}(\omega) \frac{\partial x^*}{\partial \mu}(\lambda\xi, \mu) + \right.$$

$$+\frac{\partial f}{\partial u_3}(\omega) \left[\int_0^\xi (\frac{\partial K_1}{\partial v}(\alpha) \frac{\partial x^*}{\partial \mu}(s, \mu) + \frac{\partial K_1}{\partial w}(\alpha) \frac{\partial x^*}{\partial \mu}(\lambda s, \mu))ds \right] +$$

$$+\frac{\partial f}{\partial u_4}(\omega) \left[\int_0^1 (\frac{\partial K_2}{\partial v}(\alpha) \frac{\partial x^*}{\partial \mu}(s, \mu) + \frac{\partial K_2}{\partial w}(\alpha) \frac{\partial x^*}{\partial \mu}(\lambda s, \mu))ds \right] +$$

$$+\frac{\partial f}{\partial \mu}(\omega)\Big\} d\xi$$

Here $\omega = (\xi, u_1, u_2, u_3, u_4, \mu)$, where $u_1 := x^*(\xi, \mu)$, $u_2 := x^*(\lambda\xi, \mu)$, $u_3 := \int_0^\xi K_1(\xi, s, x^*(s,\mu), x^*(\lambda s, \mu))ds$,$u_4 := \int_0^1 K_2(\xi, s, x^*(s,\mu), x^*(\lambda s, \mu))ds$ and $\alpha = (\xi, s, v, w)$,where $v := x^*(s,\mu)$, $w := x^*(\lambda s, \mu)$.

Consider the operators

$$T_1 : C_L(I \times J, X) \to C_L(I \times J, X),$$

$$\begin{aligned} T_1(x)(t,\mu) \ : \ &= \int_0^t f\left(\xi, x(\xi,\mu), x(\lambda\xi,\mu), \int_0^\xi K_1(\xi, s, x(s,\mu), x(\lambda s,\mu))ds\ ,\right. \\ &\left. \int_0^1 K_2(\xi, s, x(s,\mu), x(\lambda s,\mu))ds, \mu\right) d\xi, \end{aligned}$$

and

$$T_2 : C_L(I \times J, X) \times C_L(I \times J, X) \to C_L(I \times J, X)$$

$$T_2(x,z)(t,\mu) := \int_0^t \Big\{ \frac{\partial f}{\partial u_1}(\widetilde{\omega})z(\xi,\mu) + \frac{\partial f}{\partial u_2}(\widetilde{\omega})z(\lambda\xi,\mu) +$$

$$+\frac{\partial f}{\partial u_3}(\widetilde{\omega}) \left[\int_0^\xi (\frac{\partial K_1}{\partial v}(\widetilde{\alpha})z(s,\mu) + \frac{\partial K_1}{\partial w}(\widetilde{\alpha})z(\lambda s,\mu))ds\right] +$$

$$+\frac{\partial f}{\partial u_4}(\widetilde{\omega}) \left[\int_0^1 (\frac{\partial K_2}{\partial v}(\widetilde{\alpha})z(s,\mu) + \frac{\partial K_2}{\partial w}(\widetilde{\alpha})z(\lambda s,\mu))ds\right] +$$

$$+\frac{\partial f}{\partial \mu}(\widetilde{\omega})\Big\} d\xi$$

Here

$$\widetilde{\omega} := \left(\xi, x(\xi,\mu), x(\lambda\xi,\mu), \int_0^\xi K_1(\xi, s, x(s,\mu), x(\lambda s,\mu))ds,\right.$$

$$\left.\int_0^1 K_2(\xi, s, x(s,\mu), x(\lambda s,\mu))ds, \mu\right)$$

and $\widetilde{\alpha} := (\xi, s, x(s,\mu), x(\lambda s,\mu))$.

We have

$$\begin{aligned} ||T_2(x,z_1) - T_2(x,z_2)|| \ \leq \ & \int_0^t || \frac{\partial f}{\partial u_1}(\widetilde{\omega})(z_1(\xi,\mu) - z_2(\xi,\mu)) + \\ & + \frac{\partial f}{\partial u_2}(\widetilde{\omega})(z_1(\lambda\xi,\mu) - z_2(\lambda\xi,\mu)) + \end{aligned}$$

$$+\frac{\partial f}{\partial u_3}(\widetilde{\omega})\left[\int_0^{\xi}(\frac{\partial K_1}{\partial v}(\widetilde{\alpha})(z_1(s,\mu)-z_2(s,\mu))+\right.$$
$$\left.+\frac{\partial K_1}{\partial w}(\widetilde{\alpha})(z_1(\lambda s,\mu)-z_2(\lambda s,\mu)))ds\right]+$$
$$+\frac{\partial f}{\partial u_4}(\widetilde{\omega})\left[\int_0^{1}(\frac{\partial K_2}{\partial v}(\widetilde{\alpha})(z_1(s,\mu)-\right.$$
$$\left.-z_2(s,\mu))+\frac{\partial K_2}{\partial w}(\widetilde{\alpha})(z_1(\lambda s,\mu)-z_2(\lambda s,\mu))ds\right]||d\xi\leq$$
$$\leq(2Q+4QV)\ ||(z_1-z_2||,$$

for all $z_1,z_2\in C_L(I\times J,X)$ and all $x\in C_L(I\times J,X)$.

The condition (D_8) insures us that T_2 is a contraction with respect to the second argument.

If we take the operator

$$T:C_L(I\times J,X)\times C_L(I\times J,X)\to C_L(I\times J,X)\times C_L(I\times J,X),T=(T_1,T_2)$$

then we are in the conditions of Theorem 1.3. From this theorem we obtain that T is a Picard operator and the sequences $(x_n)_{n\in N}$, $(z_n)_{n\in N}$, where

$$x_{n+1}(t,\mu)\ :\ =\int_0^t f\left(\xi,x_n(\xi,\mu),x_n(\lambda\xi,\mu),\int_0^{\xi}K_1(\xi,s,x_n(s,\mu),x_n(\lambda s,\mu))ds,\right.$$
$$\left.\int_0^1 K_2(\xi,s,x_n(s,\mu),x_n(\lambda s,\mu))ds,\mu\right)d\xi,$$

and

$$z_{n+1}(t,\mu):=\int_0^t\left\{\frac{\partial f}{\partial u_1}(\widetilde{\omega}_n)z_n(\xi,\mu)+\frac{\partial f}{\partial u_2}(\widetilde{\omega}_n)z_n(\lambda\xi,\mu)+\right.$$
$$+\frac{\partial f}{\partial u_3}(\widetilde{\omega}_n)\left[\int_0^{\xi}(\frac{\partial K_1}{\partial v}(\widetilde{\alpha}_n)z_n(s,\mu)+\frac{\partial K_1}{\partial w}(\widetilde{\alpha}_n)z_n(\lambda s,\mu))ds\right]+$$
$$+\frac{\partial f}{\partial u_4}(\widetilde{\omega}_n)\left[\int_0^{1}(\frac{\partial K_2}{\partial v}(\widetilde{\alpha}_n)z_n(s,\mu)+\frac{\partial K_2}{\partial w}(\widetilde{\alpha}_n)z_n(\lambda s,\mu))ds\right]+$$
$$\left.+\frac{\partial f}{\partial \mu}(\widetilde{\omega}_n)\right\}d\xi$$

converge uniformly (with respect to (t,μ)) to $(x^*,z^*)\in F_T$, for all $x_0,z_0\in C_L(I\times J,X)$.

Here $\widetilde{\omega}_n$ and $\widetilde{\alpha}_n$ are the corresponding arguments:

$$\widetilde{\omega}_n\ :\ =(\xi,x_n(\xi,\mu),x_n(\lambda\xi,\mu),\int_0^{\xi}K_1(\xi,s,x_n(s,\mu),x_n(\lambda s,\mu))ds,$$
$$\int_0^1 K_2(\xi,s,x_n(s,\mu),x_n(\lambda s,\mu))ds,\mu)$$

and $\widetilde{\alpha}_n := (\xi, s, x_n(s,\mu), x_n(\lambda s,\mu))$.

If we take $x_0 = 0$, $z_0 = 0$, then $z_1(t,\mu) = \frac{\partial x_1}{\partial \mu}(t,\mu)$. By mathematical induction method we have that $z_n(t,\mu) = \frac{\partial x_n}{\partial \mu}(t,\mu)$. Thus $(x_n)_{n\in N}$ converges uniformly to x^* as $n \to \infty$ and $(\frac{\partial x_n}{\partial \mu})_{n\in\mathbb{N}}$ converges uniformly to z^* as $n \to \infty$. Using a Weierstrass argument, we conclude that $\frac{\partial x^*}{\partial \mu}$ exists and $\frac{\partial x^*}{\partial \mu} = z^*$. $\square$

References

[1] BEAR, J., *Dynamics of Fluids in Porous Media*, Elsevier, New York, 1972

[2] BENCHOHRA, M., *On initial value problem for an integro-differential equation,* Anal. Univ. "A.I. Cuza" Iasi, s. I a, Mat. XLV (1999), 297-304

[3] BYKOV, J.V., *Theory of Integro-Differential Equations*, Kirgia Gosudar Univ., Frunze, 1957 (in Russian)

[4] BYSZEWSKI, L., *Differential and Functional-Differential Problems with Nonlocal Conditions*, Kracow Univ. Techn. Monograph, vol.184,,1995 (in Polish)

[5] BYSZEWSKI, L. and LAKSHMIKANTHAM, V., *Theorem about the existence and uniqueness of a solution of a nonlocal abstract Cauchy problem in a Banach space,* Appl. Anal. 40(1991), 11-19

[6] BYSZEWSKI, L. and TABOR A., *An application of the Kuratowski measure of noncompactness to an investigation of the existence of solutions of an abstract integrodifferential problem,* Nonlin. Studies 6(1999), 111-122

[7] CONSTANTIN, A., *Topological transversality: Application to an integro-differential equation*, J. Math. Anal. Appl. 197(1996), 855-863

[8] CORDUNEANU, C., *Integral Equations and Applications,*Cambridge University Press, 1990

[9] FILIPPOVA, O. and SMARDA, Z., *Singular initial problem for Fredholm-Volterra integro-differential equations, Aplimat* - J. Appl. Math., 2(2009), no. 2, 69-74

[10] GUO, DAJUN, *Solutions of nonlinear integro-differential equations of mixed type in Banach spaces*, J. Appl. Math. Simul. 2(1989), 1-11

[11] HIRSCH, M.W. and PUGH, C.C., *Stable manifolds and hyperbolic sets,* Proc. Symp. in Pure Math, 14(1970), 133-163

[12] LAKSHMIKANTHAM, V. and LEELA, S., *Nonlinear Differential Equations in Abstract Spaces,* Pergamon Press, 1981

[13] MUREŞAN, V., *Differential Equations with Affine Modification of the Argument*, Transilvania Press, Cluj-Napoca, 1997 (in Romanian)

[14] MUREŞAN, V., *Functional-Integral Equations,* Mediamira, Cluj-Napoca, 2003

[15] MUREŞAN, V., *Existence, uniqueness and data dependence for the solutions of some integro-differential equations of mixed type in Banach space*, J. Anal. and Appl. 23(2004), no. 1, 205-216

[16] PACHPATTE, D.G. and PACHPATTE B.G., *On certain boundary value problems for nonlinear integro-differential equations,* Acta Math. Sci. 14(1994), no. 2, 226-231

[17] RUS, I.A., *Metrical Fixed Point Theorems*, Univ. of Cluj-Napoca, 1979
[18] RUS, I.A., *Picard mappings: results and problems,* Seminar on Fixed Point Theory, Babeş-Bolyai University, Faculty of Mathematics, Preprint 6(1987), 55-64
[19] RUS, I.A., *Weakly Picard mappings*, Comment. Math. Univ. Carolinae 34(1993), 769-763
[20] RUS, I.A., *Differential Equations, Integral Equations and Dynamical Systems,* Transilvania Press, Cluj- Napoca, 1996 (in Romanian)
[21] RUS, I.A., *A fibre generalized contraction theorem and applications,* Mathematica, Tome 41(64) (1999), no.1, 85-90
[22] SMARDA, Z., *On an initial value problem for singular integro-differential equations,* Demonstratio Math., XXXV(2002), no. 4, 803-811

Current address

Viorica Mureşan
Technical University of Cluj-Napoca
Faculty of Automation and Computer Science
Department of Mathematics
e-mail: vmuresan@math.utcluj.ro

Journal of Applied Mathematics

volume III (2010), number II

THE SOLOW-SWAN GROWTH MODEL WITH BOUNDED POPULATION

PŘIBYLOVÁ Lenka, (CZ), VALENTA Petr, (CZ)

Abstract. In this paper we tried to show that constant rate of population growth in classic theories of economic growth can be replaced by general bounded growth. It is obvious that labor force cannot grow exponentially in infinite time horizon because of the environmental carrying capacity. We have shown that the long run capital/labor ratio is greater in the labor bounded case than in the standard model, we also compared the short run capital/labor ratio results. As in the standard economy model, the only way how to increase the economy level in the long run is to increase the level of technology.

Key words and phrases. Solow-Swan model, population growth, stability

Mathematics Subject Classification. 91B62.

1 Introduction

The main purpose of this short paper is to introduce general growth of population instead of the constant growth rate into standard neoclassical theory of economic growth, which is more realistic and more suitable for example in the long run. In the short run the population and labor can be estimated exponentially (with constant rate), but it is quite unrealistic to expect it in the long run horizon. The population and labor tends to carrying capacity of the environment. We try to reformulate simple Solow-Swan model of economic growth using this general assumptions to labor dynamics and concretely with generalized logistic growth equation that covers most of commonly used growth equations, following the paper of Accinelli [1] with Richards population growth law. We analyse the stability of the model and compare the long run equilibrium with the short run estimation with constant labor growth rate. We also describe the solution of this modified model with Cobb-Douglas production function.

2 The Solow-Swan model

Assumptions are common

1. we consider the production function $F(K,L) \in C^2$ with properties:

   (a) it is linearly homogenous of degree one (constant returns to scale):

   $$F(\lambda F, \lambda L) = \lambda F(K,L), \ \forall \lambda, K, L > 0$$

   (b) it satisfies Inada conditions: the marginal product of capital and labor approaches infinity as capital or labor goes to 0 and approaches 0 as capital or labor goes to infinity:

   $$\lim_{K\to 0}(F_K) = \lim_{L\to 0}(F_L) \ = \infty, \quad \lim_{K\to\infty}(F_K) = \lim_{L\to\infty}(F_L) = 0. \tag{1}$$

   (c) if $K = 0$ or $L = 0 \Rightarrow F(K,L) = 0$

   (d) $\frac{\partial F}{\partial K} > 0$, $\frac{\partial F}{\partial L} > 0$, $\frac{\partial^2 F}{\partial K^2} < 0$, $\frac{\partial^2 F}{\partial L^2} < 0$

2. the capital stock changes equally to the gross investment $I = s \cdot F(K,L)$ ($s \in (0,1)$ is propensity to save) minus the capital depreciation δK

   $$\dot{K} = s \cdot F(K,L) - \delta K, \tag{2}$$

3. the labor force $L(t)$ satisfies the following properties: $L(0) = L_0 \in (0, L_\infty)$, $\dot{L}(t) > 0$, $\lim_{t\to\infty} L(t) = L_\infty$ (population is strictly increasing and bounded).

Notice that the growth rate $n(t) = \frac{\dot{L}(t)}{L(t)} > 0$ and

$$\lim_{t\to\infty} n(t) = \lim_{t\to\infty} \frac{\dot{L}(t)}{L(t)} = 0.$$

Once the population reaches the level of L_∞, which is the carrying capacity, the population cannot grow. The population growth rate is therefore equal zero. Generally, the labor dynamics can be described by an equation

$$\dot{L} = n(t)L$$

with asymptotically stable equilibrium L_∞.

3 Analysis of the reformulated model

If $k = \frac{K}{L}$ is the capital per worker then $f(k) = F(\frac{K}{L}, 1) = F(k,1)$ is the production function in the intensive form. Since

$$\frac{\dot{k}}{k} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L},$$

we get from (2) reformulated Solow-Swan model in the intensive form

$$\begin{aligned} \dot{k} = &\quad s \cdot f(k) - (\delta + n(t)) \cdot k, \\ \dot{L} = &\qquad\qquad n(t)L \end{aligned} \tag{3}$$

Note that in the standard Solow-Swan model the labor force grows exponentially, that is $n(t) = n > 0$ is constant and the equation of the motion is $\dot{k} = s \cdot f(k) - (\delta + n) \cdot k$. In this case there is a non zero globally asymptotically stable steady state $\tilde{k}$ of the first equation, that satisfies

$$\frac{sf(\tilde{k})}{\tilde{k}} = \delta + n.$$

In contrast, in the modified model, the second equation gives necessary conditions for equilibrium

$$\frac{sf(k)}{k} = \delta + n(t) \quad \text{and} \quad n(t) = 0.$$

Consequently we have the following statement for the capital dynamics:

Theorem 3.1 *The capital steady state $\hat{k}$ of the equation* (3) *satisfies*

$$\frac{sf(\hat{k})}{\hat{k}} = \delta.$$

Conditions to the production function in the intensive form guarantee existence of exactly one non-trivial equilibrium and the steady state is globally asymptotically stable.

Proof. The long-run dynamics of (3) is described by the dynamics on the attractor $L = L_\infty$, that is by the equation $\dot{k} = sf(k) - \delta k$. Consequently

$$\frac{sf(\hat{k})}{\hat{k}} = \delta$$

is satisfied for the steady state, while the left hand side is decreasing since conditions to the production function in the intensive form imply that the capital share $\frac{kf'(k)}{f(k)} \in (0, 1)$ for $k > 0$. The Inada conditions then guarantee existence and uniqueness of the non-zero equilibrium. The eigenvalue is negative, since $\lambda = sf'(\hat{k}) - \delta = s\left(f'(\hat{k}) - \frac{f(\hat{k})}{\hat{k}}\right) < 0$, the non-zero steady state is asymptotically stable, the zero steady state is unstable.

Corollary 3.2 *The long run equilibrium $\tilde{k}$ of the standard Solow-Swan model with constant labor growth rate $n > 0$ is less then the long run equilibrium $\hat{k}$ of the Solow-Swan model* (3) *with bounded labor growth.*

Proof. Conditions to the production function imply that the function $\frac{sf(k)}{k}$ is decreasing. Since from equilibrium conditions we have $\frac{sf(\tilde{k})}{\tilde{k}} = \delta + n > \frac{sf(\hat{k})}{\hat{k}}$, for the long run equilibria we get $\tilde{k} < \hat{k}$.

Example 3.3 *When the production function is of the Cobb-Douglas type*

$$F(K,L) = AK^{\alpha}L^{1-\alpha},\ 0 < \alpha < 1,$$

where A is the level of technology and we expect general non-constant growth, then the equation of motion for the modified Solow-Swan model is:

$$\dot{k} = s \cdot Ak^{\alpha} - (\delta + n(t)) \cdot k. \tag{4}$$

Equation (4) *is a Bernoulli type equation that can be transformed by the change of variables*

$$x = k^{1-\alpha}$$

into the linear equation

$$\dot{x} = a(t)x + b, \tag{5}$$

where

$$a(t) = -(1-\alpha)(\delta + n(t)) = -(1-\alpha)\left(\delta + \frac{\dot{L}(t)}{L(t)}\right), \tag{6}$$

$$b = s \cdot A \cdot (1-\alpha) > 0. \tag{7}$$

The solution of the linear differential equation (5) *is given by*

$$x(t) = e^{C(t)}(x_0 + \int_0^t be^{-C(\tau)}\ d\tau),$$

where $x_0 = \left(\frac{K_0}{L_0}\right)^{1-\alpha}$ *and*

$$C(t) = \int_0^t a(\tau)\ d\tau = -(1-\alpha)\int_0^t \delta + \frac{\dot{L}(\tau)}{L(\tau)}\ d\tau = -(1-\alpha)\delta t - (1-\alpha)\ln\frac{L(t)}{L_0}.$$

Consequently

$$\lim_{t\to\infty} C(t) = -\infty,$$

since $L(t)$ is bounded and $\alpha \in (0,1)$. The solution (3.3) *is globally asymptotically stable. As for horizontal asymptote, it is necessary that $\dot{x}(t) = 0$ for $t \to \infty$, i.e. $a(t)x(t) + b = 0$ for $t \to \infty$, in this case*

$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} -\frac{b}{a(t)} = \lim_{t\to\infty} \frac{sA(1-\alpha)}{(1-\alpha)(\delta + n(t))} = \frac{sA}{\delta}.$$

So all the solutions of (5) *have asymptote $x = \frac{sA}{\delta}$ for $t \to \infty$. Now we can substitute this back to $x = k^{1-\alpha}$ and get the value of capital/labor ratio (and capital) in the long run:*

$$\hat{k} = \left(\frac{sA}{\delta}\right)^{\frac{1}{1-\alpha}} \quad \Rightarrow \quad \widehat{K} = \left(\frac{sA}{\delta}\right)^{\frac{1}{1-\alpha}} L_{\infty}. \tag{8}$$

This is not a steady state value, because it is not the solution of equation (4) but it is a value to which k tends in a long run. In comparison with the model with constant population growth rate n, in which the long run value of capital/labor ratio is $\tilde{k}$:

$$\tilde{k} = \left(\frac{sA}{\delta + n}\right)^{\frac{1}{1-\alpha}} < \left(\frac{sA}{\delta}\right)^{\frac{1}{1-\alpha}} = \hat{k}. \tag{9}$$

Obviously using logistic growth gives also greater output/labor ratio in the long run. It is clear that when taking L_∞ constant, the only possible way how to increase long run capital/labor ratio is to increase the level of technology (in whole article arbitrarily given). This, according to properties of production function, will also ensure greater output and output per capita.

Comparing results (8) and (9) with the Theorem 1 and Corollary 2, we see they are the same.

4 Generalized logistic growth of the labor force

Following [3] we define the generalized logistic growth dynamics as

$$\dot{L} = rL^a \left(1 - \left(\frac{L}{L_\infty}\right)^b\right)^c, \tag{10}$$

where $a, b, c > 0$ are real positive parameters, $r > 0$ is the intrinsic growth rate per capita and L_∞ is a carrying capacity.

You can see that this equation for various parameters a, b, c covers a wide range of standard population growth equations as Richards (the Richards growth of labor in Solow-Swan model was already studied in [1] with corresponding results), Gompertz, Smith's and others (See [3]). Generally it has a sigmoid or concave shape with the carrying capacity L_∞. It is necessary to mention that this equation does not have analytic solution for arbitrarily chosen parameters a, b, c, but it is solvable for specific parameters (for further analysis see [3]). The relative population growth rate is obviously

$$n(t) = \frac{\dot{L}}{L} = rL^{a-1}\left(1 - \left(\frac{L}{L_\infty}\right)^b\right)^c.$$

With respect to the parameter a we should distinguish two cases:

1. for $a > 1$ we can find the maximum of $n(t)$ at L^* (that implies also existence of an inflection point L_{inf} of $L(t)$) that satisfies:

$$L^* = L_\infty \left(1 + \frac{bc}{a-1}\right)^{-\frac{1}{b}}, \; 0 < L^* < L_\infty, \tag{11}$$

while $n(t)$ is increasing to

$$n_{\max}(t) = rL_\infty^{a-1}\left(\frac{a-1}{a-1+bc}\right)^{\frac{a-1}{b}}\left(\frac{bc}{a-1+bc}\right)^c,$$

then decreasing to zero. The inflection point of $L(t)$ is given by

$$L_{\inf} = L_\infty \left(1 + \frac{bc}{a}\right)^{-\frac{1}{b}} > L^*,\ 0 < L^* < L_{\inf} < L_\infty. \tag{12}$$

2. for $a \leq 1$ $\dot{n}(t) < 0$, which means that relative population growth rate $n(t)$ is strictly decreasing for any $L \in (0, L_\infty)$, tending to zero level.

From the practical point of view this is very useful information. The dynamics of the labor equation is much more slower than the capital one (that's probably why the constant estimate of $n(t)$ is used in the standard economic theory). But if the growth rate of labor $n(t)$, $\dot{n}(t)$ respectively, is measured and viewed as a function of t, $n_{\max}$ can be estimated. We may distinguish the case before $n_{\max}$ as a case where use of the constant labor growth rate estimation will give higher $\tilde{k}$ in the short run and after $n_{\max}$ and for $a \leq 1$, where the $\tilde{k}$ will be lower. The long run value of capital/labor ratio is the maximum level never reached - a trend. Population (and labor) has already passed the inflection point ($L_{\inf} > L^*$) at both less and more developed regions (see [4]), consequently the short run capital/labor ratio estimated with the constant labor growth rate will be lower than the short and long run estimates of the reformulated model with bounded labor.

5 Conclusions

In this paper we tried to show that constant rate of population growth in classic theories of economic growth can be replaced by more general growth. It is obvious that labor force cannot grow exponentially in infinite time horizon because of the environmental carrying capacity. We have shown that the long run capital/labor ratio is greater in the labor bounded case than in the standard model, we also compared the short run capital/labor ratio results. As in the standard economy model, the only way how to increase the economy level in the long run is to increase the level of technology.

References

[1] ACCINELLI, E., BRIDAY, J.G.: *Re-formulation of the Solow economic growth model with the Richards population growth law.* Economic Working Paper at WUSTL, no. 0508006, 2005.

[2] BARRO, R.J., SALA-I-MARTIN, X.: *Economic growth*, McGraw-Hill, Inc., 1995.

[3] TSOULARIS, A., WALLACE, J.: *Analysis of logistic growth models.* Mathematical Biosciences, Vol. 179, pp. 2155., 2002.

[4] *World Population Prospects: The 2004 Revision, Volume III: Analytical Report*, United Nations Department of Economic and Social Affairs/Population Division 21, http://www.un.org/esa/population/publications/WPP2004/WPP2004_ Volume3.htm

Current address

RNDr. Lenka Přibylová, Ph.D.
Department of Mathematics and Statistics, Faculty of Science, Masaryk University Brno, Kotlářská 2, 611 37 Brno, Czech Republic, tel. number: +420549494432,
e-mail: pribylova@math.muni.cz

Bc. Petr Valenta
student of Department of Mathematics and Statistics, Faculty of Science, Masaryk University Brno, Kotlářská 2, 611 37 Brno, Czech Republic
e-mail: 175412@mail.muni.cz

Journal of Applied Mathematics

volume III (2010), number II

ASYMPTOTIC PROPERTIES OF DELAYED EXPONENTIAL OF MATRIX

SVOBODA Zdeněk, (CZ)

Abstract. Investigation of structure of the linear system of differential equations with constant delay is based on the concept of the delayed exponential of matrix. The purpose of this contribution is to find the matrix such that exponential of this has the same asymptotic properties. This problem is solved in a special case.

Key words and phrases. delayed equation, exponential matrix,

Mathematics Subject Classification. Primary 34K06; Secondary 34K25.

1 Introduction

For the investigation of the structure of solution of the systems of linear differential equations with constant delay and a constant matrix is important the concept of delayed exponential of matrix which is defined in [2]. Moreover the utilization of this concept is given also in mentioned paper and the obtained results are analogous as for systems of ordinary linear differential equations with the constant matrix. These results are obtained by "step by step" method and due to it the definition of delayed exponential matrix is given according to intervals. The aim of this contribution is to find the exponential matrix which has the asymptotical behavior like delayed exponential matrix. At the first we recall terms and the main results obtained in [2]. The delayed exponential of matrix is defined as follows:

$$e_r^{Bt} = \begin{cases} 0, & -\infty < t < -r; \\ I, & -r \le t < 0; \\ I + B\frac{t}{1!} + B^2\frac{(t-r)^2}{2!} \cdots + B^k\frac{(t-(k-1)r)^k}{k!}, & (k-1)r \le t < kr. \end{cases} \tag{1}$$

The defined matrix is together with an initial condition $e_r^{Bt} \equiv I$ for $-r \le t \le 0$ the solution of the equation

$$\dot{x}(t) = Bx(t-r). \tag{2}$$

Let the matrices A, B be permutable, i.e. $AB = BA$, then the equation $\dot{x}(t) = Ax(t) + Bx(t-r)$ has a solution satisfying the initial condition $x(t) = \varphi(t)$ for $-r \le t \le 0$ in the form

$$x(t) = e^{A(t-r)} e_r^{B_1(t-r)} \varphi(-r) + \int_{-r}^{0} e^{A(t-r-s)} e_r^{B_1(t-r-s)} e^{Ar} [\dot{\varphi}(s) - A\varphi(s)] \mathrm{d}s, \text{ where } B_1 = e^{-Ar} B.$$

The structure of solutions is in [2] studied also for the equation

$$\dot{x}(t) = Ax(t) + Bx(t-r) + f(t).$$

2 Asymptotic form

Our motivation for the next consideration are results (published in [4] and [3]) for one delayed differential equation with a constant delay

$$\dot{x}(t) = x(t-r). \tag{3}$$

In these papers it is shown that of value of unbounded solution at nr over $\exp(\lambda nr)$ has finite limit for $n \to \infty$, where λ is positive solution of semicharacteristic equation

$$\lambda = \exp(-r\lambda),$$

of the equation (3). It means that the function $\exp(\lambda t)$ is a solution of the equation (3). It is possible to interpret the positive solution λ as the function of the delay r, so

$$\lambda(r) = \left(\limsup_{n\to\infty} \sqrt[n]{A_n(r)} \right)^{-1}, \quad \text{where} \quad A_n(r) = \sum_{m=0}^{n} \frac{(n-m)^m}{m!} r^m.$$

This function is for small r (i.e. $|r| < 1/e$) analytical and has the form

$$\lambda(r) = 1 + \sum_{i=1}^{n} \frac{(-1)^n (n+1)^{n-1}}{n!} r^i.$$

In what follows, we assume that there exists a constant matrix such that the exponential of the matrix e^{Ct} has the same asymptotic properties as the matrix e_r^{Bt}. Moreover, we suppose that $\lim\limits_{t\to\infty} (e^{Ct} - e_r^{Bt}) = 0$ and that for the derivative of these matrices the analogous assertion holds too. It means that for $t = nr$ we obtain the relation $e_r^{Brn} \sim e^{Crn}$. Furthermore, we assume that for the matrices

$$e_r^{Brn} = I + \frac{Brn}{1!} + \frac{B^2 r^2 (n-1)^2}{2!} \cdots + \frac{B^n r^n}{n!}$$

there is a constant matrix C such that there exists the limit

$$\lim_{n\to\infty} \frac{e_r^{Br(n+1)}}{e_r^{Brn}} = e^{Cr}. \tag{4}$$

Lemma 2.1 *Let the constant C has at least one characteristic number with positive real part, then the exponential of matrix C i.e. e^{Ct} is the solution of the matrix equation (2).*

Proof. By using the relation (2) for differentiation of function e_r^{Bt} we obtain for a difference of derivatives of functions e_r^{Bt}, e^{Ct} the limit

$$\lim_{t\to\infty}\left(B\frac{e_r^{B(t-r)}}{e^{C(t-r)}} - Ce^{Cr}\right)e^{C(t-r)} = 0.$$

Therefore $\lim\limits_{t\to\infty} e^{Ct} \neq 0$ and $\lim\limits_{t\to\infty} e_r^{B(t-r)}\left(e^{C(t-r)}\right)^{-1} = 1$ then the matrix C is a solution of the equation $B - Ce^{Cr} = 0$ which is possible to rewrite in the form of so called semicharacteristic equation

$$C = Be^{-Cr} \tag{5}$$

and the exponential of the matrix C is a matrix solution of the equation (2).

3 Combinal identity

In 1826 A. Cauchy brought in "Exercises de Mathmatique" on the page 53 the following formula

$$\frac{(x+\alpha+n)^n - (x+n)^n}{\alpha} = \sum_{\nu=0}^{n-1}\binom{n}{\nu}(\alpha+n-\nu)^{n-\nu-1}(x+\nu)^{\nu},$$

which is known as Cauchy's formula (for more details see [1, page 274]). We want to derive the similar identity by using the well known Abel's extension of binomial theorem. This theorem can be for $\alpha \neq 0$ rewrite into the form

$$(x+\alpha)^n = \alpha\sum_{j=0}^{n}\binom{n}{j}(x-\beta j)^{i-j}(\alpha+\beta j)^{j-1}.$$

So, after the substitution $x = x - k$, $\alpha = a$, $n = k$ we have the identity

$$(x-k+a)^k = a\sum_{j=0}^{k}\binom{k}{j}(x-k-\beta j)^{k-j}(a+\beta j)^{j-1}.$$

If we put $\beta = -1$ and rearrange the sum by $j = k - j$ we get also

$$(x-k+a)^k = a\left(\sum_{j=0}^{k-1}(n-k+k-j)^j(a-k+j)^{k-j-1} + \binom{k}{k}(x-k)^k a^{-1}\right).$$

After simple adaptation of this formula we obtain the modification of the Cauchy's formula in the form:

$$\frac{(x+a-k)^k - (x-k)^k}{a} = \sum_{j=0}^{k-1}\binom{k}{j}(x-j)^j(a-k+j)^{k-j-1}. \tag{6}$$

4 Main result

In this section we study the system (2) with matrix B such that the Jordan canonical form is diagonal with real numbers λ_i, i.e. there is regular matrix D such that

$$B = D \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} D^{-1}.$$

Theorem 4.1 *Let the canonical form of matrix B is a diagonal matrix with real number λ_i satisfying the condition $|\lambda_i r| < \dfrac{1}{e}$. Then the sequence of matrices in (4) is convergent and the exponential matrix e^{Cr} has the form*

$$e^{Cr} = D \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \Lambda_n \end{pmatrix} D^{-1}, \tag{7}$$

where numbers Λ_i depend on $\lambda_i r$ and the Λ_i are analytic functions of $\lambda_i r$:

$$\Lambda_i = 1 + \lambda_i r + \sum_{i=2}^{\infty} \frac{(1-i)^{i-1}}{i!} (\lambda_i r)^i.$$

Proof.: First we note that the k-power of the matrix B is possible to write as

$$B^k = D \begin{pmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n^k \end{pmatrix} D^{-1}$$

and for e_r^{Bt} we obtain

$$e_r^{Brn} = D \begin{pmatrix} \sum\limits_{m=0}^{n} \frac{(n-m)^m}{m!}(r\lambda_1)^m & 0 & \cdots & 0 \\ 0 & \sum\limits_{m=0}^{n} \frac{(n-m)^m}{m!}(r\lambda_2)^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sum\limits_{m=0}^{n} \frac{(n-m)^m}{m!}(r\lambda_n)^m \end{pmatrix} D^{-1}.$$

Second we obtain $\left(e_r^{Br(n+1)}\right)^{-1}$ in the form

$$e_r^{-Br(n+1)} = D \begin{pmatrix} \ddots & \cdots & 0 \\ 0 & \left(\sum\limits_{m=0}^{n} \frac{(n-m)^m}{m!}(r\lambda_1)^m\right)^{-1} & 0 \\ 0 & \cdots & \ddots \end{pmatrix} D^{-1}$$

and finally using the notation for one equation we get the limit

$$\lim_{n\to\infty} \frac{e_r^{Br(n+1)}}{e_r^{Brn}} = D \begin{pmatrix} \ddots & \cdots & 0 \\ 0 & \lim\limits_{n\to\infty} \frac{A_{n+1}(\lambda_i r)}{A_n(\lambda_i r)} & 0 \\ 0 & \cdots & \ddots \end{pmatrix} D^{-1}.$$

The properties of the fraction in the diagonal can be studied by using the modification of known Cauchy's formula (6).

We express the fraction $\dfrac{A_{n+1}(\lambda_i r)}{A_n(\lambda_i r)}$ as a power series with respect to $(r\lambda_i)$ and for the first $n+1$ terms this expansion has the form $E_n(x) = 1 + (r\lambda_i) + \sum_{k=2}^{n} \dfrac{(1-k)^{k-1}}{k!}(r\lambda_i)^k$. By the notation $0^0 = 1$ we obtain $E_n(x) = \sum_{k=0}^{n} \dfrac{(1-k)^{k-1}}{k!}(r\lambda_i)^k$. This fact can be proved as follows: We rewrite the product $E_n(\lambda_i r)A_n(\lambda_i r)$ as polynomial in $(\lambda_i r)$ and for $0 \le k \le n+1$ we obtain:

$$E_n(\lambda_i r)A_n(\lambda_i r) = \cdots + (r\lambda_i)^k \sum_{j=0}^{k} \frac{(n-j)^j}{j!} \frac{(1-k+j)^{k-j-1}}{(k-j)!} + \ldots .$$

If we compare the coefficients with the same power of $(r\lambda_i)^k$, we obtain for $x = n$, $a = 1$ the identity (6):

$$(n+1-k)^k = \sum_{j=0}^{k-1} \binom{k}{j} (n-j)^j (1-k+j)^{k-j-1} + (n-k)^k.$$

Terms containing the power $(\lambda_i r)^k$ can be omitted for $k > n+1$ and for enough small $\lambda_i r$ $(\lambda_i r < 1/e)$. The assertion is proved as a limit of this equality.

Theorem 4.2 *Let the assumptions of Theorem 4.1 are satisfied, then the matrix*

$$C = D \begin{pmatrix} \widehat{\Lambda_1} & 0 & \cdots & 0 \\ 0 & \widehat{\Lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \widehat{\Lambda_n} \end{pmatrix} D^{-1}$$

is the solution of the equation (5), where

$$\widehat{\Lambda}_i = \frac{\lambda_i}{\Lambda_i} = \lambda_i \sum_{n=0}^{\infty} \frac{(-1)^n (n+1)^{n-1}}{n!} (\lambda_i r)^n \tag{8}$$

and moreover the next equality holds:

$$\lim_{n\to\infty} (e^{Crn} - e_r^{Brn}) = 0.$$

Proof. From relation (5) we obtain the matrix C substituting e^{-Cr} by the inverse matrix from the Theorem 4.1:

$$e^{-Cr} = D \begin{pmatrix} (\Lambda_1)^{-1} & 0 & \cdots & 0 \\ 0 & (\Lambda_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & (\Lambda_n)^{-1} \end{pmatrix} D^{-1}.$$

It means $\widehat{\Lambda}_i = \dfrac{\lambda_i}{\Lambda_i}$. Applying the same technique as in proof of the Theorem 4.1 we obtain the relation $\dfrac{\lambda_i}{\Lambda_i} = \lambda_i \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n (n+1)^{n-1}}{n!} x^n$, therefore we omit the calculation.

Remark 4.3 *Let us note that there exists the hypothesis that the limit* $\lim\limits_{n\to\infty} \dfrac{A_{n+1}(\lambda_i r)}{A_n(\lambda_i r)}$ *converges for the product* $\lambda_i r$ *satisfying* $\lambda_i r > -1/e$ *and we may also formulate the analogous problem for the system of delayed linear differential equations with constant coefficients as a modification of Theorem 4.1 and Theorem 4.2 where the assumption* $|\lambda_i r| < 1/e$ *is changed by* $\lambda_i r > -1/e$.

Acknowledgement

The paper was supported by grant Czech Grant Agency (Prague) no. 201/08/0469 and by the Czech Ministry of Education in the frame of projects MSM002160503

References

[1] J. KAUCKÝ *Kombinatorické identity*, VEDA vydavatel'stvo SAV Bratislava, 1973.

[2] D. YA KHUSAINOV, G. V. SHUKLIN Linear autonomous time-delay system with permutation matrices solving. *Studies of the University of Žilina, Mathematical Series* **Vol 16** (2003), 1-8.

[3] SVOBODA Z. *Remark on monotone solutions of one differential equation with constant delay.* In Proceedings of 5th interantional coference Aplimat. 5th International Conference APLIMAT. Bratislava: Faculty of Mechanical Engineering Slovak University of Technology in Bratislava, 2006, s. 335 - 340, ISBN 80-967305-5-X

[4] SVOBODA Z. *Asymptotic estimate of unbounded solutions of one retarded functional differential equation.* In Sborník příspěvků 5. konference o matematice a fyzice na vysokých školách technických s mezinárodníúčastí, 2007, s. 351 - 356, ISBN 978-80-7231-274-0

Current address

Zdeněk Svoboda, RNDr. CSc.
UMAT FEKT VUT v Brně, Technická 8, 61600 Brno, Česká republika,
tel. 420541142830
e-mail: svobodaz@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

VOLTERRA MATRIX INTEGRO-DIFFERENTIAL EQUATIONS

ŠMARDA Zdeněk, (CZ)

Abstract. In the paper conditions of the existence and uniqueness of solutions for Volterra linear matrix integro-differential equations are determined. The obtained results are proved using of the Banach fixed point theorem.

Key words and phrases. Volterra linear matrix integrodifferential equation, Banach fixed point theorem.

Mathematics Subject Classification. 45J05.

1 Introduction

Consider the following initial problem

$$x'(t) = A(t)x(t) + \int_0^t K(t,u)x(u)du + f(t), \quad x(0) = x_0, \tag{1}$$

where $A(t)$ and $K(t,u)$ are $n \times n$ continuous matrices for $t \in R^+$ and $(t,s) \in R^+ \times R^+$, $f \in C[R^+, R^n]$.

Grossman and Miller [6] defined the matrix function $R(t,s)$, called the *resolvent*, and used it for determinination of a solution of (1) in the form

$$x(t) = R(t,0)x_0 + \int_0^t R(t,s)f(s)ds.$$

They formally defined $R(t,s)$, $0 \le s \le t < \infty$

$$R(t,s) = I + \int_s^t R(t,u)\Psi(u,s)du, \tag{2}$$

where I is the identity matrix and

$$\Psi(t,s) = A(t) + \int_s^t K(t,v)dv. \tag{3}$$

They proved that $R(t,s)$ exists and is continuous for $0 \leq s \leq t$ and satisfies the equation

$$\frac{\partial}{\partial s}R(t,s) = -R(t,s)A(s) - \int_s^t R(t,u)K(u,s)du, \quad R(t,t) = I \tag{4}$$

on interval $[0,t]$ for each $t > 0$.

In 1979 Becker [1] obtained results for (1) by means of the principal matrix solution $Y(t,s)$ of the homogeneous Volterra equation

$$x'(t) = A(t)x(t) + \int_0^t K(t,u)x(u)du \tag{5}$$

Its definition looks exactly like the definition of the principal matrix solution of the homogeneous vector differential equation

$$x'(t) = A(t)x(t)$$

that is given by Hale [9]: $Z(t,s)$ is a matrix solution of () with columns that are linearly indendent such that $Z(s,s) = I$. Using $Z(t,s)$ instead of $R(t,s)$, the variation of parameters formula

$$x(t) = Z(t,0)x_0 + \int_0^t Z(t,s)f(s)ds \tag{6}$$

for (1) is a natural extension of the variation of parameters formula for the nonhomogeneous vector differential equation

$$x'(t) = A(t)x(t) + f(t).$$

The principal matrix version of the resolvent equation (4), namely,

$$\frac{\partial}{\partial s}Z(t,s) = A(t)Z(t,s) + \int_s^t K(t,u)Z(u,s)du, \quad Z(s,s) = I \tag{7}$$

has been instrumental in a number of papers for obtaining results that might not have otherwise been obtained with (4) alone.

The principal matrix solution $Z(t,s)$, the variation of parameters formula (6), and the principal matrix equation (7) are used and cited in papers Becker [2],[3], Burton [4],[5], Raffoul [10], Hino and Muramaki [7],[8] and Zhang [12].

We can also use the classical variation of parameters formula for linear differential systems to obtain an integral equation for the solutions of (1). Fot this purpose, let $Y(t)$ be a fundamental matrix solution of the equation $x'(t) = A(t)x(t)$. Now any solution of (1) with the initial function ϕ on $[t_0,\tau]$, $t_0 \geq 0$ is given

$$x(t,\tau,\phi) = Y(t)Y^{-1}(\tau)\phi(\tau) + \int_\tau^t Y(t)Y^{-1}(s)\left[\int_{t_0}^s K(s,u)x(u)du + f(s)\right]ds.$$

2 Preliminaries

Consider the operator equation

$$u = Tu, \quad u \in X, \tag{8}$$

where X is a complete metric space. Solve (8) by means of the following iteration method:

$$u_{n+1} = Tu_n, \quad n = 0, 1, \ldots, \tag{9}$$

where $u_0 \in X$. Each solution of (8) is called a *fixed point* of the operator T.

Theorem 2.1 (Banach fixed point theorem [11]). *Let (X, d) be a complete metric space $M \subseteq X$ and $T : M \to M$ be a map satisfying*

$$d(Tx, Ty) \leq kd(x, y), \quad for\ all\ x, y \in M, \tag{10}$$

where $0 \leq k < 1$ is a constant. Then, the following hold true:

(i) Existence and uniqueness. *Equation (8) has exactly one fixed point $u \in M$.*

(ii) Convergence of the iteration method. *For each given $u_0 \in M$ the sequence (u_n) constructed by (9) converges to the unique solution u of equation (8).*

(iii) Error estimates. *For all $n = 0, 1, \ldots$ we have so-called a priori error estimate*

$$d(u_n, u) \leq k^n(1 - k)^{-1}d(u_1, u_0), \tag{11}$$

and the so-called a posteriori error estimate

$$d(u_{n+1}, u) \leq k(1 - k)^{-1}d(u_{n+1}, u_n). \tag{12}$$

(iv) Rate of convergence. *For all $n = 0, 1, \ldots$ we have*

$$d(u_{n+1}, u) \leq kd(u_n, u).$$

Let $|.|$ be any vector norm in R^n. Let $|.|$ also denote the matrix norm induced by the vector norm, that is, for an $n \times n$ matrix A

$$|A| = \sup\{|Ax| : |x| \leq 1\}.$$

Let $C[a, b]$ be the vector space of continuous functions $\varphi : [a, b] \to R^n$. For a fixed real number r, let $|.|_r$ be the norm on $C[a, b]$ that is defined as follows: for $\varphi \in C[a, b]$,

$$|\varphi|_r := \sup\{|\varphi(t)|e^{-r(t-a)} : a \leq t \leq b\}.$$

Let d_r denote the induced norm metric, that is, for $\varphi, \eta \in C[a, b]$,

$$d_r(\varphi, \eta) := |\varphi - \eta|_r = \sup\{|\varphi(t) - \eta(t)|e^{-r(t-a)} : a \leq t \leq b\}. \tag{13}$$

The space $C[a, b]$ with the metric d_r is complete which we denorte by $(C[a, b], d_r)$.

Definition 2.2 *Let $x_0 \in R^n$. A solution of (1) on the interval $[0, T)$, $0 < T \leq \infty)$ with the initial value x_0 at $t = 0$ is a differentiable function $x : [0, T) \to R^n$ that satisfies (1) on $[0, T)$ and the initial condition $x(0) = x_0$.*

3 Main results

Consider

$$x'(t) = A(t)x(t) + \int_s^t K(t,u)x(u)du + f(t), \quad x(s) = x_0. \tag{14}$$

on the interval $[s, \infty)$. Integrating (14) from s to t and replacing $x(s)$ with x_0 we get

$$x(t) = x_0 + \int_s^t A(v)x(v)dv + \int_s^t \int_s^v K(v,u)x(u)dudv + \int_s^t f(v)dv. \tag{15}$$

Interchangin the order of integration in (15) we have

$$x(t) = x_0 + \int_s^t \left[A(u) + \int_u^t K(u,v)dv \right] x(u)du + \int_s^t f(u)du. \tag{16}$$

This shows that a differentiable function $x(t)$ that satisfies (14) and the initial conditions $x(s) = x_0$ also satisfies integral equation (16).
We want to solve initial problem (14) by means of the following iteration method

$$x_{n+1}(t) = x_0 + \int_s^t \left[A(u) + \int_u^t K(u,v)dv \right] x_n(u)du + \int_s^t f(u)du, \quad n = 0, 1, \dots. \tag{17}$$

Define the operator P by

$$(P\varphi)(t) := x_0 + \int_s^t \left[A(u) + \int_u^t K(u,v)dv \right] x(u)du + \int_s^t f(u)du. \tag{18}$$

Theorem 3.1 *Suppose that:*

(i) $M = \{\varphi \in C[s,T] : \varphi(s) = x_0,\ T > s\}$ *with the metric* d_r *is the complete metric space.*

(ii) The operator $P : M \to M$ *is k-contractive, where* $0 \le k < 1$.

Then the following hold true:

Initial value problem (14) *has a unique solution* $x(t),\ t \in [s, \infty)$.

The sequence (x_n) *constructed by (17) converges to* $y(x)$.

For all $n = 0, 1, \dots$ *we get the following error estimates:*

$$||x_n - x|| \le k^n (1-k)^{-1} ||x_1 - x_0||,$$

$$||x_{n+1} - x|| \le k(1-k)^{-1} ||x_{n+1} - x_n||.$$

Proof. With respect to Theorem 2.1 it is sufficient to proved that P is a contraction mapping on M. For any $\varphi, \eta \in M$ we obtain

$$\begin{aligned} |(P\varphi)(t) - (P\eta)(t)| &= \left| \int_s^t \left[A(u) + \int_u^t K(v,u)dv \right] (\varphi(u) - \eta(u))du \right| \\ &\leq \int_s^t \left[|A(u)| + \int_u^t |K(v,u)|dv \right] |\varphi(u) - \eta(u)|du. \end{aligned}$$

Since $A(t)$ and $K(t,u)$ are continuous for $s \leq u \leq t \leq T$, there is an $r > 1$ such that

$$|A(u)| + \int_u^t |K(v,u)|dv \leq r - 1.$$

For such an r

$$|(P\varphi)(t) - (P\eta)(t)| \leq \int_s^t (r-1)|\varphi(u) - \eta(u)|du.$$

Hence

$$\begin{aligned} |(P\varphi)(t) - (P\eta)(t)|e^{-r(t-s)} &\leq \int_s^t (r-1)e^{-r(t-s)+r(u-s)}|\varphi(u) - \eta(u)|e^{-r(u-s)}du. \\ &\leq |\varphi - \eta|_r \int_s^t (r-1)e^{-r(t-u)}du \leq \frac{r-1}{r}|\varphi - \eta|_r. \end{aligned} \tag{19}$$

Put

$$k = \frac{r-1}{r}.$$

Then from (19) we get

$$d_r(P\varphi, P\eta) \leq k d_r(\varphi, \eta).$$

The assertions of Theorem 3.1. follow now from Theorem 2.1.

Acknowledgement

This research has been supported by the Czech Ministry of Education in the frames of MSM002160503 Research Intention MIKROSYN New Trends in Microelectronic Systems and Nanotechnologies and MSM0021630529 Research Intention Inteligent Systems in Automatization.

References

[1] BECKER, L.C., *Stability considerations for Volterra integro-differential equations,* Ph.D. dissertation, Southern Illinois University , Carbondale, IL, 1979.

[2] BECKER, L.C., *Fuction bounds for solutions of Volterra equations and exponential asymptotic stability,* Nonlinear Anal. **28** (2008), 729-739.

[3] BECKER, L.C., *Principal matrix solutions and variation of parameters for a Volterra Integro-differential equation and its adjoint,* Electronic Journal of Qualitative Theory of Differential Equations, **14** (2006), 1-22.

[4] BURTON, T.A., *Fixed points, Volterra equations, and Becker's resolvent,* Acta Math. Hungar. **108** (2005), 261-281.

[5] BURTON, T.A., *Integral equations, Volterra equations, and the Remarkable resolvent¿ Contractions,* J. Qualitative Theory of Diff. Equ. **2** (2006), 1-17.

[6] GROSSMAN, S.I., MILLER, R.K., *Perturbation Theory for Volterra Integrodifferential Systems,* J. Differential Equations **8** (1970), 457-474.

[7] HINO,Y., MURAMAKAMI, S., *Stability properties of Linear Volterra equations,* J. Differential Equations **89** (1991), 121-137.

[8] HINO,Y., MURAMAKAMI, S., *Stabilities in linear integrodifferential equations,* lecture Notes in Num. Appl. Anal. **15** (1996), 31-46.

[9] HALE, J.K., *Ordinary Differential Equations*, John Wiley Sons, New York, 1969.

[10] Raffoul, Y.N., *Stability in neutral Nonlinear Differential Equations with Functional Delayes Using Fixed Point Theory,* Math. Comput. Modelling **40** (2004), 691-700.

[11] ZEIDLER, E., *Applied Functional Analysis,* Applied Mathematical Sciences 108, Springer-Verlag New York Berlin Heildelberg,1999.

[12] ZHANG, B., *Asymptotic stabilty criteria and integrability properties of the resolvent of Volterra and functional equations,* Funkcialaj Ekvacioj **40** (1997), 335-351.

Current address

Doc. RNDr. Zdeněk Šmarda, CSc.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology, Technická 8, 616 00 BRNO
e-mail: smarda@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

NEW ASPECTS IN PARAMETER IDENTIFICATION

AL ZOUKRA Kristine, (DE), SCHMIDT Werner H., (DE)

Abstract. We presented a paper on parameter identification in biological problems described by ODEs with parameters which are to find in order to a best approximation of measured data in APLIMAT 2007. The approximation can be in the least square norm or in the max-norm. We want to add some new results and ideas on this topic. The identification is modeled as an optimization problem as well as an optimal control problem.

1 Motivation

Many processes in natural sciences are described by a differential equation or a system of differential equations for instance the spreading of an illness or the reaction of several substances. These differential equations often include parameters, which influence the behaviour of the equation, but which are unknown. Scientists then make experiments trying to get informations about the time-dependent behaviour of their model. A special meaning for biomathematicians then turns into the determination of the parameters, which have to be interpreted biologically. In order to calculate the parameters we take the measured data from experiments.

In the diploma thesis [5] an enzyme-kinetical model has been investigated. The examined reaction in that case was the transformation of alcohol and NAD+ to acetaldehyde, NADH and H+ effected by the enzyme alcoholdehydrogenase (ADH). With the help of the reagent semicarbonite the backward-reaction could be suppressed so that we got an irreversible reaction as considered in studies of Michaelis and Menten.

$$E + S \rightleftharpoons C \rightarrow E + P$$

$$ADH + alcohol + NAD^{+} \rightleftharpoons Complex \rightarrow ADH + acetaldehyde + NADH + H^{+}$$

From the equation of reaction results a system of ordinary differential equations consisting of four ODEB4s for the four substances enzyme E, substrate S, complex C and product

P, together with three unknown parameters k_1, k_2, k_{-1} , the rate constants. With suitable transformations the system can be reduced to a system of two ODEB4s with three unknown parameters whereby two initial conditions are included.

$$\begin{aligned} \frac{dS}{dt} &= -k_1 E_0 S(t) + (k_1 S(t) + k_{-1}) C(t) \\ \frac{dP}{dt} &= k_2 (S_0 - S(t) - P(t)). \\ & S(0) = S_0, \quad P(0) = 0. \end{aligned}$$

But as it appears frequently in applications not all substances could be measured in the time course. Only the concentration of NADH could be recorded with a spectrophotometer that measures the extinction. The aim was to reconstruct the parameters from that information.

2 The Problem

The enzyme-kinetical problem desribed above is a special case of the topic we are discussing in this paper. First of all the particularity is the measurement of only one substance, the amount of NADH, from what we were able to conclude to the amount of NAD+ in the time period. That implied the difficulty that we have no information about the product acetaldehyde. The other hurdle was the stiffness of the system. This property is very often for chemical reactions. For solving the system of ODE's we used a method from Hairer and Wanner, see [3] .

In a more general view the process is modelled by an ODE

$$\dot{x} = f(x, P_1, \ldots, P_K) \text{ in } [0, T], x(t) \in \mathbb{R}^n \tag{1}$$

with (unknown) parameters $P_1, \ldots, P_K$. At time points $0 = t_0 \leq t_1 \leq \ldots \leq t_M = T$ we have corresponding measured data $m_1, \ldots, m_M$ from experiments.

We want to determine the parameters such that the solutions of (1) minimizes a cost functional

$$\|(x(t_1) - m_1, \ldots x(t_M) - m_M)\|_{\mathbb{R}^{nM}}$$

We discuss three possibilities to choose the norms: Euclidian norm, Tschebyschev norm, 1 norm, and obtain three different extremal problems:

$$\sum_{i=1}^{M} \|x(t_i, P_1, \ldots, P_K) - m_i\|_{\mathbb{R}^n}^2 = \min_{P_1,\ldots,P_K}$$

and

$$\max_{i=1,\ldots,M} \|x(t_i, P_1, \ldots, P_K) - m_i\|_\infty = \min_{P_1,\ldots,P_K}$$

and

$$\sum_{i=1}^{M} \|x(t_i, P_1, \ldots, P_M) - m_i\|_{\mathbb{R}^n} = \min_{P_1,\ldots,P_K}$$

respectively.

Without any loss of generality and only for simplicity we consider the case n=1 in the next subsection.

2.1 The Linearization

The usual approach to solve such kinds of problems is to combine an ordinary differential equation solver with an iterative optimization method. As optimization methods Gauss-Newton-methods are common, for the ODE solver the properties of the ODE-system have to be payed attention to. Systems from chemical reactions for instance are often stiff. A multiple shooting method is recommended to prevent the space out of the solution, see [1].
We need the derivatives of the state x with respect to all P_j. Instead of calculating the partial derivatives numerically for instance as

$$\frac{\partial x}{\partial P_j}(t,P) = \frac{x(t,P_1,\ldots,P_j+\Delta P_j,\ldots,P_K) - x(t,P_1,\ldots,P_K)}{\Delta P_j}$$

we recommend to solve the partial derivatives as additional ordinary equations. The idea is to bring in a new function

$$y_j(t,P) = \frac{\partial x}{\partial P_j}.$$

Differenciating the function with respect to time t and permuting the derivates gives

$$\begin{aligned} \dot{y}_j(t,P) &= \frac{d}{dt}\frac{\partial x}{\partial P_j} = \frac{\partial}{\partial P_j} h(t,x,P) \\ &= h_x(t,x,P)\cdot y_j(t,P) + h_{P_j}(t,x,P) \end{aligned}$$

with the initial value $y_j(0,P) = \frac{\partial x}{\partial P_j}(0,P)$, that means $y_j(0,P) = 0$ if the initial values of x are independent of the parameters.

3 The Optimal Control Attempt

We now consider the function $f(P)$ as extended Mayer-type costfunctional with constant control $P = (P_1,\ldots,P_K)$

$$f(P) = g(t_1, x(t_1),\ldots,t_M,x(t_M)) = \sum_{i=1}^{M}(x(t_i) - m_i)^2,$$

with $0 = t_0 \le t_1 \le \ldots \le t_M = T$. The state equation $x(t)$ is given as

$$\dot{x} = h(t,x,P), \quad x(t_0) = x_0.$$

As the controls are constant the ordinary Maximum-Principle can not be applied. We use necessary optimality conditions derived in [6]. We define the Hamilton-Function $H : \mathbb{R}^n \times \mathbb{R}^K \times \mathbb{R}^n \to \mathbb{R}$ as the righthandside of the state equation multiplied by the adjoint function $\lambda : \mathbb{R} \to \mathbb{R}^n$

$$H(x,P,\lambda) = \lambda \cdot h(t,x,P).$$

If $\hat{P} = (\hat{P}_1, \ldots, \hat{P}_K)$ is optimal then there exists a solution $\hat{\lambda}(t)$ of the adjoint equation

$$\dot{\lambda}(t) = -\lambda(t) \cdot h_x(t, x, P)$$

with jumps

$$\begin{aligned} \lambda(t_i + 0) &= \lambda(t_i) + g_{x_i}(\hat{x}(t_1), \ldots, \hat{x}(t_M)), \quad i = 1, \ldots, M, \\ &= \lambda(t_i) + 2(x(t_i) - m_i), \end{aligned}$$

such that

$$\int_0^T H_P(\hat{x}(t), \hat{P}, \hat{\lambda}(t))dt = 0$$

which is equivalent to

$$\int_0^T \hat{\lambda}(t) \cdot h_P(t, \hat{x}(t), \hat{P})dt = 0.$$

We use this necessary optimality conditions to obtain the following recipe:

1. Choose initial parameter vector $P^0 = (P_1^0, \ldots, P_K^0)$.
2. Solve the state equation with the current P to obtain $x(t; P)$.
3. Solve the adjoint equation with the current P to obtain $\lambda(t; P)$.
4. Compute $H(x, P, \lambda)$ where λ and h_P are solved for the current P
5. Define $G(P) = \int_0^T H_P(\hat{x}(t), \hat{P}, \hat{\lambda}(t))dt = 0$.
6. Improve $G(P)$ by applying an adaptive Newton‘s method (see [2]).

In case of the Tschebyschev norm the problem is more difficult, as the functional is not differentiable and has to be transformed, which includes the occurence of inequalities. But the necessary optimality conditions can be used to prove optimal parameters received by linearization as shown in the next section.

4 Necessary Optimality Conditions

We consider the parameter identification in the sense of the Tschebyschev norm. Then we have the optimal control problem:

$$\min_P \max_{i=1,\ldots,M} |x(t_i; P) - m_i|$$

with respect to

$$\dot{x} = h(t, x(t), P), \quad P \in \mathbb{R}^K.$$

The problem will be transformed into a classical one. We define the constant function $z(t) \geq \max_{i=1,\ldots,M} |x(t_i, P) - m_i|$. Then we can prove:

$$z(T) = \min!$$

such that

$$\begin{aligned}
&\dot{x} = h(t, x(t), P), \quad x(t_0) = x_0 \\
&\dot{z} = 0 \\
&z(T) - x(t_i) + m_i \geq 0, \\
&z(T) + x(t_i) - m_i \geq 0, \quad i = 1, \ldots, M.
\end{aligned}$$

Define $H(t, x, z, P, \lambda) = \lambda \cdot h(t, x(t), P)$. Assume $\hat{P}$ is the optimal parameter (vector). Then there exist multipliers $\alpha_i \geq 0$, $\beta_i \geq 0$, $i = 1, \ldots, M$, such that

$$\int_0^T H_P(\hat{x}(t), \hat{z}(t), \hat{P}, \hat{\lambda}(t))dt = 0,$$

where $\hat{\lambda}(\cdot)$ is the solution of the adjoint equation

$$\dot{\lambda}(t) = -\lambda(t) \cdot h_x(t, \hat{x}(t), \hat{P}).$$

$\lambda(t)$ has discontinuities at the time points t_i with jumps $\beta_i - \alpha_i$. The complimentary conditions

$$\begin{aligned}
\alpha_i(z(T) - x(t_i) + m_i) &= 0 \text{ and} \\
\beta_i(z(T) + x(t_i) - m_i) &= 0 \text{ are valid.}
\end{aligned}$$

This results are not useful to find the optimal parameters as we do not know the multipliers α_i and β_i. But they can be used to prove the optimality of a given parameter vector.
The necessary optimality conditions can be assigned to the case of the 1-norm in the cost functional $g(x(t_1), \ldots, x(t_M)) = \sum_{i=1}^M |x(t_i) - m_i|$. We have to introduce M additional state variables $z_i(t)$ with $\dot{z}_i(t) = 0$. We then get the new cost functional $\sum_{i=1}^M z_i(T) = \text{min!}$ and $2 \cdot M$ inequality coonstraints

$$\begin{aligned}
z_i &\geq x(t_i) - m_i), \\
z_i &\geq -(x(t_i) - m_i), \quad i = 1, \ldots, M
\end{aligned}$$

and we can prove with the necessary optimality conditions a derived P.

References

[1] H. G. BOCK, E. K., J. P. SCHLÖDER, *Numerical Methods for Parameter Estimation in Nonlinear Differential Algebraic Equations*. GAMM-Mitt. **30**, No. 3, 376-408, 2007

[2] P. DEUFLHARD, *Numerische Mathematik 1 - Eine algorithmisch orientierte Einführung*. de Gruyter Lehrbuch, Berlin, 2002

[3] E. HAIRER AND G. WANNER, *Stiff and differential-algebraic problems*, Springer Berlin, 1991

[4] J. D. MURRAY *Mathematical Biology - Reaction Kinetics*

[5] K. SCHREIBER *Untersuchungen zu Enzymkinetischen Modellen - Theorie und Mathematische Lösungsverfahren*. Diploma thesis, 2006

[6] W.H.SCHMIDT. Optimalittsbedingungen verschiedener Aufgaben von Integralprozessen in Banachrumen und das Iterationsverfahren von Chernousko. Habilitation. Greifswald, 1989

Current address

Werner H. Schmidt
Ernst- Moritz- Arndt University
Institute of Mathematics and Computer Sciences
Robert-Blum-Str. 2
17489 Greifswald
Germany
e-mail: wschmidt@uni-greifswald.de

Kristine Al Zoukra
Ernst- Moritz- Arndt University
Institute of Mathematics and Computer Sciences
Robert-Blum-Str. 2
17489 Greifswald
Germany

Journal of Applied Mathematics

volume III (2010), number II

APPLICATION OF NON-EUKLIDIAN METRICS IN DISCRETE EVENT SIMULATION

KINDLER Eugene, (CZ), KŘIVÝ Ivan, (CZ)

Abstract. The paper concerns application of non-Euclidian metrics in discrete event simulation, namely in case a pseudo-graphic animation is be very promptly implemented contrary to curved motions of the objects mapped at the display.

Key words. discrete event simulation, animation, transportation, metrics, curved space

Mathematics Subject Classification: Primary 00A72; Secondary 51K99

1 Role of Animation in Simulation

Simulation experiment is a certain image of what could happen in the world. Thus, animation of a simulation experiment is a suitable technique for displaying certain results, and – in parallel – for the verification of the computer model (frequently strange motions at the display tell much more on an error occurrence than tables of results). When simulation experiments are iterated into a so called simulation study [1,2], e.g. for optimizing a designed system, the study is a certain image of a physically creative activity in which one creates a certain entity P_1 of the real world, observes it, then – instructed by that watching – destroys it and replaces it by a different entity P_2, observes it, *etc*. until being satisfied during observing an entity P_k. Note that the entities should exist and be observed all in the same world time and thus the process of their changing does not correspond to any visible or imaginable by the humans living in real time and therefore, the animation of a very simulation study does not help at all.
The animation by means of fine multicolor graphics became very popular in computer games and animated cartoons generated at digital systems. It is of no use in computer simulation of all sorts (digital, analog, hybrid, real time, *etc*.), but its position is double-edged in that domain: as it was mentioned above, in some situation, the animation serves as good technique of displaying results, but in the other situation the simulation is extremely slowed down/delayed when joined with the animation. The conclusion is that the animation is suitable at the starting works on a simulation project (debugging the computer model, verifying it and possibly determining its validity domain). But in longer simulation experiments and/or in sophisticated simulation studies demanding

extremely long sequences of simulation experiments, the animation is not of use and should be eliminated in all its factors.

This phenomenon leads to extremes that are not agreeable for the constructors of simulation models. There are special simulation software products with a rather flexible apparatus for animation. When using them, the constructor of a simulation model directly describes the computing process as a complement of what should happen at the display. In general, such a software product is always specialized to certain class of systems, so that it is difficult or even impossible to simulate the other ones. But also in case the system that is to be simulated belongs to the class for which the software product was designed, there are the following obstacles. In that case, the constructor's work is rather easy but problems arise when the animation should be eliminated; often it happens that the animation has to be present during the computing that realizes the simulation model, but is hidden "inside the computer intestines" and only their end results are shown at the display. The demands on the rate of simulation experiments are then violated.

The mentioned software products are designed especially for the situations when a customer demands a very prompt reaction of an operator, i.e. demands him to construct a running simulation model as soon as possible and demonstrates what it really runs. Naturally, in a great number of cases, the customer anticipates the model to be later fast and included in sophisticated simulation studies.

In such a situation, there is no time enough to prepare a rather independent animation based on (fine) computer graphics, while applying a special animation software product mentioned above carries a real danger of fatal slowing down the following simulation studies.

One of the techniques to clear the mentioned difficulty consists in applying pseudo-graphics, i.e. when animating the experiment only by means of alphanumerically structured display. For the model constructor, such a technique is of no use, particularly because of discovering a lot of programming errors. But a certain fine obstacle should be solved for it. The obstacle and its solving are described in the following sections.

2 Inaccuracy of the Alphanumeric Display Grid

The animation at the alphanumerically structured display in not as beautiful as that using full graphics but is a satisfying and simple tool used by the professionals in simulation. Among the aspects mentioned in the preceding section, it extremely satisfies a (psychologically conditioned) demand that the animation should give only such information that is necessary for the understanding of the simulation model behavior.

Nevertheless, pseudo-graphics has an essential drawback, which consists in rounding space coordinates to a small set of values that are proportional to (i,j), where e.g. i passes through the integers 1 to 50 (line indexes) and j through the integers 1 to 80 (column indexes). Although for the human eyes such a rounding is not important it can introduce serious deviations into the computing process and thus into the results (data collected during the simulation experiments). Although the deviations may seem rather small (in percentage relations) they can essentially violate the information carried by the simulation, namely when it is to optimize some parameters or to anticipate some conflicts in the simulated system. Let the places of a simulation model, the coordinates of which are exactly proportional to the coordinates related to the places at the display, be called c-places, while the corresponding places at the display be called d-places.

Note that in a rather superficial view, the rounding errors may seem irrelevant, because one could suppose the computing to pass in the most exact arithmetic of the applied computer while the animation rounding would exist as the terminal phase which has no back influence on the previously passing exact computation. Nevertheless, in general that is often violated; the reason is a consequence of the offers to the operator to give some data directly through the animated scene at the display (e.g. to complete a traffic network by a – sometimes almost straight – way defined either by dragging a mouse or by giving display coordinates of some of its places). For example, a straight way so determined may be represented by a set of *c*-places that approximate the way from both sides and the sum of the distances between the mutually neighboring *c*-places can rather differ from the intended length of the way (namely in case the direction of the way image at the display is neither horizontal nor vertical).

In any case, the objective is to make the programming of animation simple and both animation itself and the data arisen during the simulation experimenting clear as much as possible.

3 Solving the Problem

The problem can be solved by introducing special metrics that differs from the Euclidian one, proportional to that applied among the images appearing at the display. Note that such a technique can be simply applied to discrete event simulation, while it should be more elaborated in case of continuous systems simulation.

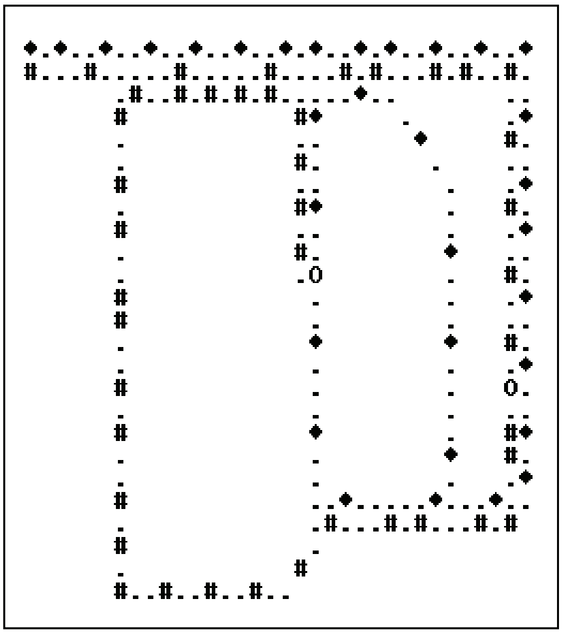


Fig. 1. A snapshot of the animation

If an entity E of the simulation model is introduced by an operator in a form of a set of *c*-places, the operator should declare also the extent (length,…) of E as a real number, independently of his drawing at the display. The simulation model then introduces a metrics into E in a form of an algorithm for computing the distances between mutually neighboring *c*-places so that the declaration expressed by the operator is respected. For example, if E is a way, the length of which is declared by the operator as D, the distances between the *c*-places related to the *d*-places that represent E at the display are computed as $D/(n-1)$ where n is the number of the *d*-places that represent E at the display. Note that such a technique has a very large spectrum of possible applic-ation, as the semantics of modern programming languages does not respect the axiom of extensionality. (In other words: in contrast to the conventional geometry, two pairs $<q_1,r_1>$ and $<q_2,r_2>$, where q_1 and q_2 have the same values and so r_1 and r_2, can have different distances D_i between their members q_i and r_i,, namely in case the pairs belong to the internal comput-er representations of different entities E_1 and E_2.) In Fig. 1 an illustration (a snapshot of the display) of a rather simple animation is presented. Along the ways

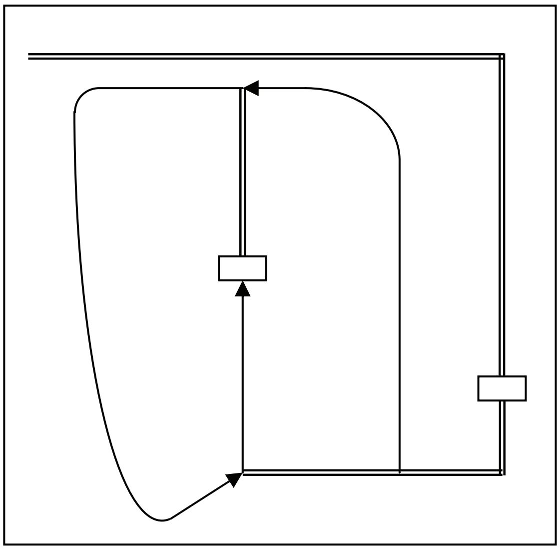


Fig. 2. Ground plan of the ways

displayed by means of their *d*-places represented by points, two sorts of objects (represented by # and romb) move and are served at places represented by letter *O*. Note that the real form of the ways is outlined in Fig. 2, where the double line represents a way where the objects can move in both orientations without conflicts.

4 Conclusion

The metrics can change in time; that can be applied for more detailed simulation of systems where the moving objects have a rather great freedom of motion at some places and an object moving in one orientation can limit (but not disable) the inverse motion of another object.
Until recently, the authors have had no occasion and free time to analyze the implications of the described sort of the non-Euclidian metrics for the whole (two-dimensional) domains. It would be interesting to study e.g. what frontiers of such applications are given by commonly respected axioms introduced for metric spaces.

Acknowledgement

This work was supported by the institutional research scheme MSM6198899701 of the Czech Ministry of Education, Youth and Sport.

References

[1.] STRAUSS, J. C., et al.: *The SCi continuous system simulation language (CSSL)*. Simulation, Vol. 9, No. 6, pp. 281-303, 1967.
[2.] KINDLER, E.: *Norma SCi pro simulační terminologii* (Simulation terminology SCi standard – in Czech). In: Simulace systémů a vědeckotechnické výpočty. House of Technology, Ostrava, pp. 38-43, 1985

Current address

Prof. RNDr. PhDr. Eugene Kindler, CSc.,
professor emeritus of the University of Ostrava,
tel. (private) **420 220 801 945,
e-mail ekindler@centrum.cz

Prof. RNDr. Ing. Ivan Křivý, CSc.,
University of Ostrava,
30. dubna 22,
701 03 Ostrava 1, tel: 597 092 177,
e-mail: ivan.krivy@osu.cz

Journal of Applied Mathematics

volume III (2010), number II

STABILITY ANALYSIS OF STATE-SPACE MODELS IN MATLAB

KVAPIL David, (CZ)

Abstract. Stability analysis is a very important task in the mathematical control theory and its practical application. Stability and performance are two of the fundamental issues in the design, analysis and evaluation of control systems. In engineering practice the control systems are designed so that stability is preserved in various classes of uncertainties – this property is known as the robust stability.
During physics modelling and identification of the heat distribution and consumption in COMSOL® we build state-space models of elements of heat plant and of distribution network. We need to check or guarantee the stability of system matrices and polynomials. We present several own created m-files in MATLAB®. These procedures are useful for stability analysis of state matrices and polynomials. We demonstrate their using in concrete examples.

Key words. Mathematical Control Theory, Stability, Robustness, State-Space Model, MATLAB

Mathematics Subject Classification: Primary 93D05, 93D09; Secondary 93A30.

1 Classic Approach

Consider continuous time (CT) or discrete time (DT), respectively, linear time-invariant (LTI) state-space system

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad\qquad x(k+1) = Ax(k) + Bu(k)$$
$$y(t) = Cx(t) + Du(t) \qquad\qquad y(k) = Cx(k) + Du(k)$$

where $t \in \mathbf{R}$, $k \in \mathbf{Z}$, with diagonalizable matrix A. Consider homogenous (undriven, $u \equiv 0$) system with unique equilibrium point at $x = 0$, provided A has no eigenvalues at 0 or at 1, respectively, in CT or DT case. Then

$$\dot{x}(t) = \mathrm{e}^{At}x(0) = V\begin{pmatrix} \mathrm{e}^{\lambda_1 t} & & \\ & \ddots & \\ & & \mathrm{e}^{\lambda_n t} \end{pmatrix} Wx(0), \qquad x(k) = A^k x(0) = V\begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} Wx(0).$$

System is asymptotically stable iff (i.e. if and only if) (see [4.][5.][7.][8.][11.])

$$\mathrm{Re}\left(\lambda_i\right)<0 \quad \text{or} \quad \left|\lambda_i\right|<1,$$

where $i=1,...,n$, respectively, for CT or DT case. For $\mathrm{Re}\left(\lambda_i\right)=0$ or $\left|\lambda_i\right|=1$ is the system on stability domain boundary. Matrix A is called Hurwitz or Schur (convergent) matrix, respectively, if all their eigenvalues have negative real part (for CT case) or their absolute values are less than 1 (for DT case). Similarly for polynomials, all their roots must lie in the open left complex half plane or in the open unit disc.

For stability analysis we can use direct Lyapunov method. [7.][8.] For undriven LTI system with Hurwitz matrix A consider quadratic Lyapunov function

$$V\left(x\right)=x^T Px,\ x\in \mathbf{R}^n,$$

where P is a symmetric matrix. Then $V\left(x\right)$ is a positive definite function iff all eigenvalues of P are positive.

The Lyapunov function of CT LTI system is

$$\dot{V}\left(x\right)=\dot{x}^T Px+x^T P\dot{x}=x^T A^T Px+x^T PAx=x^T\left(A^T P+PA\right)x=-x^T Qx$$

where $Q=-\left(A^T P+PA\right)$ is a symmetric matrix. If $Q\geq 0$, then the equilibrium point at the origin is stable in the sense of Lyapunov; if $Q>0$, then it's globally asymptotically stable. In practice, the Lyapunov equation $A^T P+PA=-Q$ is most often solved using properly chosen symmetric matrix $Q>0$ and then the positive definite solution P is found.

The Lyapunov function for DT LTI system is

$$\Delta V\left(x\right)=V\left(Ax\right)-V\left(x\right)=x^T A^T PAx-x^T Px$$

and then the Lyapunov equation is $A^T PA-P=-Q$.

Input-output stability [4.] is defined for a system with input signal u and output signal y that is obtained from input through the action of an arbitrary operator H, so $y=H\left(u\right)$. This system is l_p-stable, $p\in\{\,1,2,\infty\}$, if there exists a finite $C\in\mathbf{R}$ such that

$$\left\|y\right\|_p\leq C\left\|u\right\|_p \quad \text{for } \forall u.$$

The next two own created m-files in MATLAB – **StabilSpoj.m**, **StabilDisk.m** – find out all eigenvalues of the given matrix and determines about its stability for continuous-time case or discrete-time case, respectively. For example, if we have two matrices

```
>> A = [1 2 1; -3 -2 -1; 1 2 -1]; B = [-1 -2 1; 3 2 1; 1 -2 -1];
```

then we obtain

```
>> [lam] = StabilSpoj(A)
lam =
 -0.240803845501574 + 2.282705433037461i
 -0.240803845501574 - 2.282705433037461i
 -1.518392308996851
```

and

```
>> [lam] = StabilDiskr(B)
!!! Pozor, nestabilni vlastni cislo: !!!
lam =
   1.000000000000001 + 2.645751311064593i
   1.000000000000001 - 2.645751311064593i
  -2.000000000000000
```

Consider now a real polynomial of degree n: $p(s) = p_0 + p_1 s + p_2 s^2 + \ldots + p_n s^n$. The even and odd parts of a polynomial $p(s)$ are defined as

$$p^{even}(s) := p_0 + p_2 s^2 + p_4 s^4 + \ldots, \qquad p^{odd}(s) := p_1 s + p_3 s^3 + p_5 s^5 + \ldots.$$

Define

$$P^e(\omega) := p^{even}(j\omega) = p_0 - p_2\omega^2 + p_4\omega^4 - \ldots, \quad P^o(\omega) := \tfrac{1}{j\omega} p^{odd}(j\omega) = p_1 - p_3\omega^2 + p_5\omega^4 - \ldots.$$

$P^e(\omega)$ and $P^o(\omega)$ are both polynomials in ω^2 and their root sets are symmetric with the respect to the origin of the complex plane. If the polynomial $p(s)$ is even, then $n = 2m$ and

$$P^e(\omega) = p_0 - p_2\omega^2 + p_4\omega^4 - \ldots + (-1)^m p_{2m}\omega^{2m}, \quad P^o(\omega) = p_1 - p_3\omega^2 + p_5\omega^4 - \ldots + (-1)^{m-1} p_{2m-1}\omega^{2m-2},$$

if $p(s)$ is odd, then $n = 2m+1$ and

$$P^e(\omega) = p_0 - p_2\omega^2 + p_4\omega^4 - \ldots + (-1)^m p_{2m}\omega^{2m}, \quad P^o(\omega) = p_1 - p_3\omega^2 + p_5\omega^4 - \ldots + (-1)^m p_{2m+1}\omega^{2m}.$$

We say that $p(s)$ satisfies the interlacing property if

a) p_{2m} and p_{2m-1} (p_{2m+1} and p_{2m} for odd degree) have the same sign,

b) all the roots of $P^e(\omega)$ and $P^o(\omega)$ are real and distinct and the m positive roots of $P^e(\omega)$ together with the $m-1$ (m for odd degree) positive roots of $P^o(\omega)$ interlace in the following manner: $0 < \omega_{e,1} < \omega_{o,1} < \omega_{e,2} < \ldots < \omega_{e,m-1} < \omega_{o,m-1} < \omega_{e,m}$
($0 < \omega_{e,1} < \omega_{o,1} < \omega_{e,2} < \ldots < \omega_{e,m-1} < \omega_{o,m-1} < \omega_{e,m} < \omega_{o,m}$ for odd degree).

It is true that polynomial $p(s)$ is Hurwitz iff satisfies the interlacing property. [1.]

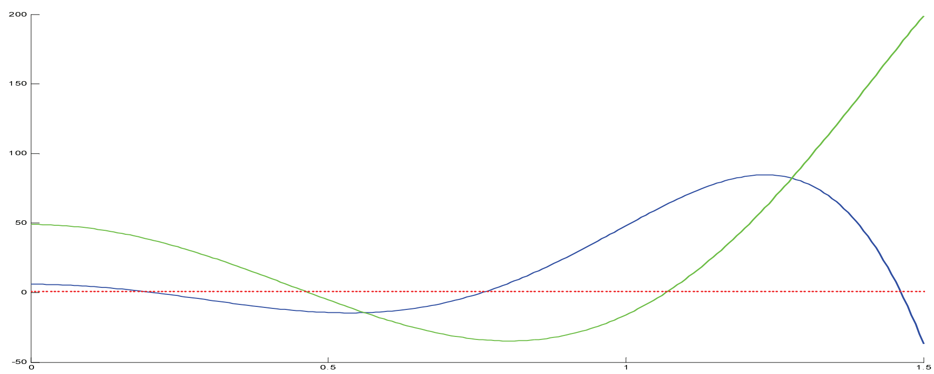


Fig.1: Interlacing Property

We can create MATLAB m-file **prolozeni.m** for visual check of interlacing property of the given polynomial p; it shows the plot of even and odd part on chosen interval $[a,b]$. For polynomial $p(s) = s^9 + 11s^8 + 52s^7 + 145s^6 + 266s^5 + 331s^4 + 280s^3 + 155s^2 + 49s + 6$ we obtain

```
>> [ZN,MatKor]=prolozeni(p,0,1.5)
Lichy stupen polynomu
Stejna znamenka pilotnich prvku
ZN =
   1
MatKor =
```

```
0.467310293305617   0.206141479511966
1.067193472969362   0.758664022543542
2.060665492162864   1.462642769866516
6.811488597247398   3.228688418057410
```

Let now $p(s)$ be a real polynomial with positive coefficients $p(s) = p_0 + p_1 s + ... + p_n s^n$, $p_i > 0$ for $i = 0,1,...,n$. With $\mu = p_n / p_{n-1}$, we define a polynomial q of degree $n-1$

$$q(s) = p_{n-1}s^{n-1} + \left(p_{n-2} - \mu p_{n-3}\right)s^{n-2} + p_{n-3}s^{n-3} + \left(p_{n-4} - \mu p_{n-5}\right)s^{n-4} + \dots .$$

If $p(s)$ has all its coefficients positive, then $p(s)$ is stable iff $q(s)$ is stable. [1.]

This lemma shows how the stability of polynomial $p(s)$ can be checked by successively reducing its degree as follows:

1. Set $p^0(s) = p(s)$.
2. Verify that all the coefficients of $p^i(s)$ are positive.
3. Construct $p^{i+1}(s) = q(s)$.
4. Go back to step 2 until you either find that any 2 is violated (then $p(s)$ is not Hurwitz) or until you reach the polynomial which is of degree 2 and in which case condition 2 is also sufficient (then $p(s)$ is Hurwitz).

We created MATLAB m-file **HurwStabTest.m** which implements the previous algorithm.

```
>> [q] = HurwStabTest(p)
q =
   59.348073576796203  30.397036437968922   6.000000000000000
Reseni: Polynom je stabilni
```

2 Alternative Approach

If we have large-scale systems, where exists very big matrices and many eigenvalues, we can solve this stability problem by application of the next theorem and using Linear Matrix Inequalities (LMI), without computing eigenvalues.

Let $A \in \mathbf{R}^{n\times n}$. The following statements are equivalent: [11.]

1. A is Hurwitz.
2. For each $Q \in \mathbf{R}^{n\times n}$, there is the unique solution P of the Lyapunov matrix equation $A^T P + PA = Q$; and if $Q < 0$ then $P > 0$.
3. There is some $P > 0$ such that $A^T P + PA < 0$.
4. There is some $P > 0$ such that $V(x) := x^T P x$ is a Lyapunov function for the system $\dot{x} = Ax$.

A linear matrix inequality (LMI) has a form [2.]

$$F(x) := F_0 + \sum_{i=1}^{m} x_i F_i > 0,$$

where $x = (x_1, ..., x_m)$ is the variable and the symmetric matrices $F_i = F_i^T \in \mathbb{R}^{n\times n}$, $i = 0,1,...,m$, are given. The inequality symbol means that $F(x)$ is positive definite. The nonstrict LMI has a form $F(x) \geq 0$. The LMI is a convex constraint on x (i.e. the set $\{ x \mid F(x) > 0\}$ is convex).

We can create MATLAB m-files **LjapC.m** (for CT case) and **LjapD.m** (for DT case) and apply an elementary algebraic test to determine whether the matrix is Hurwitz. Given a real matrix A, set up the linear matrix equation $A^T X + XA = -I$ and solve for symmetric X. If no solution exists, then A is not Hurwitz. If a solution X is found, we check whether it is positive definite. If $X > 0$, then A is Hurwitz. If not, either the solution is not unique or the unique solution is not positive definite, so we know that A is not Hurwitz.

3 Robust Stability

In dynamic models of real systems we usually need to consider some uncertainties that occur because of the uncertain physical parameters and the unstructured or structured dynamic uncertainty. We understand the uncertainty as a discrepancy between the mathematical model and the real object. Reasons for uncertainty may be different. Models of uncertainty can be divided into parametric, dynamic unstructured, structured and mixed. The dynamic uncertainty is usually expressed by the transfer of uncertainty $W(\omega)\Delta$, where Δ is arbitrary stable transfer function satisfying $\|\Delta\|_\infty = \sup_\omega |\Delta(j\omega)| \le 1$ and the stable proper rational weighting term W is used to represent any information about the accuracy of the nominal plant model varies as a function of frequency. Additive or multiplicative characterization of uncertainty, respectively, is [5.]

$$G(s) = G_n(s) + W_a(s)\Delta_a, \qquad G(s) = \left(1 + W_m(s)\Delta_m\right)G_n(s),$$

where G_n is the nominal model transfer. Consider a feedback system with an open-loop transfer function

$$L(s) = \left[G_n(s) + W(s)\Delta(s)\right]K(s) \quad \text{or} \quad L(s) = \left(1 + W(s)\Delta\right)G_n(s)K(s)$$

in additive uncertainty case or in multiplicative uncertainty case, respectively, where $L_n(s) = G_n(s)K(s)$ is the nominal open-loop transfer function.

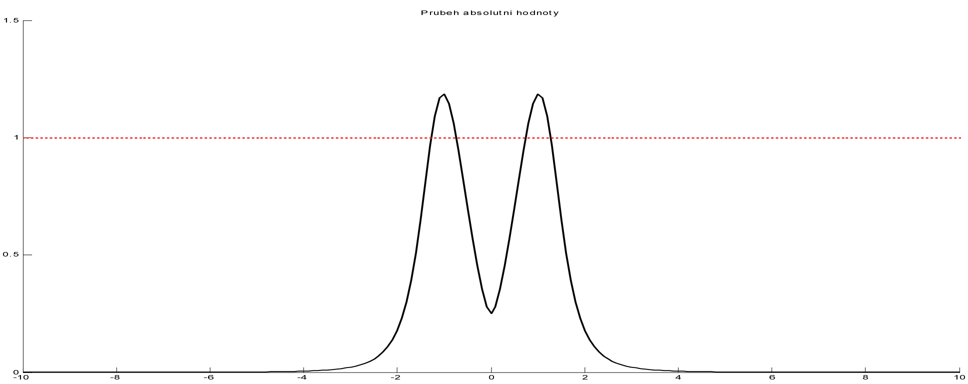


Fig.2: Multiplicative Robust Stability

We realize stability analysis using the Nyquist criterion (see [6.][14.]): A robust stability can be guaranteed if the system is stable for the nominal plant G_n and (for additive or multiplicative uncertainty case, respectively)

$$\left|\frac{W(j\omega)K(j\omega)}{1+G_n(j\omega)K(j\omega)}\right|<1 \quad \text{or} \quad \left|\frac{W(j\omega)G_n(j\omega)K(j\omega)}{1+G_n(j\omega)K(j\omega)}\right|<1 \quad \text{for } \forall\omega\in\mathbf{R}.$$

It is used in the following MATLAB procedures **AditRob.m** and **MultRob.m**. We created them for determination and visual test of stability for additive and multiplicative structures of uncertainties, respectively. Consider transfer functions

$$W=\frac{a(s)}{b(s)},\ K=\frac{c(s)}{d(s)},\ G=\frac{e(s)}{f(s)}$$

where a,b,c,d,e,f are polynomials of the same degree. Then for the input

```
>> a=[0 5 -1];b=[1 -1 2];c=[0 0 1];d=[1 0 -2];e=[0 0 2];f=[2 2 3];
```

we obtain (see Fig.2)

```
>> MultRob(a,b,c,d,e,f)
Neni stabilni
```

4 Parametric Uncertainty

The vector of real indeterminate (perturbative) parameters $q\in\mathbf{R}^k$ is used for the description of a system with a parametric uncertainty. If the parameter q is bounded by the given set Q we speak about a family of systems. The family of polynomials has a form $P(s,q)=\{p(s,q):q\in Q\}$ and we assume that it has invariant degree and is continuous with respect to q on a fixed interval. There exist several possible structures of parameters. [1.]

Let S be an open set in the complex plane representing the stability region and let ∂S denote its boundary. Suppose $\delta_1(s)$ and $\delta_2(s)$ are real polynomials of the degree n. Let

$$\delta_\lambda(s):=\lambda\delta_1(s)+(1-\lambda)\delta_2(s)$$

and consider the following one-parametric family of polynomials

$$\left[\delta_1(s),\delta_2(s)\right]=\left\{\delta_\lambda(s):\ \lambda\in[0,1]\right\}$$

called a segment of polynomials. We say that the segment is stable iff every polynomial in the segment is stable.

Bounded Phase Lemma says: [1.] Let $\delta_1(s)$ and $\delta_2(s)$ be stable with respect to S and assume that degree of $\delta_\lambda(s)$ is n for all $\lambda\in[0,1]$. Then the following are equivalent:

1. The segment $\left[\delta_1(s),\delta_2(s)\right]$ is stable with respect to S.
2. $\delta_\lambda(s^*)\neq 0$ for all $s^*\in\partial S$; $\lambda\in[0,1]$.
3. $\left|\phi_{\delta_1}(s^*)-\phi_{\delta_2}(s^*)\right|\neq\pi$ for all $s^*\in\partial S$.
4. The complex plane plot of $\delta_1(s^*)/\delta_2(s^*)$, for $s^*\in\partial S$, does not cut the negative real axis.

In creation of m-file **SegmentTest.m** we use the part 3 of the previous theorem and construct visual test of stability of polynomial segment. For two polynomials $p1(s)=s^3+4s^2+3s+1$ and $p2(s)=s^3+6s^2+4s+2$ we obtain

```
>> [r1,r2] = SegmentTest(p1,p2);
```

and we can see (Fig.3) that our polynomials create stable segment. If the difference between phases reaches π, the m-file warns us.

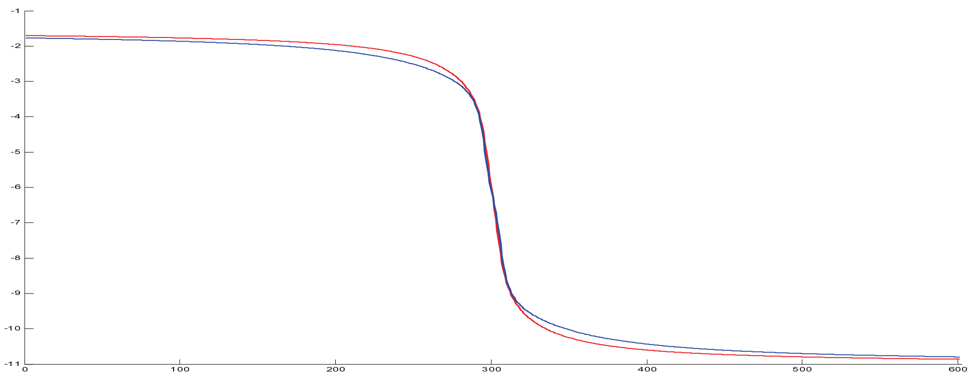


Fig.3: Phase difference of the endpoints of a stable segment

Segment Lemma says: [1.] Let $\delta_1(s)$ and $\delta_2(s)$ be the real Hurwitz polynomials of degree n with leading coefficients of the same sign. Then the line segment of polynomials $[\delta_1(s),\delta_2(s)]$ is Hurwitz stable iff there exist no real $\omega > 0$ such that

$$\delta_1^e(\omega)\delta_2^o(\omega)-\delta_2^e(\omega)\delta_1^o(\omega)=0, \quad \delta_1^e(\omega)\delta_2^e(\omega)\le 0, \quad \delta_1^o(\omega)\delta_2^o(\omega)\le 0.$$

This is used in m-file **SegmentLemma.m** where we test the stability of our two polynomials $p1(s)$ and $p2(s)$. This procedure returns the polynomial $P = p_1^e p_2^o - p_2^e p_1^o$ and its roots.

```
>> [P,rP] = SegmentLemma(p1,p2)
Pilotni prvky obou polynomu maji stejne znamenko
Neexistuji realne koreny polynomu P = P1even*P2odd - P2even*P1odd
P =
   -2     0     3     0    -2
rP =
  0.935414346693486 + 0.353553390593273i
  0.935414346693486 - 0.353553390593273i
 -0.935414346693485 + 0.353553390593274i
 -0.935414346693485 - 0.353553390593274i
Reseni: Usecka polynomu je stabilni
```

Let now have given a stability region in the complex plane and a nominal stable polynomial. We want to find the largest region in the coefficient space around the nominal polynomial where the stability property is maintained. Now we describe the procedure to determine the maximal stability region in the space of coefficients of a polynomial. [1.] Let the stability region S be any given open set of the complex plane $\mathbf{C}$, ∂S its boundary and U^0 the interior of the closed set $U = \mathbf{C} - S$. Assume that these three sets S, ∂S and U^0 are nonempty. For any given n, the set $\mathcal{P}_n$ of real polynomials of degree less than or equal to n is a vector space of dimension $n+1$. Let $\|\,.\,\|$ be an arbitrary norm defined on $\mathcal{P}_n$. The open balls induced by this norm or hypersphere, respectively, are of the form

$$B(P_0(s),r)=\{P(s)\in P_n : \|P(s)-P_0(s)\|<r\}, \quad S(P_0(s),r)=\{P(s)\in P_n : \|P(s)-P_0(s)\|=r\}.$$

For the given polynomial $\delta(s)$ of degree n with all its roots in S, there exists a positive real number ε such that every polynomial contained in $B(\delta(s),\varepsilon)$ is of degree n and has all its roots in S. Thus let $d^0(.)$ denotes the degree of a polynomial, we have the following property

$$\|\beta(s)-\delta(s)\|<\varepsilon \;\Rightarrow\; d^0(\beta(s))=n \;\wedge\; \beta(s) \text{ has all its roots in } S. \qquad (*)$$

For the stable polynomial $\delta(s)$ consider the subset of all positive real numbers having property (*)

$$R_\delta := \{t : t>0,\ t \text{ satisfies property (*)}\}.$$

R_δ is in fact an interval $(0,\rho(\delta)]$ where $\rho(\delta)=\sup\limits_{t\in R_\delta} t$; $\rho(\delta)$ is finite and satisfies property (*).

It can be shown (see [1.]) that for the given polynomial $\delta(s)$, of degree n, having all its roots in S, there exists a positive real number $\rho(\delta)$ such that:

a) Every polynomial contained in $B(\delta,\rho)$ has all its roots in S and is of degree n.
b) At least one polynomial on the hypersphere $S(\delta,\rho)$ has one of its roots in ∂S or is of degree less than n.
c) However, no polynomial lying on the hypersphere can ever have a root in U^0.

Consider on $\mathcal{P}_n$ for polynomials $p(s)=p_0+p_1s+...+p_ns^n$ and $r(s)=r_0+r_1s+...+r_ns^n$ the usual inner product and associated Euclidian norm

$$\langle p(s),r(s)\rangle = p_0r_0+p_1r_1+...+p_nr_n=\sum_{i=0}^{n}p_ir_i, \qquad \|p(s)\|_2^2=\langle p(s),p(s)\rangle = p_0^2+p_1^2+...+p_n^2.$$

Let Δ_0 be the subset of elements $p(s)$ on $\mathcal{P}_n$ such that $p(0)=0$. Dually, let Δ_n be the subset of all elements $p(s)$ on $\mathcal{P}_n$ that are of degree less than n. For each real $\omega\ge 0$ we can consider the subset Δ_ω of all elements of $\mathcal{P}_n$ which are divisible by $s^2+\omega^2$. For the given stable polynomial $\delta(s)=\delta_0+\delta_1s+...+\delta_ns^n$ we denote d_0, d_n and d_ω the distances (orthogonal projection) from $\delta(s)$ to the subspaces Δ_0, Δ_n and Δ_ω respectively. Finally let us define $d_{\min} := \inf\limits_{\omega\ge 0} d_\omega$.

It can be shown (see [1.]) that the radius of the largest stability hypersphere around a stable polynomial $\delta(s)$ is given by

$$\rho(s)=\min(d_0,d_n,d_{\min}).$$

It is easy to prove that $d_0=|\delta_0|$ and $d_n=|\delta_n|$. The main problem is to compute $d_{\min}$; but d_ω can be obtained in closed-form for any degree n. [1.] For arbitrary stable polynomial $\delta(s)$ of degree n, consider separation $\delta(s)=\delta^{even}(s)+\delta^{odd}(s)$. Then the distance $d_{\min}$ between $\delta(s)$ and Δ_ω is given by:

for $n=2p$: $$d_\omega^2=\frac{\left[\delta^e(\omega)\right]^2}{1+\omega^4+...+\omega^{4p}}+\frac{\left[\delta^o(\omega)\right]^2}{1+\omega^4+...+\omega^{4(p-1)}},$$

for $n=2p+1$: $$d_\omega^2=\frac{\left[\delta^e(\omega)\right]^2+\left[\delta^o(\omega)\right]^2}{1+\omega^4+...+\omega^{4p}}.$$

Having this expressions for d_ω, the next step is to find $d_{\min}$. It can be shown (see [1.]) that we do not need to minimalize on the infinity interval but we can confine to the interval $[0,1]$. It is true that

$$d_{\min}^2 = \min\left(\inf_{\omega\in[0,1]} d_\omega^2, \inf_{\omega\in[0,1]} d_{1/\omega}^2\right).$$

Supposing we have a procedure $\mathrm{DMIN}(\delta)$ which takes the vector coefficients δ as input and returns the minimum of d_ω^2 over $[0,1]$. Then the following algorithm will compute $d_{\min}$:

1. Set $\delta = (\delta_0, \delta_1, ..., \delta_{n-1}, \delta_n)$.
2. First call: d_1=DMIN(δ).
3. Switch: set $\delta = (\delta_n, \delta_{n-1}, ..., \delta_1, \delta_0)$.
4. Second call: d_2=DMIN(δ).
5. $d_{\min}$=min(d_1, d_2).

The previous algorithm was used in creation of m-file **StabBall.m** which finds out the stability range of the given stable polynomial. For our polynomial $p(s)$ we obtain

```
>> [rho] = StabBall(p)
rho =
      1
```

Consider now the set $P(s)$ of real polynomials of the degree n of the form $p(s) = q_0 + q_1 s + ... + q_n s^n$ where the coefficients lie within the given ranges,

$$q_0 \in \left[q_0^-, q_0^+\right],\ q_1 \in \left[q_1^-, q_1^+\right],\ \ldots,\ q_n \in \left[q_n^-, q_n^+\right].$$

Kharitonov's theorem says (see [1.]) that every polynomial in the family $P(s)$ is Hurwitz iff the following four (Kharitonov) polynomials are Hurwitz:

$$K_1(s) = q_0^- + q_1^- s + q_2^+ s^2 + q_3^+ s^3 + q_4^- s^4 + q_5^- s^5 + q_6^+ s^6 + \ldots,$$
$$K_2(s) = q_0^- + q_1^+ s + q_2^+ s^2 + q_3^- s^3 + q_4^- s^4 + q_5^+ s^5 + q_6^+ s^6 + \ldots,$$
$$K_3(s) = q_0^+ + q_1^- s + q_2^- s^2 + q_3^+ s^3 + q_4^+ s^4 + q_5^- s^5 + q_6^- s^6 + \ldots,$$
$$K_4(s) = q_0^+ + q_1^+ s + q_2^- s^2 + q_3^- s^3 + q_4^+ s^4 + q_5^+ s^5 + q_6^- s^6 + \ldots.$$

We create m-file **ParamRobu.m** for determination of the stability of the polynomials from the family $P(s)$. Input is $(n+1)\times 2$-matrix where i-th row corresponds with the coefficient $q_{(n+1)-i}$.

```
>> mat = [0.5 1.5; 3 4; 1 2; 0.5 1];
```

This means that we have the polynomial $p(s) = [0.5\ ;\ 1.5]s^3 + [3\ ;\ 4]s^2 + [1\ ;\ 2]s + [0.5\ ;\ 1]$. Outputs are Kharitonov polynomials and their eigenvalues.

```
>> [K1,K2,K3,K4,rK1,rK2,rK3,rK4] = ParamRobu(mat);
K1 stabilni
K2 stabilni
K3 stabilni
K4 stabilni
```

For example, the polynomial K1 and its eigenvalues:

```
>> K1
K1 =
   1.500000000000   4.000000000000   1.000000000000   0.500000000000
```

```
>> rK1
rK1 =
 -2.450096589027711
 -0.108285038819478 + 0.352595247452171i
 -0.108285038819478 - 0.352595247452171i
```

Conclusion

Stability is very important for the subsequent analysis of practical systems, such as structured singular values or Hankel analysis. Primarily the last three own created m-files in MATLAB® – the test of stability of a polynomial segment, the determination of a sphere of stability and the usage of Kharitonov Theorem – could be very useful for many practical tasks and real situations. This advanced and perspective approach could play a significant role in the applied mathematical control theory.

References

[1.] BHATTACHARYYA, S.P. – CHAPELLAT, H. – KEEL, L.H.: *Robust Control: The Parametric Approach*, Prentice Hall, 1995.
[2.] BOYD, S. – El GHAOUI, L. – FERON, E. – BALAKRISHNAN, V.: *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, 1994.
[3.] HAVLENA, V. – ŠTECHA, J.: *Moderní teorie řízení*, skriptum ČVUT, Praha, 1999.
[4.] HEIJ, C. – RAN, A. – van SCHAGEN, F.: *Introduction to Mathematical Systems Theory*, Birkhauser Verlag, Basel, 2007.
[5.] IOANNOU, P.I. – SUN J.: *Robust Adaptive Control*, Prentice Hall, 1996.
[6.] KALAS, J.: *Analýza v komplexním oboru*, skriptum MU, Brno, 2006.
[7.] KALAS, J. – Ráb, M.: *Obyčejné diferenciální rovnice*, skriptum MU, Brno, 1995.
[8.] KURZWEIL, J.: *Obyčejné diferenciální rovnice*, SNTL, Praha, 1978.
[9.] PÍRKO, Z. – Veit, J.: *Laplaceova transformace*, SNTL, Praha, 1972.
[10.] REKTORYS, K.: *Přehled užité matematiky*, Prometheus, Praha, 1995.
[11.] SONTAG, E.D.: *Mathematical Control Theory*, Springer Verlag, 1998.
[12.] ŠTECHA, J. – HAVLENA, V.: *Teorie dynamických systémů*, ČVUT, Praha, 2005.
[13.] STREJC, V.: *Stavová teorie lineárního diskrétního řízení*, Academia, Praha, 1978.
[14.] ŠVARC, I. – ŠEDA, M. – VÍTEČKOVÁ, M.: *Automatické řízení*, CERM, Brno, 2007.

Current address

David Kvapil, RNDr.
UNIS, a.s.
Department of Optimisation and Advanced Control
Jundrovská 33, 624 00 Brno, CZ
+420 541 515 390
e-mail: dkvapil@unis.cz

Journal of Applied Mathematics

volume III (2010), number II

MATHEMATICAL MODEL FOR SPATIAL DATA PROPERTIES ASSESSMENT

TALHOFER Václav, (CZ), HOŠKOVÁ Šárka, (CZ),
KRATOCHVÍL Vlastimil, (CZ), HOFMANN Alois, (CZ)

Abstract. When employing spatial data and information in decision-making processes, complex knowledge of their values is the prerequisite for assessing the credibility and accuracy of decisions made. By implementing the methods of value analysis and mathematical modeling it is possible to create an assessment system of spatial data complex usability. Based on input characteristics of the used spatial data and databases, quality characteristics and their changes can be calculated with the help of analytical methods. By comparing costs necessary for different variants of enhancement or for adjustment of database quality it is possible to optimize both the total usability and the costs put in securing the required.

Key words. spatial data, GIS, quality assessment, utility value, mathematical modeling

Mathematics Subject Classification: Primary: 00A71, Secondary 00A72.

1 Introduction

Rather extensive databases of area-localized data utilized in a number of fields are created in the Czech Republic. Data model objects and phenomena of both natural and social character (water courses, settlement structure, atmospheric pressure, etc.). The created and utilized data always encompass a position element, which localizes objects and phenomena in a given reference coordinate system, and a thematic element, which describes qualities of the given objects and phenomena (e.g. the speed of a water course, number of inhabitants in a settlement, actual atmospheric pressure readings). The actual data may then be of both geographic and non-geographic character (data on water courses contrasted with data on the transported cargo). The following text therefore uses predominantly the general term “spatial data” or “spatial information”. Basic localization databases are created by state administration bodies (Czech Office for Surveying, Mapping and Cadastre - COSMC, Army of the Czech Republic – ACR) and are intended for activities related to state functions, including command and control systems implemented in armed forces and in crisis management of individual components of the Integrated Rescue System (IRS). Spatial data are used not only for basic orientation in space but also as data for solving tasks connected with actual decisions, e.g. geographic impact on combat and non-combat army activities in given environments, predictions of landscape damage under extreme meteorological conditions or emergencies, in cases of military threats to the state, etc. In a number of tasks the source data combine and based on mathematically or procedurally described processes, new data are created.

2 Quality and Reliability of Spatial Data Concept

Users of both source and newly created data should always obtain the actual spatial information and information on its quality. When assessing the quality, we may draw on the general scheme of quality components which assesses the production-technological aspects which affect technical functionality and the operational, security as well as reliability aspects related to the given utilization of a product or service (see Figure 1).

When assessing spatial data and information it is necessary to modify the initial understanding of quality with regard to spatial determination of modeled objects and phenomena. The existing systems of spatial data assessment differ in relation to whether they assess technical parameters of data or technological impacts at play during their collection or whether they assess the resulting utility value determined by the quality of used information. Assessed technical parameters of data include the positional accuracy of information which is frequently determined by the mean positioning error or mean error in each co-ordinate axes. Thematic accuracy is also assessed. However, the actual assessment may be further complicated by the fact that thematic information may differ and that not all the assessed parameters may be known for a given object.

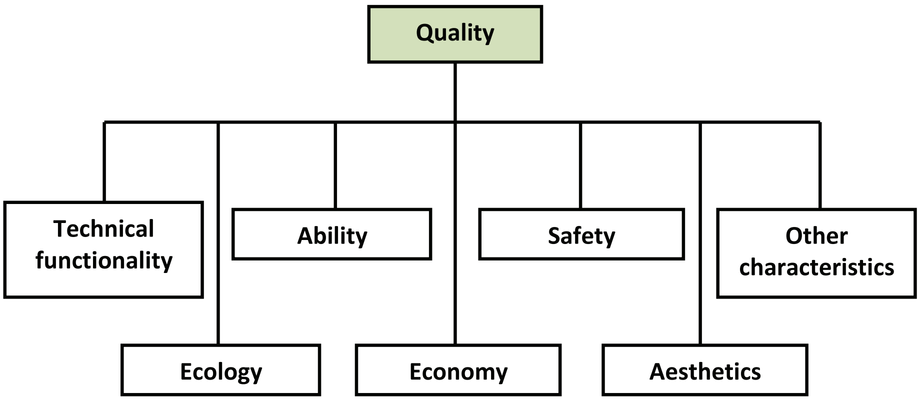


Fig.1: General components of quality assessment

Generally, when formulating the issue of spatial data quality assessment and the resulting geo-spatial information, one must draw on recommendations issued by international organizations, such as ISO, OGC and DGIWG which consistently deal with the development of geo-information science, as well as follow the INSPIRE (2) directive. These organizations and pools develop quality assessment systems for geo-spatial data. For example, according to the Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards (1), it is necessary to assess quality as a complex issue which encompasses both production and customers/users (see Fig.2).

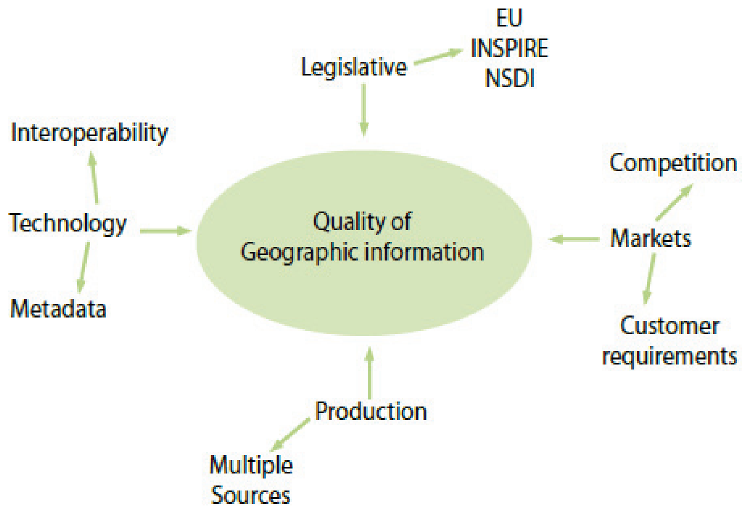


Figure 2: Reasons for Implementing Geographic Information Quality Standards
(according to Jacobsson & Giversen, 2007)

Technical functionality of spatial data is affected primarily by technological processes of geo-spatial data preparation and production, which are determined by the used systems of data collecting and processing, the formats of recorded data, etc. Data included in a single database do not necessarily have to be collected by a single organization and with the help of a single technology. On the contrary, databases defined by a single conception are frequently created under various national and international co-operations but in reality these conceptions may be adjusted by production organizations to comply with standards on the one hand and to take into account the production organization's technical and technological conditions on the other hand. This trend is apparent particularly in international projects (e.g. in cross-border spaces or such global projects as Vector Smart Map, Multinational Geospatial Co-Production Program and others) whose approaches to defining and modeling objects differ. An example of such different understanding of the significance of communications may be seen in the conditions of Central Europe and Afghanistan.

The produced data are intended for consumption, which means they are to be used in concrete spatial analyses, planning and management. From the user's perspective, systems for data evaluation are also highly important, particularly for their feature of utility value assessment (3). Technical functionality can be assessed generally without knowing the particular task, methodology or spatial information use. Other quality components, particularly reliability, must be assessed with relation to the given implementation in a given process. The paper (5) lists a definition of characteristics and quality parameters with regard to information on communication network which draws on ISO 19113 but is adapted to given purposes. According to (3), it is possible to assess the quality of spatial data and information according to the following criteria (see Table 1):

Main characteristics - main criteria	Sub-criteria characteristics	Definition	Quality parameter
Data model content – k_1	Complexity of conceptual landscape model	Concord of conceptual model and user requirements.	Percentage of incomplete information - k_{11}
	Compliance of required resolution of geometric and thematic data – k_{12}	Concord of required geometric and thematic resolution.	Percentage of objects without required level of geometric resolution – k_{121}
			Percentage of objects without required level of thematic resolution – k_{122}
Technical functionality – k_2	Transparency of data sources and methods for secondary data derivation - k_{21}	Transparency of source materials on primary data collection	Level of availability of information about used sources - k_{211}
		Transparency of used methods and model for secondary data derivation	Level of availability of information about used methods – k_{212}
	Position accuracy – k_{22}	Compliance with declared horizontal accuracy	Percentage of objects with unsatisfied conditions of declared horizontal accuracy – k_{221}
		Compliance with declared vertical accuracy	Percentage of objects with unsatisfied conditions of declared vertical accuracy – k_{222}
	Thematical accuracy – k_{23}	Compliance with declared accuracy of thematical data	Percentage of objects with unsatisfied conditions of declared thematical accuracy – k_{23}
	Logical consistency – k_{24}	Degree of adherence of geographic data (data structure, their features, attributes and relationships) to the models and schemas (conceptual model, conceptual schema, application schema and data model)	Percentage of objects with topological inconsistence – k_{241}
			Percentage of objects with thematical inconsistence – k_{242}
			with thematical inconsistence – k_{243}
	Data completeness – k_{25}	Degree of adherence of the entirety of geographic data (features, their attributes and relationships) to the entirety of the modelled part of landscape	Percentage of missing objects or objects there are surplus – k_{251}
			Percentage of incomplete thematical properties of objects – k_{252}
Up-to-dataness – k_3	Degree of adherence geographic data to the time changing in the landscape – k_3	Value of the time function describing process of the landscape changing	Number of changes
			Interval from the time of the last up-grade
			…
Landscape importance – k_4	Value of inverse distance to objects of interest – k_4	Landscape importance for subserved task or functions	Geographic location
			Entering corridors

			Amount and characteristics of obstacles
			Industrial centres
			Built-up areas
			Dislocation of defence systems
			…
Techniques of application and safety – k_5	Data standardization – k_{51}	Declared standards adherence	Percentage of non observance of standards
	Independency on application software– k_{52}	Degree of independency on application software	
	Data protection – k_{53}	Degree of the data protection system and its level	Degree of data access protection - k_{531}
			Degree of copyright protection - k_{532}
			Degree of data damage protection - k_{533}
Ability – k_6	Availability – k_6	Degree to which geographic data are failure rate available at a certain place and at a defined time	Percentage of failure rate

Table 1: Criteria of spatial data and information quality

3 Mathematical Model of Quality Assessment

From the previous text it follows that spatial data can be regarded as a product, which has its own purpose. However, for product evaluation purposes it is not sufficient to assess the product's quality but also the degree of meeting user's needs. In practice, method of value analysis (4), which objectifies the specification of meeting these needs both for products and services, is frequently used. Value analysis methods thus can be used for evaluating final products as well as for designing new products to assess individual variants. The analysis does not evaluate purely the degree of meeting consumer needs but also the costs inherent in securing them (financial, capacity, temporal and other costs).

If value analysis is implemented in assessing the usefulness of spatial databases, it is necessary to be aware of all the relevant data characteristics as well as of relevant cost items. Should the publication's recommendations (1) be fully accepted, the listed data characteristics should always constitute a part of meta-information on given used data and files in future. In case of inferred data, the characteristics should be automatically generated in the process of analyses.

Data quality and their utility value will not be fully revealed until they are implemented. In this case it is necessary to assess not only their technical characteristics but also the reliability of their security (see Table 1). In processes which combine spatial data with another services, such as positioning services provided by global satellite-based navigation systems (GNSS, GPS) or inertial navigation systems (INS), it is necessary to asses both the reliability of the data themselves but also the reliability of securing the provided service.

4 Functionality of spatial data database

For the analytical solution itself we may draw on combined methods applied from value analysis theory, reliability theory and contemporary mathematical, particularly from the fields of probability, mathematical statistics, differential calculus and applied theory of fuzzy logic.

With regard to the application of value analysis theory, we may assess the utility value as a degree of digital spatial data database functionality *(°F),* which can be expressed by the following aggregate function:

$$°F = p_3k_3p_4k_4p_6k_6\left(p_1k_1 + p_2k_2 + p_5k_5\right)$$

where variable k_i expresses the main quality criteria and p_i represents the weights of individual criteria. As a rule, the main criteria are expressed as a set of partial criteria, which also have their own weights (for more refer to (3)).

When assessing the utility of used part of a database, the ideal level of quality must be defined at first. This ideal level then works as a comparative etalon for expressing the level of meeting the individual criteria in the given assessed part of spatial database. Upon implementing the comparative etalon, the level of meeting the individual criteria may be assessed and consequently also the total utility value, or the degree of user function *F*.

The level of meeting individual criteria can be generally expressed by the following formula:

$$u_s = \frac{k_s}{k_s^*},$$

- k_s represents the value of meeting the *s*'th partial criterion,
- k_s* is the degree of meeting the *s*'th partial criterion or criteria of its sub-group under the comparative etalon.

For example, the criterion of "*technical quality of database*" will reach the ideal value if all the relevant data are faultless, accurate and consistent. Individual parameters of partial criteria must have the following values:

- $\alpha_{211} = \alpha_{212} = 0$, where α_{211} and α_{212} are percentage ratios of unknown or incomplete information on used source materials and methods of creating derived information;
- $100\frac{n_{22i}}{n} \geq 100 - h_s$
- n is the declared number of all objects and phenomena recorded in the database,
- n_{22i} is the number of objects and phenomena in the database whose horizontal and vertical accuracy meets the criteria of a given category,
- h_s is the selected level of reliability in %;
- $100\frac{n_{23}}{n} \geq 100 - h_s$
- n is the declared number of all objects and phenomena in the database,
- n_{23} is the number of objects and phenomena in the database whose attribute accuracy meets the criteria of a given category;
- h_s is the selected level of reliability in %;

- $n_{24i} = n$, where n_{24i} represents the numbers of objects and phenomena in the database which are consistent with respect to individual tested topology characteristics;
- $n_{25i} = n$, where n_{251} represents the number of actually accomplished objects and phenomena in the database and n_{252} is the number of objects and phenomena with fully complemented attributes.

Ideal values of individual partial criteria will then always equal 100 and the level of meeting criterion k_2 for the used section of the database (x^{th} recordable unit, e.g. a map sheet, used part of the database from a given area, etc.) can be calculated according to the following relation:

$$u_2^x = \frac{p_{21} - p_{21}\frac{\sum_{i=1}^{2} p_{21i}\frac{\alpha_{21i}^x}{100}}{\sum_{i=1}^{2} p_{21i}} + p_{22}\frac{\sum_{i=1}^{2} p_{22i}\left(\frac{n_{22i}^x}{n^x} + \frac{h_s}{100}\right)}{\sum_{i=1}^{2} p_{22i}} + p_{23}\left(\frac{n_{23}^x}{n^x} + \frac{h_s}{100}\right) + p_{24}\frac{\sum_{i=1}^{3} p_{24i}\frac{n_{24i}^x}{n^x}}{\sum_{i=1}^{3} p_{24i}} + p_{25}\frac{\sum_{i=1}^{2} p_{25i}\frac{n_{25i}^x}{n^x}}{\sum_{i=1}^{2} p_{25i}}}{\sum_{i=1}^{5} p_{2i}}$$

Reliability level expresses the degree of an individual service's availability in a given time and space for a specific task or used service. Unlike other quality components, it is essential to always focus on reliability in terms of its specific use. For instance, for navigation systems it is necessary to draw on both correct geospatial data and the available GPS service, which enables to determine the immediate position of a GPS receiver. If the data are incomplete or incorrect, or if GPS signal is not available, navigation is interrupted or terminated. Other quality components can be dealt with either generally, i.e. for general use, or with regard to particular cases. Reliability level can be expressed by the following formula:

$$u_6^x = p_6 \frac{\alpha_6^x}{100}$$

where $\alpha_6/100$ is a percentage expression of the degree of unreliability for the availability of a given service or system as a whole. In this case, the entire used system is concerned. Should the reliability of individual system components be considered (e.g. GIS thematic layers, such as the layer of water bodies, settlements or road network), possibly also that of complementary services (reception of GPS signal), the ensuing formula shall be more complex.

The total individual utility value (individual functionality) of a database's used part is defined by the aggregate function:

$$U^x = {}^\circ F^x = p_3 u_3^x p_4 u_4^x p_6 u_6^x (p_1 u_1^x + p_2 u_2^x + p_5 u_5^x).$$

5 Change in Functionality of Spatial Database

Owing to the fact that spatial databases can never be ideal, it is recommendable to assess the impact of the aggregate function's individual components on changes in the database. In this case we may use derivatives of the U^x *(°F)* function according to individual variables, which express the levels of meeting the given criteria. Generally, the impact of changes in meeting the main i^{th} criterion can be expressed in the following way:

$$d°F = \frac{dU}{du_i}$$

However, the degrees of meeting the main criteria are represented by functions of more variables. In order to express the value du_i, it is possible to employ two methods depending on the required information structure. If the *impact of individual variables* on the total individual utility value should be assessed upon an assumption that all other variables are constant, derivations of function U must be expressed in the following way:

$$d°F = \frac{dU}{du_i}\frac{du_i}{dx}$$

where x is one of the given variables.

In practice, a situation may arise that a number of factors may be changed at a time, e.g. the technical quality of database is changed in all its parameters – used methods of secondary data inference are improved, localization and attribute accuracy and data complexity are enhanced and simultaneously data are placed in a geo-database accessible to authorized users where all topological, thematic and temporal relations are well treated. In this case it is recommended to express the value of du_i as the total differential of all the variables.

Through mathematical modeling it is then possible to solve tasks of the following types:

- how a change in a given partial parameter or several parameters of a database is reflected in its total usability;
- which parameters need to be changed to achieve the required product functionality;
- which parameters may be "degraded" owing to the fact that the product's functionality is unnecessarily high.

If economic calculations (e.g. by expressing investment value – financial, personnel or temporal) are added in the assessment system, it is possible to determine the optimal degree of product functionality for achieving the highest possible functionality while minimizing costs necessary for securing it.

6 Relative cost efficiency

Data base functionality degree is comparable to the cost necessary for provisions – direct material, direct wages, other direct expenditures (HW, SW, amortisation, costs for co-operations, tax and social payments etc.), research and development cost, overhead cost and others. Functionality and cost imply *relative cost efficiency (RCE)* calculated as follows:

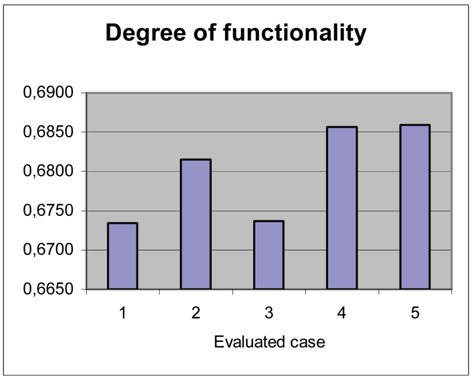
$$RCE = \frac{°F}{\sum_{i=1}^{n} N_i}.$$

It is possible to find the most suitable option using RCE. The presented model functionality is shown in the following table and diagrams (Table 2 and Figure 2).

In the conditions of the Military Geographic and Hydro Meteorological Institute of the Army of the Czech Republic the working time and expenses for database up-dating are precisely specified. Next example is based on its standards.
In the initial stage, the database degree of functionality is 0.5238. In cases 1 to 5, there are various attitudes to improve its properties – more database update (case 1), increased stored features amount (case 2), completing all missing features (case 3), completing all missing thematic properties (case 4) and completing all missing features and thematic properties (case 5). The cases 4 and 5 proved as the most functional ones. But if expenses are calculated, case 3 is the most effective output.
The described model doesn't bring absolute solution, but it can represent a useful tool for DGI utility value assessment as well as for finding economic ways how to increase this value even under personnel or financial restrictions.

Case	initial	1	2	3	4	5
	T=5, α_{11}=20, n_{251} = 99, n_{252} = 50	T=1, α_{11}=20, n_{251} = 99, n_{252} = 50, difficulty class 3	T=1, α_{11}=15, n_{251} = 99, n_{252} = 50, difficulty class 4	T=1, α_{11}=20, n_{251} = 100, n_{252} = 50, difficulty class 3	T=1, α_{11}=20, n_{251} = 99, n_{252} = 100, difficulty class 4	T=1, α_{11}=20, n_{251} = 100, n_{252} = 100, difficulty class 4
°*F*	0.5238	0.6734	0.6815	0.6737	0.6856	0.6859
RCE (currency unit)		2.8878	2.4965	2.8889	2.5116	2.5126
ΔRCE (currency unit)			0.3913	-0.0011	0.3762	0.3752

Table 2: Model of RCE calculation



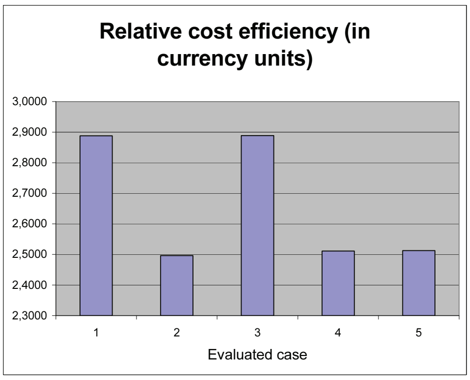


Figure 2: DGI functionality and RCE comparing

The presented process is applicable to evaluation of present products as well as planned products. When this model is used for a present product, it is possible to optimise its characteristics. In the case of a planned product, it is possible to assess various variants.

7 Conclusion

The proposed solution aims to enhance the efficiency of activities related to the use of non-homogenous data and information in command and control systems to provide operation bodies not only with their own databases but also with a relevant base concerning the quality and credibility of the used data. This information allows them to draw on such data in their decision making and possibly to adjust their decisions correspondingly.

Acknowledgement

The paper was written under the research project "Evaluation of Integrated Digital Spatial Data Reliability" supported by the Grant Agency of the Czech Republic (project code 205/09/198) which solves a complex method of quality assessment for digital spatial data and its integration in order to determine the degree of reliability of the resulting data and information.

References

[1] JACOBSSON, A., & GIVERSEN, J. (2007). *Eurogeographics.* Získáno 2009, from http://www.eurogeographics.org/documents/Guidelines_ISO19100_Quality.pdf
[2] KONEČNÝ, M., KUBÍČEK, P., & STANĚK, K. (1998). Mezinárodní standardizační aktivity v oblasti prostorově orientovaných informačních systémů. *Sborník příspěvků konference Geomatika 98* (pp. 35 - 41). Praha: CAGI.
[3] TALHOFER, V. (2002). *Možnosti zdokonalení užitnosti digitálních geografických dat* (Thesis, published in). Brno: Vojenská akademie v Brně, Fakulta vojensko technická druhů vojsk,
[4] http://user.unob.cz/talhofer/Uzitna%20hodnota%20digitalnich%20geografických%20dat.pdf.
[5] VLČEK, R. (1983). *Příručka hodnotové analýzy.* Praha: SNTL - Nakladatelství technické literatury.
[6] WILTSCHKO, T., & KAUFMANN, T. (2005). *Euroroads.* (U. Sandgren, Editor) Získáno 2009, from http://www.euroroads.org/php/Reports/D2.3.pdf

Current address

Václav Talhofer, COL Assoc. Prof., Ing., CSc.
University of Defense, Faculty of Military Technology,
Department of Military Geography and Meteorology, Kounicova 65, 61200 Brno, Czech Republic
e-mail: vaclav.talhofer@unob.cz

Šárka Hošková, Assoc. Prof., Dr., Ph.D.
University of Defense, Faculty of Military Technology,
Department of Mathematics and Physics, Kounicova 65, 61200 Brno, Czech Republic
e-mail: sarka.hoskova@unob.cz

Vlastimil Kratochvíl, Assoc. Prof., Ing., CSc.
University of Defense, Faculty of Military Technology,
Department of Military Geography and Meteorology, Kounicova 65, 61200 Brno, Czech Republic
e-mail: vlastimil.kratochvil@unob.cz

Alois Hofmann, Ing., CSc.
University of Defense, Faculty of Military Technology, Department of
Military Geography and Meteorology, Kounicova 65, 61200 Brno, Czech Republic
e-mail: alois.hofmann@unob.cz

Journal of Applied Mathematics

volume III (2010), number II

REALIZABITY OF THE ENDOMORPHISM MONOID OF A SEMI-CASCADE FORMED BY SOLUTION SPACES OF LINEAR ORDINARY N-TH ORDER DIFFERENTIAL EQUATIONS

CHVALINA Jan, (CZ), CHVALINOVÁ Ludmila, (CZ)

Abstract. In this contribution there is solved a certain modification of the so called "realization problem" discussed between C. J. Ewerett, J. von Neumann, E. Teller and S. M. Ulam in the year 1948 coming from the field of the Einstein's special relativity theory. More precisely we construct a certain minimal extension of the action of the aditive monoid of all non-negative integers on the phase set formed by n-dimensional solution spaces of n-th order linear homogeneous ordinary differetnial equations allowing to endowe the mentional phase set with a structure of an extensive join space, good-endomorphism monoid of which coincides with the endomorphism monoid of the obtained semi-cascade.

Key words and phrases. Linear n-th order ordinary differential operator, semi-cascade, solution space of n-th order linear ordinary homogeneous differential equation.

Mathematics Subject Classification. Primary 34A30, 47D03, 47E05; Secondary 20N20.

This contribution is motivated by one classical realization problem following — according to considerations due to R. Z. Domiaty [8] — from one discussion between Cornelius J. Ewerett, John von Neumann, Edward Teller and Stanislaw M. Ulam in the year 1948 which core is lying in the Einstein's special relativity theory. In general, the classical realization problem can be simply formulated in this way:

Given a concrete category $\mathcal{C}$, a set X and a group G of permutations of the set X. Does there exists an object $(X, \Xi) \in \mathcal{C}$ such that the automorphism group $\mathrm{Aut}(X, \Xi) = G$? ([8, 9]). A certain importand impuls cames from the relativity theory as the question whether it is possible to change locally euclidean topologies in mathematical models of space-time by metrics or more generaly by topologies under the supposition of preservation of corresponding homeomorphism groups.

The above mentioned problem can be considered as a question in the sense Felix Klein's "Erlangener Program" and, clearly, it can be modified for any concrete category or for arbitrary pairs of concrete categories. It is to be noted as a certain specification of the above realization problem — consists in the question under which conditions one mathematical structure can be substituted by the other one with the same carrier such that actual monoids of mappings carying morphisms of different categories coincide. More concretely in the book [13] there is solved the following problem on pages 40–84:

Find a characterization of a set transformation $f\colon X \to X$ (or in another words a characterization of a mono-unary algebra (X, f)) such that there exists a quasi-ordering $\leq$ on X with the property

$$C(f) = SI(X, \leq)$$

(or $\mathrm{End}(X, f) = SI(X, \leq)$), where $C(f) = \{g\colon X \to X \mid g \circ f = f \circ g\}$ is the centralizer of f (i.e. the endomorphism monoid of the mono-unary algebra (X, f)) within the full transformation monoid of the set X and $SI(X, \leq)$ is the monoid of all strongly isotone self maps of the quasi-ordered set $(X, \leq)$, i.e. such mappings $f\colon (X, \leq) \to (X, \leq)$ that for an arbitrary pair of elements $[x, y] \in X \times Y$ we have $f(x) \leq y$ if and only if there is an element $x' \in X$ with property $x \leq x'$ and $f(x') = y$. Denoting by $[x)_{\leq} = \{y \in X \mid x \leq y\}$ i.e. the principal end generated by the element $x \in X$ it can be easily shown that f is a strongly isotone self maps of $(X, \leq)$ if and only if $f\big([x)_{\leq}\big) = \big[f(x)\big)_{\leq}$ for any element $x \in X$. This concept is motivated by investigations of Saul Aaron Kripke — [21, 22, 47] and the answer of the above formulated question is contained in the below presented theorem. It is to be noted that professor Kripke has made fundamental contributions to a variety areas of logic, and his name is attached to a corresponding variety of objects and results.

Kripke semantics (also known as relational semantics or frame semantics) is a formal semantics for non-classical logic systems created in the late 1950s and early 1960s by Saul A. Kripke. A Kripke frame or modal frame is a pair $\langle W, R\rangle$, where W is a non-empty set, and R is a binary relation on W. Elements of W are called nodes or worlds, and the relation R is known as the accessibility relation. This is a binary relation between possible words which has very powerful uses in both the formal/theoretic aspects of modal logic as well as in its applications to thinks like epistemology and value theory ([47]). As in the classical model theory, there are methods for constructing a new Kripke model from other models.

The natural homomorphisms in Kripke semantics are called p-morphisms (or pseudo-epimorphisms, but the latter term is rarely used). A p-morphism of Kripke frames $\langle W, R\rangle$ and $\langle W', R'\rangle$ is a mapping $f\colon W \to W'$ such that f preserves the accessibility relation, i.e. xRy implies $f(x)R'f(y)$, and whenever $f(x)R'y$ there is a node $y' \in W$ such that xRy' and $f(y') = y$. Notice that p-morphisms are special kind of so called bisimulations – [47].

In monography [13] chapt. I, § 3 p-morphisms are called strongly isotone mappings or strong homomorphisms and such mapping can be characterized ([13], Proposition 3.3) as mappings satisfying the condition:

For any $x \in W$, there holds $R'\big(f(x)\big) = f\big(R(x)\big)$. In words – the f-image of principal R-end generating by the node x equals to the R'-end generated by the image $f(x)$.

Let us recall a realization theorem which is crucial for the main result of this contribution. By $SI(A,p)$ we denote the moniod of all strong endomorphisms of the quasi-ordered set (A,p), i.e. $\varphi \in SI(A,p)$ whenever $\varphi(p(x)) = p(\varphi(x))$ for any $x \in A$.

Let $A \neq \emptyset$ be an infinite set and $f\colon A \to A$ be a mapping. Thus (A,f) is an infinite mono-unary algebra, i.e. a unar – [44, 45]. It can be shown easily — see [13], p.23, that the relation $\sim_f$ on A defined by $x \sim_f y$ if and only if there exists a pair of nonnegative integers $m, n \in \mathbb{N}_0$ such that $f^m(x) = f^n(y)$ is an equivalence (called also the Kuratowski-Whyburn equivalence in literature). Here, as usually, $\mathbb{N}$ stands for the set of positive integers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and $\mathbb{Z}$ denotes the set of all integers. The classes of this equivalence are called orbits of f. The transformation f is connected if it has the only orbit. The following notation is overtaken from [13].

Let f have only one orbit; for the transformation f we introduce realization types denoted $\mathrm{ret}\,(f)$ and defined as there follows:

The transformation f is of realization type

1. $\mathrm{ret} = \tau_1$ if the set X has one element, i.e., (X,f) is a loop;
2. $\mathrm{ret} = \tau_2$ if (X,f) is two elements cyclic unar, i.e., $X = \{a,b\}$, $f(a) = b$, $f(b) = a$;
3. $\mathrm{ret} = \tau_3$ if f is a constant mapping and $\operatorname{card} X \geq 2$;
4. $\mathrm{ret} = \tau_4$ if $(X,f) \cong (\mathbb{Z}, v)$ where v is an unary operation $v\colon \mathbb{Z} \to \mathbb{Z}$ defined as follows: $v\colon (m) = m+1$ for each $m \in \mathbb{Z}$;
5. $\mathrm{ret} = \tau_5$ if f is an acyclic surjection which is not a bijection;
6. $\mathrm{ret} = \tau_6$ if $\big(f(X), f \restriction f(X)\big) \cong (\mathbb{Z}, v)$ and $f(X) \neq X$;
7. $\mathrm{ret} = \tau_7$ if $\big(f(X), f \restriction f(X)\big) \cong (\mathbb{N}, v)$ and $f^{-1}\big(f(x)\big) = X \backslash f(X)$ for any $x \in X \backslash f(X)$.

If the transformation f is not of any type $\tau_1, \ldots, \tau_7$ we put $\mathrm{ret}(f) = \tau_0$. Let the transformation f is not connected and $(X,f) = \sum\limits_{\alpha \in A} (X_\alpha, f_\alpha)$ is its orbital decomposition. Then we set $\mathrm{ret}(f) = \sum\limits_{i=0}^{7} \kappa_i \tau_i$, where κ_i is a cardinal number of all components (X_α, f_α) of unar (X,f), for which $\mathrm{ret}(f_\alpha) = \tau_i$ or $\kappa_i = 0$ if $\big\{\alpha \in A; \mathrm{ret}(f_\alpha) = \tau_i\big\} = \emptyset$.

Example 1 *Let $p\colon \mathbb{R} \to \mathbb{R}$, $q\colon \mathbb{R} \to \mathbb{R}$ be functions defined as follows: $p(x) = 2x+1$, $q(x) = x^2$ for all $x \in \mathbb{R}$. It is easy to verify that* $\mathrm{ret}(p) = 2^{\aleph_0}\tau_4 = c\tau_4$, $\mathrm{ret}(q) = \tau_1 + \tau_3 + 2^{\aleph_0}\tau_6 = \tau_1 + \tau_3 + c\tau_3$ *holds.*

Example 2 *As usually, let us denote by $\mathbb{R}_2[x]$ the three-dimensional space of all real polynomials of at most second degree. Define a linear first-order differential operator $L\colon \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ by $L(f) = \frac{df}{dx} + f$ for any polynomial $f \in \mathbb{R}_2[x]$. On pages 187, 188 in the book [13] there is verified that* $\mathrm{ret}(L) = 2^{\aleph_0}\tau_1 + 2^{\aleph_0}\tau_4 = c\tau_1 + c\tau_4$. *Moreover, it follows from analysis of the orbital structure of the operator $L\colon \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ that it is conjugated to the function $\varphi\colon \mathbb{R} \to \mathbb{R}$ defined by $\varphi(x) = x$ for $x \in \mathbb{R}$, $x < 0$ and $\varphi(x) = x^2$, $x \in \mathbb{R}$, $x \geq 0$, i.e. mono-unary algebras $(\mathbb{R}_2[x], L)$, $(\mathbb{R}, \varphi)$ are isomorphic. Thus* $\mathrm{ret}(L) = \mathrm{ret}(\varphi)$. *Notice, that literature devoted to this field is [44, 45, 46].*

For an unar (A, f) we define a binary relation $p_f \subset A \times A$ in this way: For $x, y \in A$ there is $[x, y] \in p_f$ whenever there exists a non-negative integer $n \neq 1$ such that $y = f^n(x)$. It is easy to see that p is a reflexive and transitive relation on A, i.e., a quasi-ordering. Moreover the relation p_f is antisymmetric, i.e. an ordering on A if and only if the transformation $f\colon A \to A$ has at most one-element cycles.

The following theorem is contained in [13] Theorem 5.1 and Theorem 7.1.

Theorem 3 *Let $f\colon A \to A$ be a mapping of the realization type* $\mathrm{ret}(f) = \sum\limits_{i=0}^{7} \kappa_i \tau_i$. *Let T be a tolerance relation on the set $\mathbb{N}(8)$ $(= \{1, 2, \ldots, 7\}$, which is a reflexive and symmetrical cover of the binary relation*

$$\big\{[n, 1]; n \in \mathbb{N}(8)\big\} \cup \big\{[2, 3]\big\} \cup \big\{[4, m]; m = 5, 6, 7\big\}\,.$$

The following conditions are equivalent:

1° *The coefficient $\kappa_0 = 0$ and for each pair of non-zero coefficient κ_i, κ_j from the combination $\sum\limits_{i=0}^{7} \kappa_i \tau_i$ we have $[i, j] \in T$.*

2° *There holds* $\mathrm{End}\,(A, f) = SI(A, p_f)$.

3° *There exists a preorder $p \subset A \times A$ (i.e. a reflexive and transitive binary relation p on the set A) such that* $\mathrm{End}\,(A, f) = SI(A, p)$.

4° *There exists a binary hyperoperation $*\colon A \times A \to \mathcal{P}(A)$ such that $(A, *)$ is a commutative extensive hypergroup with the property* $\mathrm{End}\,(A, f) = \mathrm{Gend}(A, *)$.

Now recall some basic concepts from the hyperstructure theory and some facts about linear ordinary second order differential operators.

Hypergroups and in particular join spaces play an important role in theories of various mathematical structures and their applications. The concept of a join space was introduced by Walter Prenowitz and used by him and James Jantosciak to reconstruct several branches of geometry — [30], [31], [32]. The other fields of applications of join spaces are lattices, graphs, ordered sets and automata. Noncommutative join spaces form an interesting subclass of the class of transposition hypergroups which satisfies a postulated property of transposition [30], [31]. More precisely, if H is a set, $\mathcal{P}(H)$ is the family of all subsets of H then a mapping $*\colon H \times H \to \mathcal{P}(H)$ is called a hyperoperation or join operation in H and the pair $(H, *)$ is said to be a hypergroupoid. The join operation is extended to subsets of H in a natural way, so that for $\emptyset \neq A \subset H$, $\emptyset \neq B \subset H$ the hyperproduct $A * B$ is given by $A * B = \bigcup\{a * b; a \in A, b \in B\}$. The relational notation $A \approx B$ (read A meets B) is used to assert that the sets A and B have nonempty intersection.

In H two hypercompositions right extension "/" and left extension "\" each being an inverse to $*$ are defined by $a/b = \{x; a \in x * b\}$ and $b\backslash a = \{x, a \in b * x\}$. Hence $x \approx a/b$ if and only if $a \approx x * b$ and $x \approx b\backslash a$ if and only if $a \approx b * x$.

Now a hypergroupoid $(H, *)$ is called a hypergroup if it satisfies these axioms:

1. $a * (b * c) = (a * b) * c$ for all $a, b, c \in H$ (Associativity),

2. $a * H = H = H * a$ for all $a \in H$ (Reproduction).

Moreover a hypergroup $(H, *)$ is called a transposition hypergroup or a noncommutative join space if

3. $b\backslash a \approx c/d$ implies $a * d \approx b * c$ for all $a, b, c, d \in H$ (Transposition).

Notice that the hypergroups $(H, *)$ is said to be extensive if $\{a, b\} \subset a * b$ for any pair $a, b \in H$.

By a *quasi-ordered semigroup* we mean a triple $(G, \bullet, \leq)$, where $(G, \bullet)$ is a semigroup and binary relation $\leq$ is a quasi-ordering (i.e is reflexive and transitive) on the set G such that for any triple $x, y, z \in G$ with the property $x \leq y$ also $x \bullet z \leq y \bullet z$ and $z \bullet x \leq z \bullet y$ hold. By an *ordered* (*semi*) *group* we mean (as usually) a triple $(G, \bullet, \leq)$, where $(G, \bullet)$ is a (semi)group and $\leq$ is a reflexive, anti-symmetrical and transitive binary relation on G such that for any triple $x,\ y,\ z \in G$ property $x \leq y$ also $x \bullet z \leq y \bullet z$ and $z \bullet x \leq z \bullet y$ hold. By an inclusion homomorphism we mean a mapping $f\colon (G, \bullet_G) \to (H, \bullet_H)$ such that $f(a \bullet_G b) \subset f(z) \bullet_H f(b)$ for all pairs $a,\ b \in G$ and by $\mathrm{Gend}(H, *)$ we mean the monoid of all good endomorphisms of the hypergroupoid $(H, *)$, i.e $\varphi \in \mathrm{Gend}(H, *)$ if and only if $\varphi(a * b) = \varphi(a) * \varphi(b)$. If the equalities hold instead of inclusions the corresponding morphism is termed as a good homomorphism. A bijective good homomorphism is an isomorphism.

Application of algebraic topological and geometrical methods of investigation of ordinary differential equations and thier transformations belongs to characteristic approaches of the school founded by Professor Otakar Borůvka — [5], [38] – [43]. The outstanding representative of the mentioned school Professor František Neuman wrote in his paper [40]: "Algebraic, topological and geometrical tools together with the methods of the theory of dynamical systems and functional equations make it possible to deal with problems concerning global properties of solutions by contrast to the previous local investigations and isolated results." Influence of mentioned ideas is a certain motivating factor of our investigations.

So, we consider linear ordinary differential operators of the form

$$L_n = \sum_{k=0}^{n} p_k(x) D^k\,,$$

where $D_k = \frac{d^k}{dx^k}$, $p_k(x)$ is a continuous function on some open interval $J \subset \mathbb{R}$, $k = 0, 1, \ldots, n-1$, $p_n(x) \equiv 1$, created equations $L_n(y) = 0$ which are linear homogeneous ordinary differential equations of the form

$$y^{(n)}(x) + \sum_{k=0}^{n-1} p_k(x)\, y^{(k)}(x) = 0\,.$$

As usually, $\mathbb{R}$ stands for the set of all reals, $J \subset \mathbb{R}$ is an open interval (bounded or unbounded) of real numbers, $\mathbb{C}^k(J)$ is the ring (with respect to usual addition and multiplication of functions) of all real functions with continuous derivatives up to the order $k \geq 0$ including. We write $\mathbb{C}(J)$

instead of $\mathbb{C}^0(J)$. For a positive integer $n \geq 2$ we denote by $\mathbb{A}_n$ the set of all linear homogeneous differential equations of the n-th order with continuous real coefficients on J, i.e.

$$y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_0(x)y = 0\,,$$

(cf. [38] – [43]), where $p_k \in \mathbb{C}(J)$, $k = 0, 1, \ldots, n-1$, $p_o(x) > 0$ for any $x \in J$ (this is not essential restriction). Denote $L(p_0, \ldots, p_{n-1})\colon \mathbb{C}^n(J) \to \mathbb{C}^n(J)$ the above defined linear operator defined by

$$L(p_0, \ldots, p_{n-1})(y) = y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_0(x)y$$

and put

$$\mathbb{LA}_n(J) = \big\{L(p_0, \ldots, p_{n-1}); p_k \in \mathbb{C}(J), p_0 > 0\big\}\,.$$

Further $\mathbb{N}_0(n) = \{0, 1, \ldots, n-1\}$ and δ_{ij} stands for the Kronecker δ, $\overline{\delta_{ij}} = 1 - \delta_{ij}$. For any $m \in \mathbb{N}_0(n)$ we denote by $\mathbb{LA}_n(J)_m$ the set of all linear differential operators of the n-th order $L_0(p_0, \ldots, p_{n-1})\colon \mathbb{C}^n(J) \to \mathbb{C}(J)$, where $p_k \in \mathbb{C}(J)$ for any $k \in \mathbb{N}_0(n)$, $p_m \in \mathbb{C}_+(J)$, (i.e. $p_m(x) > 0$ for each $x \in J$). Using the vector notation $\vec{p}(x) = \big(p_0(x), \ldots, p_{n-1}(x)\big)$, $x \in J$ we can write $L_n(\vec{p})y = y^{(n)} + \big(\vec{p}(x), (y, y', \ldots, y^{(n-1)})\big)$ (i.e. a scalar product).

We define a binary operation "$\circ_m$" and a binary relation "$\leq_m$" on the set $\mathbb{LA}_n(J)_m$ in this way:

For arbitrary pair $L(\vec{p})$, $L(\vec{q}) \in \mathbb{LA}_n(J)_m$, $\vec{p} = (p_0, \ldots, p_{n-1})$, $\vec{q} = (q_0, \ldots, q_{n-1})$ we put $L(\vec{p}) \circ_m L(\vec{q}) = L(\vec{u})$, $\vec{u} = (u_0, \ldots, u_{n-1})$, where

$$u_k(x) = p_m(x)\, q_k(x) + (1 - \overline{\delta_{km}})\, p_k(x), \quad x \in J$$

and $L(\vec{p}) \leq_m L(\vec{q})$ whenever $p_k(x) \leq q_k(x)$, $k \in \mathbb{N}_0(n)$, $p_m(x) = q_m(x)$, $x \in J$.

Evidently, $\big(\mathbb{LA}_n(J)_m, \leq_m\big)$ is an ordered set. The paper [17] contains a sketch of the proof of the following lemma.

Lemma 4 *The triad* $(\mathbb{LA}_n(J)_m, \circ_m \leq_m)$ *is an ordered (noncommutative) group.*

Denote by $\mathcal{P}\big(\mathbb{LA}_n(J)_m\big)^*$ the power set of $\mathbb{LA}_n(J)_m$ consisting of all nonempty subsets of the last set and define a binary hyperoperation

$$*_m\colon \mathbb{LA}_n(J)_m \times \mathbb{LA}_n(J)_m \to \mathcal{P}\big(\mathbb{LA}_n(J)_m\big)^*$$

by the rule

$$L(\vec{p}) *_m L(\vec{q}) = \big\{L(\vec{u}); L(\vec{p}) \circ_m L(\vec{q}) \leq_m L(\vec{u})\big\}$$

for all pairs $L(\vec{p})$, $L(\vec{q}) \in \mathbb{LA}_n(J)_m$. More in detail, if $\vec{u} = (u_0, \ldots, u_{n-1})$, $\vec{p} = (p_0, \ldots, p_{n-1})$, $\vec{q} = (q_0, \ldots, q_{n-1})$, then $p_m(x)q_m(x) = u_m(x)$, $p_m(x)q_k(x) + p_k(x) \leq u_k(x)$ if $k \neq m$, $x \in J$. From results of [17] or from [13], Theorems 1.3 and 1.4 it follows that $\big(\mathbb{LA}_n(J)_m, *_m\big)$ is a (noncommutative) hypergroup.

Moreover, there holds the following Theorem [17] p. 283.

Theorem 5 *Let $J \subset \mathbb{R}$ be an open interval, $2 \leq n$ be a positive integer. Let*

$$\mathbb{LA}_n(J)_m = \left\{ L(p_0, \dots, p_{n-1}); p_k \in \mathcal{C}(J), p_m > 0 \right\}.$$

Then the hypergroup $(\mathbb{LA}_n(J)_m, *_m)$ is a transposition hypergroup (i.e a non-commutative join space) of linear ordinary differential operators of the n-th order.

By a semi-cascade is usually considered an action of the monoid $(\mathbb{N}_0, +)$ (called a phase semigroup) on a set X (called a phase set or a phase space) of this discrete dynamical system i.e. it is a triad $\big(X, (\mathbb{N}_0, +), \delta\big)$, where $\delta\colon X \times \mathbb{N}_0 \to X$ is a transition or evolution function satisfying the usual conditions:

1. $\delta(x, 0) = x$,
2. $\delta(\delta(x, m), n) = \delta(x, m + n)$,

for all $x \in X$ and m, $n \in \mathbb{N}_0$. If $f\colon X \to X$ is an arbitrary transformation then the triad $\big(X, (\mathbb{N}_0, +), \delta_f\big)$, where $\delta_f(x, m) = f^m(x)$ is called the semi-cascade determined by the unar (X, f). Notice that a semi-cascade is a certain modification of the concept of an algebraic space considered in [5].

We are going to construct a semi-cascade from the system of solution spaces of linear homogeneous ordinary differential equations. It is to be noted that from the general theory of ordinary linear differential equations there is known that there exists one-to-one correspondence between the set $\mathbb{LA}_n(J)$ and the system $\mathbb{VA}_n(J)$ of all n-dimensional solution spaces of differential equations $L(p_0, \dots, p_{n-1})y = 0$, $L(p_0, \dots, p_{n-1}) \in \mathbb{LA}_n(J)$, i.e.

$$y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_0(x)y = 0\,,$$

(cf. also results in the monography [39]). This correspondence can be used for the defining of binary hyperoperation on the system $\mathbb{VA}_n(J)$ such that this system endowed with the mentioned hyperoperation is a noncommutative semihypergroup (in particular the system $\mathbb{VA}_n(J)_m$ corresponding to the hypergroup of n-order differential operators $\mathbb{LA}_n(J)_m$ creates a transposition hypergroup, i.e. a noncommutative join space). These multistructures are constructed in papers [17, 25].

Now choose an arbitrary n-tuple $\Phi = [\varphi_1, \dots, \varphi_n] \in \mathbb{C}(J) \times \cdots \times \mathbb{C}(J) = \mathbb{C}(J)^n$ of linearly independent functions and denote by $V_0(\varphi_1, \dots, \varphi_n)$ the (at most) n-dimensional linear space with the above base $\varphi_1, \dots, \varphi_n$ over the field $\mathbb{R}$. Denote by $\mathbb{V}_1^\Phi\mathbb{A}_n$ the system of all n-dimensional solution spaces $V(p_0, \dots, p_{n-1})$ of differential equations $L(p_0, \dots, p_{n-1})y = 0$, where $p_k \in V_0(\varphi_1, \dots, \varphi_n)$, $k = 0, 1, \dots, n-1$, i.e.

$$\mathbb{V}_1^\Phi\mathbb{A}_n = \left\{ V^{(1)}(p_0, \dots, p_{n-1}); p_k \in V_0(\varphi_1, \dots, \varphi_n), k = 0, 1, \dots, n-1 \right\}.$$

Further, let $\mathbb{V}_m^\Phi\mathbb{A}_n$ be defined. Then define

$$\begin{aligned}\mathbb{V}_{m+1}^\Phi\mathbb{A}_n = \Big\{ V(q_0, \dots, q_{n-1}); q_k \in V^{(m)}(u_0, \dots, u_{n-1}), k = 0, 1, \dots, n-1,\\ V^{(m)}(u_0, \dots, u_{n-1}) \in \mathbb{V}_m^\Phi\mathbb{A}_n \Big\}.\end{aligned}$$

Put $\mathbb{T}^{\Phi}\mathbb{A}_n = \big\{V_0(p_0,\dots,p_{n-1})\big\} \underset{m\in\mathbb{N}}{\cup} \mathbb{V}^{\Phi}_m\mathbb{A}_n$ and consider a sequence of n-dimensional vector spaces of continuous functions

$$V_0(p_0,\dots,p_{n-1}) = V^{\Phi}_0(J), V^{\Phi}_1(J_1),\dots,V^{\Phi}_m(J_m),\dots$$

where $\{J_k; k\in\mathbb{N}_0\}$ is a sequence of open intervals $J_0 = J$, $J_{k+1}\subseteq J_k$, $J_{k+1}\neq J_k$ such that $\bigcap\limits_{k\in\mathbb{N}_0} J_k$ is an open interval, and the space $V^{\Phi}_{m+1}(J_{m+1})$ is formed by all continuous functions $f\colon J_{m+1}\to\mathbb{R}$ which are restrictions of functions $g\in V^{\Phi}_m(J_m)$, $m\in\mathbb{N}_0$. Further denote

$$E\mathbb{T}^{\Phi}\mathbb{A}_n = \mathbb{T}^{\Phi}\mathbb{A}_n \cup \{V^{\Phi}_m(J_m); n\in\mathbb{N}_0\}\,.$$

Define

$$\begin{gathered} F_{\Phi}\colon E\mathbb{T}^{\Phi}\mathbb{A}_n \to E\mathbb{T}^{\Phi}\mathbb{A}_n \quad \text{by} \\ F_{\Phi}(U) = V \quad \text{for all} \quad U\in\mathbb{T}^{\Phi}\mathbb{A}_n, \quad V\neq V_0(p_0,\dots,p_{n-1})\,, \end{gathered}$$

where $V\in\mathbb{V}^{\Phi}_m\mathbb{A}_n$ is such a space that for some n-tuple of functions $[u_0,\dots,u_{n-1}]$, $u_k\in\mathbb{V}^{\Phi}_m\mathbb{A}_n$ the space V is the solution space of the differential equation $L(u_0,\dots,u_{n-1})y = 0$, i.e. of the equation

$$y^{(n)} + u_{n-1}(x)y^{(n-1)} + \cdots + u_1(x)y' + u_0(x)y = 0\,.$$

Defining further

$$F_{\Phi}\big(V_0(p_0,\dots,p_{n-1})\big) = F_{\Phi}\big(V^{\Phi}_0(J)\big) = V^{\Phi}_1(J_1)$$

and

$$F_{\Phi}\big(V^{\Phi}_m(J_m)\big) = V^{\Phi}_{m+1}(J_m)\,,$$

then we obtain a mono-unary algebra (i.e. a unar) $(E\mathbb{T}^{\Phi}\mathbb{A}_n, F_{\Phi})$ with the below described "realization" property.

Further, for any n-tuple $\Phi = [\varphi_1,\dots,\varphi_n]\in\mathbb{C}(J)^n$ we construct the tree $E\mathbb{T}^{\Phi}\mathbb{A}_n$ and the mono-unary algebra $(E\mathbb{T}^{\Phi}\mathbb{A}_n, F_{\Phi})$. Put

$$(E\mathbb{T}^{\Phi}\mathbb{A}_n, F) = \sum_{\Phi\in\mathbb{C}(J)^n} (E\mathbb{T}^{\Phi}\mathbb{A}_n, F_{\Phi})\,,$$

i.e. $E\mathbb{T}^{\Phi}\mathbb{A}_n = \bigcup\limits_{\Phi\in\mathbb{C}(J)^n} E\mathbb{T}^{\Phi}\mathbb{A}_n$ with disjoint summands on the right-hand side of the equality and F_{Φ} is the restriction of the mapping $F\colon E\mathbb{T}\mathbb{A}_n\to E\mathbb{T}\mathbb{A}_n$ onto $E\mathbb{T}^{\Phi}\mathbb{A}_n$. As above we construct the semi-cascade $\big(E\mathbb{T}\mathbb{A}_n, (\mathbb{N}_0,+), \delta_F\big)$ by the putting

$$\delta_F(V,0) = V \quad \text{and} \quad \delta_F(V,m) = F^m(V)$$

for any linear space $V\in E\mathbb{T}\mathbb{A}_n$ and any integer $m\in\mathbb{N}$. Denoting by End $\big(E\mathbb{T}\mathbb{A}_n, (\mathbb{N}_0,+), \delta_F\big)$ the endomorphism monoid of the semi-cascade $\big(E\mathbb{T}\mathbb{A}_n, (\mathbb{N}_0,+), \delta_F\big)$ we get the following theorem. (Notice that a mapping $h\colon E\mathbb{T}\mathbb{A}_n\to E\mathbb{T}\mathbb{A}_n$ is an endomorphism of the semi-cascade $(E\mathbb{T}\mathbb{A}_n, (\mathbb{N}_0,+), \delta_F)$ if for any pair $[V,m]\in E\mathbb{T}\mathbb{A}_n\times\mathbb{N}_0$ there holds $h\big(\delta_F(V,m)\big) = \delta_F\big(h(V),m\big)$.)

Theorem 6 *There exists a binary hyperoperation "$*$" on the phase set $E\mathbb{TA}_n$ of the semi-cascade $(E\mathbb{TA}_n, (\mathbb{N}_0, +), \delta_F)$ with the following properties:*

1° *The hypergroupoid $(E\mathbb{TA}_n, *)$ is an extensive commutative transposition hypergroup, i.e. an extensive join space.*

2° $\mathrm{End}(E\mathbb{TA}_n, (\mathbb{N}_0, +), \delta_F) = \mathrm{Gend}(E\mathbb{TA}_n, *)$ (*it is the monoid of all good endomorphisms of the hypergroup* $(E\mathbb{TA}_n, *)$.)

Proof. Let $J \subset \mathbb{R}$ be an open interval. Since

$$\operatorname{card} \mathcal{C}(J) = \operatorname{card} \mathcal{C}(J)^n = 2^{\aleph_0} = c$$

and $\mathrm{ret}(E\mathbb{T}^\Phi\mathbb{A}_n, F_\Phi) = \tau_5$ for any n-tuple $\Phi = [p_0, p_1, \dots, p_{n-1}] \in \mathcal{C}(J)^n$ of functions we have

$$\mathrm{ret}(E\mathbb{TA}_n, F) = \mathrm{ret}\Big(\sum_{\Phi \in \mathcal{C}(J)^n} (E\mathbb{T}^\Phi\mathbb{A}_n, F_\Phi)\Big) = c \cdot \tau_5\,.$$

Then by Theorem 3 there exists a binary hyperoperation "$*$" on the set $E\mathbb{TA}_n$ such that $(E\mathbb{TA}_n, *)$ is an extensive commutative hypergroup with the property

$$\mathrm{End}(E\mathbb{TA}_n, F) = \mathrm{Gend}(E\mathbb{TA}_n, *)\,.$$

Evidently $\mathrm{End}(E\mathbb{TA}_n, F) = \mathrm{End}\big(E\mathbb{TA}_n, (\mathbb{N}_0, +), \delta_F\big)$ and from the proof of Theorem 5.1, p. 85–86 [13], i.e. Theorem 3 there follows that we can define $V_1 * V_2 = \big\{F^k(V_1); k = 2, 3, \dots\big\} \cup \big\{F^k(V_2); k = 2, 3, \dots\big\} \cup \{V_1, V_2\}$. By Theorem 6.1, p. 182, [13] the hypergroup $(E\mathbb{TA}_n, *)$ is a join space. $\square$

Remark 7 *If we consider a non-prolongated semi-cascade, say $\big(\mathbb{T}^\Phi\mathbb{A}_n, (\mathbb{N}_0, +), \delta_{F_\Phi}\big)$, where*

$$\Phi = [\varphi_1, \dots, \varphi_n] \in \mathcal{C}(J)^n, \mathbb{T}^\Phi\mathbb{A}_n = \bigcup_{m \in \mathbb{N}_0} \mathbb{V}_m\mathbb{A}_n\,,$$

$$V_0(\varphi_1, \dots, \varphi_n) = \Big\{\sum_{k=1}^{n} \lambda_k\varphi_k; [\lambda_1, \dots, \lambda_n] \in \mathbb{R}^n\Big\}\,,$$

and the mapping $F\colon \mathbb{T}^\Phi\mathbb{A}_n \to \mathbb{T}^\Phi\mathbb{A}_n$ is defined similarly as above with the only difference that the space $V_0(\varphi_1, \dots, \varphi_n)$ is its fixed point, i.e. $F\big(V_0(\varphi_1, \dots, \varphi_n)\big) = V_0(\varphi_1, \dots, \varphi_n)$ then it is not difficult to show (which also there follows immediately from Theorem 3) that for any binary hyperoperation "$\circ$" on the set $\mathbb{T}^\Phi\mathbb{A}_n$ monoids

$$\mathrm{End}\big(\mathbb{T}^\Phi\mathbb{A}_n, (\mathbb{N}_0, +), \delta_{F_\Phi}\big), \quad \mathrm{Gend}(\mathbb{T}^\Phi\mathbb{A}_n, \circ)$$

are different. Of course, here as above $\delta_{F_\Phi}(V, m) = F_\Phi^m(V)$ for any pair $[V, m] \in \mathbb{T}^\Phi\mathbb{A}_n \times \mathbb{N}_0$.

Remark 8 *Semi-cascades or cascades (actions of the group of all integres on phase spaces) are special types of infinite automata called quasi-automata or automata without outputs. This is why in References are papers belonging to the mentioned field — [1, 2, 11, 15, 16, 18, 19, 20, 23, 24, 25, 27, 29, 33, 34, 35, 36, 37].*

Acknowledgement

The first author was supported by the Council of Czech Government grant No. MSM0021630529.

References

[1] ASHRAFI, A. R., GHORBANI, M.: *A note on the general action of a hypergroup on a set.* In Abstracts of the 9th Int. Conference on AHA, Babolsar, 2005.

[2] BAVEL, Z.: *The source as a tool in automata.* In: Information and Control, Vol. 18 (1971), 140–155.

[3] BERÁNEK, J., CHVALINA, J.: *From groups of linear functions to noncommutative transposition hypergroups.* In: Dept. Math. Report Series Vol. 7, Univ. South Bohemia, České Budějovice, 1999, 1–10.

[4] BERÁNEK, J., CHVALINA, J.: *On a non-commutative transposition hypergroup of linear spaces of smooth functions of the dimension two.* Acta Math. 11, Fac. Nat. Sci., Konstantine the Philosopher Univ. Nitra 2008, 17–23 (in Czech).

[5] BORŮVKA, O.: *Algebraic spaces and their realization by differential equations I, II.* Text of "Seminář o diferenciálních rovnicích", PřF UJEP, 1988, 35pp (in Czech).

[6] CORSINI, P.: *Prolegomena of Hypergroup Theory.* Tricesimo: Aviani editore 1993.

[7] CORSINI, P., LEOREANU, V.: *Applications of Hyperstructure Theory.* Dordrecht–Boston–London, Kluwer Academic Publishers, 2003.

[8] DOMIATY, R. Z.: *Remarks to the problem of defining a topology by its homomorphism group.* General Topology and its Relations to Modern Analysis and Algebra V, Proc. of the Fourth Prague Topological Symposium (J. Novák, ed.). Soc. Czechoslovak Mathematicians and Physicists Prague 1977.

[9] EWERETT, C. J., ULAM, S. M.: *On the problem of determination of mathematical structures by their endomorphisms.* Bull. Amer. Math. Soc. 54 (1948), 646.

[10] GÉCSEK, F., PEÁK, I.: *Algebraic Theory of Automata.* Budapest, Akadémiai Kiadó, 1972.

[11] HOŠKOVÁ, Š.: *Discrete transformation hypergroups.* Proc. of 4th International Conference Aplimat, Bratislava (2005), 275–279.

[12] HOŠKOVÁ, Š., CHVALINA, J.: *Join spaces of first-order linear partial differential operators with compatible proximity induced by a congruence on their group.* Mathematical and Computer Modelling in Science and Engineering. Proc. Internat. Conf. Czech. Tech. Univ. in Prague, 2003, 166–170.

[13] CHVALINA, J.: *Functional Graphs, Quasi Ordered Sets and Commutative Hypergroups.* Brno, Masaryk University, 1995 (in Czech).

[14] CHVALINA, J., CHVALINOVÁ, L.: *Join spaces of linear ordinary differential operators of the second order.* Colloqium on Differential and Difference Equations, CDDE 2002. In: Folia Fac. Sci. Nat. Univ. Masarykianae Brunesis, Mathematica 13, MU Brno, 2003, 77–86.

[15] CHVALINA, J., CHVALINOVÁ, L.: *State hypergroups of automata.* Acta Mathematica et Informatica Universitatis Ostraviensis 4, 1996, 105–120.

[16] CHVALINA, J., CHVALINOVÁ, L.: *Transposition hypergroups formed by transformation operators on rings of differentiable functions.* Italian J. Pure Appl. Math., No. 15 – January 2004, 93–106.

[17] CHVALINA, J., CHVALINOVÁ, L.: *Modelling of join spaces by n-th order linear ordinary differential operators*. Proc. of the 4th International Conference Aplimat, Dept. Math. FME Slovak Univ. of Technology, Bratislava (2005), 279–284.

[18] CHVALINA, J., HOŠKOVÁ, Š.: *Action of hypergroups of the first-order partial differential operators*. 6th International Conference Aplimat, Dept. Math. FME Slovak Univ. of Technology, Bratislava (2007), 177–184.

[19] CHVALINA, J. *Infinite automata with phase hypergroups of various operators.* Proc. 10th Internat. Congress on AHA, Univ. of Defense, Brno 2008, 1–14 (in print).

[20] CHVALINA, J., MOUČKA, J.: *Actions of join spaces of continuous functions on hypergroups of second-order linear differential operators*. Proc. 6-th Math. Workshop with International Attendance, FCI Univ. of Tech. Brno 2007, CD-ROM, 10 pp.

[21] CHVALINA, J., MOUČKA, J.: *Kripke-type compatibility of transformations of finite sets of alternatives with cognitive and incomplete preferences*. Proc. XXVI Internat. Colloquium on the Management of Educational Process, Univ. of Defense Brno 2008, CD-ROM 7pp.

[22] CHVALINA, J., MOUČKA, J.: *Kripke-type compatibility of transformations of general sets of alternatives with cognitive and incomplete preferences*. Proc. XXVII Internat. Colloquium on the Management of Educational Process, Univ. of Defense, Brno 2009, CD-ROM 6pp.

[23] CHVALINA, J., MOUČKA, J., NOVÁK, M.: *Actions of centralizer semihypergroups of certain transformation operators on rings of functions of a complex variable.* XXV. Internat. Colloquium on the Management of Educational Process, Proc. of Contributions, University of Defense, Brno, 2007.

[24] CHVALINA, J., MOUČKA, J., VÉMOLOVÁ, R.: *Functorial passage from quasi-automata to multiautomata*. XXIV Internat. Colloquium on the Management of Educational Process, Proc. of Contributions, University of Defense, Brno, 2006, 8pp.

[25] CHVALINA, J., CHVALINOVÁ, L.: *Actions of centralizer hypergroups of n-th order linear differential operators on rings of smooth functions*. J. Appl. Math. Aplimat, vol. 1, n. 1 (2008), 45–53.

[26] CHVALINA, J., CHVALINOVÁ, L., NOVÁK, M.: *On certain tree of two-dimensional function spaces growing from a linear space of continuous functions*. Proc. XXVII Internat. Colloquium on the Management of Educational Process, Univ. of Defense Brno 2009, CD-ROM 11pp.

[27] CHVALINA, J., NOVÁK, M.: *On Levine's decomposition of the power phase set of a semihypergroup act.* 5th International Conf. Aplimat, Part II, Dept. Math. FME Slovak Univ. of Technology, Bratislava 2006, 89–94.

[28] CHVALINA, J., NOVÁK, M.: *Minimal countable extension of a certain cascade with the ordered phase-set and strongly isotone endomorphisms*. 6-th Conf. on Math. and Physics at Technical Universities, Proc. of Contributions. Univ. of Defense, Brno 2009, 131–138.

[29] CHVALINA, J., RAČKOVÁ, P.: *Join spaces of smooth functions and their actions on transposition hypergroups of second order linear differential operators*. J. Appl. Math. Aplimat (STU Bratislava) **1** No. 1 (2008), 55-63.

[30] JANTOSCIAK, J.: *A brief survey of the theory of join spaces.* (M. Stefanescu ed.) Proc. 5th Internat. Congr. on AHA, Jasi, Rumania 1993, Hadronic Press, Palm Harbor USA (1994), 1–12.

[31] JANTOSCIAK, J.: *Transposition in hypergroups.* Algebraic Hyperstructures and Appl. Proc. 6th Internat. Congrress Prague 1966, Alexanroupolis, Democritus Univ. of Thrace Press, 1977, 77–84.

[32] JANTOSCIAK, J.: *Transposition hypergroups: noncommutative join spaces.* J. Algebra 187 (1997), 97–119.

[33] MADANSHEKAF, A., REZA ASHRAFI, A.: *Generalized action of a hypergroup on a set.* Italian J. Pure and Appl. Math., No. 3 (1998), 127–135.

[34] MASSOUROS, G. G.: *An automaton during its operation.* Algebraic Hyperstructures and Appl. Proc. 5th Internat. Congress Iaşi 1993, Hadronic Press, Palm Harbor USA (1994), 267–276.

[35] MASSOUROS, G. G.: *Hypercompositional structures in the theory of the languages and automata.* Analele Ştiinţifice ale Universităţii "Al I. Cuza" Iaşi , Tom. III, Informatică, 1994, 65–73.

[36] MASSOUROS, G. G., MITTAS, J.: *Languages — automata and hypercompositional structures.* Proc. of 4th Internat. Congress AHA (1990), World Scientific, 137–147.

[37] MOUČKA, J.: *Applied Structures for Discrete Modelling Processes.* Habilitation Thesis VVŠ PV Vyškov 2002 (in Czech).

[38] NEUMAN, F.: *From local to global investigations of linear differential equations of the n-th order.* Jahrbuch Überblicke Mathematik 1984, 55–80.

[39] NEUMAN, F.: *Global Properties of Linear Ordinary Differential Equations.* Mathematics and its Applications, East European Series 52, Kluwer Academic Publishers (with Academia Praha) Dordrecht–Boston–London 1991.

[40] NEUMAN, F.: *Ordinary differential equations – a survey of the global theory.* Equadiff 6, Proc. Internat. Conf. Differential Equations and their Appl., Brno 1985, 59–70.

[41] NEUMAN, F.: *Global theory of ordinary linear homogeneous differential equations in the real domain.* I Aequationes Math. 33 (1987), 123–149, II ibid. 34 (1987), 1–22.

[42] NEUMAN, F.: *Transformation theory of linear ordinary differential equations — from local to global investigations.* Arch. Math. (Brno) 33/1–2 (1997), 65–74.

[43] NEUMAN, F.: *Algebraic aspects of transformations with an application to differential equations.* Nonlinear Analysis 40 (2000), 505–511.

[44] NOVOTNÝ, M.: *Abbildung von Mengen.* Pacific J. Math. 13 (1963), 1359–1369.

[45] NOVOTNÝ, M.: *Mono-unary algebras in the work of Czechoslovak mathematicians.* Arch. Math. (Brno) 26(1990), 155-164.

[46] TARGONSKI, Gy.: *Topics in Iteration Theory.* Vandenhoeck et Ruprecht, Göttingen-Zürich 1981.

[47] Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Kripke-semantics, 1–9, http://en.wikipedia.org/wiki/Kripke-structure

Current address

Chvalina Jan, Prof., RNDr., DrSc.
Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Mathematics, Technická 8, 616 00 Brno, Czech Republic, e-mail: chvalina@feec.vutbr.cz

Chvalinová Ludmila, RNDr., CSc.
Brno University of Technology, Faculty of Mechanical Engineering, Department of Mathematics, Technická 2, 616 69 Brno, Czech Republic, e-mail: chvalinova@fme.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

SOLVABILITY CONCEPTS FOR INTERVAL SYSTEMS IN MAX-PLUS ALGEBRA

MYŠKOVÁ Helena, (SK)

Abstract. In the algebraic structure $(B, \oplus, \otimes)$, where

$$B = R \cup \{-\infty\},\ a \oplus b = \max\{a, b\},\ a \otimes b = a + b,$$

the notation $\boldsymbol{A} \otimes x = \boldsymbol{b}$ represents an interval system of linear max-plus equations, where $\boldsymbol{A} = \langle \underline{A}, \overline{A} \rangle$ and $\boldsymbol{b} = \langle \underline{b}, \overline{b} \rangle$ are given matrix interval and vector interval, respectively. Several types of solvability of interval systems are known. We summarize knowledge about them and deal with relations among particular solvability concepts. In conclusions, we describe the set of all solvability concepts by Hasse diagram.

Key words and phrases. max-plus algebra, interval system, solvability concepts

Mathematics Subject Classification. 15A06, 65G30.

1 Introduction

The last decades have seen a lot of attention given to study of simple systems of linear equations in the form $A \otimes x = b$, where A is a matrix, b and x are vectors of suitable dimensions, and one or both of classical addition, and multiplication operations are replaced by maximum and/or minimum. If addition and multiplication are replaced by maximum and addition, respectively, we call this algebraic structure the *max-plus algebra*. One of questions, which we can deal with in the max-plus algebra, is solving systems of linear equations. Systems of linear equations over the max-plus algebra are used in several branches of applied mathematics. They can assist in modelling and analysis of discrete event systems. Among interesting real-life applications let us mention, e.g., a large scale model of Dutch railway network or synchronizing traffic lights in Delft [9].

However, when the matrix and vector entries are estimated incorrectly, the obtained theoretical results may become useless in practice, due to imprecise results. A possible method of restoring solvability is to replace matrix A and vector b by a matrix interval and a vector interval. Then we talk about an interval system of linear equations. The theory of interval computations and in particular of interval systems in the classical algebra is already quite developed, see, e.g., the monograph [5] or [10, 11]. In the max-plus algebra, interval systems of linear equations have been studied by K. Cechlárová and R. A. Cuninghame-Green [1, 3]. They dealt with the weak, strong and tolerance solvability. An extension of their work are our papers [6] -[8].

2 Preliminaries

Let $(B, \oplus, \otimes)$ be an algebraic structure with two binary operations. $(B, \oplus, \otimes)$ is called the *max-plus algebra*, if

$$B = R \cup \{\varepsilon\},\ a \oplus b = \max\{a, b\},\ a \otimes b = a + b,$$

where $\varepsilon = -\infty$.

Let m,n be given positive integers. Denote by M, N the sets of indices $\{1, 2, \dots, m\}$, $\{1, 2, \dots, n\}$, respectively. The set of all $m \times n$ *matrices* over B is denoted by $B(m, n)$ and the set of all column *n-vectors* over B by $B(n)$.

If we multiply a matrix $A \in B(m, n)$ by some vector $x \in B(n)$ we get $[A \otimes x]_i = \max_{j \in N}\{a_{ij} + x_j\}$. We shall consider the *ordering* $\leq$ on the sets $B(m, n)$ and $B(n)$ defined as follows:

- for $A, B \in B(m, n):\ A \leq B$ if $a_{ij} \leq b_{ij}$ for all $i \in M,\ j \in N$,
- for $x, y \in B(n):\ x \leq y$ if $x_j \leq y_j$ for all $j \in N$.

It is easy to see that for each $A, B \in B(m, n)$ and for each $x, y \in B(n)$ holds:

$$\text{if } A \leq B \text{ and } x \leq y, \text{ then } A \otimes x \leq B \otimes y.$$

We call this property the monotonicity of the operation $\otimes$.

For the given matrix interval $\boldsymbol{A} = \langle \underline{A}, \overline{A} \rangle$ with $\underline{A}, \overline{A} \in B(m, n),\ \underline{A} \leq \overline{A}$ and the given vector interval $\boldsymbol{b} = \langle \underline{b}, \overline{b} \rangle$ with $\underline{b}, \overline{b} \in B(m),\ \underline{b} \leq \overline{b}$ the notation

$$\boldsymbol{A} \otimes x = \boldsymbol{b} \tag{1}$$

represents the set of all systems of linear max-plus equations of the form

$$A \otimes x = b \tag{2}$$

such that $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$.

The set $\boldsymbol{A} \otimes x = \boldsymbol{b}$ will be called an *interval system* of max-plus linear equations. Each system of the form (2) is said to be a *subsystem* of system (1), if $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$. We say, that an interval system has a *constant matrix* if $\underline{A} = \overline{A}$ and has a *constant right-hand side*, if $\underline{b} = \overline{b}$. A subsystem is called *extremal*, if each equation has the form $[\underline{A} \otimes x]_i = \overline{b}_i$ (*LU equation*) or $[\overline{A} \otimes x]_i = \underline{b}_i$ (*UL equation*).

The crucial role for the solvability of (2) plays the so called *principal solution*. Before its definition, we add some conditions.

At first, we shall suppose that $b_i > \varepsilon$ for all $i \in M$ in (2). To justify this assumption, we show how to get rid of ε-s. Namely, denote by $M_0 = \{i \in M;\ b_i = \varepsilon\}$. Then any solution x of (2) has $x_j = \varepsilon$ for each $j \in N_0$, where $N_0 = \{j \in N; a_{ij} \neq \varepsilon$ for some $i \in M_0\}$. Therefore it is possible to omit the equations with indices from M_0 and columns of A with indices from N_0 and the solutions of the original and reduced systems correspond to each other by setting $x_j = \varepsilon$ for $j \in N_0$ in the former.

Secondly, we shall suppose that$A$ does not contain a column with full ε-s. Namely, denote by $N_1 = \{j \in N;\ a_{ij} = \varepsilon$ for each $i \in M\}$. Then x_j can be arbitrary for each $j \in N_1$. Therefore it is possible to omit the columns of A with indices from N_1 and the solutions of the original and reduced systems correspond to each other by setting $x_j = x,\ x$ for all $j \in N_1$, where x is an arbitrary element from B.

By now, we can use the definition of the principal solution as follows:

$$x_j^*(A, b) = \min_{i \in M}\{b_i - a_{ij}\} \tag{3}$$

for each $j \in N$. The following assertions describe the importance of the principal solution for the solvability of (2).

Lemma 2.1 [4, 13] *Let $A \in B(m, n)$ and $b \in B(m)$ be given.*

i) If $A \otimes x = b$ for $x \in B(n)$, then $x \leq x^*(A, b)$.

ii) $A \otimes x^*(A, b) \leq b$.

Theorem 2.2 [3, 4] *Let $A \in B(m, n)$ and $b \in B(m)$ be given. Then the system $A \otimes x = b$ is solvable if and only if $x^*(A, b)$ is its solution.*

To use the above arguments we shall suppose for interval system (1) that

- $\underline{b}_i \neq \varepsilon$ for each $i \in M$,
- for each $j \in N$ there exists $i \in M$ such that $\underline{a}_{ij} \neq \varepsilon$.

3 Solvability concepts

We shall consider over the solvability of interval system on the ground of the solvability of its subsystems. If we ask for the solvability of at least one subsystem we say about the *weak solvability* which has been studied by K. Cechlárová [1]. K. Cechlárová and R. A. Cuninghame-Green [3] dealt with *the strong solvability* which requires solvability of all subsystems. Another possibility studied in [1] is the *tolerance solvability* which asks for the existence a vector $x \in B(n)$ such that for each $A \in \boldsymbol{A}$ the product $A \otimes x$ belongs $\boldsymbol{b}$. In this way we can define various solvability concepts. Table 1 contains the list of all up to now defined types of a solvability in the max-plus algebra. Some of them were defined in the classical algebra [10].

Table 1

Solvability concept	**Condition**
Weak solvability	$(\exists x \in B(n))(\exists A \in \boldsymbol{A})(\exists b \in \boldsymbol{b}): A \otimes x = b$
Strong solvability	$(\forall A \in \boldsymbol{A})(\forall b \in \boldsymbol{b})(\exists x \in B(n)): A \otimes x = b$
Tolerance solvability	$(\exists x \in B(n))(\forall A \in \boldsymbol{A})(\exists b \in \boldsymbol{b}): A \otimes x = b$
Weak tolerance solvability	$(\forall A \in \boldsymbol{A})(\exists x \in B(n))(\exists b \in \boldsymbol{b}): A \otimes x = b$
Control solvability	$(\exists x \in B(n))(\forall b \in \boldsymbol{b})(\exists A \in \boldsymbol{A}): A \otimes x = b$
Weak control solvability	$(\forall b \in \boldsymbol{b})(\exists x \in B(n))(\exists A \in \boldsymbol{A}): A \otimes x = b$
Universal solvability	$(\exists x \in B(n))(\forall b \in \boldsymbol{b})(\forall A \in \boldsymbol{A}): A \otimes x = b$
Weak universal solvability	$(\forall b \in \boldsymbol{b})(\exists x \in B(n))(\forall A \in \boldsymbol{A}): A \otimes x = b$
T4 solvability	$(\exists b \in \boldsymbol{b})(\exists x \in B(n))(\forall A \in \boldsymbol{A}): A \otimes x = b$
T5 solvability	$(\exists b \in \boldsymbol{b})(\forall A \in \boldsymbol{A})(\exists x \in B(n)): A \otimes x = b$

These solvability concepts exclusive of the weak, strong and tolerance solvability are our definitions, studied in the papers [6, 7, 8].

3.1 Weak, strong and tolerance solvability

K. Cechlárová and R. A. Cuninghame-Green gave necessary and sufficient conditions for the weak, strong and tolerance solvability.

Theorem 3.1 [1] *Interval system (1) is weakly solvable if and only if*

$$\overline{A} \otimes x^*(\underline{A}, \overline{b}) \geq \underline{b}. \tag{4}$$

Theorem 3.2 [2] *Interval system (1) is strongly solvable if and only if all its extremal subsystems with exactly one LU equation are solvable.*

Theorem 3.3 [1] *Interval system (1) is tolerance solvable if and only if*

$$\underline{A} \otimes x^*(\overline{A}, \overline{b}) \geq \underline{b}. \tag{5}$$

3.2 Universal, weak universal and weak tolerance solvability

These solvability concepts have been studied by H. Myšková [6, 7]. The following lemma gives a necessary condition for the universal solvability.

Lemma 3.4 [6] *If interval system (1) is universally solvable then $\underline{b} = \overline{b}$.*

Theorem 3.5 [6] *Interval system (1) with a constant right-hand side $b = \underline{b} = \overline{b}$ is universally solvable if and only if*

$$\underline{A} \otimes x^*(\overline{A}, b) = b, \tag{6}$$

and in this case $x^*(\overline{A}, b)$ is the maximum universal solution.

To formulate a necessary and sufficient condition for the weak universal and weak tolerance solvability we assign for each $p \in M$ the matrix $A^{(p)}$ defined as follows:

$$a_{ij}^{(p)} = \begin{cases} \underline{a}_{ij} & \text{for } i = p,\ j \in N, \\ \overline{a}_{ij} & \text{for } i \in M,\ i \neq p,\ j \in N. \end{cases} \tag{7}$$

Theorem 3.6 [7] *Interval system (1) is weakly universally solvable if and only if*

$$\underline{A} \otimes x^*(\overline{A}, b^{(p)}) = b^{(p)} \tag{8}$$

for each $p \in M$.

Theorem 3.7 [6] *Interval system (1) is weakly tolerance solvable if and only if*

$$A^{(p)} \otimes x^*(A^{(p)}, \overline{b}) \geq \underline{b} \tag{9}$$

for each $p \in M$.

3.3 Control and weak control solvability

Control solvability over the classical algebra were defined and studied by S. P. Shary [12].

Theorem 3.8 [7] *Interval system (1) is control solvable if and only if*

$$\overline{A} \otimes x^*(\underline{A}, \underline{b}) \geq \overline{b}. \tag{10}$$

For each $p \in M$ denote by $b^{(p)}$ the vector with the following entries

$$b_i^{(p)} = \begin{cases} \overline{b}_i & for\ i = p, \\ \underline{b}_i & for\ i \neq p,\ i \in M. \end{cases} \tag{11}$$

Theorem 3.9 [7] *Interval system (1) is weakly control solvable if and only if*

$$\overline{A} \otimes x^*(\underline{A}, b^{(p)}) \geq b^{(p)} \tag{12}$$

for each $p \in M$.

All above given necessary and sufficient conditions can be verified in a polynomial time. The residual solvability concepts – T4 and T5 solvability can be tested using pseudopolynomial algorithms.

3.4 T4 and T5 solvability

We dealt with T4 and T5 solvability in [8]. Some assertions adduced there are improved in this subsection.

Definition 3.10 *A vector $b \in \boldsymbol{b}$ is called a* T4-vector *of interval system (1) if there exists $x \in B(n)$ such that $A \otimes x = b$ for each $A \in \boldsymbol{A}$.*

By definition, T4 solvability of interval system (1) means that there exists a vector $b \in \boldsymbol{b}$ such that b is T4-vector of (1).

Lemma 3.11 [8] *A vector $b \in \boldsymbol{b}$ is a T4-vector of interval system (1) if and only if*

$$\underline{A} \otimes x^*(\overline{A}, b) = b. \tag{13}$$

Lemma 3.11 implies the following necessary, but not sufficient condition for the T4 solvability. Denote $M_j = \{i \in M : \underline{a}_{ij} = \overline{a}_{ij}\}$ for each $j \in N$.

Lemma 3.12 *If interval system (1) is T4 solvable, then $\bigcup\limits_{j \in N} M_j = M$.*

Proof. Suppose that $\bigcup\limits_{j \in N} M_j \neq M$, i.e., there exists $r \in M$ such that $\underline{a}_{rj} < \overline{a}_{rj}$ for each $j \in N$. Let $b \in \boldsymbol{b}$ be an arbitrary vector. Then

$$[\underline{A} \otimes x^*(\overline{A}, b)]_r = \max_{j \in N}\{\underline{a}_{rj} + x^*_j(\overline{A}, b)\} = \max_{j \in N}\{\underline{a}_{rj} + \min_{i \in M}\{b_i - \overline{a}_{ij}\}\} \leq$$

$$\max_{j \in N}\{\underline{a}_{rj} + b_r - \overline{a}_{rj}\} \leq b_r + \max_{j \in N}\{\underline{a}_{rj} - \overline{a}_{rj}\} < b_r.$$

As equality (13) is not fulfilled in the r-th row, the vector b is not T4-vector of (1). By reason of the vector b was chosen arbitrarily, there does not exist a T4-vector of interval system (1), so it is not T4 solvable.

Lemma 3.11 does not give the method for finding a T4-vector. For this reason, we define a sequence $\{c^{(k)}\}_{k=0}^{\infty}$ as follows:

$$c^{(k)} = \begin{cases} \overline{b} & \text{for } k = 0, \\ \underline{A} \otimes x^*(\overline{A}, c^{(k-1)}) & \text{for } k \geq 1. \end{cases} \tag{14}$$

Lemma 3.13 [8] *Let $b \in \boldsymbol{b}$ be a T4-vector of interval system (1). Then*

i) the sequence $\{c^{(k)}\}_{k=0}^{\infty}$ is nonincreasing,

ii) for each nonnegative integer k the inequality $b \leq c^{(k)}$ is satisfied.

Theorem 3.14 [8] *Interval system (1) is T4 solvable if and only if there exists a positive integer l such that $c^{(l)} \in \boldsymbol{b}$ and $c^{(l+1)} = c^{(l)}$.*

The previous theorem implies the following algorithm:

Algorithm 1
Input: $\boldsymbol{A}$, $\boldsymbol{b}$
Output: YES, if the given interval system is T4 solvable and NOT, if it is not T4 solvable.

Step 1. If $\bigcup_{j\in N} M_i \neq M$, then write NOT, go to END.
Step 2. $c^{(0)} = \overline{b},\ k = 0$.
Step 3. $c^{(k+1)} = \underline{A} \otimes x^*(\overline{A}, c^{(k)})$.
Step 4. If $\underline{b} \leq c^{(k+1)}$ does not hold, then write NOT, go to END.
Step 5. If $c^{(k+1)} = c^{(k)}$, then write YES, $c^* = c^{(k)}$, go to END.
Step 6. $k = k + 1$, go to *Step 3*.
END

Corollary 3.15 *If interval system (1) is T4 solvable, then the vector c^* given by Algorithm 1 is the maximum T4-vector of (1).*

Now, we shall deal with the computational complexity of this algorithm. The most time-consuming is Step 3 which requires $O(mn)$ operations. The question which arise is the number of repetitions of the loop 3–6 till the algorithm gives answer. This number is bounded by the number of different vectors $b^{(k)}$. Suppose that elements of matrices $\underline{A}, \overline{A}$ and vectors $\underline{b}, \overline{b}$ are integers. As the sequence $b^{(k)}$ is nonicreasing, the number of repetitions of the loop 3 – 6 is bounded by $K \cdot m$, where $K = \max_{i\in M}\{\overline{b}_i - \underline{b}_i\}$. Hence the complexity of Algorithm 1 is $O(K \cdot m^2 n)$. So Algorithm 1 is pseudopolynomial.

Definition 3.16 *A vector $b \in \boldsymbol{b}$ is called a T5-vector of interval system (1) if for each $A \in \boldsymbol{A}$ system $A \otimes x = b$ is solvable.*

Realise that the T5 solvability of interval system (1) is equivalent to the existence of a T5-vector of (1).

Lemma 3.17 [8] *A vector $b \in \boldsymbol{b}$ is a T5-vector of interval system (1) if and only if*

$$A^{(k)} \otimes x^*(A^{(k)}, b) = b \tag{15}$$

holds for each $k \in M$.

To suggest an algorithm for T5 solvability, we define the sequence $\{d^{(k)}\}_{k=1}^{\infty}$ as follows:

$$d_i^{(k)} = \begin{cases} \overline{b}_i & \text{for} \quad k = 0, \\ \min\limits_{r\in M}\{[A^{(r)} \otimes x^*(A^{(r)}, d^{(k-1)})]_i\} & \text{for} \quad k \geq 1, \end{cases} \tag{16}$$

for each $i \in M$.

Lemma 3.18 *[8] Let $b \in \boldsymbol{b}$ be a T5-vector of interval system (1). Then*

i) the sequence $\{d^{(k)}\}_{k=1}^{\infty}$ is nonincreasing, *ii) for each nonnegative integer k the inequality $b \leq d^{(k)}$ is satisfied.*

Theorem 3.19 [8] *Interval system (1) is T5 solvable if and only if there exists a positive integer l such that $d^{(l+1)} \in \boldsymbol{b}$ and $d^{(l+1)} = d^{(l)}$.*

Theorem 3.19 gives the following algorithm.

Algorithm 2
Input: $\boldsymbol{A}$,$\boldsymbol{b}$
Output: YES, if the given interval system in the max-plus algebra is T5 solvable and NOT, if it is not T5 solvable.

Step 1. $d^{(0)} = \overline{b},\ k = 0$.
Step 2. For each $i \in M$ compute $d_i^{(k+1)} = \min\limits_{r\in M}\{[A^{(r)} \otimes x^*(A^{(r)}, d^{(k)})]_i\}$.
Step 3. If $\underline{b} \leq d^{(k+1)}$ does not hold then write NOT, go to END.
Step 4. If $d^{(k+1)} = d^{(k)}$ then write YES, $d^* = d^{(k)}$, go to END.
Step 5. $k = k + 1$, go to *Step 2.*
END

Corollary 3.20 *If interval system (1) is T5 solvable, then the vector d^* given by Algorithm 2 is the maximum T5-vector of (1).*

Similarly as Algorithm 1, Algorithm 2 is pseudopolynomial, too.

Example 3.21 *Examine all solvability concepts for the given interval system $\boldsymbol{A} \otimes x = b$ with*

$$\boldsymbol{A} = \begin{pmatrix} \langle 3,10\rangle & \langle 4,7\rangle & \langle 8,10\rangle \\ \langle 5,7\rangle & \langle 5,9\rangle & \langle 7,10\rangle \\ \langle 6,8\rangle & \langle 7,10\rangle & \langle 4,7\rangle \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} \langle 4,7\rangle \\ \langle 3,6\rangle \\ \langle 4,6\rangle \end{pmatrix}.$$

At first, we check the universal solvability and the weak solvability.

*The given interval system is **not universally solvable**, because $\underline{b} \neq \overline{b}$ (necessary condition for the universal solvability is not fulfilled).*

*As $x^*(\underline{A}, \overline{b}) = (0, -1, -1)^T$, we have $\overline{A} \otimes x^*(\underline{A}, \overline{b}) = (10, 9, 9)^T \geq \underline{b}$ which implies that given interval system is **weakly solvable**.*

*To verify the control solvability we compute $x^*(\underline{A}, \underline{b}) = (-2, -3, -4)^T$, then $\overline{A} \otimes x^*(\underline{A}, \underline{b}) = (8, 6, 7)^T \geq \overline{b}$, which means that the given interval system is **control solvable**. Then this is **weakly control solvable**, too.*

*By Step 1 of Algorithm 2 we can deduce that the given interval system is **not T4 solvable**. Then it is **not weakly universally solvable**.*

*For checking the tolerance solvability caculate $x^*(\overline{A}, \overline{b}) = (-3, -4, -4)^T$, $\underline{A} \otimes x^*(\overline{A}, \overline{b}) = (4, 3, 3)^T$. As inequality (5) is not fulfilled, the given interval system is **not tolerance solvable**.*

We shall need matrices

$$A^{(1)} = \begin{pmatrix} 3 & 4 & 8 \\ 7 & 9 & 10 \\ 8 & 10 & 7 \end{pmatrix}, \ A^{(2)} = \begin{pmatrix} 10 & 7 & 10 \\ 5 & 5 & 7 \\ 8 & 10 & 7 \end{pmatrix}, \ A^{(3)} = \begin{pmatrix} 10 & 7 & 10 \\ 7 & 9 & 10 \\ 6 & 7 & 4 \end{pmatrix}.$$

Because of

$$A^{(1)} \otimes x^*(A^{(1)}, \overline{b}) = A^{(1)} \otimes (-2, -2, -4)^T = (4, 6, 6)^T \geq \underline{b},$$

$$A^{(2)} \otimes x^*(A^{(2)}, \overline{b}) = A^{(2)} \otimes (-3, -4, -3)^T = (7, 4, 6)^T \geq \underline{b},$$

$$A^{(3)} \otimes x^*(A^{(3)}, \overline{b}) = A^{(3)} \otimes (-3, -3, -4)^T = (7, 6, 4)^T \geq \underline{b},$$

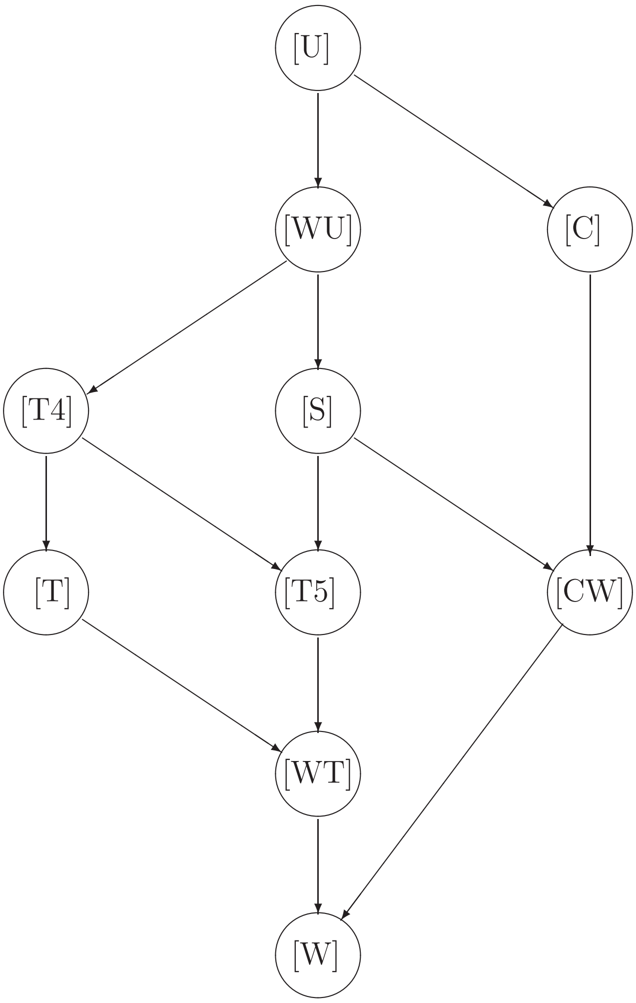
the given interval system is ***weakly tolerance solvable****.*

We can use the previous products for checking the T5 solvability. We get $d^{(1)} = (4, 4, 4)$. *As* $A^{(1)} \otimes x^*(A^{(1)}, d^{(1)}) = A^{(1)} \otimes (-4, -6, -6)^T = (2, 4, 4)^T$, *using Algorithm 2 we get* $d_1^{(2)} \leq 2 < \underline{b}_1$, *which follows that the given interval system is* ***not T5 solvable****. Then this is* ***not strongly solvable****.*

Result*: The given interval system is weakly, control, weakly control and weakly tolerance solvable.*

4 Graphical representation

Definitions of solvability concepts imply relations among them. For example, if interval system (1) is control solvable, then it is weak control solvable, or the weak universal solvability of (1) implies the weak control solvability of (1). It is easy to see, that universal solvability implies all other solvability concepts and on the other hand, the weak solvability follows from all solvability concepts. For each solvability concept S_i denote by $[S_i]$ the set of all assertions which are equivalent to the solvability concept S_i. Let us define on the set $[S]$ of all sets $[S_i]$ a relation $\mathcal{R}$ such that $[S_i]\mathcal{R}[S_j]$ if and only if S_j implies S_i. Realise, that if $[S_i]\mathcal{R}[S_j]$ then for each assetions $A_i \in [S_i]$ and $A_j \in [S_j]$ holds A_j implies A_i. It is easy to see that relation $\mathcal{R}$ is reflexive and transitive. Antisymmetry follows from the the fact that if S_j implies S_i and S_i implies S_j then S_i and S_j are equivalent which follows that $[S_i] = [S_j]$. In consequence of these properties of relation $\mathcal{R}$ the set $[S]$ of all $[S_i]$ with relation $\mathcal{R}$ is a partially ordered set. So we can describe its by Hasse diagram, see Figure 1.



Used abbreviations:

W – weak solvability
WU – weak universal solvability
WT – weak tolerance solvability
CW – weak control solvability
T – tolerance solvability
C – control solvability
S – strong solvability
T4 – T4 solvability
T5 – T5 solvability
U – universal solvability

Figure 1

For all pairs of solvability concepts such that S_j implies S_i we can prove that S_i does not imply S_j. This means that there arent solvability concepts S_i and S_j defined in this paper such that $[S_i] = [S_j]$. For example tolerance solvability implies weak tolerance solvability, but contrariwise it does not hold. In Example 3.21 we have given the interval system which is weakly tolerance solvable, but it is not tolerance solvable. For each pair of incomparable sets $[S_i]$, $[S_j]$ we can find an interval system such that S_i holds, but S_j does not hold and another interval system for which S_j holds, but S_i does not hold. For example, the strong solvability and the T4 solvability are incomparable in the sense of the relation R.

References

[1] K. CECHLÁROVÁ: Solutions of interval systems in max-plus algebra. Proc. SOR 2001, Eds.V.Rupnik, L.Zadnik-stirn, S.Drobne, Preddvor, Slovenia, 321–326.

[2] K.CECHLÁROVÁ, R. A. CUNINGHAME-GREEN: Interval systems of max-plus separable linear equations. Linear Algebra and its Applications 340 (2002), 215–224.
[3] R. A. CUNINGHAME-GREEN, K. CECHLÁROVÁ: Residuation in fuzzy Algebra and some applications, Fuzzy sets and systems 71 (1995), 227–239.
[4] R. A. CUNINGHAME-GREEN: Minimax Algebra. Lecture notes in Economics and Mathematical systems 1966, Springer, Berlin 1979.
[5] J. KREINOVICH, A. LAKEYEV, J. ROHN, P. KAHL: Computational complexity of feasibility of data processing and interval computations, Kluwer, Dordrecht 1998.
[6] H. MYŠKOVÁ: Interval systems of max-separable linear equations. Linear Algebra and its Applications 403 (2005), 263–272.
[7] H. MYŠKOVÁ: Control solvability of interval systems of max-separable linear equations. Linear Algebra and its Applications 416 (2006), 215–223.
[8] H. MYŠKOVÁ: Solvability concepts of interval systems of linear equations in max-plus and max-min algebra. Submitted to Linear Algebra and its Applications, in reviewing process.
[9] G. J. OLSDER ET AL.: Course Notes: Max-algebra Aproach to Discrete Event Systems. Algebres Max-Plus et Applications an Informatique et Automatique, INRIA, 1998, 147–196.
[10] J.ROHN: Systems of interval linear equations and inequalities (rectangular case), Institut of Computer Science, Academy of Sciences of the Czech Republic, Technical report No. 875, September 17, 2002.
[11] J. ROHN: Complexity of some linear problems with interval data. Reliable computing 3 (1997), 315–323.
[12] S. P. SHARY: On controlled solution set of interval algebraic systems. Interval computations 6 (1992), 66–75.
[13] K. ZIMMERMANN: Extremální algebra. Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha 1976.

Current address

RNDr. Helena Myšková, PhD.
Department of Mathematics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, B. Nemcovej 32, 042 00 Košice, Slovakia, helena.myskova@tuke.sk

Journal of Applied Mathematics

Aplimat

volume III (2010), number II

THE NOTION OF SUBHYPERSTRUCTURE OF "ENDS LEMMA" – BASED HYPERSTRUCTURES

NOVÁK Michal, (CZ)

Abstract. The article deals with hyperstructure theory. There exists a way of creating semi-hypergroups and hypergroups (or rather transposition hypergroups) from partially/quasi–ordered semigroups and groups. Even though it has been widely used by some authors, properties of hyperstructures created in this way have not yet been comprehensively studied. In this article the concept of subhyperstructure of such hyperstructures is discussed. The article may be regarded as a sequel to an earlier article of mine which discusses the issue of identities and inverses of "Ends lemma"–based hyperstructures.

Key words and phrases. hyperstructure, subhyperstructure, partially ordered group, quasi-ordered group

Mathematics Subject Classification. Primary 20N20; Secondary 06A06.

1 Motivation

A number of articles and contributions in the hyperstructure theory (especially by Czech authors such as Chvalina, Chvalinová, Hošková – e.g. in [5, 6], Račková – e.g. in [9], Moučka or Novák) make use of the construction first used in [3] as Theorems 1.3 and 1.4 (chapter IV), pp. 146–147. Using these results known as the "Ends lemma" (or "Ending lemma") we can form hyperstructures from quasi / partially ordered structures. Even though the lemma has been widely used, its possibilities and limits have been comprehensively studied for the first time only in [7]. This article can be regarded as a sequel to [7]. It deals with substructures of both single–valued and multi–valued structures. The question it answers is very simple: *What is the relation (if any) between the substructures of the underlying single–valued structure and the subhyperstructures of the associated hyperstructure?* First of all, however, the article discusses the question of whether the converse of the "Ends lemma" holds.

2 Preliminaries

Recall first some basic definitions and ideas from the hyperstructures theory. A *hypergrupoid* is a pair $(H, \bullet)$, where $H \neq 0$ and $\bullet : H \times H \to \mathcal{P}^*(H)$ is a binary hyperoperation on H. Symbol $\mathcal{P}^*(H)$ denotes the system of all nonempty subsets of H. If the associativity axiom $a \bullet (b \bullet c) = (a \bullet b) \bullet c$ holds for all $a, b, c \in H$, then the pair $(H, \bullet)$ is called a *semihypergroup*. If moreover the reproduction axiom: for any element $a \in H$ equalities $a \bullet H = H = H \bullet a$ hold, is satisfied, then the pair $(H, \bullet)$ is called a *hypergroup*. A hypergroup $(H, \bullet)$ is called a *transposition hypergroup* if it satisfies the following transposition axiom: For all $a, b, c, d \in H$ the relation $b \backslash a \approx c/d$ implies $a \bullet d \approx b \bullet c$, where $X \approx Y$ for $X, Y \subseteq H$ means $X \cap Y \neq \emptyset$. Sets $b \backslash a = \{x \in H; a \in b \bullet x\}$ and $c/d = \{x \in H; c \in x \bullet d\}$ are called *left* and *right extensions*, or *fractions*, respectively. A commutative transposition hypergroup is called a *join space*.

Let G be a nonempty subset of H. $(G, \bullet)$ is called a *subhypergroupoid* of $(H, \bullet)$ (or *multiplicatively closed*) if $G \bullet G \subseteq G$. If $(G, \bullet)$ is moreover a (semi)hypergroup, $(G, \bullet)$ is called a sub(semi)hypergroup of $(H, \bullet)$.

An element of $e \in H$, where $(H, \bullet)$ is a hyperstructure, is called an *identity* if for $\forall x \in H$ there holds $x \bullet e \ni x \in e \bullet x$. If for $\forall x \in H$ there holds $x \bullet e = \{x\} = e \bullet x$, then $e \in H$ is called a *scalar identity*.

As far as the theory of ordered structures is concerned, we need to recall that by a *quasi-ordered (semi)group* we mean a triple $(G, \cdot, \leq)$, where $(G, \cdot)$ is a (semi)group and $\leq$ is a reflexive and transitive binary relation on G such that for any triple $x, y, z \in G$ with the property $x \leq y$ also $x \cdot z \leq y \cdot z$ and $z \cdot x \leq z \cdot y$ hold. We call the semigroup *partially ordered*[1] if the relation $\leq$ is moreover antisymmetric. Further, $[a)_{\leq} = \{x \in G; a \leq x\}$ is a *principal end* generated by $a \in G$.

We are going to examine the "Ends lemma", which has the form of the following Theorems:

Theorem 2.1 *([3], Theorem 1.3, p. 146) Let $(S, \cdot, \leq)$ be a partially ordered semigroup. Binary hyperoperation $* : S \times S \to \mathcal{P}'(S)$ defined by*

$$a * b = [a \cdot b)_{\leq}$$

is associative. The semi-hypergroup $(S, *)$ is commutative if and only if the semigroup $(S, \cdot)$ is commutative.

Proof. Later in the text the proof of this Theorem will be referred to, let me therefore include its main part here. Suppose $a, b, c \in S$ arbitrary. First of all, it is useful to show that the following equality holds:

$$\bigcup_{t \in [b \cdot c)_{\leq}} [a \cdot t)_{\leq} = \bigcup_{x \in [a \cdot b)_{\leq}} [x \cdot c)_{\leq}.$$

Suppose therefore an abitrary $s \in \bigcup_{t \in [b \cdot c)_{\leq}} [a \cdot t)_{\leq}$. This means that $s \geq a \cdot t_0$ for a suitable $t_0 \in S$, $t_0 \geq b \cdot c$. Then $a \cdot t_0 \geq a \cdot (b \cdot c) = (a \cdot b) \cdot c$ and if we set $x_0 = a \cdot b$, we get that $x_0 \cdot c \leq s$,

[1]In fact, the term *ordered* is often used in the Czech environment by authors such as Chvalina, Račková or even myself instead of the correct English term *partially ordered*.

$x_0 \in [a \cdot b)_{\leq}$, i.e. $s \in [x_0 \cdot c)_{\leq} \subseteq \bigcup_{x \in [a \cdot b)_{\leq}} [x \cdot c)_{\leq}$. The other inclusion may be proved in the analogous way.

Now we get that

$$a * (b * c) = \bigcup_{t \in b * c} a * t = \bigcup_{t \in [b \cdot c)_{\leq}} [a \cdot t)_{\leq} = \bigcup_{x \in [a \cdot b)_{\leq}} [x \cdot c)_{\leq} = \bigcup_{x \in a * b} x * c = (a * b) * c,$$

which completes the proof of associativity.

In accordance with [7], the hyperstructure $(S, *)$ constructed in this way will further on be called the *associated* hyperstructure to the structure $(S, \cdot)$ or *an "Ends lemma"–based hyperstructure*. Instead of S the carrier set will be denoted by H.

Theorem 2.2 *([3], Theorem 1.4, p. 147) Let $(S, \cdot, \leq)$ be a partially ordered semigroup. The following conditions are equivalent:*

1^0 *For any pair $a, b \in S$ there exists a pair $c, c' \in S$ such that $b \cdot c \leq a$ and $c' \cdot b \leq a$*

2^0 *The associated semi-hypergroup $(S, *)$ is a hypergroup.*

Remark 2.3 *If $(S, \cdot, \leq)$ is a partially ordered group, then if we take $c = b^{-1} \cdot a$ and $c' = a \cdot b^{-1}$, then condition 1^0 is valid. Therefore, if $(S, \cdot, \leq)$ is a partially ordered group, then its associated hyperstructure is a hypergroup.*

Remark 2.4 *The wording of the above Theorems is the exact translation of theorems from [3]. The respective proofs, however, do not change in any way, if we regard* quasi-ordered *structures instead of* partially ordered *ones as the anti-symmetry of the relation $\leq$ is not needed (with the exception of the $\Leftarrow$ implication of the part on commutativity, which does not hold in this case). The often quoted version of the "Ends lemma" is therefore the version assuming quasi–ordered structures.*

Remark 2.5 *In their article [4] Chvalina and Moučka explore the approach to defining hyperoperations in similar way as the "Ends lemma" suggests – further hyperoperations using the (quasi–) ordering are defined and studied there.*

The following theorem extending the "Ends lemma" was proved by Račková in her Ph.D. thesis. The proof can be also found in [9]. Notice that if $(H, \cdot)$ is commutative, then $(H, *)$ is a join space.

Theorem 2.6 *(Theorem 4, [9]) Let $(H, \cdot, \leq)$ be a quasi-ordered group and $(H, *)$ be the associated hypergroupoid. Then $(H, *)$ is the transposition hypergroup.*

Finally, notice the main result of article [7] and especially its immediate corollary – the fact that $(H, *)$ is not a canonical hypergroup.

Theorem 2.7 *(Theorem 3.1 [7]) Let $(H, \cdot, \leq)$ be a non-trivial quasi-ordered group, where the relation $\leq$ is not the identity relation, and let $(H, *)$ be its associated transposition hypergroup. Then $(H, *)$ does not have a scalar identity.*

3 New results

The "Ends lemma" gives a way to create hyperstructures from ordered structures. Can it be reversed? Can we say that if a suitably (i.e. in the "Ends lemma" way) defined hyperoperation is associative, the underlying single–valued operation is associative too? Surprisingly, this question has not long been answered (or rather, asked).

Theorem 3.1 *Let $(H,\cdot)$ be a non-trivial groupoid and $\leq$ a binary ordering on H such that for an arbitrary pair of elements $a,b \in H$, $a \leq b$, and for arbitrary $c \in H$ there holds $c \cdot a \leq c \cdot b$ and $a \cdot c \leq b \cdot c$. Further define a hyperoperation $* : H \times H \rightarrow \mathcal{P}^*(H)$ for an arbitrary pair of elements $a,b \in H$ by $a * b = [a \cdot b)_{\leq} = \{x \in H; a \cdot b \leq x\}$. Then if the hyperoperation $*$ is associative, then the single–valued operation $\cdot$ is associative too. Furthermore, if there exists an element $e \in H$ such that for $\forall a \in H$ there holds $a * e = e * a = [a)_{\leq}$, then this element e is the identity of the semigroup $(H,\cdot)$.*

Proof.

1. If the hyperoperation $*$ defined in the theorem is associative, then the fact that an arbitrary element $x \in (a * b) * c$ implies that $x \in a * (b * c)$. Conversely, the fact that an arbitrary element $y \in a * (b * c)$ implies that $y \in (a * b) * c$.

   (a) If there holds $x \in (a * b) * c$, then there exists an element $x_1 \in a * b$ such that $x \in x_1 * c$, i.e. there exists an element $x_1 \in H$ such that $a \cdot b \leq x_1$ and $x_1 \cdot c \leq x$. Thanks to the assumed properties of the relation $\leq$ we get that $(a \cdot b) \cdot c \leq x_1 \cdot c \leq x$, i.e. $(a \cdot b) \cdot c \leq x$, which means that $x \in [(a \cdot b) \cdot c)_{\leq}$.

   (b) Furthermore, we know that $x \in a * (b * c)$, i.e. by analogous reasoning we get that $x \in [a \cdot (b \cdot c))_{\leq}$.

   Since x is an arbitrary element of H and since the same reasoning holds for the arbitrary above mentioned $y \in H$, we get that $[(a \cdot b) \cdot c)_{\leq} = [a \cdot (b \cdot c))_{\leq}$. However, on condition of antisymmetry of the relation $\leq$ this implies that $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, which means that the operation $\cdot$ is associative.

2. If there exists an element $e \in H$ such that for $\forall a \in H$ there holds $a * e = e * a = [a)_{\leq}$ then there for $\forall a \in H$ holds $[a \cdot e)_{\leq} = [e \cdot a)_{\leq} = [a)_{\leq}$, which on condition of antisymmetry of the relation $\leq$ means that $a \cdot e = e \cdot a = a$, i.e. e is the identity of $(H,\cdot)$. Obviously, the element satisfying the condition of the theorem is unique.

Remark 3.2 *If the relation $\leq$ is not antisymmetric, the above theorem is not true. This is caused by the fact that only for antisymmetric relations $\leq$ there holds that $[a)_{\leq} = [b)_{\leq}$ implies that $a = b$. Indeed, suppose a simple two element set $M = \{a,b\}$ where the relation $\leq$ is defined as $a \leq a, a \leq b, b \leq a, b \leq b$. This reflexive and transitive relation $\leq$ is obviously not antisymmetric and there holds $[a)_{\leq} = [b)_{\leq}$ yet $a \neq b$.*

Let me now focus on the issue of subhyperstructures of "Ends lemma" based hyperstructures and of substructures of the respective single–valued structures. First of all, however, we must clarify the concept of a principal end generated by an element, which lies in the subset in question.

Suppose a hyperstructure $(H, *)$ associated to a quasi / partially ordered semigroup $(H, \cdot, \leq)$ and a non-empty subset G of H. For an arbitrary element $g \in G$ we may write

$$[a)_{\leq_G} \quad = \quad \{x \in G : a \leq x\} \tag{1}$$

as well as

$$[a)_{\leq_H} \quad = \quad \{x \in H : a \leq x\}. \tag{2}$$

Given this notation we may distinguish between $(G, *_G)$ based on the hyperoperation $*_G$ such that for an arbitrary pair of elements $a, b \in G$ we set

$$a *_G b = [a \cdot b)_{\leq_G} = \{x \in G : a \cdot b \leq x\}$$

and $(G, *_H)$, where $a *_H b$ is defined by

$$a *_H b = [a \cdot b)_{\leq_H} = \{x \in H : a \cdot b \leq x\}.$$

Obviously, properties of $(G, *_G)$ and $(G, *_H)$ will not be the same.

Since the notation $*_H$ reflects the idea of an "end generated by an element" better, I will start with examination of subhyperstructures of "Ends lemma"–based hyperstructures in this case. Instead of $*_H$ and $\leq_H$ the usual notation $*$ and $\leq$ is going to be used.

It will be useful to utilize the concept of an *upper set*. In the following definition I use the term *upper end* of a set in order to visually relate the concept to the "Ends lemma". Furthermore, identifying the elements which "spoil" the property of being an upper set / end of a set will be useful.

Definition 3.3 *Let $(H, \cdot, \leq)$ be a partially ordered semigroup and let G be a nonempty subset of H. If for an arbitrary element $g \in G$ there holds $[g)_{\leq} \subseteq G$, we call G an upper end of H. If there exists an element $g \in G$ such that there exists an element $x \in H \setminus G$ such that $g \leq x$ (i.e. $x \in [g)_{\leq}$), we say that G is not an upper end of H because of the element x.*

3.1 Subhyperstructures and upper ends

First of all, the issue of (hyper) groupoids must be clarified.

Lemma 3.4 *Let $(H, *)$ be an associated semihypergroup of a partially ordered semigroup $(H, \cdot, \leq)$ and $G \subseteq H$ nonempty. If it exists, denote u the identity of $(H, \cdot)$. Further suppose that $(G, \cdot)$ is a subgroupoid of $(H, \cdot)$.*

1. *If G is an upper end of H, then $(G, *)$ is a subhypergroupoid of $(H, *)$.*
2. *If G is not an upper end of H and there holds $u \in G$, then $(G, *)$ is not a subhypergroupoid of $(H, *)$.*

3. *The statement in part 2 holds even in case that $u \notin G$ (or u does not exist) yet for some $a, b \in G$ there holds that $a \cdot b = c$, where $c \in G$ is such that there exists an element x_i because of which G is not an upper end of H such that $c \leq x_i$.*

4. *On simultaneous validity of conditions that*

   (a) u does not exist or $u \notin G$

   (b) G is not an upper end of H because of elements $x_i, i \in I$

   (c) for every $a, b, c \in G$ there holds $a \cdot b = c$ and all the triples are such that for no x_i there holds $c \leq x_i$

   the couple $(G, *)$ is a subhypergroupoid of $(H, *)$.

Proof.

1. Since $\cdot$ is an operation on G, for an arbitrary pair $a, b \in G$ there holds $a \cdot b = c$, where $c \in G$. Thus $a * b = [a \cdot b)_{\leq} = [c)_{\leq}$, which is a subset of G because G is an upper end of H. Therefore we have that $G * G \subseteq G$, which means that $(G, *)$ is a subhypergroupoid of $(H, *)$.

2. If G is not an upper end of H, then there exists an element $g \in G$ such that there exists an element $x \in H \setminus G$ such that $g \leq x$. If furthermore $u \in G$, then if we consider the above mentioned element g, then $g * u = [g \cdot u)_{\leq} = [g)_{\leq} \not\subseteq G$ (because of the element x, the existence of which is assumed), which means that $G * G \not\subseteq G$, i.e. $(G, *)$ is not a subhypergroupoid of $(H, *)$.

3. Obvious since $a * b = [a \cdot b)_{\leq} = [c)_{\leq}$, for which there by definition holds $[c)_{\leq} \not\subseteq G$, i.e. $(G, *)$ is not a subhypergroupoid of $(H, *)$. Elements a, b, c have the meaning defined in part 3.

4. In this case for $\forall a, b \in G$ we have that $a * b = [a \cdot b)_{\leq} = [c)_{\leq}$, where $c \in G$ is such that $[c)_{\leq} \subset G$, i.e. we have that $G * G \subset G$, which means that $(G, *)$ is a subhypergroupoid of $(H, *)$.

Remark 3.5 *In fact parts 2 and 3 of the above Lemma may be written as one. Yet they are included separately because of uniqueness of the element u. Instead of $c \leq x_i$ we could write $c < x_i$ because we suppose $c \in G$ while $x_i \notin G$, which means that c and x_i cannot be equal. Finally notice that if $\cdot$ is not an operation on G, then $(G, *)$ is not a subhypergroupoid of $(H, *)$. Indeed, in this case there exists a triple a, b, c, where $a, b \in G$ while $c \notin G$, such that $a \cdot b = c$. This means that $a * b = [a \cdot b)_{\leq} = [c)_{\leq}$. However since the relation $\leq$ is reflexive and $c \notin G$, we get that $[c)_{\leq} \not\subseteq G$, i.e. $G * G \not\subseteq G$.*

Example 3.6 *The set $\mathbb{N} \setminus \{1, 2, 3, 4, 5, 7, 9\} \subset \mathbb{N}$ with the operation $+$ and the usual ordering of numbers is an example of a set constructed under Lemma 3.4, part 4. Indeed, $\mathbb{N} \setminus \{1, 2, 3, 4, 5, 7, 9\} = \{6, 8, 10, 11, 12, 13, 14 \ldots\}$ is not an upper end of $\mathbb{N}$ because of elements 7 and 9 (since e.g. $7 \in [6)_{\leq}$ but $7 \notin \mathbb{N} \setminus \{1, 2, 3, 4, 5, 7, 9\}$). Yet for no couple $a, b \in \{6, 8, 10, 11, 12, 13, 14 \ldots\}$ there holds $a + b \leq 7$ or $a + b \leq 9$.*

Lemma 3.4 gives a complete description of an arbitrary subset of an arbitrary "Ends lemma"–based hyperstructure. Since subsemihypergroup, subhypergroup and other concepts are defined as special classes of subhypergroupoids, the lemma gives a complete list of candidates for various types of subhyperstructures of "Ends lemma"–based hyperstructures. Let me now examine the case of subsemihypergroups.

Theorem 3.7 *Let $(H,*)$ be the associated semihypergroup of a partially ordered semigroup $(H,\cdot,\leq)$. Suppose that G is either an upper end of H or such a subset of H that assumptions of Lemma 3.4, part 4 are fulfilled. Then*

1. *$(G,\cdot)$ is a subsemigroup of $(H,\cdot)$ if and only if $(G,*)$ is a subsemihypergroup of $(H,*)$.*

If furthermore $(H,\cdot)$ is a monoid, then

2. *$(G,\cdot)$ is a submonoid of $(H,\cdot)$ if and only if there exists an element $u\in G$ such that for $\forall g\in G$ there holds $g*u=u*g=[g)_{\leq}$.*

Proof. Suppose that $(H,*)$ is the associated semihypergroup of a partially ordered semigroup $(H,\cdot)$ and G is a nonempty subset of H.

1. **"⇒"** The fact that $(G,*)$ is a subhypergroupoid of $(H,*)$ follows from Lemma 3.4, parts 1 and 4 respectively. For both types of G, the associativity of $(G,*)$ follows from the first part of the "Ends lemma", Theorem 2.1 – notice that the proof may be applied without any changes even when G is not an upper end of H.

   "⇐" Suppose that $(G,*)$ is a subsemihypergroup of $(H,*)$. First we have to prove that G is closed with respect to the operation $\cdot$ of H. Yet for arbitrary elements $a,b\in G$ the fact that $a*b\subseteq G$ implies that $[a\cdot b)_{\leq}\subseteq G$, i.e. any element $x\in H$ such that $a\cdot b\leq x$ belongs to G. Since the relation $\leq$ is reflexive, we get that $a\cdot b\in G$. As a result $(G,\cdot)$ is a groupoid. The fact that it is associative is granted by the reasoning of the proof of Theorem 3.1, part 1. Altogether we get that $(G,\cdot)$ is a subsemigroup of $(H,\cdot)$.

2. **"⇒"** Denote u the identity of $(H,\cdot)$. If $(G,\cdot)$ is a submonoid of $(H,\cdot)$, then $u\in G$ and obviously the statement holds for this identity.

   "⇐" Cf. part 2 of the proof of Theorem 3.1, which may be literally repeated.

From this proof, part 1 "⇐", we directly get an obvious statement equivalent to the one included in Remark 3.5. Notice that its validity does not depend on the fact whether G is an upper end of H.

Corollary 3.8 *Let $(H,*)$ be the associated hypergroupoid of a quasi–ordered groupoid $(H,\cdot,\leq)$ and G a nonempty subset of H. If $(G,*)$ is a subhypergroupoid of $(H,*)$, then $(G,\cdot)$ is a subgroupoid of $(H,\cdot)$.*

The issue of subhypergroups seems to be a bit more complicated.

Theorem 3.9 *Let $(H, *)$ be the associated semihypergroup of a partially ordered semigroup $(H, \cdot, \leq)$. Suppose that G is an upper end of H. If $(G, \cdot)$ is a subgroup of $(H, \cdot)$, then $(G, *)$ is a subhypergroup of $(H, *)$.*

Proof. Since we assume that $(G, \cdot)$ is a subgroup of $(H, \cdot)$, we have that for an arbitrary $a, b \in G$ there holds $a \cdot b^{-1} \in G$, $b^{-1} \cdot a \in G$. Therefore if elements $c = b^{-1} \cdot a$ and $c' = a \cdot b^{-1}$ are regarded, Theorem 2.2 may be directly applied (or rather, its proof literally copied) since the relations $\leq$ of Theorem 2.2 and of the above theorem are identical.

As follows from the proof of Theorem 2.6 the subhypergroup is a transposition hypergroup or (if it is commutative) a join space (i.e. the proof can be directly applied). Unfortunatelly, due to Theorem 2.7 such a subhypergroup is not a canonical hypergroup.

Remark 3.10 *Notice that $(G, *)$, where G is such as defined in the assumptions of Lemma 3.4, part 4, can never be a subhypergroup of $(H, *)$. In this case the inclusion $G \subseteq a * G$ of the reproduction axiom is problematic. Indeed, suppose an arbitrary element $a \in G$ and any element $g \in G$ for which there holds $g \leq x_i$, where x_i is an arbitrary of those elements because of which G is not an upper end of H. In other words, g is such that there holds $[g)_\leq \not\subseteq G$. Then we have that $a * g = [a \cdot g]_\leq = [b)_\leq$ and thanks to the assumption of Lemma 3.4, part 4, $g \notin [b)_\leq$, which means that $G \not\subseteq a * G$.*

Remark 3.11 *Also notice that Theorem 3.9 holds for quasi–ordered groups as well as the antisymmetry of relation $\leq$ is not needed in Theorem 3.7, part 1 $\Rightarrow$.*

Proposition 3.12 *Let $(H, *)$ be the associated semihypergroup of a partially ordered semigroup $(H, \cdot, \leq)$ and $G \subseteq H$ nonempty. If $(G, *)$ is a subhypergroup of $(H, *)$, then $(G, \cdot)$ is a subsemigroup of $(H, \cdot)$ and G is an upper end of H such that for any pair $a, b \in G$ there exists a pair $c, c' \in G$ such that $b \cdot c \leq a$ and $c' \cdot b \leq a$.*

Proof. Thanks to Lemma 3.4, Remark 3.5, Theorem 3.9 and Remark 3.10 it is obvious that all "Ends lemma"–based subhypergroups $(G, *)$ of $(H, *)$ are such that G is an upper end of H. Since every hypergroup is a semihypergroup, we get that $(G, *)$ is a subsemihypergroup of $(H, *)$. Yet according to Theorem 3.7, part 1, $(G, \cdot)$ is in this case a subsemigroup of $(H, \cdot)$. The proposition for the arbitrary pair $a, b \in G$ is a copy of condition 1^0 of Theorem 2.2.

What remains to be proved is whether $(G, \cdot)$ in the above proposition is a subgroup of $(H, \cdot)$. This is still an open question. Notice that in Proposition 3.12 there need not at all be $u \in G$, where u is the identity of $(H, \cdot)$.

Remark 3.13 *Suppose that $(H, \cdot, \leq)$ is a partially ordered group and G is a non-empty subset of H. If $(G, \cdot)$ is simultaneously a subgroup of $(H, \cdot)$ and an upper end of H, then notice the following:*

If we take an arbitrary $x \in H$ such that $x < g$, where $g \in G$ is arbitrary, then $x < g$ implies $u < x^{-1} \cdot g$ and since $(G, \cdot)$ is a subgroup of $(H, \cdot)$, which is a group, and simultaneously G is an upper end of H, we get that $x^{-1} \cdot g \in G$. Yet since $g \in G$, there is also $x^{-1} \in G$, which implies that $x \in G$.

As a result we get that if $(H, \cdot, \leq)$ is a linear ordered group, there do not exist any proper subhypergroups associated to subgroups of $(H, \cdot)$ because there are no proper subgroups $(G, \cdot)$ of $(H, \cdot)$, where G is an upper end of H. Theorem 3.9 is thus of no practical use for linear ordered groups. Also cf. Remark 3.10, which states that upper end are the only candidates for subhypergroups.

However, if $(H, \cdot)$ is a monoid only, then $x < g$ does not imply $u < x^{-1} \cdot g$ and consequently $x \in G$ because x need not have the inverse element.

Remark 3.14 *The issue of commutativity has already been discussed in the original "Ends lemma", namely in Theorem 2.1. For a partially ordered semigroup $(H, \cdot, \leq)$ there holds that the hyperoperation $*$ is commutative if and only if the single valued operation $\cdot$ is commutative. However, if $\leq$ is not antisymmetric, i.e. if it is only a quasi–ordering, the proof of Theorem 2.1 as included in [3] cannot be repeated. Indeed, suppose that for an arbitrary pair $a, b \in H$ there holds $a * b = b * a$. This means that $[a \cdot b)_{\leq} = [b \cdot a)_{\leq}$, from which there follows $a \cdot b = b \cdot a$ only on condition of antisymmetry of the relation $\leq$.*

Example 3.15 *Let $H = \{a, b, c\}$ and define operation $\cdot$ on H by the following table:*

$\cdot$	a	b	c
a	a	a	a
b	a	b	c
c	a	a	a

Define that for an arbitrary pair $x, y \in H$ there holds $x \leq y$. It can be easily verified that $(H, \cdot, \leq)$ is a quasi–ordered semigroup (and that $\leq$ is not antisymmetric). Further define the hyperoperation in the usual "Ends lemma" way, i.e. for an arbitrary pair $x, y \in H$ define $x * y = [x \cdot y)_{\leq}$. Thanks to the definition of the relation $\leq$ we get that $[a)_{\leq} = [b)_{\leq} = [c)_{\leq}$, which means that for an arbitrary pair of elements $x, y \in H$ there holds $x * y = y * x$. Yet despite the fact that the hyperoperation $*$ is commutative, the single valued operation $\cdot$ is not commutative.

3.2 Other cases

If we define the end generated by an element $a \in G$, where $G \subseteq H$, as $[a)_{\leq_G} = \{x \in G : a \leq x\}$, problems of "holes" caused by elements $x \in H \setminus G$ (cf. Definition 3.3) will not come up. Technically speaking there are two distinct hyperoperations in the following theorem: $*$ and $*_G$. Therefore, the (hyper)structures on G are not called sub(hyper)structures. As far as commutativity of the below mentioned hyperstructures is concerned, Remark 3.14 is applicable.

Theorem 3.16 *Let $(H, *)$ be the associated semihypergroup of a partially ordered semigroup $(H, \cdot, \leq)$. Further, let $G \subseteq H$ be non-empty and such that $(G, \cdot)$ is a subgroupoid of $(H, \cdot)$ and the relation $\leq_G$ be a restriction of $\leq$ on G, i.e. for arbitrary elements $a, b \in G$ let $a \leq b \Rightarrow a \leq_G b$. Finally – if it exists – denote u the identity of $(H, \cdot)$ and define a new hyperoperation $*_G : G \times G \to P^*(G)$ for arbitrary elements $a, b \in G$ by*

$$a *_G b = [a \cdot b)_{\leq_G} = \{x \in G; a \cdot b \leq_G x\}.$$

Then

1. *$(G, \cdot)$ is a semigroup if and only if $(G, *_G)$ is a semihypergroup.*

2. *$(G, \cdot)$ is a monoid if and only if $(G, *_G)$ is a semihypergroup and $u \in G$.*

3. *If $(G, \cdot)$ is a group, then $(G, *)$ is a transposition hypergroup.*

4. *If $(G, *)$ is a hypergroup, then $(G, \cdot)$ is a semigroup such that for any pair $a, b \in G$ there exists a pair $c, c' \in G$ such that $b \cdot c \leq a$ and $c' \cdot b \leq a$.*

Proof. The theorem is a simple corollary to the "Ends lemma", Theorem 3.1 and Theorem 2.2.

Remark 3.17 *The fact that G is closed with respect to $\cdot$ is again essential: suppose a triple a, b, c such that $a, b \in G$ and $c \in H \setminus G$. If now $a *_G b$ was constructed, we would get $a *_G b = [a \cdot b)_{\leq_G} = [c)_{\leq_G}$, which is difficult to be assigned with any sense since due to reflexivity of $\leq_G$ there must hold $c \in [c)_{\leq_G}$, i.e. $c \in \{x \in G; c \leq x\}$ yet we suppose that $c \notin G$.*

Remark 3.18 *If $(H, *)$ is quasi–ordered, then only the "$\Rightarrow$" implications hold: the fact that $(G, \cdot)$ is a semigroup (monoid) implies the fact that $(G, *_G)$ is a subsemihypergroup (and $u \in G$). Cf. Remark 3.2 for a counterexample of the "$\Leftarrow$" implications.*

Acknowledgement

The paper was supported by the Council of Czech Government grant No. MSM0021630529.

References

[1] CORSINI, P.: *Prolegomena of Hypergroup Theory.* Tricesimo: Aviani Editore 1993.

[2] CORSINI, P., LEOREANU, V.: *Applications of Hyperstructure Theory.* Dodrech – Boston – London: Kluwer Academic Publishers, 2003.

[3] CHVALINA, J.: *Functional Graphs, Quasi-ordered Sets and Commutative Hypergroups.* Brno: Masaryk University, 1995 (in Czech).

[4] CHVALINA, J., MOUČKA.: *When certain basic hyperoperations on lattices coincide.* In 5. Matematický workshop s mezinárodní účastí. Brno: FAST VUT, 2006. (CD-ROM).

[5] HOŠKOVÁ, Š.: *Binary hyperstructures determined by relational and transformation systems. Habilitation Thesis.* Faculty of Sciences, University of Ostrava. 2008

[6] HOŠKOVÁ, Š.: *Order Hypergroups—The unique square root condition for quasi-order hypergroups.* Set-valued Mathematics and Applications, **1** (2008), No. 1, Global Research Publications, pp. 1–7.

[7] NOVÁK, M.: *Identities and inverses of "Ends lemma"–based hyperstructures.* Submitted to Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Mathematica, submitted 2009-07-01.

[8] NOVÁK, M.: *Identities of some join spaces of certain transformation operators.* In XXVII International Colloquium, Proc. of Contributions, Brno: University of Defence, 2009.

[9] RAČKOVÁ, P.: *Hypergroups of symmetric matrices.* In Proc. 10^{th} International Congress Algebraic Hyperstructures and Applications (AHA 2008). Brno: University of Defence, 2009, pp. 267–272.

[10] VOUGIOUKLIS, Th..: *Representations of hypergroups by generalized permutations.* Algebra Universalis **29** (1992), pp. 172–183.

Current address

RNDr. Michal Novák, Ph.D.
Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 8, 616 00 Brno, Czech Republic, tel: +420-541143135
email: novakm@feec.vutbr.cz

Journal of Applied Mathematics

volume III (2010), number II

SOME QUANTITIES RELATED TO THE DISTANCE MATRIX OF A SPECIAL TYPE OF DIGRAPHS

SEIBERT Jaroslav, (CZ)

Abstract. The distance matrix of a digraph is a square matrix which has as its entries the lengths of the shortest oriented path between each pair of vertices. It is interesting to investigate such digraphs G_n , which have the companion matrix as their adjacency matrix. The Wiener index of a graph is defined as the sum of the all entries of the distance matrix. In this contribution we calculate the Wiener index and the determinant of the distance matrix for a special type of the graphs G_n.
Key words. Digraph, distance matrix, companion matrix, Wiener index, determinant
Mathematics Subject Classification: Primary 05C12, 05C20, Secondary 11C20, 11C08.

1. Introduction

The graphs in this contribution are directed weighted graphs without loops. The distance matrix is one of the most useful matrices, which characterize the structure of graphs.

Definition 1
Let G be a weighted digraph with n vertices $v_1, v_2, \ldots, v_n$. Then the distance matrix of G is defined as the $n \times n$ matrix $D(G) = D = (d_{ij})$, where

$$\begin{aligned} d_{ij} &= \text{ the distance from the vertex } v_i \text{ to the vertex } v_j\,, \\ &= 0, \text{ if } i = j\,, \\ &= \infty\,, \text{ if no path from } v_i \text{ to } v_j \text{ exists.} \end{aligned}$$

The distance from v_i to v_j is the weight of the shortest path from v_i to v_j. The sum of all (nondiagonal) entries of the distance matrix is called the Wiener index $W(G)$ of a graph G.
The distance polynomial of G is defined as $P(G; x) = \det(x\,I - D)$, where I is the unit matrix of the size $n \times n$.

We will only use strongly connected digraphs which have the companion matrices as their adjacency matrices. The companion matrix can be defined, e.g. [3], as the $n \times n$ square matrix

$$C_n = C_n\left(u_1, u_2, \ldots, u_n\right) = \begin{pmatrix} u_1 & u_2 & u_3 & \ldots & u_{n-1} & u_n \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ & & & \ldots\ldots & & \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

where $u_1, u_2, \ldots, u_n$ are nonnegative integers for our purpose. The corresponding digraphs can be denoted by $G_n(u_1, u_2, ..., u_n)$.

In [4] we found some results for the digraph $G_n(0, 0, ..., 0, 1)$ which is a cycle with n vertices.

Theorem 1 ([4], Theorem 6)
For a cycle G_n with $n \geq 2$ vertices the following statements hold.

1.

$$\det D(G_n) = (-1)^{n-1} \binom{n}{2} n^{n-2},$$

2.

$$P(G_n; x) = \left(x - \binom{n}{2} \right) \prod_{j=2}^{n} \left(x - \frac{\varepsilon_j}{(1-\varepsilon_j)^2}(1 - n\varepsilon_j{}^{n-1} + (n-1)\varepsilon_j{}^n \right),$$

where ε_j, $j = 1, ..., n$, are the n-th roots of unity,

3. the matrix $D(G_n)$ has the eigenvalues

$$x_1 = \binom{n}{2}, x_j = \frac{\varepsilon_j}{(1-\varepsilon_j)^2}(1 - n\varepsilon_j{}^{n-1} + (n-1)\varepsilon_j{}^n),$$

for $2 \leq j \leq n$.

In [5] we investigated the digraphs $G_n(0, 1, ..., 1)$ and we obtained the following results.

Theorem 2
Let $G_n = G_n(0, 1, ..., 1)$ be an digraph of the given type with n vertices. Then

a) ([5], Theorem 2)
The Wiener index is given by the relation

$$W(G_n) = 2\binom{n+1}{3}$$

for any integer $n \geq 2$.

b) ([5], Theorem 3)

For the determinant of the distance matrix of a graph G_n the following recurrence holds
$\det D_n = (-1)^{n+1}(n-1)! - n \det D_{n-1}$, where $n \geq 2$, and $\det D_1 = 0$.

c) ([5], Corollary)
For any positive integer n the relation $\det D_n = (-1)^{n+1} a_n$ holds, where a_n is the n-th generalized Stirling number.

d) ([5], Theorem 4)
A recursive formula for the distance polynomial $P(G_n;x)$ has the form

$$P(G_n;x) = (x+n)\, P(G_{n-1};x) + x \det B_{n-1} - (x+1)(x+2)...(x+n-1)$$

for any integer $n \geq 2$ and $P(G_1;x) = x$. The corresponding sequence of the determinants of B_n satisfies the recurrence $\det B_n = x \det B_{n-1} - (x+1)(x+2)...(x+n-1)$, with $\det B_1 = -1$.

2. The main results

In this paper we are concerned with the digraphs $G_n = G_n(0, 1, 2, \ldots, n-1)$ corresponding to the companion matrix $C_n(0, 1, 2, \ldots, n-1)$. Let $\{v_1, v_2, \ldots, v_n\}$ be the vertex set of G_n, then the arc (v_1, v_j), $j = 2, 3, \ldots, n$, has the weight $j-1$ and the arc (v_j, v_{j-1}) has the weight 1.

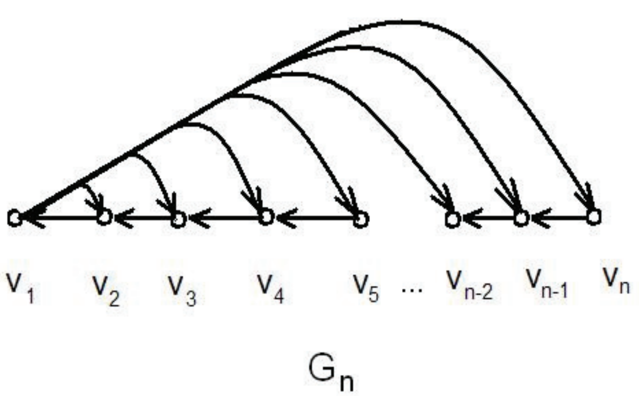


Fig. 1.

It is easy to see that distance matrix D_n of such digraph G_n has the form

$$D_n = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & n-2 & n-1 \\ 1 & 0 & 3 & 4 & \cdots & n-1 & n \\ 2 & 1 & 0 & 5 & \cdots & n & n+1 \\ 3 & 2 & 1 & 0 & \cdots & n+1 & n+2 \\ & & & \cdots\cdots & & & \\ n-2 & n-3 & n-4 & n-5 & \cdots & 0 & 2n-3 \\ n-1 & n-2 & n-3 & n-4 & \cdots & 1 & 0 \end{pmatrix}.$$

Theorem 3
The Wiener index of the above mentioned digraph G_n has the value

$$W(G_n) = 4\binom{n+1}{3} - 2\binom{n}{2}$$

for any integer $n \geq 2$.

Proof
We can write with respect to the definition of the Wiener index

$$W(G_n) = \sum_{k=1}^{n-1}\sum_{i=1}^{k} i + \sum_{k=1}^{n-1}\sum_{i=k}^{2k-1} i = \sum_{k=1}^{n-1}\binom{k+1}{2} + \sum_{k=1}^{n-1}\frac{k}{2}(3k-1) = \binom{n+1}{3} + 3\sum_{k=1}^{n-1}\binom{k+1}{2} - 2\sum_{k=1}^{n-1}\binom{k}{1} =$$

$$= \binom{n+1}{3} + 3\binom{n+1}{3} - 2\binom{n}{2} = 4\binom{n+1}{3} - 2\binom{n}{2}$$

by using simple combinatorial identities.

To derive a formula for the determinant of the distance matrix D_n of the digraphs G_n we will use the well – known statement.

Proposition
Let A_1, A_2, B be the $n \times n$ square matrices such that

$$A_1 = \begin{pmatrix} a_{11}\dots a_{1n} \\ \dots\dots \\ a_{i1}\dots a_{in} \\ \dots\dots \\ a_{n1}\dots a_{nn} \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_{11}\dots a_{1n} \\ \dots\dots \\ b_{i1}\dots b_{in} \\ \dots\dots \\ a_{n1}\dots a_{nn} \end{pmatrix}, \quad B = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \dots\dots & \\ a_{i1}+b_{i1} & \dots & a_{in}+b_{in} \\ & \dots\dots & \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Then $\det B = \det A_1 + \det A_2$ and the statement holds for an arbitrary i-th row (or column), where $i = 1, 2, \dots, n$.

Theorem 4
For any integer $n \geq 2$ the following recurrence relation holds

$$\det D_n = 2\,(1-n)\det D_{n-1} + (-1)^{n-1}\,2^{n-2}\,(n-1)!$$

with $D_1 = 0$.

Proof
Let A_n be the $n \times n$ matrix obtained from D_n such that one is subtracted from the each entries of the last row of D_n. Then we have successively for any $n \geq 2$

$$\det A_n = \begin{vmatrix} 0 & 1 & 2 & 3 & \dots & n-2 & n-1 \\ 1 & 0 & 3 & 4 & \dots & n-1 & n \\ 2 & 1 & 0 & 5 & \dots & n & n+1 \\ 3 & 2 & 1 & 0 & \dots & n+1 & n+2 \\ & & & \dots\dots & & & \\ n-2 & n-3 & n-4 & n-5 & \dots & 0 & 2n-3 \\ n-2 & n-3 & n-4 & n-5 & \dots & 0 & -1 \end{vmatrix} = \det D_n + \begin{vmatrix} 0 & 1 & 2 & 3 & \dots & n-2 & n-1 \\ 1 & 0 & 3 & 4 & \dots & n-1 & n \\ 2 & 1 & 0 & 5 & \dots & n & n+1 \\ 3 & 2 & 1 & 0 & \dots & n+1 & n+2 \\ & & & \dots\dots & & & \\ n-2 & n-3 & n-4 & n-5 & \dots & 0 & 2n-3 \\ -1 & -1 & -1 & -1 & \dots & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + \begin{vmatrix} 0 & 1 & 2 & \dots & n-3 & n-2 & n-1 \\ 1 & 0 & 3 & \dots & n-2 & n-1 & n \\ 1 & 1 & 0 & \dots & n-1 & n & n+1 \\ & & & \dots\dots & & & \\ n-3 & n-4 & n-5 & \dots & 0 & 2n-5 & 2n-4 \\ n-3 & n-4 & n-5 & \dots & 0 & -1 & 2n-4 \\ -1 & -1 & -1 & \dots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + \begin{vmatrix} 0 & 1 & 2 & \dots & n-3 & n-2 & n-1 \\ 1 & 0 & 3 & \dots & n-2 & n-1 & n \\ 2 & 1 & 0 & \dots & n-1 & n & n+1 \\ & & & \dots\dots & & & \\ n-3 & n-4 & n-5 & \dots & 0 & 2n-5 & 2n-4 \\ 0 & 0 & 0 & \dots & 0 & 4-2n & 0 \\ -1 & -1 & -1 & \dots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (4-2n) \begin{vmatrix} 0 & 1 & 2 & \dots & n-4 & n-3 & n-1 \\ 1 & 0 & 3 & \dots & n-3 & n-2 & n \\ 2 & 1 & 0 & \dots & n-2 & n-1 & n+1 \\ & & & \dots\dots & & & \\ n-4 & n-5 & n-6 & \dots & 0 & 2n-7 & 2n-5 \\ n-4 & n-5 & n-6 & \dots & 0 & -1 & 2n-5 \\ -1 & -1 & -1 & \dots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (4-2n)\begin{vmatrix} 0 & 1 & 2 & \ldots & n-4 & n-3 & n-1 \\ 1 & 0 & 3 & \ldots & n-3 & n-2 & n \\ 2 & 1 & 0 & \ldots & n-2 & n-1 & n+1 \\ & & & \ldots\ldots & & & \\ n-4 & n-5 & n-6 & \ldots & 0 & 2n-7 & 2n-5 \\ 0 & 0 & 0 & \ldots & 0 & 6-2n & 0 \\ -1 & -1 & -1 & \ldots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (4-2n)(6-2n)\begin{vmatrix} 0 & 1 & 2 & \ldots & n-5 & n-4 & n-1 \\ 1 & 0 & 3 & \ldots & n-4 & n-3 & n \\ 2 & 1 & 0 & \ldots & n-3 & n-2 & n+1 \\ & & & \ldots\ldots & & & \\ n-5 & n-6 & n-7 & \ldots & 0 & 2n-9 & 2n-6 \\ n-5 & n-6 & n-7 & \ldots & 0 & -1 & 2n-6 \\ -1 & -1 & -1 & \ldots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (4-2n)(6-2n)\begin{vmatrix} 0 & 1 & 2 & \ldots & n-5 & n-4 & n-1 \\ 1 & 0 & 3 & \ldots & n-4 & n-3 & n \\ 2 & 1 & 0 & \ldots & n-3 & n-2 & n+1 \\ & & & \ldots\ldots & & & \\ n-5 & n-6 & n-7 & \ldots & 0 & 2n-9 & 2n-6 \\ 0 & 0 & 0 & \ldots & 0 & 8-2n & 0 \\ -1 & -1 & -1 & \ldots & -1 & -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (4-2n)(6-2n)(8-2n)\begin{vmatrix} 0 & 1 & \ldots & n-5 & n-1 \\ 1 & 0 & \ldots & n-6 & n \\ & & \ldots\ldots & & \\ n-5 & n-6 & \ldots & 0 & 2n-6 \\ -1 & -1 & \ldots & -1 & -1 \end{vmatrix} =$$

$$= \ldots = \det D_n + (4-2n)(6-2n)(8-2n)\ldots(-2)\begin{vmatrix} 0 & n-1 \\ -1 & -1 \end{vmatrix} =$$

$$= \det D_n + (n-1)\prod_{i=2}^{n-1}(2i-2n) = \det D_n + (n-1)(-1)^{n-2}2^{n-2}(n-2)! = \det D_n + (-1)^{n-2}2^{n-2}(n-1)!$$

It is also possible to calculate det A_n by subtracting the (n-1)-st row from the n-th row. Then

$$\det A_n = \begin{vmatrix} 0 & 1 & 2 & 3 & \dots & n-2 & n-1 \\ 1 & 0 & 3 & 4 & \dots & n-1 & n \\ 2 & 1 & 0 & 5 & \dots & n & n+1 \\ 3 & 2 & 1 & 0 & \dots & n+1 & n+2 \\ & & & \dots\dots & & & \\ n-2 & n-3 & n-4 & n-5 & \dots & 0 & 2n-3 \\ 0 & 0 & 0 & 0 & \dots & 0 & 2-2n \end{vmatrix} = (2-2n)\det D_{n-1}.$$

The recurrence for det D_n immediately follows after comparing the both obtained expressions.

Theorem 5
For any positive integer n the identity

$$\det D_n = (-1)^{n-1}(n-1)\,(n-1)!\,2^{n-2}$$

holds, where det $D_1 = 0$.

Proof
It can be done by induction on n. First, it is easy to see that det $D_2 = -1$. Now, suppose that the formula is true for an arbitrary integer $n \geq 2$. Then

$$\det D_{n+1} = -2n\det D_n + (-1)^n n!\,2^{n-1} = -2n\,(-1)^{n-1}(n-1)\,(n-1)!\,2^{n-2} + (-1)^n n!\,2^{n-1} =$$
$$= (-1)^n n!\,2^{n-1}(n-1+1) = (-1)^n n \cdot n!\,2^{n-1}$$

which completes the proof.

The sequence from Theorem 5 is A 014479 in Neil Sloane´s On–line Encyclopedia of Integer Sequences [6]. Its exponential generating function is $f(x) = \dfrac{1+2x}{(1-2x)^3}$.

3. Concluding remarks

In [4] we created the computer program to calculate some quantities which are related to the distance matrix of a digraph. We used Floyd's algorithm for calculation of the distance matrix of a weighted digraph. After small arrangement the created program works for undirected graphs, too. Then the distance matrix can be used to compute another useful invariants of a graph G, that are related to the center of the graph. These invariants are for example the eccentricity of a vertex in G, the radius and the diameter of a graph G .

References

[1] BANG–JENSEN, J., GUTIN, G.: *Digraphs, Theory, Algorithms and Applications.* Springer – Verlag, London, 2001.

[2] CHEN, W.Y.C., LOUCK, J. D.: *The combinatorial power of the companion matrix.* Linear Algebra and its Applications, 232 (1996), pp. 261 – 278.

[3] LANCASTER, P., TISMENSKI, M.: *The Theory of Matrices.* Academic, 1985.
[4] SEIBERT, J., TROJOVSKÝ, P.: *The distance matrix of a graph and computation of it and some related quantities.* Proceedings of the 4th International Conference APLIMAT 2005, Bratislava, pp. 351 – 356.
[5] SEIBERT, J., TROJOVSKÝ, P.: *Some properties of directed graphs corresponding to the companion matrix APLIMAT–Journal of Applied Mathematics, Volume II (2009),* No 1, pp. 55 – 64.
[6] SLOANE, N. J. A.: "The On–Line Encyclopedia of Integer Sequences." http://www.research.att.com/~njas/sequences/

Current address

Jaroslav Seibert
Institute of Mathematics,
Faculty of Economics and Administration,
University of Pardubice, Studentska 95, 532 10 Pardubice, Czech republic,
tel. number +420466036019,
e-mail jaroslav.seibert@upce.cz

Journal of Applied Mathematics

volume III (2010), number II

GIANT BODIES PIECE BY PIECE

SZALAY István, (HU)

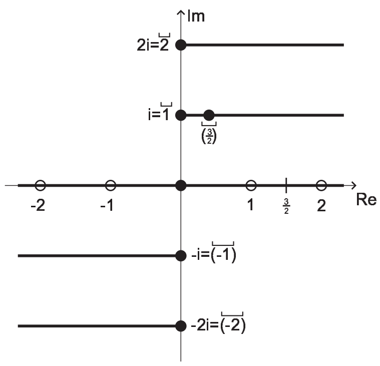
Abstract. A three - dimensional set is called giant body if its diameter is not finite. Giant bodies are described by exploded numbers. The ordered field of exploded numbers is isomorphic with the ordered field of real numbers such that the set of real numbers is a proper subset of the set of exploded numbers. In this paper two kinds of examples are given for giant bodies: super- balls and super-octaeders.

1 Introduction

The c-explosion of real numbers was introduced in [1]. In this paper we use the case $c = 1$, only. For the sake of simplicity we repeat the most important facts, illustrated by the complex model of exploded real numbers which had already been introduced in [2].
For any real number x we say that its exploded is the complex number

$$\overset{\sqcup}{x} = (\operatorname{sgn} x) \cdot (\operatorname{area\,th}\{|x|\} + i \cdot [|x|]); x \in R \tag{1.1}$$

where $[x]$ is the greatest integer number which is less than or equal to x and $\{x\} = x - [x]$. The set of exploded numbers is a proper subset of the set of complex numbers:



Denoting the set of exploded numbers by $\overset{\sqcup}{R}$ the figure shows that $R \subset \overset{\sqcup}{R}$. This model is called complex model because the set $\overset{\sqcup}{R}$ is a subset of the set of complex numbers C. In this model $\overset{\sqcup}{x} < \overset{\sqcup}{y}$ if $\mathrm{Im}\, \overset{\sqcup}{x} < \mathrm{Im}\, \overset{\sqcup}{y}$ or if $\mathrm{Im}\, \overset{\sqcup}{x} = \mathrm{Im}\, \overset{\sqcup}{y}$ then $\mathrm{Re}\, \overset{\sqcup}{x} < \mathrm{Re}\, \overset{\sqcup}{y}$.
In the complex model of exploded numbers the compression is

$$\underset{\sqcup}{u} = \mathrm{Im}\, u + \mathrm{th}\, \mathrm{Re}\, u; \quad u \in \overset{\sqcup}{R}. \tag{1.2}$$

By (1.1) and (1.2) we have that for any $x \in R$ the first inversion identity

$$\underset{\sqcup}{(\overset{\sqcup}{x})} = x \tag{1.3}$$

holds. Conversely, for any $u \in \overset{\sqcup}{R}$ the second inversion identity

$$\overset{\sqcup}{(\underset{\sqcup}{u})} = u \tag{1.4}$$

is valid. Moreover, we have

Theorem of uniqueness: For any $x, y \in R$, $\overset{\sqcup}{x} = \overset{\sqcup}{y} \Longleftrightarrow x = y$.

Theorem of ordering: For any $x, y \in R$, $\overset{\sqcup}{x} < \overset{\sqcup}{y} \Longleftrightarrow x < y$.

Definition of super - addition: $\overset{\sqcup}{x} \oplus \overset{\sqcup}{y} = \overset{\sqcup}{(x+y)}$.

Definition of super -multiplication: $\overset{\sqcup}{x} \otimes \overset{\sqcup}{y} = \overset{\sqcup}{(x \cdot y)}$.

Property of monotonity of super - addition: $\forall u, v, w \in \overset{\sqcup}{R}$,

$$u < v \Rightarrow u \oplus w < v \oplus w.$$

Property of monotonity of super - multiplication: $\forall u, v, w \in \overset{\sqcup}{R}$

$$(u < v) \wedge (0 < w) \Rightarrow u \otimes w < v \otimes w.$$

So, $(\overset{\sqcup}{R}, \oplus, \otimes)$ is an ordered field which is isomorphic with the ordered field $(R, +, \cdot)$. Other super - operations:

Super - subtraction: $\overset{\sqcup}{x} \ominus \overset{\sqcup}{y} = \overset{\sqcup}{(x-y)}; x, y \in R$.

Super - division: $\overset{\sqcup}{x} \odot \overset{\sqcup}{y} = \overset{\sqcup}{(x:y)} = \overset{\sqcup}{\left(\frac{x}{y}\right)}; y \in R, y \neq 0$.

2 The explosion and compression of R^3

Let $X = (x, y, z)$ be an arbitrary point of our familiar three - dimensional space R^3. We say that its exploded is $\overset{\sqcup}{X} = (\overset{\sqcup}{x}, \overset{\sqcup}{y}, \overset{\sqcup}{z})$. The set $\overset{\sqcup}{R^3} = \{\overset{\sqcup}{X} : X \in R^3\}$ is called exploded three -

dimensional space. If $|x| < 1$ then by (1.1)

$$\overset{\sqcup}{x} = (\operatorname{sgn} x) \cdot (\operatorname{area\,th}\{|x|\} + i[|x|]) = \operatorname{area\,th} x \in R; \quad \operatorname{area\,th} x = \frac{1}{2} \ln \frac{1+x}{1-x}. \tag{2.1}$$

Hence if X is a point of the open cube $\underset{\sqcup}{R^3} = \{(x,y,z) \in R^3 := 1 < x < 1;\ -1 < y < 1;\ -1 < z < 1\}$ then $\overset{\sqcup}{X} \in R^3$. But if only one of the coordinates of $X \in R^3$ has that its absolute value is greater than or equal to 1 then $\overset{\sqcup}{X} \notin R^3$. Of course, $\overset{\sqcup}{X} \in \overset{\sqcup}{R^3}$ but it is invisible in the familiar three-dimensional space R^3. Considering an $U = (u,v,w) \in \overset{\sqcup}{R^3}$ its compressed is $\underset{\sqcup}{U} = (\underset{\sqcup}{u}, \underset{\sqcup}{v}, \underset{\sqcup}{w}) \in R^3$ and by (1.3) and (1.4) we have the inversion identities

$$\underset{\sqcup}{(\overset{\sqcup}{X})} = X; \quad X \in R^3 \tag{2.2}$$

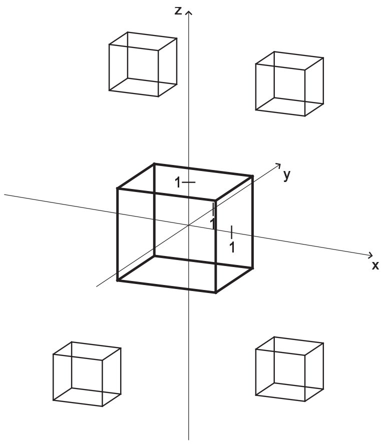
and

$$\overset{\sqcup}{(\underset{\sqcup}{U})} = U; \quad U \in \overset{\sqcup}{R^3}. \tag{2.3}$$

If $u \in R$ then (1.2) yields

$$\underset{\sqcup}{u} = \operatorname{th} u = \frac{e^u - e^{-u}}{e^u + e^{-u}}; \quad u \in R. \tag{2.4}$$

Hence, we have that $U \in R^3$, then $\underset{\sqcup}{U} = (\operatorname{th} u, \operatorname{th} v, \operatorname{th} w) \in \underset{\sqcup}{R^3}$. So, the central open cube can be considered as a compressed model of R^3. (See [3].) Similarly, itself R^3 can be considered as a compressed model of $\overset{\sqcup}{R^3}$.



For example, in the case of the closed set

$$L = \{U(u,v,w) \in \overset{\sqcup}{R^3} : u = t, v = t, w = t;\ \overset{\sqcup}{1} \le t \le \overset{\sqcup}{1}\} \tag{2.5}$$

using (2.4) we have that its compressed set

$$\underset{\sqcup}{L} = \{X(x,y,z) \in R^3 : x = \underset{\sqcup}{t}, y = \underset{\sqcup}{t}, z = \underset{\sqcup}{t}; -1 \le \underset{\sqcup}{t} \le 1\}. \tag{2.6}$$

(2.6) shows that it is not situated in the central open cube $\underbracket{R^3}$, because its points $(-1,-1,-1) \notin \underbracket{R^3}$ and $(1,1,1) \notin \underbracket{R^3}$. Consequently by (2.2) and (2.3) we obtain that L is not situated in our familiar three - dimensional space R^3. Really $(\overbracket{-1}, \overbracket{-1}, \overbracket{-1}); (\overbracket{1}, \overbracket{1}, \overbracket{1}) \notin R^3$. Following the analogy of $\underbracket{R^3} \subset R^3$, by the relation $R^3 \subset \overbracket{R^3}$ we can say that

$$R^3 = \{(x,y,z) \in \overbracket{R^3} : \overbracket{-1} < x < \overbracket{1}; \overbracket{-1} < y < \overbracket{1}; \overbracket{-1} < z < \overbracket{1}\} \tag{2.7}$$

is the central open cube of the exploded three - dimensional space $\overbracket{R^3}$. Now, we can see again that the end - points of L (given by (2.5)) are not elements of R^3. Considering $U_0 = (u_0, v_0, w_0) \in R^3$ such that $U_0 \neq O = (0,0,0)$ the open cube $R^3_{U_0} = \{(u,v,w) \in \overbracket{R^3} :$

$$\overbracket{-1} \circledast u_0 < u < \overbracket{1} \circledast u_0; \overbracket{-1} \circledast v_0 < v < \overbracket{1} \circledast v_0; \overbracket{-1} \circledast w_0 < w < \overbracket{1} \circledast w_0\}$$

is another three - dimensional space, different from our familiar three - dimensional space R^3. $R^3_{U_0}$ is called quasi - familiar three-dimensional space. The connection between three - dimensional spaces $R^3_{U_0}$ and R^3 is given by the super-shift transformation

$$\begin{aligned} u &= x \circledast u_0 \\ v &= y \circledast v_0 \\ w &= z \circledast w_0 \end{aligned}$$

Of course, it is possible that $R^3 \cap R^3_{U_0}$ is empty. If $R^3 \cap R^3_{U_0}$ is not empty we can say that the quasi - familiar three - dimensional space $R^3_{U_0}$ is a partially familiar three-dimensional space, too. We remark that at the transformation using super-addition is essential because in the case of the simple shift transformation

$$\begin{aligned} u &= x + u_0 \\ v &= y + v_0 \\ w &= z + w_0 \end{aligned}$$

the space R^3 does not change.

3 The concept of giant body

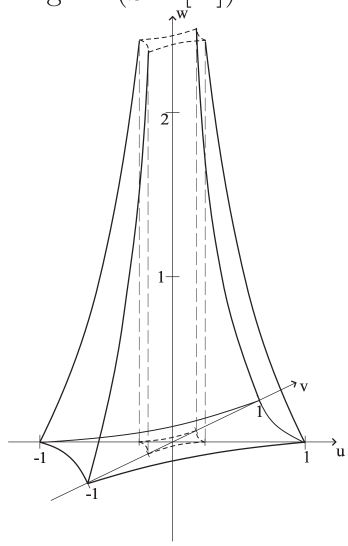
A set $H \subseteq \overbracket{R^3}$ is called giant body if there is not a quasi - familiar three - dimensional space $R^3_{U_0}$ such that $H \subseteq R^3_{U_0}$. Clearly, the greatest giant body is $\overbracket{R^3}$. On the other hand, for any $U_0 = (u_0, v_0, w_0) \in \overbracket{R^3}$, the set $R^3_{U_0}$ is not a giant body. The set L (given by (2.5)) is giant body. Considering a set $H \subseteq \overline{R^3}$, the set $H_{box} = H \cap R^3$ is called the box - phenomenon of H. Of course, it is possible that H_{box} is empty. It is not true that if H_{box} is unbounded then H is a giant body. Let us consider the box - phenomenon Λ_{box} of the super - pyramid Λ. Its description is (discussed in[4]) (see next figure left)

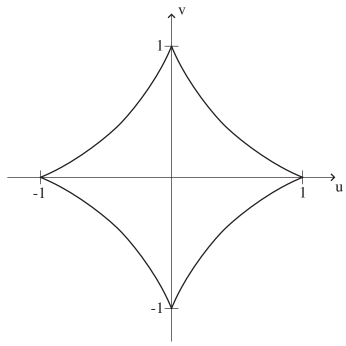
$$\Lambda = \{(u,v,w) \in \overline{R^3} : 0 \leq w \leq \overline{1,5} \otimes (\overline{1} \ominus ((|u| \odot 1) \oplus (|v| \odot 1))); (u,v) \in B\}, \quad (3.1)$$

where basis

$$B = \{(u,v,0) : |u| \oplus |v| \leq 1; u,v \in \overline{R}\} \quad (3.2)$$

is showed by the figure (See [4.])





If the points $(u,v,0)$ satisfy the equation $|u| \oplus |v| = \operatorname{area\,th} \frac{\operatorname{th} 1}{3}$ then by (3.2) $(u,v,0) \in B$ is obtained. Moreover, (3.1) yields that $w = \overline{1}$. Hence, the set Λ_{box} does not have any upper bound in R^3. On the other hand, the peak-point $(0,0,\overline{1,5}) \notin \Lambda_{box}$. Where is it? To answer this question we consider the open cube

$$R^3_{\overline{(0,0,\ 0,75)}} = \{(u,v,w) \in \overline{R^3} : \quad (3.3)$$

$$\overline{-1} < u < \overline{1}; \ \overline{-1} < v < \overline{1}; \ \overline{-1} \oplus \overline{0,75} < w < \overline{1} \oplus \overline{0,75}\}.$$

Clearly

$$R^3 \cap R^3_{\overline{(0,0,\ 0,75)}} = \{(x,y,z) \in R^3 : \ \overline{-1} < x < \overline{1}; \ \overline{-1} < y < \overline{1}; \ \overline{-0,25} < z < \overline{1}\}$$

so, $R^3_{\overline{(0,0,\ 0,75)}}$ is a partially - familiar three - dimensional space. We can see that the peak-point $(0,0,\overline{1,5}) \notin R^3 \cap R^3_{\overline{(0,0,\ 0,75)}}$ but $(0,0,\overline{1,5}) \in R^3_{\overline{(0,0,\ 0,75)}}$. Let us use another coordinate - system

$[\rho, \sigma, \tau]$

$$\begin{aligned} \rho &= u \\ \sigma &= v \\ \tau &= w \ominus \overline{0,75} \end{aligned} \tag{3.4}$$

instead of the coordinate - system $[u, v, w]$. The origo in this new coordinate - system is

$$0_* = (u = 0, v = 0, w = \overline{0,75}\,) = (\rho = 0, \sigma = 0, \tau = 0)$$

and by (3.3) the space $R^3_{(0,0,\,\overline{0,75}\,)}$ has the form

$$R^3_{0_*} = \{(\rho, \sigma, \tau) \in \overline{R^3} : \overline{-1} < \rho < \overset{\sqcup}{1};\ \overline{-1} < \sigma < \overset{\sqcup}{1};\ \overline{-1} < \tau < \overset{\sqcup}{1}\}.$$

Moreover, (3.1) and (3.4) yield a new description of super - pyramid Λ.

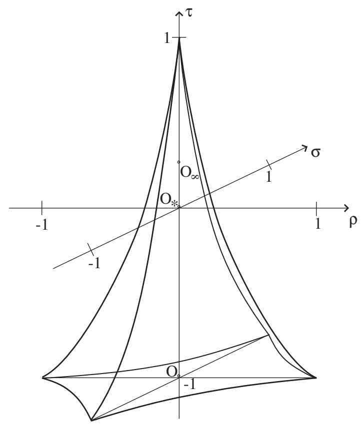
$\Lambda = \{(\rho, \sigma, \tau) \in \overline{R^3} :$

$$\overline{-0,75} \leq \tau \leq \overline{0,75} \ominus (\overline{1,5} \otimes ((|\rho| \oslash 1) \oplus (|\sigma| \oslash 1))); (\rho, \sigma) \in B\}$$

where

$$B = \{(\rho, \sigma, \overline{-0,75}\,) : |\rho| \oplus |\sigma| \leq 1; \rho, \sigma \in \overset{\sqcup}{R}\}.$$

Clearly $\Lambda \subset R^3_{0_*}$, so is not a giant body. We can see it in one piece.

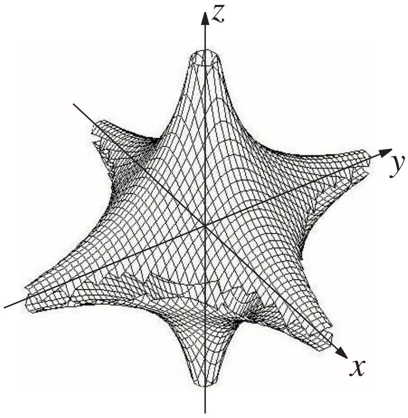


4 Giant balls

A typical giant body is the super ball (discussed in [5])

$$G_0(\overset{\sqcup}{1}) = \{(u, v, w) \in \overline{R^3} : (u \otimes u) \oplus (v \otimes v) \oplus (w \otimes w) \leq \overset{\sqcup}{1}\} \tag{4.1}$$

having the box - phenomenon



By this box - phenomenon it seems that $G_0(\overset{\sqcup}{1})$ has $0_\infty = (0, 0, \overset{\sqcup}{1})$ "upper", $0_{-\infty} = (0, 0, \overset{\sqcup}{-1})$ "lower" and other $((\overset{\sqcup}{-1}, 0, 0); (0, \overset{\sqcup}{-1}, 0); (\overset{\sqcup}{1}, 0, 0); (0, \overset{\sqcup}{1}, 0))$ peak-points. It is not true. For example, in the partially familiar three-dimensional space

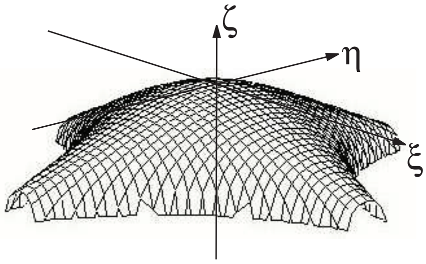
$R^3_{0_\infty} = \{(u, v, w) \in \overset{\sqcup}{R^3}$:

$$\overset{\sqcup}{-1} < u < \overset{\sqcup}{1};\ \overset{\sqcup}{-1} < v < \overset{\sqcup}{1};\ \overset{\sqcup}{-1} \oplus \overset{\sqcup}{1} = 0 < w < \overset{\sqcup}{2} = \overset{\sqcup}{1} \oplus \overset{\sqcup}{1}\} \tag{4.2}$$

using the coordinate - system $[\xi, \eta, \zeta]$

$$\begin{aligned} \xi &= u \\ \eta &= v \\ \zeta &= w \ominus \overset{\sqcup}{1} \end{aligned} \tag{4.3}$$

instead of the coordinate - system $[u, v, w]$, we can see, that $G_0(\overset{\sqcup}{1})$ in the neighborhood of 0_∞ is smooth. By (4.3) the new origo is the point $0_\infty = (u = 0, v = 0, w = \overset{\sqcup}{1}) = (\xi = 0, \zeta = 0, \rho = 0)$.



Considering the partially familiar three - dimensional space

$$R^3_{(0,0,\overset{\sqcup}{0,5})} = \{(u, v, w) \in \overset{\sqcup}{R^3} : \overset{\sqcup}{-1} < u < \overset{\sqcup}{1};$$

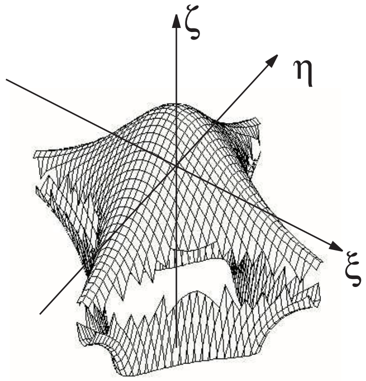
$$\overset{\vdash\dashv}{-1} < v < \overset{\sqcup}{1};\ \overset{\vdash\dashv}{-1} \oplus \overset{\vdash\dashv}{0,5} = \overset{\vdash\dashv}{-0,5} < w < \overset{\vdash\dashv}{1,5} = \overset{\sqcup}{1} \oplus \overset{\vdash\dashv}{0,5}\} \tag{4.4}$$

with the coordinate - system $[\xi, \eta, \zeta]$

$$\begin{aligned} \xi &= u \\ \eta &= v \\ \zeta &= w \ominus \overset{\vdash\dashv}{0,5} \end{aligned} \tag{4.5}$$

instead of coordinate - system $[u, v, w]$, we can see the point $0_\infty = (u = 0, v = 0, w = \overset{\sqcup}{1}) =$ $(\xi = 0, \eta = 0, \zeta = \overset{\vdash\dashv}{0,5})$ in the "upper" position again, such that the partially familiar three - dimensional space $R^3_{\underset{(0,0,\ 0,5\ )}{\sqcup}}$ contains the "original" origo $0 = (u = 0, v = 0, w = 0) =$ $= (\xi = 0, \eta = 0, \zeta = \overset{\vdash\dashv}{-0,5})$ which is the origo of our familiar three -dimensional space. By (4.5) the new origo is

$$0_{\underset{(\frac{1}{2})}{\sqcup}} = (u = 0, v = 0, w = \overset{\vdash\dashv}{0,5}) = (\xi = 0, \eta = 0, \zeta = 0).$$



If $r > \overset{\sqcup}{1}$ then the super ball

$$G_0(r) = \{(u, v, w) \in \overset{\vdash\dashv}{R^3} : (u \otimes u) \oplus (v \otimes v) \oplus (w \otimes w) \le r \otimes r\}$$

is a giant body again, such that $G_0(\overset{\sqcup}{1}) \subset G_0(r)$. If $r \ge \overset{\vdash\dashv}{\sqrt{3}}$, then $R^3 \subset G_0(r)$ and $G_0(r)_{box} = R^3$.

5 Super-octaeders

Exploding a familiar octaeder

$$H = \{(x, y, z) \in R^3 : \frac{|x|}{a} + \frac{|y|}{b} + \frac{|z|}{c} \le 1; 0 < a, b, c \in R\}$$

the set $\overset{\sqcup}{H} = \{(\overset{\sqcup}{x}, \overset{\sqcup}{y}, \overset{\sqcup}{z}) \in \overset{\sqcup}{R^3} : (x, y, z) \in H\}$ is called super-octaeder. By the computation

$$\overset{\sqcup}{\frac{|x|}{a} + \frac{|y|}{b} + \frac{|z|}{c}} = \overset{\sqcup}{\left(\frac{|x|}{a}\right)} \oplus \overset{\sqcup}{\left(\frac{|y|}{b}\right)} \oplus \overset{\sqcup}{\left(\frac{|z|}{c}\right)} =$$

$$= (|\overset{\sqcup}{x}| \oslash \overset{\sqcup}{a}) \oplus (|\overset{\sqcup}{y}| \oslash \overset{\sqcup}{b}) \oplus (|\overset{\sqcup}{z}| \oslash \overset{\sqcup}{c})$$

with $u = \overset{\sqcup}{x}; v = \overset{\sqcup}{y}; w = \overset{\sqcup}{z}$ and using the identity $\overset{\sqcup}{|x|} = |\overset{\sqcup}{x}|$; $x \in R$,
$\overset{\sqcup}{H} = \{(u, v, w) \in \overset{\sqcup}{R^3}$:

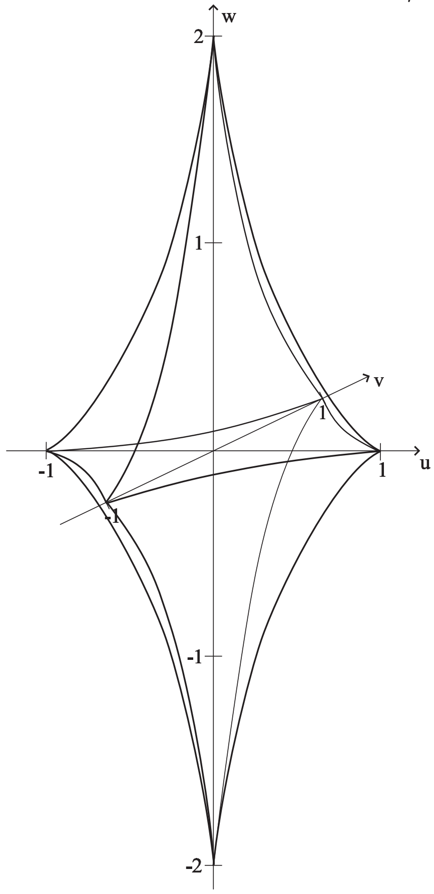
$$(|u| \oslash \overset{\sqcup}{a}) \oplus (|v| \oslash \overset{\sqcup}{b}) \oplus (|w| \oslash \overset{\sqcup}{c}) \leq \overset{\sqcup}{1}; 0 < a, b, c \in R\} \tag{5.1}$$

is obtained. We consider the special case $a = \underset{\sqcup}{1}$, $b = \underset{\sqcup}{1}$, $\overset{\sqcup}{c} = \mu (\in \overset{\sqcup}{R})$, only. Now (5.1) with (1.4) gives that the description of our super-octaeder is

$$\overset{\sqcup}{H}_\mu = \{(u, v, w) \in \overset{\sqcup}{R^3} : (|u| \oslash 1) \oplus (|v| \oslash 1) \oplus (|w| \oslash \mu) \leq \overset{\sqcup}{1}; 0 < \mu \in \overset{\sqcup}{R}\}. \tag{5.2}$$

Writing $w = 0$, by (5.2) gives that the basis B (see (3.2)) is a subset of $\overset{\sqcup}{H}_\mu$.

If $\mu < \overset{\sqcup}{1}$ then the super octaeder H_μ is small and $\overset{\sqcup}{H}_{box} = \overset{\sqcup}{H}_\mu$. For example in the case $\mu = 2$.



A familiar octaeder H_μ ($\mu \in R$) is not able to grow out of the familiar three-dimensional space. (The "sky" would be the highness $\mu = \overset{\sqcup}{1}$.) The super-octaeder is able to do that. Let us consider the case $\mu = \overset{\sqcup}{3}$. From (5.2)

$$\overset{\sqcup}{H}_{\overset{\sqcup}{3}} = \{(u,v,w) \in \overset{\sqcup}{R^3} : (|u| \odot 1) \oplus (|v| \odot 1) \oplus (|w| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}\} \tag{5.3}$$

is reduced. Let us begin the description of $\overset{\sqcup}{H}_{\overset{\sqcup}{3}\,box}$. Now $w \in R$ so (5.3) gives

$$\overset{\sqcup}{H}_{\overset{\sqcup}{3}\,box} = \{(u,v,w) \in R^3 : (|u| \odot 1) \oplus (|v| \odot 1) \oplus (|w| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}\}. \tag{5.4}$$

By (1.3) and (2.4) we can write

$$(|u| \odot 1) \oplus (|v| \odot 1) \oplus (|w| \odot \overset{\sqcup}{3}) = \overline{\underline{(|u| \odot 1)} + \underline{(|v| \odot 1)} + \underline{(|w| \odot \overset{\sqcup}{3})}} =$$

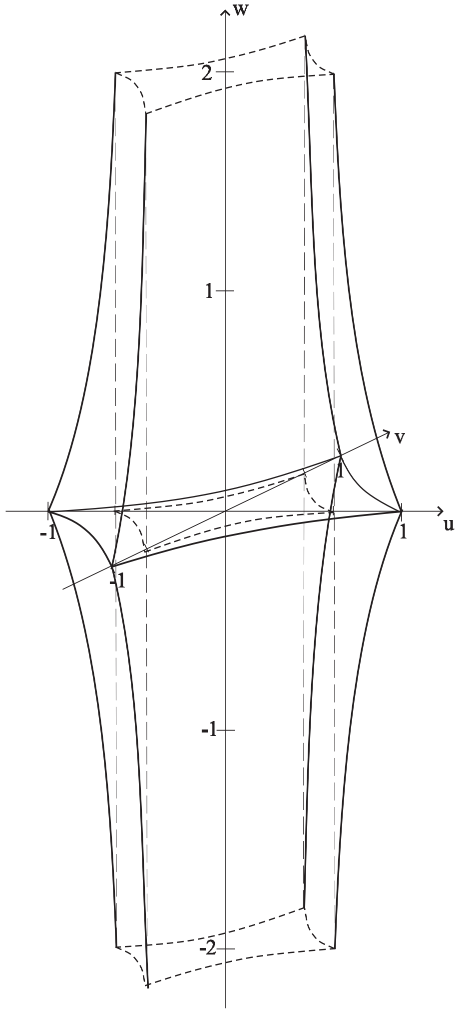
$$= \overline{((\frac{\overline{\underline{|u|}}}{\underset{\square}{1}})) + ((\frac{\overline{\underline{|v|}}}{\underset{\square}{1}})) + ((\frac{\overline{\underline{|w|}}}{3}))} = \overline{(\frac{\underline{|u|}}{\underset{\square}{1}}) + (\frac{\underline{|v|}}{\underset{\square}{1}}) + (\frac{\underline{|w|}}{3})} = \overline{(\frac{\operatorname{th}|u|}{\operatorname{th}1} + \frac{\operatorname{th}|v|}{\operatorname{th}1} + \frac{\operatorname{th}|w|}{3})} =$$

$$= \overline{(\frac{|\operatorname{th}u|}{\operatorname{th}1} + \frac{|\operatorname{th}v|}{\operatorname{th}1} + \frac{|\operatorname{th}w|}{3})}$$

and (5.4) has the form

$$\overset{\sqcup}{H}_{\overset{\sqcup}{3}\,box} = \{(u,v,w) \in R^3 : \frac{|\operatorname{th}u|}{\operatorname{th}1} + \frac{|\operatorname{th}v|}{\operatorname{th}1} + \frac{|\operatorname{th}w|}{3} \leq 1\}. \tag{5.5}$$

Moreover, $\overset{\sqcup}{H}_{\overset{\sqcup}{3}\,box}$ is illustrated by the figure:



The super-octaeder $\overset{\sqcup}{H}_{\underset{3}{\sqcup}}$ is continued in the partially familiar three-dimensional space $R^3_{O_\infty}$. (See (4.2).) Using the coordinate - system (4.3), by (5.3) we have

$$\overset{\sqcup}{H}_{\underset{3}{\sqcup}} = \{(\zeta,\eta,\zeta) \in \overset{\sqcup}{R^3} : (|\xi| \odot 1) \oplus (|\eta| \odot 1) \oplus (|\zeta \oplus \overset{\sqcup}{1}| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}\}.$$

Hence, (5.6) $\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_{0_x} = \{(\xi,\eta,\zeta) \in \overset{\sqcup}{R^3}$:

$$(|\xi| \odot 1) \oplus (|\eta| \odot 1) \oplus (|\zeta \oplus \overset{\sqcup}{1}| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}; \xi,\eta,\zeta \in R\}. \tag{5.6}$$

Moreover, by (1.3) , (2.4) and the identity $\overset{\sqcup}{|x|} = |\overset{\sqcup}{x}|$; $x \in R$, we have the computation

$$(|\xi| \odot 1) \oplus (|\eta| \odot 1) \oplus (|\zeta \oplus \overset{\sqcup}{1}| : \overset{\sqcup}{3}) =$$

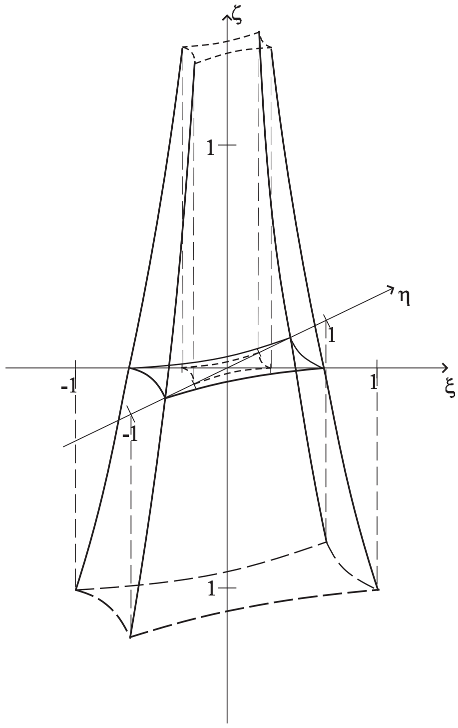
$$= \overline{\lfloor|\zeta \odot 1\rfloor + \lfloor|\eta| \odot 1\rfloor + \lfloor|\zeta \oplus \overset{\sqcup}{1}| \odot \overset{\sqcup}{3}\rfloor} = \left(\left(\frac{\underline{|\xi|}}{\underset{\sqcup}{1}}\right)\right) + \left(\left(\frac{\underline{|\eta|}}{\underset{\sqcup}{1}}\right)\right) + \left(\left(\frac{\underline{|\zeta \oplus \overset{\sqcup}{1}|}}{3}\right)\right) =$$

$$= \overline{\frac{\underline{|\xi|}}{\underset{\sqcup}{1}} + \frac{\underline{|\eta|}}{\underset{\sqcup}{1}} + \frac{\underline{|\zeta \oplus \overset{\sqcup}{1}|}}{3}} = \overline{\frac{|\operatorname{th}\xi|}{\operatorname{th}1} + \frac{|\operatorname{th}\eta|}{\operatorname{th}1} + \frac{\operatorname{th}\zeta+1}{3}}$$

So, (5.6) has the form

$$\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_{0\infty} = \left\{(\xi,\eta,\zeta) \in \overset{\sqcup}{R^3} : \frac{|\operatorname{th}\xi|}{\operatorname{th}1} + \frac{|\operatorname{th}\eta|}{\operatorname{th}1} + \frac{\operatorname{th}\zeta+1}{3} \le 1; \xi,\eta,\zeta \in R\right\} \tag{5.7}$$

illustrated by the figure:

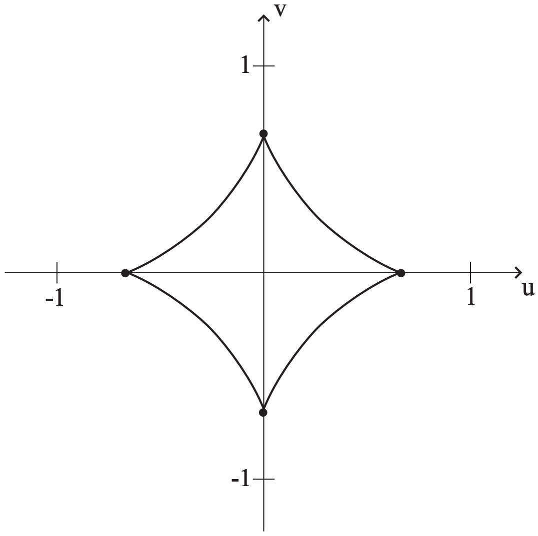


This figure shows the joint part $\overset{\sqcup}{H}_{\underset{3\,box}{\sqcup}} \cap (\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_{O\infty})$ which is a subset of

$$R^3 \cap R^3_{O\infty} = \{(u,v,w) \in \overset{\sqcup}{R^3} : \overline{-1} < u < \overset{\sqcup}{1}; \overline{-1} < v < \overset{\sqcup}{1}; 0 < w < \overset{\sqcup}{1}\}$$

Especially interesting is a new basis

$$B_\infty = \{(u, v, \overset{\sqcup}{1}) = |u| \oplus |v| \le \operatorname{area}\operatorname{th}\left(\frac{2\operatorname{th}1}{3}\right) \approx 0,5596658124\}.$$



B_∞ is invisible in our familiar three - dimensional space R^3, but $B_\infty \subset R^3_{0_\infty}$. The super-octaeder $\overset{\sqcup\!\sqcup}{H}_{\overset{\sqcup}{3}}$ is continued in the quasi- familiar three-dimensional space R^3_Ω, where $\Omega = (0, 0, \overset{\sqcup}{3}) \in \overset{\sqcup\!\sqcup}{R^3}$. Moreover,

$R^3_\Omega = \{(u, v, w) \in \overset{\sqcup\!\sqcup}{R^3}$:

$$\overset{\sqcup\!\sqcup}{-1} < u < \overset{\sqcup}{1};\ \overset{\sqcup\!\sqcup}{-1} < v < \overset{\sqcup}{1};\ \overset{\sqcup\!\sqcup}{-1} \oplus \overset{\sqcup}{3} = \overset{\sqcup}{2} < w < \overset{\sqcup}{4} = \overset{\sqcup}{1} \oplus \overset{\sqcup}{3}\}. \tag{5.8}$$

using the coordinate - system $[\alpha, \beta, \gamma]$

$$\begin{aligned} \alpha &= u \\ \beta &= v \\ \gamma &= w \ominus \overset{\sqcup}{3} \end{aligned} \tag{5.9}$$

instead of the coordinate - system $[u, v, w]$, (5.3) yields

$$\overset{\sqcup\!\sqcup}{H}_{\overset{\sqcup}{3}} = \{(\alpha, \beta, \gamma) \in \overset{\sqcup\!\sqcup}{R^3} : (|\alpha| \odot 1) \oplus (|\beta| \odot 1) \oplus (|\gamma \oplus \overset{\sqcup}{3}| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}\} \tag{5.10}$$

Hence, (5.11) $\overset{\sqcup\!\sqcup}{H}_{\overset{\sqcup}{3}} \cap R^3_\Omega = \{(\alpha, \beta, \gamma) \in \overset{\sqcup\!\sqcup}{R^3}$:

$$(|\alpha| \odot 1) \oplus (|\beta \ \odot 1) \oplus (|\gamma \oplus \overset{\sqcup}{3}| \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}; \alpha, \beta, \gamma \in R\}.$$

Moreover, by (1.3) , (2.4) and the identity $\overset{\sqcup\!\sqcup}{|x|} = |\overset{\sqcup}{x}|; x \in R$, we have the computation

$$(|\alpha| \odot 1) \oplus (|\beta| \odot 1) \oplus (|\gamma \oplus \overset{\sqcup}{3}| \odot \overset{\sqcup}{3}) =$$

$$= |\alpha \odot 1| + |\beta| \odot 1 + |\gamma \oplus \overset{\sqcup}{1}| \odot \overset{\sqcup}{3} = \left(\left(\frac{|\alpha|}{1}\right)\right) + \left(\left(\frac{|\beta|}{1}\right)\right) + \left(\left(\frac{|\gamma \oplus \overset{\sqcup}{1}|}{3}\right)\right) =$$

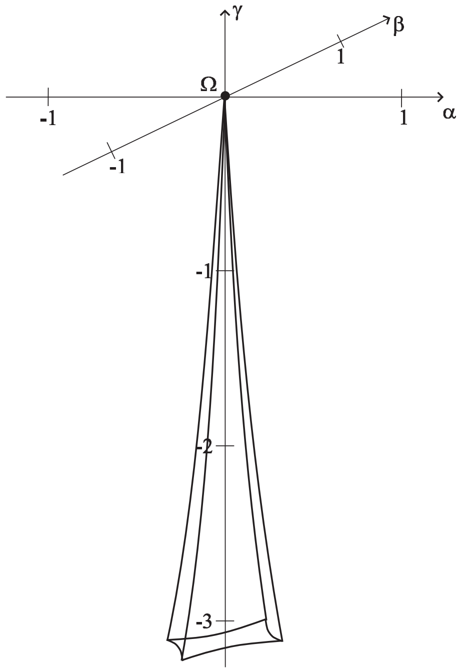
$$= \left(\frac{|\alpha|}{1}\right) + \left(\frac{|\beta|}{1}\right) + \frac{|\gamma \oplus \overset{\sqcup}{3}|}{3} = \frac{|\alpha|}{1} + \frac{|\beta|}{1} + \frac{|y+3|}{3} = \frac{|\alpha|}{1} + \frac{|\beta|}{1} + \frac{|\gamma+3|}{3} =$$

$$\frac{\operatorname{th}|\alpha|}{\operatorname{th}1} + \frac{\operatorname{th}|\beta|}{\operatorname{th}1} + \frac{|\operatorname{th}\gamma+3|}{3} = \frac{|\operatorname{th}\alpha|}{\operatorname{th}1} + \frac{|\operatorname{th}\beta|}{\operatorname{th}1} + \frac{\operatorname{th}\gamma+3}{3}$$

So, (5.11) has the form

$$\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_\Omega = \left\{(\alpha,\beta,\gamma) \in \overset{\sqcup}{R^3} : \frac{|\operatorname{th}\alpha|}{\operatorname{th}1} + \frac{|\operatorname{th}\beta|}{\operatorname{th}1} + \frac{\operatorname{th}\gamma}{3} \le 0; \alpha,\beta,\gamma \in R\right\} \tag{5.12}$$

illustrated by the figure:

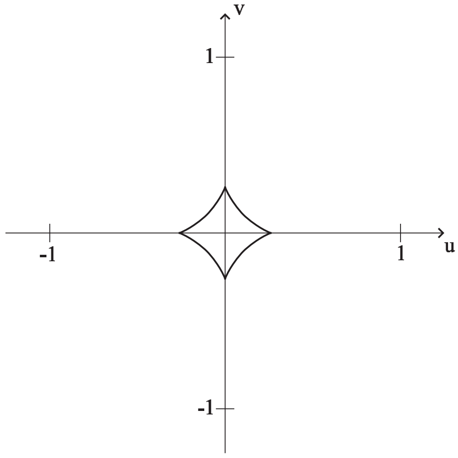


So, for the upper part of $\overset{\sqcup}{H}_{\underset{3}{\sqcup}}$, that is for the

$$\overset{\sqcup}{H}^{upper}_{\underset{3}{\sqcup}} = \{(u,v,w) \in \overset{\sqcup}{R^3} : (|u| \odot 1) \oplus (|v| \odot 1) \oplus (w \odot \overset{\sqcup}{3}) \le \overset{\sqcup}{1}; w \ge 0\}$$

it is "almost true", that $\overset{\sqcup}{H}^{upper}_{\underset{3}{\sqcup}}$ is the union $B \cup (\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_{0_\infty}) \cup (\overset{\sqcup}{H}_{\underset{3}{\sqcup}} \cap R^3_\Omega)$ because $R^3_{0_\infty} \cap R^3_\Omega$ is empty so there exists a gap

$$G = \left\{(u, v, \overset{\sqcup}{2}) : |u| \oplus |v| \le \operatorname{area}\operatorname{th}\left(\frac{\operatorname{th}1}{3}\right) \approx 0,2595394547\right\}$$



between $R^3_{0_\infty}$ and R^3_Ω. So, we have

$$\overline{H}{}^{upper}_{\overset{\sqcup}{3}} = B \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{0_\infty}) \cup G \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_\Omega). \tag{5.13}$$

Denoting $-0_\infty = (0, 0, \overline{-1})$ and $-\Omega = (0, 0, \overline{-3})$ with

$$G_- = \left\{(u, v, \overline{-2}) = |u| \oplus |v| \leq \operatorname{area\,th}\left(\frac{\operatorname{th} 1}{3}\right) \approx 0,2595394547\right\}$$

for

$$\overline{H}{}^{lover}_{\overset{\sqcup}{3}} = \left\{(u, v, w) \in \overline{R^3} : (|u| \odot 1) \oplus (|v| \odot 1) \ominus (w \odot \overset{\sqcup}{3}) \leq \overset{\sqcup}{1}; w < 0\right\}$$

so,

$$\overline{H}{}^{lover}_{\overset{\sqcup}{3}} = (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{-0_\infty}) \cup G_- \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{-\Omega} \tag{5.14}$$

is obtained.

Finally, having that $\overline{H}_{\overset{\sqcup}{3}} = \overline{H}{}^{upper}_{\overset{\sqcup}{3}} \cup \overline{H}{}^{lover}_{\overset{\sqcup}{3}}$, using (5.13) and (5.14) we can see the giant body $\overline{H}_{\overset{\sqcup}{3}}$ piece by piece is

$$\overline{H}_{\overset{\sqcup}{3}} = (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{-\Omega}) \cup G_- \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{-0_\infty}) \cup B \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_{0_\infty}) \cup G \cup (\overline{H}_{\overset{\sqcup}{3}} \cap R^3_\Omega).$$

References

[1] I. SZALAY, On the c - explosion and c - compression of real numbers, 5 -th International Conference APLIMAT 2006., (Part II.) 173-180.

[2] I. SZALAY, A complex model of exploded real numbers, Acta Mathematica Nitra, 6(2003), 93-107.

[3] I. SZALAY, On the cube model of three -dimensional euclidean space, Acta Acad. Paed. Agresiensis, Sectio Mathematicae, 29(2002), 15-34.
[4] I. SZALAY, The super - pyramid, International Journal of Pure and Applied Mathematics, Vol.46, No. 3(2008), 337-346.
[5] I. SZALAY, On the shift - box phenomenon, 7 -th International Conference APLIMAT 2008, (Section: Algebra and applications) (2008), 153-163.

Current address

Szalay István, CSc
University of Szeged, Faculty of Teacher's Training, Department of Math., H-6723 Szeged, Hattyas u. 10, Hungary, e-mail: szalay.istvan@tvnetwork.hu