Contents

ROMAI J., v.14, no.1(2018), 1–21

ON SOLVING THE VARIATIONAL PROBLEM

Serikbay Aisagaliev, Assem Kabidoldanova

Al-Farabi Kazakh National University, Almaty, Kazakhstan

serikbai.aisagaliev@kaznu.kz, kabasem@mail.ru

Abstract A method for solving the Lagrange problem with state variable constraints for processes described by ordinary differential equations without involvement of the Lagrange principle is proposed. A necessary and sufficient condition for existence of a solution to the variational problem is obtained, an admissible control is found and an optimal solution is constructed by narrowing a set of admissible controls. The basis of the proposed method for solving the variational problem is an imbedding principle. An essence of the imbedding principle is that the original variational problem with boundary conditions and state variable constraints is replaced by equivalent free end point optimal control problem. This approach is possible due to finding a general solution of a class of the first kind Fredholm integral equations.

Keywords: imbedding principle, admissible control, optimal solution, minimizing sequence.
2010 MSC: 49J15.

1. INTRODUCTION

Calculus of variations was formed as an independent branch of mathematics in the second half of the 19th century, with regards to the works of L. Euler, J. Lagrange. One of the methods for solving problems of calculus of variations is the Lagrange principle. The Lagrange principle allows to reduce the original problem to searching an extremum of the Lagrange functional formulated by introducing auxiliary variables (the Lagrange multipliers). The Lagrange principle is an assertion about existence of the Lagrange multipliers satisfying a set of conditions in the case when the original problem has a weak local minimum. The Lagrange principle provides a necessary condition for a weak local minimum and it doesn't exclude an existence of another methods for solving problems of calculus of variations that don't involve the Lagrange functional.

The works [1]-[3] are devoted to the Lagrange principle. An unified approach to different extremal problems based on the Lagrange principle is described in [4].

The aim of this work is to develop a method for solving a problem of calculus of variations for processes described by ordinary differential equations with state variable constraints different from the known methods based on the Lagrange principle. It is a continuation of research presented in [9]-[20].

2. STATEMENT OF THE PROBLEM

Consider the following problem

$$J(u(\cdot), x_0, x_1) = \int\limits_{t_0}^{t_1} F_0(x(t), u(t), x_0, x_1, t)dt \to inf \tag{1}$$

for the dynamical system described by

$$\dot{x} = A(t)x + B(t)f(x, u, t), \ \ t \in I = [t_0, t_1], \tag{2}$$

with the boundary conditions

$$(x(t_0)) = x_0, x(t_1) = x_1) \in S_0 \times S_1 = S \subset R^{2n}, \tag{3}$$

the state variable constraints

$$x(t) \in G(t) : G(t) = \{x \in R^n / \omega(t) \leq F(x, t) \leq \varphi(t), \ \ t \in I\}, \tag{4}$$

where the control function

$$u(\cdot) \in L_2(I, R^m). \tag{5}$$

Here $A(t)$, $B(t)$ are $n \times n$, $n \times r$ matrices with piecewise-continuous elements respectively, the vector valued function $f(x, u, t) = (f_1(x, u, t), \dots, f_r(x, u, t))$ is continuous with respect to $(x, u, t) \in R^n \times R^m \times I$, and satisfies the Lipschitz condition with respect to x, i.e.

$$|f(x, u, t) - f(y, u, t)| \leq l(t)|x - y|, \tag{6}$$

$$\forall (x, u, t), (y, u, t) \in R^n \times R^m \times I,$$

and the condition

$$|f(x, u, t)| \leq c_0(|x| + |u|^2) + c_1(t), \ \ \forall (x, u, t), \tag{7}$$

where $l(t) \geq 0$, $l(t) \in L_1(I, R^1)$, $c_0 = const > 0$, $c_1(t) \geq 0$, $c_1(t) \in L_1(I, R^1)$. The vector valued function $F(x, t) = (F_1(x, t), \dots, F_s(x, t))$ is continuous with respect to the set of variables $(x, t) \in R^n \times I$. The scalar function $F_0(x, u, x_0, x_1, t)$ is defined and continuous together with its partial derivatives with respect to (x, u, x_0, x_1), $\omega(t)$, $\varphi(t)$, $t \in I$ are given $s \times 1$ continuous functions. S is a given bounded convex closed set in R^{2n}, the time instants t_0, t_1 are fixed.

Note that the differential equation (2) has an unique solution $x(t)$, $t \in I$ under the conditions (6), (7) for any control function $u(\cdot) \in L_2(I, R^m)$ and the initial condition $x(t_0) = x_0$. This solution has a derivative $\dot{x} \in L_2(I, R^n)$ and satisfies the equation (2) at almost all $t \in I$.

Definition 2.1. *The triple* $(u_*(t), x_0^*, x_1^*) \in U \times S_0 \times S_1$ *is said to be an admissible control for the problem (1) – (5), if the boundary value problem (2) – (5) has a solution. Denote by* Σ *the set of all admissible controls,* $\Sigma \subset U \times S_0 \times S_1$.

It follows from this definition that for each element of the set Σ the following assertions hold: 1) the solution $x_*(t)$, $t \in I$ to the differential equation (2), starting from the point $x_0^* \in S_0$, satisfies the condition $x_*(t_1) = x_1^* \in S_1$, $(x_0^*, x_1^*) \in S_0 \times S_1 = S$; 2) the inclusion $x_*(t) \in G(t)$, $t \in I$ holds.

The following problems are posed:

Problem 1. Find a necessary and sufficient condition for existence of a solution to the boundary value problem (2) – (5).

Note that the optimal control problem (1) – (5) has a solution if and only if the boundary value problem (2) – (5) has a solution.

Problem 2. Find an admissible control $(u_*(t), x_0^*, x_1 *^*) \in \Sigma$.

If problem 1 has a solution, then there exists an admissible control.

Problem 3. Find an optimal control $\overline{u}_*(t)\in U(t)$, a point $(\overline{x}_0^*,\ \overline{x}_1^*)\in S_0 \times S_1$ and an optimal trajectory $\overline{x}_*(t; t_0, x_0^*)$, $t \in I$, where $\overline{x}_*(t) \in G(t)$, $t \in I$, $\overline{x}_*(t_1) = \overline{x}_1^* \in S_1$, $J(\overline{u}_*(\cdot), \overline{x}_0^*, \overline{x}_1^*) = \inf J(u(\cdot), x_0, x_1)$, $\forall (u(\cdot), x_0, x_1) \in L_2(I, R^m) \times S_0 \times S_1$.

In classical calculus of variations it is assumed that a solution to the differential equation (2) belong to the space $C^1(I, R^n)$, and a control $u(t)$, $t \in I$ is considered to belong to $C(I, R^m)$, and in optimal control problems [5] a solution $x(t) \in KC^1(I, R^n)$, and a control $u(t) \in KC(I, R^m)$. In the presented paper the control $u(t)$, $t \in I$ is chosen from $L_2(I, R^m)$, and the solution $x(t)$, $t \in I$ is an absolutely continuous function in $I = [t_0, t_1]$. For this case an existence and uniqueness of a solution to the initial problem (2) are presented in [4], [6]-[8].

3. THE IMBEDDING PRINCIPLE

Consider the linear control system

$$\dot{y} = A(t)y + B(t)w(t), \quad t \in I, \tag{8}$$

$$w(\cdot) \in L_2(I, R^r), \tag{9}$$

$$y(t_0) = x_0 \in S_0, \quad y(t_1) = x_1 \in S_1. \tag{10}$$

The basis for the imbedding principle resides in the following theorems about properties of solutions to the first kind Fredholm integral equation

$$Ku := \int\limits_{t_0}^{t_1} K(t_0,t)u(t)dt = a, \tag{11}$$

where $K : L_2(I, R^r) \to R^n$, $K(t_0,t)$ is an $n \times r$ given matrix with piecewise continuous elements with respect to t at every fixed t_0, $t_0 \in \Delta_0 \subset R^1$, $t_1 \in \Delta_1 \subset R^1$, $\Delta_0 \cap \Delta_1 = \emptyset$, here the symbol $\emptyset$ denotes an empty set, $a \in R^n$ is an arbitrary given vector, $u(\cdot) \in L_2(I, R^r)$ is an unknown function.

Theorem 3.1. *For the integral equation (11) to have a solution at any fixed $a \in R^n$ it is necessary and sufficient a positive definiteness of the $n \times n$ matrix*

$$C(t_0,t_1) = \int\limits_{t_0}^{t_1} K(t_0,t)K^*(t_0,t)dt, \tag{12}$$

where the symbol $(*)$ denotes a transposition.

Theorem 3.2. *Let the matrix $C(t_0,t_1)$ be positive definite. Then a general solution to the integral equation (11) is defined by*

$u(t) = K^*(t_0,t)C^{-1}(t_0,t_1)a + v(t)-$

$$-K^*(t_0,t)C^{-1}(t_0,t_1)\int\limits_{t_0}^{t_1} K(t_0,t)v(t)dt, \ \ t \in I, \tag{13}$$

where $v(\cdot) \in L_2(I, R^r)$ is an arbitrary function, $a \in R^n$ is any given vector.

We refer the reader for the proofs of theorems 3.1, 3.2 to [9], [18].

It can be easily shown that a control $w(\cdot) \in L_2(I, R^r)$, moving the system (8) starting from any initial point x_0 to any desired final state x_1, is a solution to the integral equation

$$\int\limits_{t_0}^{t_1} \Phi(t_0,t)B(t)w(t)dt = a, \tag{14}$$

where $\Phi(t,\tau) = \lambda(t)\lambda^{-1}(\tau)$, $\lambda(t)$ is a fundamental matrix solutions to the linear homogeneous system $\dot{\rho} = A(t)\rho$, the vector

$$a = a(x_0,x_1) = \Phi(t_0,t_1)x_1 - x_0. \tag{15}$$

As it follows from (11), (14), the matrix $K(t_0,t) = \Phi(t_0,t)B(t)$. For the integral equation (12) the assertions of theorems 3.1, 3.2 are applicable. Define the following matrices and vectors by the given data of the system (8) – (10)

$$T(t_0,t_1) = \int\limits_{t_0}^{t_1} \Phi(t_0,t)B(t)B^*(t)\Phi^*(t_0,t)dt,$$

$$\Lambda_1(t,x_0,x_1) = B^*\Phi^*(t_0,t)T^{-1}(t_0,t_1)a,$$
$$N_1(t) = -B^*(t)\Phi^*(t_0,t)T^{-1}(t_0,t_1)\Phi(t_0,t_1),$$
$$\Lambda_2(t,x_0,x_1) = \Phi(t,t_0)T(t,t_1)T^{-1}(t_0,t_1)x_0 +$$
$$+\Phi(t,t_0)T(t_0,t)T^{-1}(t_0,t_1)\Phi(t_0,t_1)x_1,$$
$$N_2(t) = -\Phi(t,t_0)T(t_0,t)T^{-1}(t_0,t_1)\Phi(t_0,t_1), \quad t \in I,$$
$$T(t,t_1) = \int\limits_{t}^{t_1} \Phi(t_0,\tau)B(\tau)B^*(\tau)\Phi^*(t_0,\tau)d\tau,$$
$$T(t_0,t) = T(t_0,t_1) - T(t,t_1), \quad t \in I,$$

where the vector a is defined by (13).

Theorem 3.3. *Let the matrix $T(t_0,t_1)$ be positive definite. Then a control $w(\cdot) \in L_2(I,R^r)$ brings the trajectory of the system (8) – (10) from the initial point $x_0 \in S_0$ to the final state $x_1 \in S_1$ if and only if $w(t) \in W$,*

$$W = \{w(\cdot) \in L_2(I,R^r)/w(t) = v(t) + \Lambda_1(t) + N_{11}(t)z(t_1),\ t \in I\}, \tag{16}$$

where $v(\cdot) \in L_2(I,R^r)$ is an arbitrary function. The function $z(t) = z(t,v)$, $t \in I$ is a solution to the differential equation

$$\dot{z} = A(t)z + B(t)v(t), \quad z(t_0) = 0, \quad t \in I, \tag{17}$$

$$v(\cdot) \in L_2(I,R^r). \tag{18}$$

The solution to the differential equation (8) corresponding to the control (16) is defined by

$$y(t) = z(t) + \Lambda_2(t,x_0,x_1) + N_2(t)z(t_1,v), \quad t \in I, \tag{19}$$

where $z(t) = z(t_1,v)$, $t \in I$.

Proof. The proof of the theorem follows from theorems 3.1, 3.2. As it follows from the presented above solving the boundary value problem (8) – (10) is reduced to the integral equation (14). The integral equation (14) is a special case of (11), where $K(t_0, t) = \Phi(t_0, t)B(t)$. Further by substituting $\Phi(t_0, t)B(t)$ for $K(t_0, t)$ we obtain $C(t_0, t_1) = T(t_0, t_1)$ (see (12)). The formula (13) implies (16). The differential equation (17) with the control (18) and the relation (19) directly follows from

$$z(t, v) = \int\limits_{t_0}^{t} \Phi(t, \tau)B(\tau)v(\tau)d\tau, \;\; z(t_1, v) = \Phi(t_1, t_0) \int\limits_{t_0}^{t_1} \Phi(t_0, t)B(t)v(t)dt.$$

It can be easily seen that $y(t_0) = x_0$, $y(t_1) = x_1$. ∎

Note that the following assertions hold. 1) The set $W \subset L_2(I, R^r)$ contains all the controls $w(t)$, $t \in I$, such that the boundary value problem (8) – (10) has a solution; 2) If $w(t) \in W$, then the solution to the system (8) – (10) is defined by (19); 3) There is no any control outside the set W, for which the boundary value problem (8) – (10) has a solution; 4) Theorem 3.3 allows to reduce the boundary value problem (8) – (10) to the initial problem (17) – (19).

Lemma 3.1. *Let the matrix $T(t_0, t_1)$ be positive definite. Then the boundary value problem (2) – (5) is equivalent to the following*

$$w(t) \in W, \;\; w(t) = f(P_1y(t), u(t), x_0, x_1, t), \;\; t \in I, \tag{20}$$

$$p(t) = F(P_1y(t), t), \;\; p \in V, \tag{21}$$

$$\dot{z} = A(t)z + B(t)v(t), \;\; z(t_0) = 0, \;\; t \in I, \tag{22}$$

$$v(\cdot) \in L_2(I, R^r), \tag{23}$$

$$(x_0, x_1) \in S_0 \times S_1 = S \subset R^{2n}, \;\; u(\cdot) \in L_2(I, R^m), \tag{24}$$

where the function $y(t)$, $t \in I$ is given by (19),

$$V(t) = \{p(\cdot) \in L_2(I, R^s)/ \; \omega(t) \leq p(t) \leq \varphi(t), \;\; t \in I\}.$$

Proof. Lemma 3.1 states that the boundary value problem (2) – (5) has a solution if and only if the relations (20) – (24) hold.

Indeed, if the relations (20) – (24) hold, then $y(t) = x(t)$, $t \in I$, moreover $y(t_0) = x(t_0) = x_0$, $y(t_1) = x(t_1) = x_1$ and the inclusion (3) hold.

Let the boundary value problem (2) – (5) has a solution. This holds if and only if $f(P_1x(t), u(t), t) \in W$ by theorem 3.3. This inclusion is equivalent to (20), where $z(t)$, $t \in I$ is a solution to the differential equation (22) corresponding to the control (23). The inclusion $P_1x(t) \in G(t)$, $t \in I$ has the form (21), and the inclusions from (3) and (5) are rewritten as (24). ∎

Consider the following optimal control problem

$$I_1(u(\cdot), p(\cdot), v(\cdot), x_0, x_1) = \int\limits_{t_0}^{t_1} F_1(q(t), t)dt \to \inf \tag{25}$$

under the conditions

$$\dot{z} = A(t)z + B(t)v(t),\ z(t_0) = 0,\ t \in I, \tag{26}$$

$$v(\cdot) \in L_2(I, R^r), \tag{27}$$

$$p(t) \in V(t),\ u(\cdot) \in L_2(I, R^m),\ (x_0, x_1) \in S_0 \times S_1 = S, \tag{28}$$

where $F_1(q(t), t) = |w(t) - f(P_1y(t), u(t), x_0, x_1, t)|^2 + |p(t) - F(P_1y(t), t)|^2$, $q(t) = (z(t, v), z(t_1, v), u(t), p(t), v(t), x_0, x_1)$.

Denote

$$H = L_2(I, R^m) \times L_2(I, R^s) \times L_2(I, R^r) \times R^n \times R^n,$$

$$X = L_2(I, R^m) \times V \times L_2(I, R^r) \times S_0 \times S_1 \subset H,$$

$\theta(t) = (u(t), p(t), v(t), x_0, x_1) \in X$, $q(t) = (z(t), z(t_1), \theta(t))$.

The optimization problem (27) – (30) can be represented in the form:

$$I_1(\theta(\cdot)) = \int\limits_{t_0}^{t_1} F_1(q(t), t) \to \inf,\ \theta(\cdot) \in X \subset H.$$

Note that the following assertions hold.

1) Since the value $I_1 \geq 0$, for existence of a solution to the boundary value problem (2) – (5) it is necessary and sufficient to have $\inf I_1(\theta(t)) = 0$ under the conditions (26) – (28).

2) Reducing the original boundary value problem (2) – (5) to the free end-point optimal control problem $I_1(\theta(t)) \to \inf$, (26) – (28) is called an imbedding principle.

4. EXISTENCE OF A SOLUTION

Let the set $X_* = \{\theta_*(\cdot) \in X |\ I_1(\theta_*(\cdot)) = \inf\limits_{\theta \in X} I_1(\theta(\cdot))\}$.

Lemma 4.1. *Let the matrix $T(t_0, t_1)$ be positive definite. A necessary and sufficient condition for the boundary value problem (2) – (5) to have a solution is $\lim\limits_{n\to\infty} I_1(\theta_n) = I_{1*} = \inf\limits_{\theta\in X} I_1(\theta) = 0$, where $\{\theta_n(\cdot)\} \subset X$ is a minimizing sequence for the problem (25) – (28).*

The proof of the lemma follows from theorem 3.3 and lemma 3.1.

Theorem 4.1. *Let the matrix* $T(t_0, t_1)$ *be positive definite, the function* $F_1(q,t)$ *be defined and continuous with respect to* (q,t) *together with its partial derivatives with respect to* q *and satisfies the Lipschitz condition*

$$|F_{1q}(q+\Delta q,t) - F_{1q}(q,t)| \leq l|\Delta q|,\ t \in I, \tag{29}$$

where $l = const > 0$,

$$F_{1q}(q,t){=}(F_{1z}(q,t), F_{1z(t_1)}(q,t), F_{1u}(q,t), F_{1p}(q,t), F_{1v}(q,t), F_{1x_0}(q,t), F_{1x_1}(q,t)),$$

$q \in R^n{\times}R^n{\times}R^m{\times}R^s{\times}R^r{\times}R^n{\times}R^n$, $\Delta q = (\Delta z, \Delta z(t_1), \Delta u, \Delta p, \Delta v, \Delta x_0, \Delta x_1)$. *Then the functional (25) under the conditions (26) – (28) is continuously Frechet differentiable, the gradient*

$$I_1'(\theta) = (I_{1u}'(\theta), I_{1p}'(\theta), I_{1v}'(\theta), I_{1x_0}'(\theta), I_{1x_1}'(\theta)) \in H$$

at any point $\theta \in X$ *is computed by*

$$I_{1u}'(\theta) = F_{1u}(q(t),t),\ \ I_{1p}'(\theta) = F_{1p}(q(t),t),$$

$$I_{1v}'(\theta) = F_{1v}(q(t),t) - B^*(t)\psi(t), \qquad I_{1x_0}'(\theta) = \int\limits_{t_0}^{t_1} F_{1x_0}(q(t),t)dt, \tag{30}$$

$$I_{1x_1}'(\theta) = \int\limits_{t_0}^{t_1} F_{1x_1}(q(t),t)dt,$$

where $z(t)$, $t \in I$ *is a solution to the differential equation (26), and the function* $\psi(t)$, $t \in I$ *is a solution to the conjugate system*

$$\dot{\psi} = F_{1z}(q(t),t) - A^*(t)\psi,\ \psi(t_1) = -\int\limits_{t_0}^{t_1} F_{1z(t_1)}(q(t),t)dt. \tag{31}$$

Moreover the gradient $I_1'(\theta)$, $\theta \in X$ *satisfies the Lipschitz condition*

$$\|I_1'(\theta_1) - I_1'(\theta_2)\| \leq K\|\theta_1 - \theta_2\|,\ \forall\theta_1, \theta_2 \in X, \tag{32}$$

where $K = const > 0$.

Proof. Let $\theta(t), \theta(t) + \Delta\theta(t) \in X$, $z(t,v)$, $z(t, v+\Delta v)$, $t \in I$ are solutions to the system (26), (27). Let $z(t,v) = z(t,v) + \Delta z(t)$, $t \in I$. Then

$$|\Delta z(t)| \leq C_1\|\Delta v\|. \tag{33}$$

The increment of the functional (see (29))

$$\Delta I_1 = I_1(\theta+\Delta\theta) - I_1(\theta) = \int\limits_{t_0}^{t_1} [F_1(q(t)+\Delta q(t),t) - F_1(q(t),t)]dt =$$

$$= \int\limits_{t_0}^{t_1} [\Delta u^*(t)F_{1u}(q(t),t) + \Delta p^*(t)F_{1p}(q(t),t) + \Delta v^*(t)F_{1v}(q(t),t) +$$

$$+ \Delta x_0^* F_{1x_0}(q(t),t) + \Delta x_1^* F_{1x_1}(q(t),t) +$$

$$+ \Delta z^*(t)F_{1z}(q(t),t) + \Delta z^*(t_1)F_{1z(t_1)}(q(t),t)]dt + \sum_{i=1}^{7} R_i, \tag{34}$$

where

$$|R_1| \le l_1 \int\limits_{t_0}^{t_1} |\Delta u(t)||\Delta q(t)|dt, \qquad |R_2| \le l_2 \int\limits_{t_0}^{t_1} |\Delta p(t)||\Delta q(t)|dt,$$

$$|R_3| \le l_3 \int\limits_{t_0}^{t_1} |\Delta v(t)||\Delta q(t)|dt, \qquad |R_4| \le l_4 \int\limits_{t_0}^{t_1} |\Delta x_0||\Delta q(t)|dt,$$

$$|R_5| \le l_5 \int\limits_{t_0}^{t_1} |\Delta x_1||\Delta q(t)|dt, \qquad |R_6| \le l_6 \int\limits_{t_0}^{t_1} |\Delta z(t)||\Delta q(t)|dt,$$

$|R_7| \le l_7 \int\limits_{t_0}^{t_1} |\Delta z(t_1)||\Delta q(t)|dt$ by the Lipschitz condition (29). Note that (see (31), (33))

$$\int\limits_{t_0}^{t_1} \Delta z^*(t_1)F_{1z(t_1)}(q(t),t)dt =$$

$$= -\int\limits_{t_0}^{t_1} \Delta v^*(t)B^*(t)\psi(t)dt - \int\limits_{t_0}^{t_1} \Delta z^*(t)F_{1z}(q(t),t)dt. \tag{35}$$

From (34), (35) we have

$$\Delta I_1 = \int_{t_0}^{t_1} \{\Delta u^*(t)F_{1u}(q(t),t) + \Delta p^*(t)F_{1p}(q(t),t) + \Delta v^*(t)[F_{1v}(q(t),t)-$$

$$-B^*(t)\psi(t)]+\Delta x_0^* F_{1x_0}(q(t),t)+\Delta x_1^* F_{1x_1}(q(t),t)+\sum_{i=1}^{7} R_i =< I_1'(\theta), \Delta\theta >_H +R,$$

where $R = \sum\limits_{i=1}^{7} R_i$, $|R| \le C_3\|\Delta\theta\|^2$, $\dfrac{|R|}{\|\Delta\theta\|} \to 0$, as $\|\Delta\theta\| \to 0$. This yields (30). Let

$$\theta_1 = (u+\Delta u, p+\Delta p, v+\Delta v, x_0+\Delta x_0, x_1+\Delta x_1), \quad \theta_2 = (u,p,v,x_0,x_1) \in X.$$

As

$$|I_1'(\theta_1) - I_1'(\theta_2)|^2 \le l_{10}|\Delta q(t)|^2 + l_{11}|\Delta\psi(t)|^2 + l_{12}|\Delta\theta|^2,$$
$$|\Delta q(t)| \le l_{13}\|\Delta\theta\|, |\Delta\psi(t)| \le l_{14}\|\Delta\theta\|,$$

the estimate holds

$$\|I_1'(\theta_1) - I_1'(\theta_2)\|^2 = \int_{t_0}^{t_1} |I_1'(\theta_1) - I_1'(\theta_2)|^2 dt \le l_{15}\|\Delta\theta\|^2,$$

where $l_i = const > 0$, $i = \overline{10,15}$. This implies (32), where $K = \sqrt{l_{15}}$. ∎

Lemma 4.2. *Let the matrix $T(t_0,t_1)$ be positive definite, the function $F_1(q,t)$ be convex with respect to $q \in R^N$, $N = 4n+m+s+r$. Then the functional (25) under the conditions (26) – (28) is convex.*

Proof. Let $\theta_1, \theta_2 \in X$, $\alpha \in [0,1]$. It is not hard to prove that

$$z(t, \alpha v_1 + (1-\alpha)\overline{v}_1) = \alpha z(t, v_1) + (1-\alpha) z(t, \overline{v}_1),$$
$$\forall v_1,\ \overline{v}_1 \in L_2(I, R^r).$$

Then

$$I_1(\alpha\theta_1 + (1-\alpha)\theta_2) = \int_{t_0}^{t_1} F_1(\alpha q_1(t) + (1-\alpha) q_2(t))dt \le \alpha I_1(\theta_1) + (1-\alpha) I_1(\theta_2),$$

$$\forall \theta_1, \theta_2 \in X,\ \theta_1 = (u_1, p_1, v_1, x_0^1, x_1^1),\ \theta_2 = (\overline{u}_1, \overline{p}_1, \overline{v}_1, \overline{x}_0, \overline{x}_1).$$

∎

The free end-point optimal control problem (25) – (28) can be solved by numerical methods for solving extremal problems [21]-[23]. Let us introduce the following sets

$$U = \{u(\cdot) \in L_2(I, R^m)/\|u\| \le \beta\},$$
$$V(I, R^r) = \{v(\cdot) \in L_2(I, R^r)|\ \|v\| \le \beta\},$$

$\beta > 0$ is a sufficiently large number.

Generate the sequence $\{\theta_n\} = \{u_n, p_n, v_n, x_0^n, x_1^n\} \subset X_1$, $n = 0, 1, 2, \dots$ by the rules

$$u_{n+1} = P_U[u_n - \alpha_n I'_{1u}(\theta_n)], \quad p_{n+1} = P_V[p_n - \alpha_n I'_{1p}(\theta_n)],$$
$$v_{n+1} = P_{V_1}[v_n - \alpha_n I'_{1v}(\theta_n)], \quad x_0^{n+1} = P_{S_0}[x_0^n - \alpha_n I'_{1x_0}(\theta_n)],$$
$$x_1^{n+1} = P_{S_1}[x_1^n - \alpha_n I'_{1x_1}(\theta_n)], \quad n = 0, 1, 2, \dots, \tag{36}$$
$$0 < \varepsilon_0 \le \alpha_n \le \frac{2}{K + 2\varepsilon}, \ \varepsilon > 0,$$

here $P_\Omega[\cdot]$ is a projection of a point onto the set Ω, $K = const > 0$ from (32).

Theorem 4.2. *Let the assumptions of theorem 4.1 hold and in addition the function $F_1(q, t)$ be convex with respect to $q \in R^N$ and the sequence $\{\theta_n\} \subset X_1$ be defined by (36). Then the following assertions hold.*

1) The functional (25) attains its infimum under the conditions (26) – (28), i.e.

$$\inf_{\theta \in X_1} I_1(\theta) = I_1(\theta_*) = \min_{\theta \in X_1} I_1(\theta), \ \theta_* \in X_1;$$

2) The sequence $\{\theta_n\}{\subset}X_1$ is minimizing, $\lim\limits_{n\to\infty} I_1(\theta_n){=}I_{1*}{=}\inf\limits_{\theta\in X_1} I_1(\theta)$;

3) The sequence $\{\theta_n\} \subset X_1$ weakly converges to the point $\theta_* = (u_*, p_*, v_*, x_0^*, x_1^*) \in X_1$;

4) For the problem (2) – (5) to have a solution it is necessary and sufficient to have $\lim\limits_{n\to\infty} I_1(\theta_n) = I_{1*} = 0$;

5) The rate of convergence can be estimated as

$$0 \le I_1(\theta_n) - I_{1*} \le \frac{C_0}{n}, \ n = 1, 2, \dots, \ C_0 = const > 0. \tag{37}$$

Proof. Since the function $F_1(q, t)$, $t \in I$ is convex the functional $I_1(\theta)$, $\theta \in X_1$ is convex on the weakly bicompact set X_1 by lemma 4.2.

Consequently $I_1(\theta) \in C^1(X_1)$ is weakly semicontinuous from below on the weakly bicompact set X_1 and attains its infimum on X_1. This yields the first assertion of the theorem.

By the property of a projection of a point onto the convex and closed set X_1 and taking into account $I_1(\theta) \in C^{1,1}(X_1)$ it can be easily proved that

$$I_1(\theta_n) - I_1(\theta_{n+1}) \geq \varepsilon\|\theta_n - \theta_{n+1}\|^2,$$

$n = 0, 1, 2, \ldots$, $\varepsilon > 0$. Hence we have the following.

1) The numerical sequence $\{I_1(\theta_n)\}$ strictly decreases;

2) $\|\theta_n - \theta_{n+1}\| \to 0$ as $n \to \infty$.

Since the functional is convex and the set X_1 is bounded, the inequality holds

$$0 \leq I_1(\theta_n) - I_1(\theta_*) \leq C_1\|\theta_n - \theta_{n+1}\|, \ C_1 = const > 0, \ n = 0, 1, 2, \ldots. \quad (38)$$

Therefore taking into account $\|\theta_n - \theta_{n+1}\| \to 0$ as $n \to \infty$, we get that the sequence $\{\theta_n\}$ is minimizing $\lim\limits_{n\to\infty} I_1(\theta_n) = I_1(\theta_*) = \inf\limits_{\theta\in X_1} I_1(\theta)$.

As $\{\theta_n\} \subset X_1$, the set X_1 is weakly bicompact the sequence $\theta_n \xrightarrow{\text{weakly}} \theta_*$ as $n \to \infty$.

As it follows from lemma 4.1 if the value $I_1(\theta_*) = 0$, then the optimal control problem (1) – (5) is solvable.

The estimate (37) directly follows from the inequality (38),

$$I_1(\theta_n) - I_1(\theta_{n+1}) \geq \varepsilon\|\theta_n - \theta_{n+1}\|^2.$$

The main stages of the proof for the theorem has been presented above. Proof of the similar theorem is given in [20] in detail. ∎

The following theorem is for the case when the function $F_1(q, t)$ is non-convex with respect to q.

Theorem 4.3. *Let the assumptions of theorem 4.1 hold, the sequence $\{\theta_n\} \subset X_1$ be defined by (36). Then the following assertions hold.*

1) The values of the functional $I_1(\theta_n)$ strictly decreases, where $n = 0, 1, 2, \ldots$;

2) $\|\theta_n - \theta_{n+1}\| \to 0$ as $n \to \infty$.

The proof of the theorem follows from theorem 4.2.

The following assertions follow from the results presented above.

1) If $\theta_* = (u_*, p_*, v_*, x_0^*, x_1^*) \in X_1$ is a solution to the optimization problem (25) – (28) such that $I_1(\theta_*) = 0$, then $(u_* = u_*(t), x_0^*, x_1^*) \in \Sigma \subset U \times S_0 \times S_1$ is an admissible control;

2) The function $x_*(t; t_0, x_0^*)$, $t \in I$ is a solution to the differential equation (2), satisfies the conditions $x(t_1; t_0, x_0^*) = x_1^*$, $x_*(t; t_0, x_0^*) \in G(t)$, $t \in I$;

3) A necessary and sufficient condition for existence of a solution to the boundary value problem (2) – (5) is $I_1(\theta_*) = 0$ where $\theta_* \in X_1$ is a solution to the problem (25) – (28);

4) The value of the functional (1) for the admissible control is

$$I(u_*(\cdot), x_0^*, x_1^*) = \int\limits_{t_0}^{t_1} F_0(x_*(t), u_*(t), x_0^*, x_1^*, t)dt = \gamma_*, \tag{39}$$

where $x_*(t) = x_*(t; t_0, x_0^*)$, $t \in I$. In general case the value $I(u_*(\cdot), x_0^*, x_1^*) \neq I(\overline{u}_*, \overline{x}_0^*, \overline{x}_1^*) = \inf I(u(\cdot), x_0, x_1)$, $(u(\cdot), x_0, x_1) \in L_2(I, R^m) \times S_0 \times S_1$.

5. CONSTRUCTING AN OPTIMAL SOLUTION

Consider the optimization problem (1) – (5). Let us define the scalar function $\sigma(t)$, $t \in I$ by

$$\sigma(t) = \int\limits_{t_0}^{t} F_0(x(\tau), u(\tau), x_0, x_1, \tau)d\tau, \ t \in I.$$

Then $\dot{\sigma}(t) = F_0(x(t), u(t), x_0, x_1, t)$, $\sigma(t_0) = 0$,
$\sigma(t_1) = \gamma = I(u(\cdot), x_0, x_1) \in \Omega = \{\gamma \in R^1 | \gamma \geq \gamma_0, \ \gamma_0 > -\infty\}$,
where $\gamma = I(u(\cdot), x_0, x_1) \geq \gamma_0$, the value γ is bounded from below, in particular $\gamma_0 = 0$, if $F_0 \geq 0$.

Now the optimal control problem (1) – (5) is presented in the form (see (25))

$$\sigma(t_1) = \gamma = I(u(\cdot), x_0, x_1) \to \inf \tag{40}$$

under the conditions

$$\dot{\sigma}(t) = F_0(x(t), u(t), x_0, x_1, t), \ \sigma(t_0) = 0, \ \sigma(t_1) = \gamma, \tag{41}$$

$$\dot{x} = A(t)x + B(t)f(x, u, t), \ (x(t_0) = x_0, x(t_1) = x_1) \in S_0 \times S_1, \tag{42}$$

$$x(t) \in G(t), \ u(\cdot) \in L_2(I, R^m), \ t \in I. \tag{43}$$

Introduce the notations

$$\mu(t) = \begin{pmatrix} \sigma(t) \\ x(t) \end{pmatrix}, \ A_2(t) = \begin{pmatrix} O_{1,1} & O_{1,n} \\ O_{n,1} & A(t) \end{pmatrix}, \ B_0 = \begin{pmatrix} 1 \\ O_{n,1} \end{pmatrix},$$

$$C_0(t) = \begin{pmatrix} O_{1,r} \\ B(t) \end{pmatrix}, \ P_0 = \begin{pmatrix} 1, & O_{1,n} \end{pmatrix}, \ P_1 = \begin{pmatrix} O_{n,1}, & I_n \end{pmatrix},$$

where $P_0\mu(t_1) = \sigma(t_1)$, $P_1\mu = x$.

Then the optimal control problem (40) – (43) has the form

$$P_0\mu(t_1) = \gamma = I(u(\cdot), x_0, x_1) \to \inf, \tag{44}$$

under the conditions

$$\dot{\mu} = A_2(t)\mu + B_0 F_0(P_1\mu, u, x_0, x_1, t) + C_0(t) f(P_1\mu, u, t), \tag{45}$$

$$\mu(t_0) = \mu_0 = \begin{pmatrix} \sigma(t_0) \\ x(t_0) \end{pmatrix} = \begin{pmatrix} O_{1,1} \\ x_0 \end{pmatrix} \in O_{1,1} \times S_0 = T_0, \tag{46}$$

$$\mu(t_1) = \mu_1 = \begin{pmatrix} \sigma(t_1) \\ x(t_1) \end{pmatrix} = \begin{pmatrix} \gamma \\ x_1 \end{pmatrix} \in \Omega \times S_1 = T_1, \tag{47}$$

$$P_1\mu(t) \in G(t), u(\cdot) \in L_2(I, R^m), \tag{48}$$

here $x(t) = P_1\mu(t)$, $\sigma(t) = P_0\mu(t)$, $t \in I$, γ is defined by (44).

6. IMBEDDING PRINCIPLE

Consider the boundary value problem (45) – (48). The corresponding linear control system has the form

$$\dot{\zeta} = A_2(t)\zeta + B_0\overline{w}_1(t) + C_0(t)\overline{w}_2(t), t \in I, \tag{49}$$

$$\overline{w}_1(\cdot) \in L_2(I, R^1), \quad \overline{w}_2(\cdot) \in L_2(I, R^r), \tag{50}$$

$$\zeta(t_0) = \mu_0 \in T_0, \quad \zeta(t_1) = \mu_1 \in T_1. \tag{51}$$

Introduce the following notations

$$\overline{B}_0(t) = (B_0, C_0(t)), \overline{w}(t) = (\overline{w}_1(t), \overline{w}_2(t)), \Psi(t,\tau) = K(t)K^{-1}(\tau),$$

$$\overline{a} = \Psi(t_0, t_1)\mu_1 - \mu_0, R(t_0, t_1) = \int\limits_{t_0}^{t_1} \Psi(t_0, t)\overline{B}_0(t)\overline{B}_0^*(t)\Psi^*(t_0, t)dt,$$

$$R(t_0, t) = \int\limits_{t_0}^{t} \Psi(t_0, \tau)\overline{B}_0(\tau)\overline{B}_0^*(\tau)\Psi^*(t_0, \tau)d\tau, R(t_0, t_1) = R(t_0, t) + R(t, t_1),$$

$$\overline{\Lambda}_1(t, \mu_0, \mu_1) = \overline{B}_0^*(t)\Psi^*(t_0, t)R^{-1}(t_0, t_1)a =$$
$$= \begin{pmatrix} B_0^*\Psi^*(t_0, t)R^{-1}(t_0, t_1)\overline{a} \\ C_0^*\Psi^*(t_0, t)R^{-1}(t_0, t_1)\overline{a} \end{pmatrix} = \begin{pmatrix} \overline{\Lambda}_{11}(t, \mu_0, \mu_1) \\ \overline{\Lambda}_{12}(t, \mu_0, \mu_1) \end{pmatrix},$$

$$K_1(t) = -\overline{B}_0^*\Psi^*(t)(t_0,t)R^{-1}(t_0,t_1)\Psi(t_0,t_1) =$$
$$= \begin{pmatrix} -B_0^*\Psi^*(t_0,t)R^{-1}(t_0,t_1)\Psi(t_0,t_1) \\ -C_0^*\Psi^*(t_0,t)R^{-1}(t_0,t_1)\Psi(t_0,t_1) \end{pmatrix} = \begin{pmatrix} K_{11}(t) \\ K_{12}(t) \end{pmatrix},$$

$$\overline{\Lambda}_2(t,\mu_0,\mu_1) = \Psi(t,t_0)R(t,t_1)R^{-1}(t_0,t_1)\mu_0 + \Psi(t,t_0)R(t_0,t)R^{-1}(t_0,t_1)\Psi(t_0,t_1)\mu_1,$$
$$K_2(t) = -\Psi(t,t_0)R(t_0,t)R^{-1}(t_0,t_1)\Psi(t_0,t_1), t \in I.$$

Theorem 6.1. *Let the matrix $R(t_0,t_1)$ be positive definite. Then a control $\overline{w}(t) = (\overline{w}_1(t), \overline{w}_2(t)) \in L_2(I, R^{1+r})$ brings a trajectory of the system (49) – (51) from any initial point $\mu_0 \in R^{1+n}$ to any desired final state $\mu_1 \in R^{1+n}$ if and only if $\overline{w}_1(t) \in \overline{W}_1,\ \ \overline{w}_2(t) \in \overline{W}_2$,*

$$\overline{W}_1 = \{\overline{w}_1(\cdot) \in L_2(I,R^1)/\overline{w}_1(t) = \overline{v}_1(t) + \overline{\Lambda}_{11}(t,\mu_0,\mu_1) + K_{11}(t)\overline{z}(t_1,\overline{v}),$$
$$\forall \overline{v}_1(\cdot) \in L_2(I,R^1), t \in I\}, \quad (52)$$

$$\overline{W}_2 = \{\overline{w}_{2(\cdot)} \in L_2(I,R^r)/\overline{w}_2(t) = \overline{v}_2(t) + \overline{\Lambda}_{12}(t,\mu_0,\mu_1) + K_{12}(t)\overline{z}(t_1,\overline{v}),$$
$$\forall \overline{v}_2(\cdot) \in L_2(I,R^r), t \in I\}, \quad (53)$$

where $\overline{v}(t) = (\overline{v}_1(t), \overline{v}_2(t)), \overline{z}(t) = \overline{z}(t,\overline{v}), t \in I$ is a solution to the differential equation

$$\dot{\overline{z}} = A_2(t)\overline{z} + B_0\overline{v}_1(t) + C_0(t)\overline{v}_2(t), \overline{z}(t_0) = 0, \quad (54)$$

$$\overline{v}_1(\cdot) \in L_2(I,R^1), \overline{v}_2(\cdot) \in L_2(I,R^r). \quad (55)$$

The solution to the system (49) – (51) is defined by

$$\zeta(t) = \overline{z}(t,\overline{v}) + \overline{\Lambda}_2(t,\mu_0,\mu_1) + K_2(t)\overline{z}(t_1,\overline{v}), t \in I. \quad (56)$$

The proof of the theorem is similar to that of theorem 3.3.

Lemma 6.1. *Let the matrix $R(t_0,t_1)$ be positive definite. Then the boundary value problem (45) – (48) is equivalent to the problem*

$$\overline{w}_1(t) \in \overline{W}_1, \quad \overline{w}_1(t) = F_0(P_1\zeta, u, x_0, x_1, t), \ \ t \in I, \quad (57)$$

$$\overline{w}_2(t) \in \overline{W}_2, \quad \overline{w}_2(t) = f(P_1\zeta, u, t), \ \ t \in I, \quad (58)$$

$$p(t) \in V(t), \quad p(t) = F(P_1\zeta, t), \ \ t \in I, \quad (59)$$

$$\dot{\overline{z}} = A_2(t)\overline{z} + B_0\overline{v}_1(t) + C_0(t)\overline{v}_2(t), \quad \overline{z}(t_0) = 0, \ \ t \in I, \quad (60)$$

$$\overline{v}_1(\cdot) \in L_2(I,R^1), \ \ \overline{v}_2(\cdot) \in L_2(I,R^r), \quad (61)$$

$$(x_0, x_1) \in S_0 \times S_1, \quad u(\cdot) \in L_2(I, R^m), \quad \gamma \in \Omega, \tag{62}$$
$$V(t)=\{p(\cdot)\in L_2(I, R^s)/\omega(t) \leq p(t) \leq \varphi(t), \ \ t \in I\},$$

here $\zeta(t)$, $t \in I$ is defined by (56), $\overline{z}(t, \overline{v})$ is a solution to the system (54), (55).

The assertion of lemma 6.1 follows from theorem 6.1.

Consider the following optimal control problem

$$J_2(\overline{v}, u, p, x_0, x_1, \gamma) = \int\limits_{t_0}^{t_1} F_2(\overline{q}(t), t)dt =$$
$$= \int\limits_{t_0}^{t_1} [|\overline{w}_1(t) - F_0(P_1\zeta(t), u(t), x_0, x_1, t)|^2 + |\overline{w}_2(t) - f(P_1\zeta(t), u(t), t)|^2 +$$
$$+ |p(t) - F(P_1\zeta(t), t)|^2]dt \to \inf \tag{63}$$

under the conditions (60)- (62), where $\overline{w}_1(t) \in \overline{W}_1$, $\overline{w}_2(t) \in \overline{W}_2$, $\overline{v} = (\overline{v}_1, \overline{v}_2)$, $\overline{q}(t) = (\overline{v}_1, \overline{v}_2, u, p, x_0, x_1, \gamma, \overline{z}(t), \overline{z}(t_1)$.

Note that the optimization problem (63), (60)–(62) has been obtained on the base of (57)- (62).

Theorem 6.2. *Let the matrix $R(t_0, t_1)$ be positive definite, the derivative $\dfrac{\partial F_2(\overline{q}, t)}{\partial q}$ satisfies the Lipschitz condition. Then the following assertions hold.*

1 *The functional (63) under the conditions (60) – (62) is continuously Frechet differentiable, the gradient of the functional*

$$J_2'(\overline{\theta}) = (J_{2\overline{v}_1}'(\overline{\theta}), J_{2\overline{v}_2}'(\overline{\theta}), J_{2u}'(\overline{\theta}), J_{2p}'(\overline{\theta}), J_{2x_0}'(\overline{\theta}), J_{2x_1}'(\overline{\theta}), J_{2\gamma}'(\overline{\theta})),$$
$$\overline{\theta} = (\overline{v}_1, \overline{v}_2, u, p, x_0, x_1, \gamma) \in \overline{X},$$
$$\overline{X} = L_2(I, R^1) \times L_2(I, R^r) \times L_2(I, R^m) \times V \times S_0 \times S_1 \times \Omega,$$
$$H_1 = L_2(I, R^1) \times L_2(I, R^r) \times L_2(I, R^m) \times L_2(I, R^s) \times$$
$$\times R^n \times R^n \times R^1, \quad \overline{X} \subset H_1, \quad J_2'(\overline{\theta}) \in H_1$$

at any point $\overline{\theta} \in \overline{X}$ is calculated by

$$J_{2\overline{v}_1}'(\overline{\theta}) = \frac{\partial F_2(\overline{q}(t), t)}{\partial \overline{v}_1} - B_0^*\overline{\psi}(t), \quad J_{2\overline{v}_2}'(\overline{\theta}) = \frac{\partial F_2(\overline{q}(t), t)}{\partial \overline{v}_2} - C_0^*\overline{\psi}(t),$$
$$J_{2u}'(\overline{\theta}) = \frac{\partial F_2(\overline{q}(t), t)}{\partial u}, \quad J_{2p}'(\overline{\theta}) = \frac{\partial F_2(\overline{q}(t), t)}{\partial p},$$

$$J'_{2x_0}(\overline{\theta}) = \int\limits_{t_0}^{t_1} \frac{\partial F_2(\overline{q}(t),t)}{\partial x_0} dt, \quad J'_{2x_1}(\overline{\theta}) = \int\limits_{t_0}^{t_1} \frac{\partial F_2(\overline{q}(t),t)}{\partial x_1} dt,$$

$$J'_{2\gamma}(\overline{\theta}) = \int\limits_{t_0}^{t_1} \frac{\partial F_2(\overline{q}(t),t)}{\partial \gamma} dt,$$

here $\overline{\psi}(t)$, $t \in I$ is a solution to the conjugate system

$$\dot{\overline{\psi}} = \frac{\partial F_2(\overline{q}(t),t)}{\partial \overline{z}} - A_2^*(t)\overline{\psi}, \quad \overline{\psi}(t_1) = -\int\limits_{t_0}^{t_1} \frac{\partial F_2(\overline{q}(t),t)}{\partial \overline{z}(t_1)} dt;$$

2 The gradient $J'_2(\overline{\theta}), \overline{\theta} \in \overline{X}$ satisfies the Lipschitz condition

$$\|J'_2(\overline{\theta}_1) - J'_2(\overline{\theta}_2)\| \leq l\|\overline{\theta}_1 - \overline{\theta}_2\|, \quad \forall \overline{\theta}_1, \overline{\theta}_2 \in \overline{X}. \tag{64}$$

The proof of the theorem is similar to that of theorem 4.1.

Construct the sequence $\{\overline{\theta}_n\} = \{\overline{v}_1^n, \overline{v}_2^n, u_n, p_n, x_0^n, x_1^n, \gamma_n\} \subset \overline{X}_2$ by the rule

$$\begin{gathered}
\overline{v}_1^{n+1} = P_{\overline{V}_1}[\overline{v}_1^n - \alpha_n J'_{2\overline{v}_1}(\overline{\theta}_n)], \quad \overline{v}_2^{n+1} = P_{\overline{V}_2}[\overline{v}_2^n - \alpha_n J'_{2\overline{v}_2}(\overline{\theta}_n)], \\
u_{n+1} = P_U[u_n - \alpha_n J'_{2u}(\overline{\theta}_n)], \\
p_{n+1} = P_V[p_n - \alpha_n J'_{2p}(\overline{\theta}_n)], \qquad x_0^{n+1} = P_{S_0}[x_0^n - \alpha_n J'_{2x_0}(\overline{\theta}_n)], \\
x_1^{n+1} = P_{S_1}[x_1^n - \alpha_n J'_{2x_1}(\overline{\theta}_n)], \qquad (65) \\
\gamma_{n+1} = P_{\overline{\Omega}}[\gamma_n - \alpha_n J'_{2\gamma}(\overline{\theta}_n)], \qquad n = 0,1,2,..., \\
0 \leq \alpha_n \leq \frac{2}{l+2\varepsilon}, \quad \varepsilon > 0, \quad l = const > 0, \qquad (66)
\end{gathered}$$

where

$$\overline{V}_1 = \{\overline{v}_1(\cdot) \in L_2(I,R^1)/\|\overline{v}_1\| \leq \overline{\beta}\}, \quad \overline{V}_2 = \{\overline{v}_2(\cdot) \in L_2(I,R^r)/\|\overline{v}_2\| \leq \overline{\beta}\},$$

$$U = \{u(\cdot) \in L_2(I,R^m)/\|u\| \leq \overline{\beta}\}, \quad \overline{\Omega} = \{\gamma \in R^1/\gamma_* \leq \gamma \leq \overline{\beta}\},$$

$\overline{X}_2 = \overline{V}_1 \times \overline{V}_2 \times U \times V \times S_0 \times S_1 \times \overline{\Omega} \subset H_1$, $\overline{\beta} > 0$ is a sufficiently large number.

Theorem 6.3. *Let the assumptions of theorem 6.2 hold, $\overline{X}_1$ is a bounded convex closed set, the sequence $\{\overline{\theta}_n\} \subset \overline{X}_2$ be defined by (66). Then the following assertions hold.*

1 The numerical sequence $\{J_2(\overline{\theta}_n)\}$ *strictly decreases,* $\|\overline{\theta}_n - \overline{\theta}_{n+1}\| \to 0$ *as* $n \to \infty$.

If in addition $F_2(\overline{q}, t)$ *is a convex function with respect to* $\overline{q}$*, then the following assertions hold.*

2 An infimum of the functional (63) is attained under the conditions (60) – (62);

3 The sequence $\{\overline{\theta}_n\} \subset \overline{X}_2$ *is minimizing,* $\lim\limits_{n\to\infty} J_2(\overline{\theta}_n) = J_{2*} = \inf\limits_{\overline{\theta}\in\overline{X}_2} J_2(\overline{\theta})$;

4 The sequence $\{\overline{\theta}_n\} \subset \overline{X}_2$ *weakly converges to the point* $\overline{\theta}_* \in \overline{X}_{1*}$,

$$\overline{X}_{2*} = \{\overline{\theta}_*/J_2(\overline{\theta}_*) = J_{2*} = \inf_{\overline{\theta}\in\overline{X}_1} J_2(\overline{\theta}) = \min_{\overline{\theta}\in\overline{X}_1} J_2(\overline{\theta})\},$$

$\overline{\theta}_* = (\overline{v}_1^*, \overline{v}_2^*, \overline{u}_*, p_*, \overline{x}_0^*, \overline{x}_1^*, \overline{\gamma}_*)$;

5 If $J_2(\overline{\theta}_*) = 0$*, then the optimal control for the problem (1) – (5) is* $\overline{u}_* \in U$, $\overline{x}_0^* \in S_0$, $\overline{x}_1^* \in S_1$*, and the optimal trajectory*

$$\overline{x}_*(t) = P_1\zeta_*(t) = P_1[\overline{z}(t, \overline{v}_*) + \overline{\Lambda}_2(t, \mu_0^*, \mu_1^*) + K_2(t)\overline{z}(t_1, \overline{v}_*)], t \in I,$$

where $\overline{v}_* = (\overline{v}_1^*, \overline{v}_2^*), \mu_0^* = (O_{1,1}, \overline{x}_0^*), \mu_1^* = (\gamma_*, \overline{x}_1^*)$*, the inclusion* $\overline{x}_*(t) \in G(t)$ *and the constraint (5) hold,* $J(\overline{u}_*, \overline{x}_0^*, \overline{x}_1^*) = \overline{\gamma}_*$;

6 The convergence rate can be estimated as

$$0 \leq J_2(\overline{\theta}_n) - J_{2*} \leq \frac{\overline{c}_0}{n}, n = 1, 2, ..., \overline{c}_0 = const > 0.$$

The proof of the similar theorem is presented above.

The more evident method for solving the problem (1) – (5) is the method of sequently narrowing of the set of admissible controls.

Theorem 6.4. *Let the assumptions of theorem 6.2 hold,* $\overline{X}_3 = \overline{V}_1 \times \overline{V}_2 \times U \times V \times S_0 \times S_1$ *be a bounded convex and closed set, the sequence* $\{\overline{\theta}_n\} \subset \overline{X}_2$ *be defined by (64) except the sequence* $\{\gamma_n\} \subset \overline{\Omega}$*. Then the following assertions hold.*

1 The numerical sequence $\{J_2(\overline{\theta}_n)\}$ *strictly decreases,* $\{\overline{\theta}_n\} \subset X_3$*;*

2 $\|\overline{\theta}_n - \overline{\theta}_{n+1}\| \to 0$ *as* $n \to \infty$, $\{\overline{\theta}_n\} \subset \overline{X}_3$;

If besides the function $F_2(\overline{q}, t)$ *is convex with respect to* $\overline{q}$ *at fixed* γ*, then the following assertions hold.*

3 The sequence $\{\overline{\theta}_n\} \subset \overline{X}_3$ *at fixed* $\gamma = \gamma_*$ *is minimizing;*

4 $\overline{\theta}_n \xrightarrow{weakly} \overline{\theta}_* \in \overline{X}_3$ *as* $n \to \infty, \gamma = \gamma_*$;

5 $J_2(\overline{\theta}_*) = \underset{\overline{\theta}_n \in \overline{X}_3}{inf} J_2(\overline{\theta}_n) = \underset{\overline{\theta}_n \in \overline{X}_3}{min} J_2(\overline{\theta}_n)$;

6 The convergence rate can be estimated as

$$0 \leq J_2(\overline{\theta}_n) - J_2(\overline{\theta}_*) \leq \frac{c_1}{n}, \quad c_1 = const > 0, \;\; n = 1, 2, ..., \;\; \{\overline{\theta}_n\} \subset \overline{X}_3.$$

The proof of the theorem is based on theorem 6.3 at fixed $\gamma \in \overline{\Omega}$, $\gamma = \gamma_*$.

Let $\overline{\theta}_* \in \overline{X}_2$ is a solution to the problem (63), (60) – (62) at $\gamma = \gamma_* \in \overline{\Omega}$. The two cases are possible:

1 the value $J_2(\overline{\theta}*) > 0$;

2 the value $J_2(\overline{\theta}_*) = 0$.

Note that $J_2(\overline{\theta}) \geq 0$, $\overline{\theta} \in \overline{X}_3$.

If $J_2(\overline{\theta}_*) > 0$, then for the new value of γ one can choose $\gamma = 2\gamma_*$, and if $J_2(\overline{\theta}_*) = 0$, then the new value $\gamma = \frac{\gamma_*}{2}$. By repeatedly bisecting an interval the minimal value of the functional (1) can be found under the conditions (2) – (5).

7. CONCLUSIONS

The Lagrange problem of calculus of variations with state variable constraints for processes described by ordinary differential equations is studied. The special cases of this problem are the simplest problem, the Bolza problem, the isoperimetric problem, conditional extremum problem.

The new approach named the imbedding principle different from the known method based on the Lagrange principle is proposed. The basis for the imbedding principle is studies on the first kind Fredholm integral equation. The existence theorem and the theorem on a general solution has been proved for the first kind Fredholm integral equation.

The main results of the research are the following.

– Reducing the boundary value problem associated with the conditions in the Lagrange problem to the free end-point optimal control problem with the specific objective functional;

– A necessary and sufficient condition for existence of an admissible control;

– The method for constructing an admissible control by the limit point of minimizing sequences;

– A necessary and sufficient condition for existence of a solution to the Lagrange problem;

– The method for constructing a solution to the Lagrange problem.

The novelty of the obtained results is that it isn't required an introducing auxiliary variables as the Lagrange multiplies and consequently, studies on an existence of a saddle point for the Lagrange functional; the existence problem and constructing a solution to the Lagrange problem are solved together.

References

[1] Dzh. Bliss, *Lektsii po variatsionnomu ischisleniyu*, IL, Moscow, 1950 (in Russian).

[2] M.A. Lavrentev, L.A. Lyusternik, *Osnovyi variatsionnogo ischisleniya*, ONTI, Moscow, 1935 (in Russian).

[3] L. Yang, *Lektsii po variatsionnomu ischisleniyu i teorii optimalnogo upravleniya*, Mir, Moscow, 1974 (in Russian).

[4] V.M. Alekseev, V.M. Tihomirov, S.V. Fomin, *Optimalnoe upravlenie*, Nauka, Moscow, 1979 (in Russian).

[5] L.S. Pontryagin, V.G. Boltyanskiy, R.V. Gamkrelidze, E.F. Mischenko, *Matematicheskaya teoriya optimalnyih protsessov*, Nauka, Moscow, 1969 (in Russian).

[6] Dzh. Varga, *Optimalnoe upravlenie differentsialnyimi i funktsionalnyimi uravneniyami*, Nauka, Moscow, 1977 (in Russian).

[7] R.V. Gamkrelidze, *Osnovyi optimalnogo upravleniya*, Izd-vo Tbilisskogo n-ta, Tbilisi, 1977 (in Russian).

[8] Ya.B. Li, L. Markus, *Osnovyi teorii optimalnogo upravleniya*, Nauka, Moscow, 1972 (in Russian).

[9] S.A. Aisagaliev, *Upravlyaemost nekotoroy sistemyi differentsialnyih uravneniy*, Differentsialnyie uravnniya, **9**, 27(1991), 1475-1486 (in Russian).

[10] S.A. Aisagaliev, S.S. Aisagalieva, *Konstruktivnyiy metod resheniya zadachi upravlyaemosti dlya obyiknovennyih differentsialnyih uravneniy*, Differentsialnyie uravneniya, **4**, 29(1993), 555-567 (in Russian).

[11] S.A. Aisagaliev, *Optimalnoe upravlenie lineynyimi sistemami s zakreplennyimi kontsami traektorii i ogranichennyim upravleniem*, Differentsialnyie uravneniya, **6**, 32(1996), 1-10 (in Russian).

[12] S.A. Aisagaliev, *Upravlyaemost i optimalnoe upravlenie nelineynyih sistem*, Izvestiya RAN, Tehnicheskaya kibernetika, **3**, 1993, 96-102 (in Russian).

[13] S.A. Aisagaliev, A.A. Kabidoldanova, *Optimalnoe byistrodeystvie nelineynyih sistem s ogranicheniyami*, Differentsialnyie uravneniya i protsessyi upravleniya, **1**, 2010, 30-55 (in Russian).

[14] S.A. Aisagaliev, A.P. Belogurov, *Upravlyaemost i byistrodeystvie protsessa, opisyivaemogo parabolicheskim uravneniem s ogranichennyim upravleniem*, Sibirskiy matematicheskiy zhurnal, **1**, 53(2011), 20-37 (in Russian).

[15] S.A. Aisagaliev, *K teoriii upravlyaemosti lineynyih sistem*, AA SSSR, Avtomatika i telemehanika, **5**, 1991, 35-44 (in Russian).

[16] S.A. Aisagaliev, A.A. Kabidoldanova, *Ob optimalnom upravlenii lineynyimi sistemami s lineynyim kriteriem kachestva i ogranicheniyami*, Differentsialnyie uravneniya, **6**, 48(2012), 826-838 (in Russian).

[17] S.A. Aisagaliev, *Obschee reshenie odnogo klassa integralnyih uravneniy*, Matematicheskiy zhurnal, **4(18)**, 5(2005), 17-34 (in Russian).

[18] S.A. Aisagaliev, I.V. Sevryugin, *Upravlyaemost i byistrodeystvie protsessa. opisyivaemogo lineynoy sistemoy obyiknoennyih differentsialnyih uravneniy*, Matematicheskiy zhurnal, **2(4)**, 13(2013), 5-30 (in Russian).

[19] S.A. Aisagaliev, *Konstruktivnaya teoriya kraevyih zadach optimalnogo upravleniya*, Kazakh universiteti, Almaty, 2007 (in Russian).

[20] S.A. Aisagaliev, A.A. Kabidoldanova, *Optimalnoe upravlenie dinamicheskih sistem*, Palmarium Academic Publishing, Verlag, Germaniya, 2012 (in Russian).

[21] S.A. Aisagaliev, *K prosteyshey zadache variatsionnogo ischisleniya*, Vestnik KazNU, **4(71)**, 2011, 20-32 (in Russian).

[22] F.P. Vasilev, *Chislennyie metodyi resheniya ekstremalnyih zadach*, Nauka, Moscow, 1980 (in Russian).

[23] F.P. Vasilev, *Metodyi resheniya ekstremalnyih zadach*, Nauka, Moscow, 1981 (in Russian).

[24] N.N. Moiseev, Yu.P. Ivanilov, E.M. Stolyarov, *Metodyi optimizatsii*, Nauka, Moscow, 1978 (in Russian).

ROMAI J., v.14, no.1(2018), 23–41

COUPLES DES SOUS-CATÉGORIES CONJUGUÉES

Botnaru Dumitru

L'Université d'État de Tiraspol, Chişinău, Moldova

dumitru.botnaru@gmail.com

Abstract On définit et étudie les couples de sous-catégories coréflectives et les sous-catégories c-réflectives dans la catégorie des espaces localement convexes Hausdorff.

Si $(\mathcal{K}, \mathcal{L})$ est un couple de sous-catégories conjuguées, alors:

1. $\mathcal{K}$ et $\mathcal{L}$ sont des catégories isomorphes.
2. $\mathcal{K}$ et $\mathcal{L}$ sont des sous-catégories semi-abéliennes.
3. Le produit semi-réflexif de la sous-catégorie $\mathcal{L}$ avec toute sous-catégorie réflective est une sous-catégorie $\mathcal{L}$-semi-reflexive.
4. Le produit semi-coréflexif de la sous-catégorie $\mathcal{K}$ avec toute sous-catégorie coréflective est une sous-catégorie $\mathcal{K}$-semi-coréflexive.

Keywords: foncteure exactement à gaucghe, sous-catégories coréflective, réflective, $\mathcal{L}$-semi-réflexive.

2010 MSC: 46 M 15, 1813 30.

1. Introduction

Notons avec $\mathcal{C}_2\mathcal{V}$ la catégorie des espaces localement convexes topologiques vectoriels Hausdorff (voir [14]).

Dans cet article, on va définir plusieurs notions. Nous utiliserons les notations suivantes.

Structures de factorisation:

$(\mathcal{E}pi, \mathcal{M}_f)$ = (la classe des épimorphismes, la classe des noyaux) = (la classe des morphismes à image dense, les inclusions topologiques à image fermée);

$(\mathcal{E}_u, \mathcal{M}_p)$ = (la classe des épimorphismes universels, la classe des monomorphismes précis)=(la classe des morphismes surjectifs, la classe des inclusions topologiques);

$(\mathcal{E}_p, \mathcal{M}_u)$ = (la classe des épimorphismes précis, la classe des monomorphismes universels) (voir [4]);

$(\mathcal{E}_f, \mathcal{M}ono)$ =(la classe des conoyaux, la classe des monomorphismes)=(la classe des morphismes factoriels, la classe des morphismes injectifs).

Sous-catégories coréflectives et réflectives:

Σ = la sous-catégorie coréflective des espaces avec la plus fine topologie localement convexe [14];

$\widetilde{\mathcal{M}}$ = la sous-catégorie coréflective des espaces avec la topologie Mackey [14];

$\mathcal{S}$ = la sous-catégorie réflective des espaces avec la topologie faible [14];

Π = la sous-catégorie réflective des espaces complets avec la topologie faible [14];

$u\mathcal{N}$ = la sous-catégorie réflective des espaces ultranucléaires [8, 11];

$\mathcal{N}$ = la sous-catégorie réflective des espaces nucléaires [15];

$\mathcal{S}h$ = la sous-catégorie réflective des espaces Schwartz [13];

$i\mathcal{R}$ = la sous-catégorie réflective des espaces inductifs semi-réflexifs [2];

$s\mathcal{R}$ = la sous-catégorie réflective des espaces semi-réflexifs [14];

Γ_0 = la sous-catégorie réflective des espaces complets;

$\mathbb{K}$ la classe des sous-catégories coréflectives non nulles;

$\mathbb{R}$ la classe des sous-catégories réflectives non nulles;

$\mathbb{R}(\mathcal{A})$ la classe des sous-catégories réflectives de la catégorie $\mathcal{A}$, où $\mathcal{A} \in \mathbb{K}$, ou $\mathcal{A} \in \mathbb{R}$;

$\mathbb{K}(\mathcal{A})$ la classe des sous-catégories coréflectives de la catégorie $\mathcal{A}$, où $\mathcal{A} \in \mathbb{K}$, ou $\mathcal{A} \in \mathbb{R}$;

$\mathbb{K}(\mathcal{B})$ (respectivement: $\mathbb{R}(\mathcal{B})$) la classe des sous-catégories $\mathcal{B}$-coréflectives (respectivement: $\mathcal{B}$-réflectives), où $\mathcal{B} \subset \mathcal{C}_2\mathcal{V}$;

$\mathbb{R}_{ex}$ (respectivement: $\mathbb{R}_{ex}(\mathcal{E}_u)$) la classe des sous-catégories réflectives (respectivement: $\mathcal{E}_u$-réflectives) fermée par rapport aux extensions: $(\mathcal{E}pi \cap \mathcal{M}_p)$-facteur-objets.

1.1. Soit $\mathcal{A}$ et $\mathcal{B}$ deux classes de morphismes. Alors:

1. $\mathcal{A} \circ \mathcal{B} = \{a \cdot b | a \in \mathcal{A}, b \in \mathcal{B}$ et la composition $a \cdot b$ existe$\}$.

2. La classe $\mathcal{A}$ se nomme $\mathcal{B}$-héréditaire, si $f \cdot g \in \mathcal{A}$ et $f \in \mathcal{B}$, alors $g \in \mathcal{A}$.

2^0. La classe $\mathcal{A}$ se nomme $\mathcal{B}$-cohéréditaire, si $f \cdot g \in \mathcal{A}$ et $g \in \mathcal{B}$, alors $f \in \mathcal{A}$. La classe $\mathcal{E}pi$ est $\mathcal{M}_u$-héréditaire ([4], Lemme 2.6), la classe $\mathcal{M}_u$ est $\mathcal{E}pi$-cohéréditaire.

3. $\mathcal{A}^{\top}$ est la classe de tous les morphismes orthogonaux du dessus pour tout morphisme de $\mathcal{A}$, et $\mathcal{A}^{\urcorner} = \mathcal{A}^{\top} \cap \mathcal{E}pi$ (voir [1,4,6]).

3^0. $\mathcal{A}^{\perp}$ est la classe de tous les morphismes orthogonaux du bas pour tout morphisme de $\mathcal{A}$, et $\mathcal{A}^{\llcorner} = \mathcal{A}^{\perp} \cap \mathcal{M}ono$.

4. La classe $\mathcal{A}$ se nomme stable à gauche, si pour tout carré cartésien

$$f \cdot g' = g \cdot f'$$

avec $f \in \mathcal{A}$, il résulte que $f' \in \mathcal{A}$ aussi.

4^0. La classe stable à droite.

Dans la catégorie $\mathcal{C}_2\mathcal{V}$, les classes $\mathcal{E}_f$ et $\mathcal{E}_u$ sont stables à gauche, et les classes $\mathcal{M}_f$, $\mathcal{M}_p$ et $\mathcal{M}_u$ sont stables à droite (voir [4]).

1.2. Pour $\mathcal{M}$, classe de monomorphismes, et $\mathcal{A}$, classe d'objets (une sous-catégorie), notons par $\mathbf{S}_{\mathcal{M}}(\mathcal{A})$ la sous-catégorie pleine de tous les $\mathcal{M}$-sous-objets des objets de $\mathcal{A}$, et $P(\mathcal{A})$ la sous-catégorie pleine de tout produit des objets de $\mathcal{A}$.

Notation duale: $\mathbf{Q}_{\mathcal{E}}(\mathcal{A})$, où $\mathcal{E} \subset \mathcal{E}pi$.

1.3. *L'opération* $\lambda_{\mathcal{R}}$ (voir [1]). Soit $\mathcal{A}$ une classe d'épimorphismes de la catégorie $\mathcal{C}_2\mathcal{V}$. Notons avec $\lambda(\mathcal{A})$ la sous-catégorie pleine de tout les objets Z à propriété:

Pour tout $p : X \to Y \in \mathcal{A}$, tout morphisme $f : X \to Z$ s'exteint par p:

$$f = g \cdot p,$$

pour un g.

Si $\mathcal{L}$ est une classe d'objets ou une sous-catégorie de la catégorie $\mathcal{C}_2\mathcal{V}$ et $\mathcal{R} \in \mathbb{R}$, alors notons $\lambda_{\mathcal{R}}(\mathcal{L}) = \lambda(\mathcal{A})$, où $\mathcal{A} = \{r^X | X \in |\mathcal{L}|\}$.

L'opération $\lambda^*(\mathcal{A})$ est définie duale et $\lambda^*_{\mathcal{K}}(\mathcal{A})$, où $\mathcal{A} \subset \mathcal{M}ono$, ou $\mathcal{A}$ est une sous-catégorie de la catégorie $\mathcal{C}_2\mathcal{V}$.

1.4. Proposition. *Pour toute classe d'épimorphismes, $\mathcal{A}$ la sous-catégorie $\lambda(\mathcal{A})$ est épiréflective.*

Les résultats principaux de l'ouvrage.

On définit les couples de sous-catégories conjuguées, les sous-catégories c-coréflectives et c-réflectives (Définition 2.6). On indique les conditions pour que deux sous-catégories forment un couple de sous-catégories conjuguées (Théorème 2.5), ou pour qu'une sous-catégorie soit c-réflective (Théorème 2.7).

On établit une application bijective entre la classe $\mathbb{R}_c$ des sous-catégories c-réflectives et la classe $\mathbb{R}_e(\Pi, \Gamma_0)$ des sous-catégories réflectives qui se contiennent dans Γ_0 et a le foncteur réflecteur exactement à gauche (Théorème 2.13).

On démontre que toute classe d'objects $\mathcal{M}_p$-injectifs génère une sous-catégorie c-réflective (Théorème 3.1). La sous-catégorie des espaces ultranucléaires $u\mathcal{N}$ est la plus grande sous-catégorie c-réflective qui se contient dans la sous-catégorie des espaces nucléaires $\mathcal{N}$ (Théorème 3.8).

2. Couples de sous-catégories conjuguées

La notion de couples de sous-catégories conjuguées a été introduite par l'auteur [3], pour que diverses propriétés soient formulées dans les ouvrages sans être rigoureusement démontrées.

2.1. Soit $k : \mathcal{C}_2\mathcal{V} \to \mathcal{K}$ et $r : \mathcal{C}_2\mathcal{V} \to \mathcal{R}$ un foncteur coréflecteur et un foncteur réflecteur. Notons

$$\mu\mathcal{K} = \{m \in \mathcal{M}ono | k(m) \in \mathcal{I}so\}, \varepsilon\mathcal{R} = \{e \in \mathcal{E}pi | r(e) \in \mathcal{I}so\}.$$

1. Soit $\mathcal{R} \in \mathbb{R}$ et $p : X \to Y \in \mathcal{E}pi$. Alors $p \in \varepsilon\mathcal{R}$ si et seulement si

$$r^X = f \cdot p \tag{1}$$

pour un f.

2. Soit $b : X \to Y \in \varepsilon\mathcal{R}$ et $A \in |\mathcal{R}|$. Alors pour tout $f : X \to A$, on a

$$f = g \cdot b \tag{2}$$

pour un g.

3. a) Dans la catégorie $\mathcal{C}_2\mathcal{V}$, pour tout $\mathcal{R} \in \mathbb{R}$, le couple $(\varepsilon\mathcal{R}, (\varepsilon\mathcal{R})^{\perp})$ est une structure de factorisation de droite (voir [4]);

b) Le couple $((\varepsilon\mathcal{R}) \circ \mathcal{E}_p, ((\varepsilon\mathcal{R}) \circ \mathcal{E}_p)^{\perp})$ qui est noté $(\mathcal{P}''(\mathcal{R}), \mathcal{I}''(\mathcal{R}))$ est une structure de factorisation (voir [4]).

Lemme. 1. *Soit* $\mathcal{L} \in \mathbb{R}$. *Alors* $\varepsilon\mathcal{L} = \mathcal{L}^{\top}$.

1^0. *Soit* $\mathcal{K} \in \mathbb{K}$. *Alors* $\mu\mathcal{K} = \mathcal{K}^{\perp}$.

Démonstration. $\varepsilon\mathcal{L} \subset \mathcal{L}^{\top}$. Soit $b : X \to Y \in \varepsilon\mathcal{L},\ f : A \to B \in \mathcal{L}$ et

$$f \cdot u = v \cdot b. \tag{3}$$

Si $l^Y : Y \to lY$ est $\mathcal{L}$-réplique de Y, alors $l^Y \cdot b : X \to lY$ est $\mathcal{L}$-réplique de X. Ainsi

$$u = g \cdot l^Y \cdot b \tag{4}$$

pour un g. Des égalités écrites, il résulte que

$$v = f \cdot g \cdot l^Y. \tag{5}$$

Ainsi $b \perp f$, et $\varepsilon\mathcal{L} \subset \mathcal{L}^{\top}$.

$$\begin{tikzcd}
X \arrow[rr, "b"] \arrow[dd, "u"'] & & Y \arrow[dl, "l^Y"'] \arrow[dd, "v"] \\
& lY \arrow[dl, dashed, "g"'] & \\
A \arrow[rr, "f"] & & B
\end{tikzcd}$$

$\mathcal{L}^{\top} \subset \varepsilon\mathcal{L}$. Soit $b : X \to Y, b \perp \mathcal{L}$ et $l^X : X \to lX, l^Y \cdot Y \to lY$ $\mathcal{L}$-réplique des objets respectifs. Alors

$$l(b) \cdot l^X = l^Y \cdot b. \tag{6}$$

Puisque $b \perp l(b)$, il résulte que

$$l^X = t \cdot b, \tag{7}$$

$$l^Y = l(b) \cdot t \tag{8}$$

pour un t. Alors

$$t = h \cdot l^Y \tag{9}$$

pour un h. Des égalités écrites, on obtient

$$l(b) \cdot h \cdot l^Y = l(b) \cdot t = l^Y, \tag{10}$$

Ou

$$l(b) \cdot h = 1. \tag{11}$$

$$\begin{tikzcd}
X \arrow[r, "b"] \arrow[d, "l^X"'] & Y \arrow[d, "l^Y"] \arrow[dl, dashed, "t"'] \\
lX \arrow[r, "l(b)"] & lY \arrow[l, dashed, bend left, "h"]
\end{tikzcd}$$

De l'égalité (8), on déduit que $t \in \mathcal{M}_u$. Alors de l'égalité (7), puisque la classe $\mathcal{E}pi$ est $\mathcal{M}_u$-héréditaire, il résulte que $b \in \mathcal{E}pi$. De l'égalité (7), on obtient que $t \in \mathcal{E}pi$, et de l'égalité (9), on obtient que $h \in \mathcal{E}pi$. Alors de (11) on a $h, l(b) \in \mathcal{I}so$. Ainsi $b \in \varepsilon\mathcal{L}$.

On démontre dualement l'égalité $\mu\mathcal{K} = \mathcal{K}^{\perp}$.

2.2. Proposition. *Soit $\mathcal{K} \in \mathbb{K}$ et $\mathcal{L} \in \mathbb{R}$. Les affirmations suivantes sont équivalentes pour la catégorie $\mathcal{C}_2\mathcal{V}$.*

1. Pour tout objet $X \in |\mathcal{C}_2\mathcal{V}|$, le morphisme $l^X \cdot k^X : kX \to lX$ est $\mathcal{K}$-coréplique de lX.

2. $\varepsilon\mathcal{L} \subset \mu\mathcal{K}$.

3. $\mathcal{L}^{\top} \subset \mathcal{K}^{\perp}$.

4. $\mathcal{K} \subset \lambda^*(\varepsilon\mathcal{L})$.

Démonstration. 1 ⇒ 2. Soit $b : X \to T \in \varepsilon\mathcal{L}$, $l^Y : Y \to lY$ et $k^X : kX \to X$ $\mathcal{L}$-réplique et $\mathcal{K}$-coréplique. Alors

$$l^X = l^Y \cdot b, \tag{12}$$

et

$$k^{lX} = l^X \cdot k^X. \tag{13}$$

Ainsi

$$k^{lX} = l^Y \cdot b \cdot k^X. \tag{14}$$

De l'égalité (14), on déduit que $b \cdot k^X \in \mu\mathcal{K}$, et comme $b \in \mathcal{M}ono$, il résulte que $b \in \mu\mathcal{K}$.

$$\begin{tikzcd}
& kX \arrow[ld, "k^X"'] \arrow[rd, "k^{lX}"] & \\
X \arrow[rr, "l^X"] \arrow[rd, "b"'] & & lX \\
& Y \arrow[ru, "l^Y"'] &
\end{tikzcd}$$

$2 \Rightarrow 3$. Ainsi, de la condition 2, et la Lemme 2.1, il résulte la condition 3.

$3 \Rightarrow 2$. En vertu des démonstrations ci-dessus.

$2 \Rightarrow 4$. $\mathcal{K} \subset \lambda^*(\varepsilon\mathcal{L})$. Soit $A \in |\mathcal{K}|, b : X \to Y \in \varepsilon\mathcal{L}$, et $f : A \to Y$.

$$\begin{tikzcd}
X \arrow[rr, "b"] & & Y \\
& A \arrow[lu, dashed, "g"] \arrow[ru, "f"'] &
\end{tikzcd}$$

Puisque $b \in \varepsilon\mathcal{L} \subset \mu\mathcal{K}$, et $A \in |\mathcal{K}|$, il résulte que

$$f = b \cdot g \tag{15}$$

pour un g.

$4 \Rightarrow 1$. Soit $A \in |\mathcal{C}_2\mathcal{V}|, l^A : A \to lA$ $\mathcal{L}$-réplique de A, et $k^A : kA \to A$ et $k^{lA} : klA \to lA$ $\mathcal{K}$-corépliques des objets respectifs. Alors

$$l^A \cdot k^A = k^{lA} \cdot u \tag{16}$$

pour un u. Puisque $l^X \in \varepsilon\mathcal{L}$, et $klA \in |\lambda^*(\varepsilon\mathcal{L})|$, il résulte que

$$k^{lA} = l^A \cdot h \tag{17}$$

pour un h. Alors

$$h = k^A \cdot v \tag{18}$$

pour un v. On vérifie facilement que $u = v^{-1}$.

$$\begin{tikzcd}
 & klA \arrow[dl, dashed, bend left=10, "v"'] \arrow[d, dashed, "h"] \arrow[dr, "k^{lA}"] & \\
kA \arrow[ur, dashed, bend left=10, "u"] \arrow[r, "k^A"] & A \arrow[r, "l^A"] & lA
\end{tikzcd}$$

2.2*. **Proposition.** *Soit $\mathcal{K} \in \mathbb{K}$ et $\mathcal{L} \in \mathbb{R}$. Les affirmations suivantes sont équivalentes.*

1. Pour tout objet $X \in |\mathcal{C}_2\mathcal{V}|$, le morphisme $l^X \cdot k^X : kX \to lX$ est $\mathcal{L}$-réplique de kX.

2. $\mu\mathcal{K} \subset \varepsilon\mathcal{L}$.

3. $\mathcal{K}^\perp \subset \mathcal{L}^\top$.

4. $\mathcal{L} \subset \lambda(\mu\mathcal{K})$.

2.3. Remarque. *La condition 1 de la Proposition 2.2, et la condition 1 de la Proposition* 2.2*, *on va l'écrire dans la variante $k \cdot l = k$ et respectivement $l \cdot k = l$.*

2.4. Proposition. *Avec les notations ci-dessus, les affirmations suivantes sont équivalentes:*

1. $\mu\mathcal{K} = \varepsilon\mathcal{L}$.

2. $\mathcal{K} = \lambda^*(\varepsilon\mathcal{L})$ et $\mathcal{L} = \lambda(\mu\mathcal{K})$.

Démonstration. 1 $\Rightarrow$ 2. $\mathcal{K} = \lambda^*(\varepsilon\mathcal{L})$. Il est suffisant de démontrer que $\lambda^*(\varepsilon\mathcal{L}) \subset \mathcal{K}$. Soit $A \in |\lambda^*(\varepsilon\mathcal{L})|$, $l^A : A \to lA$ et $k^{lA} : klA \to lA$ $\mathcal{L}$-réplique et $\mathcal{K}$-coréplique des objets respectifs. Puisque $k^{lA} \in \mu\mathcal{K} = \varepsilon\mathcal{L}$ et $A \in |\lambda^*(\varepsilon\mathcal{L})|$, il résulte que

$$l^A = k^{lA} \cdot h \tag{19}$$

pour un h. De l'égalité écrite, il résulte que $h \in \varepsilon\mathcal{L} = \mu\mathcal{K}$. Donc $h \in \mathcal{I}so$.

$\mathcal{L} = \lambda(\mu\mathcal{K})$. Démonstration duale.

2 $\Rightarrow$ 1. En vertu des Propositions 2.2 et 2.2* (équivalentes des points 2 et 4).

2.5. Des propositions 2.2 et 2.2* on formulera le résultat suivant.

Théorème. *Soit $\mathcal{K} \in \mathbb{K}$ et $\mathcal{L} \in \mathbb{R}$. Les affirmations suivantes sont équivalentes:*
1. $\mu\mathcal{K} = \varepsilon\mathcal{L}$.
2. a) $k \cdot l = k$; b) $l \cdot k = l$.
3. $\mathcal{K}^{\perp} = \mathcal{L}^{\top}$.
4. a) $\mathcal{K} = \lambda^*(\varepsilon\mathcal{L})$; b) $\mathcal{L} = \lambda(\mu\mathcal{K})$).

2.6. Définition. *Soit $\mathcal{A}$ une sous-catégorie pleine de la catégorie $\mathcal{C}_2\mathcal{V}, \mathcal{K}$ (respectivement: $\mathcal{L}$) une sous-catégorie coréflective (respectivement: réflective) de la catégorie $\mathcal{A}$. $(\mathcal{K}, \mathcal{L})$ se nomme couple de sous-catégories conjuguées de la catégorie $\mathcal{A}$ si $\mathcal{A} \cap \mu\mathcal{K} = \mathcal{A} \cap \varepsilon\mathcal{L}$.*

Dans ce cas, $\mathcal{K}$ (respectivement: $\mathcal{L}$) se nomme sous-catégorie c-coréflective (respectivement: c-réflective) de la catégorie $\mathcal{A}$, et $\mathcal{K}$ et $\mathcal{L}$ se nomment conjuguées l'une à l'autre.

Utilisons les notions suivantes:

$\mathbb{P}_c(\mathcal{A})$ (respectivement: $\mathbb{P}_c$) la clase des sous-catégories conjuguées de la catégorie $\mathcal{A}$ (respectivement: $\mathcal{C}_2\mathcal{V}$).

$\mathbb{K}_c(\mathcal{A})$ (respectivement: $\mathbb{K}_c$) la classe des sous-catégories c-coréflectives de la catégorie $\mathcal{A}$ (respectivement: $\mathcal{C}_2\mathcal{V}$).

$\mathbb{R}_c(\mathcal{A})$ (respectivement: $\mathbb{R}_c$) la classe des sous-catégories c-refléctives de la catégorie $\mathcal{A}$ (respectivement: $\mathcal{C}_2\mathcal{V}$).

2.7. Théorème. *Soit $\mathcal{L} \in \mathbb{R}$. Les affirmations suivantes sont équivalentes:*
1. $\mathcal{L} \in \mathbb{R}_c$.
2. $((\varepsilon\mathcal{L})^{\top}, \varepsilon\mathcal{L})$ est une structure de factorisation de gauche.
3. La classe $\varepsilon\mathcal{L}$ est stable à gauche.
4. $\mathcal{S} \subset \mathcal{L}$ et $l(\mathcal{M}_p) \subset \mathcal{M}_p$.
5. $\mathcal{S} \subset \mathcal{L}$ et $l(\mathcal{M}_f) \subset \mathcal{M}_f$.
6. $\mathcal{S} \subset \mathcal{L}$ et le foncteur l est exactement à gauche.
7. $\mathcal{S} \subset \mathcal{L}$ et le foncteur l commute avec des carrés cartésiens.
8. $\mathcal{S} \subset \mathcal{L}$ et le foncteur l commute avec des limites projectives.
9. $\mathcal{S} \subset \mathcal{L}$ et $l(\mathcal{I}''(\mathcal{R})) \subset \mathcal{I}''(\mathcal{R})$ pour tout $\mathcal{R} \in \mathbb{R}(\mathcal{L})$.
10. $\mathcal{S} \subset \mathcal{L}$, et pour tout $\mathcal{R} \subset \mathbb{R}(\mathcal{L})$, le couple $(\mathcal{P}''(\mathcal{R}) \cap (\varepsilon\mathcal{L})^{\top}, \mathcal{I}''(\mathcal{R}) \circ (\varepsilon\mathcal{L}))$ est une structure de factorisation dans la catégorie $\mathcal{C}_2\mathcal{V}$.
11. Le foncteur l possède un adjoint à gauche.

Démonstration. On démontrera les implications suivantes

$$\begin{tikzcd}
 & 1 \arrow[r, Rightarrow] \arrow[dl, Rightarrow] & 2 \arrow[r, Rightarrow] & 3 \arrow[r, Rightarrow] & 9 \arrow[r, Rightarrow] & 4 \arrow[r, Rightarrow] & 5 \arrow[r, Leftrightarrow] & 6 \arrow[r, Leftrightarrow] & 7 \arrow[r, Leftrightarrow] & 8 \\
11 \arrow[r, Rightarrow] & 8 \arrow[r, Rightarrow] & 2 \arrow[r, Rightarrow] \arrow[ul, Rightarrow] & 10 \arrow[u, Rightarrow] & & & & & &
\end{tikzcd}$$

$1 \Rightarrow 2$. Puisque $\varepsilon\mathcal{L} = \mu\mathcal{K}$, et $((\mu\mathcal{K})^{\top}, \mu\mathcal{K})$ est une structure de factorisation de gauche (voir [4], Théorème 2.12*).

$2 \Rightarrow 3$. Evidemment.

5 ⇔ 6. Dans la catégorie $\mathcal{C}_2\mathcal{V}$, on a $\mathcal{M}_f = Ker(\mathcal{C}_2\mathcal{V})$.

6 ⇔ 7 ⇔ 8. Un foncteur réflecteur dans la catégorie $\mathcal{C}_2\mathcal{V}$ commute avec les produits ([6], Théorème 1.12), et les noyaux peuvent être construits à l'aide des produits et des carrés cartésiens, et inversement.

1 ⇔ 11. Soit $(\mathcal{K}, \mathcal{L}) \in \mathbb{P}_c$, et on va démontrer que le foncteur $k : \mathcal{C}_2\mathcal{V} \to \mathcal{C}_2\mathcal{V}$ est un adjoint à gauche du foncteur $l : \mathcal{C}_2\mathcal{V} \to \mathcal{C}_2\mathcal{V}$. Soit X, Y deux objets de la catégorie $\mathcal{C}_2\mathcal{V}$, et on établira les isomorphismes fonctoriels (voir [9], cap. 1, §1)

$$Hom(kX, Y) \overset{\varphi}{\to} Hom(X, lY) \overset{\psi}{\to} Hom(kX, Y),$$

en notant pour $f : kX \to Y$ et $g : X \to lY$

$$\varphi(f) = l(f) \cdot l^X, \psi(g) = k^Y \cdot k(g).$$

$$\begin{tikzcd}
klX{=}kX \arrow[r, "k^X"] \arrow[d, shift right, "k(f)"'] \arrow[d, shift left, "k(g)"] \arrow[dr, "f"] & X \arrow[r, "l^X"] \arrow[dr, "g"] & lX{=}lkX \arrow[d, shift right, "l(f)"'] \arrow[d, shift left, "l(g)"] \\
klY{=}kY \arrow[r, "k^Y"] & Y \arrow[r, "l^Y"] & lY{=}lkY
\end{tikzcd}$$

On a

$$\psi\varphi(f) = \psi(l(f){\cdot}l^X) = k^Y{\cdot}k(l(f){\cdot}l^X) = k^Y{\cdot}kl(f){\cdot}k(l^X) = k^Y{\cdot}k(f){\cdot}1 = k^Y{\cdot}k(f) = f,$$

$$\varphi\psi(f) = \varphi(k^Y{\cdot}k((g)) = l(k^Y{\cdot}k(g)){\cdot}l^X = l(k^Y){\cdot}lk(g){\cdot}l^Y = 1{\cdot}l(g){\cdot}l^X = l(g){\cdot}l^X = g.$$

11 ⇒ 8. En vertu du Théorème P. Freyd (voir [9], Théorème 3.13). Un foncteur qui possède un adjoint à gauche commute avec les limites projectives.

9 ⇒ 4. Puisque $\mathcal{M}_p = \mathcal{J}''(\mathcal{S})$, et $\mathcal{S} \subset \mathcal{L}$.

4 ⇒ 5. Soit $m : X \to Y \in \mathcal{M}_f$. Examinons le carré commutatif

$$l(m) \cdot l^X = l^Y \cdot m. \tag{20}$$

Alors $l(m) \in \mathcal{M}_p$. Puisque $m(X)$ est un ensemble fermé dans l'espace Y, et Y et lY sont compatibles avec la même dualité, il résulte que $m(X)$ est un ensemble fermé dans lY (voir [14], cap. IV, §2, Théorème 1). Ainsi $l(m)$ a une image fermée. Donc $l(m) \in \mathcal{M}_f$.

8 ⇒ 2. Puisque le foncteur $l : \mathcal{C}_2\mathcal{V} \to \mathcal{L}$ commute avec les produits, il résulte que la classe $\varepsilon\mathcal{L}$ est fermée par rapport aux produits. Démontrons qu'elle est stable à gauche et fermée par rapport aux intersections. Soit $b : X \to Y \in \varepsilon\mathcal{L}$, et

$$b \cdot f' = f \cdot b' \tag{21}$$

est un carré cartésien. Examinons l'image de ce carré en utilisant le foncteur l :

$$l(b) \cdot l(f') = l(f) \cdot l(b') \tag{22}$$

En vertu de l'hypothèse 8, le carré (22) est cartésien et $l(b) \in \mathcal{I}so$. Donc $l(b') \in \mathcal{I}so$ aussi. Puisque $\mathcal{S} \subset \mathcal{L}$, il résulte que $\varepsilon\mathcal{L} \subset \varepsilon\mathcal{S}$. Donc $b' \in \mathcal{E}_u$ et $b' \in \varepsilon\mathcal{L}$.

$$\begin{tikzcd}
P \arrow[rr, "b'"] \arrow[dr, "l^P"] \arrow[dd, "f'"'] & & Z \arrow[dr, "l^Z"] \arrow[dd, "f"] & \\
& lP \arrow[rr, "l(b')"] \arrow[dd, "l(f')"] & & lZ \arrow[dd, "l(f)"] \\
X \arrow[rr, "b"] \arrow[dr, "l^X"'] & & Y \arrow[dr, "l^Y"] & \\
& lX \arrow[rr, "l(b)"] & & lY
\end{tikzcd}$$

Vérifions que la classe $\varepsilon\mathcal{L}$ est fermée par rapport aux intersections.

Soit $\{b_i : X \to X | i \in \mathcal{I}\}$ une famille de $(\varepsilon\mathcal{L})$-sous-objets de l'objet X, et $\{p_i : \bar{X} \to X_i | i \in \mathcal{I}\}$ et $b : \bar{X} \to X$ la limite projective du spectre respectif. Alors

$$b_i \cdot p_i = b, \ \forall i \in \mathcal{I}. \tag{23}$$

Puisque $b_i \in \mathcal{E}_u \cap \mathcal{M}_u$, il résulte que $b \in \mathcal{E}_u \cap \mathcal{M}_u$. En utilisant le foncteur l, on obtient que $l(b_i) \in \mathcal{I}so$, $\forall i \in \mathcal{I}$, et $\{l(p_i) | i \in \mathcal{I}\}$ et $l(b)$ est la limite projective du spectre respectif. Donc $l(b) \in \mathcal{I}so$.

$$\begin{tikzcd}
\bar{X} \arrow[d, "p_i"'] \arrow[dr, "b"] & & l\bar{X} \arrow[d, "l(p_i)"'] \arrow[dr, "l(b)"] & \\
X_i \arrow[r, "b_i"] & X & lX_i \arrow[r, "l(b_i)"] & lX
\end{tikzcd}$$

Les propriétés démontrées ci-dessus sont suffisantes pour que $((\varepsilon\mathcal{L})^{\top}, \varepsilon\mathcal{L})$ soit une structure de factorisation de gauche (voir [1], sec.14).

$2 \Rightarrow 1$. En factorisant conformément à la structure $((\varepsilon\mathcal{L})^{\top}, \varepsilon\mathcal{L})$ $\sigma^X : \sigma X \to X$ Σ-coréplique ou $m^X : mX \to X$ $\tilde{\mathcal{M}}$-coréplique de l'objet X, on obtient une sous-catégorie $(\varepsilon\mathcal{L})$-coréflective $\mathcal{K}$. Puisque $mlX = mX$, il résulte que $klX = kX$, et l'égalité $lkX = lX$ est déduite du fait que $l^X \cdot k^X \in \varepsilon\mathcal{L}$.

$2 \Rightarrow 10$. Examinons la structure de factorisation $(\mathcal{P}''(\mathcal{R}), \mathcal{I}''(\mathcal{R}))$, et la structure de factorisation de gauche $((\varepsilon\mathcal{L})^{\top}, \varepsilon\mathcal{L})$. Puisque $\mathcal{R} \subset \mathcal{L}$, la classe $\mathcal{P}''(\mathcal{R})$ est $(\varepsilon\mathcal{L})$-héréditaire en vertu du Théorème 3.2* [4], déduisons que $(\mathcal{P}''(\mathcal{R}) \cap (\varepsilon\mathcal{L})^{\top}, \mathcal{I}''(\mathcal{R}) \circ (\varepsilon\mathcal{L}))$ est une structure de factorisation.

$10 \Rightarrow 3$. Examinons le cas $\mathcal{R} = \mathcal{S}$. Alors $\mathcal{I}''(\mathcal{R}) = \mathcal{I}''(\mathcal{S}) = \mathcal{M}_p$ et la class $\mathcal{M}_p \circ (\varepsilon\mathcal{L})$ est stable à gauche. Démontrons que la classe $\varepsilon\mathcal{L}$ est aussi stable à gauche. Soit $b \in \varepsilon\mathcal{L}$, et

$$f \cdot b' = b \cdot f' \tag{24}$$

est un carré cartésien. Alors $b' \in \mathcal{M}_p \circ (\varepsilon\mathcal{L})$, et $b' \in \mathcal{E}_u$, puisque $\varepsilon\mathcal{L} \subset \varepsilon\mathcal{S} = \mathcal{E}_u \cap \mathcal{M}_u$. Donc $b' \in \varepsilon\mathcal{L}$.

$3 \Rightarrow 9$. Soit $\mathcal{R} \in \mathbb{R}$ et $\mathcal{R} \subset \mathcal{L}$. Notons $(\mathcal{P}, \mathcal{I}) = (\mathcal{P}''(\mathcal{R}), \mathcal{I}''(\mathcal{R}))$. Ainsi $\mathcal{P}''(\mathcal{L}) \subset \mathcal{P}$. Soit $i : X \to Y \in \mathcal{I}$. Alors

$$l(i) \cdot l^X = l^X \cdot i, \tag{25}$$

Démontrons que $\mathcal{P} \perp l(i)$. Soit $e : P \to T \in \mathcal{P}$ et

$$v \cdot e = l(i) \cdot u \tag{26}$$

Construisons les carrés cartésiens sur les morphismes v et l^X

$$v \cdot t' = l^Y \cdot v', \tag{27}$$

sur les morphismes e et t'

$$e \cdot t'' = t' \cdot e', \tag{28}$$

sur les morphismes $u \cdot t''$ et l^X

$$(u \cdot t'') \cdot t''' = \ell^X \cdot f. \tag{29}$$

La classe $\varepsilon\mathcal{L}$ est stable à gauche (hypothèse 3). Donc $t', t'', t''' \in \varepsilon\mathcal{L}$. Dans l'égalité (28) $t'' \in \varepsilon\mathcal{L} \subset \varepsilon\mathcal{R} \subset \mathcal{P}$, ou $t' \cdot e' \in \mathcal{P}$, et $t' \in \mathcal{M}_u$. La classe $\mathcal{P}$ est $\mathcal{M}_u$-héréditaire, ainsi $e' \in \mathcal{P}$ et $e', t''' \in \mathcal{P}$. On a

$$l^Y \cdot v' \cdot e' \cdot t''' = v \cdot t' \cdot e' \cdot t''' = v \cdot e \cdot t'' \cdot t''' = l(i) \cdot u \cdot t'' \cdot t''' \cdot \ell^Y \cdot i \cdot f,$$

i.e.

$$l^Y \cdot v' \cdot e' \cdot t''' = l^Y \cdot i \cdot f$$

en simplifiant l^Y, il reste

$$v' \cdot e' \cdot t''' = i \cdot f. \tag{30}$$

Dans la dernière égalité, $e' \cdot t''' \in \mathcal{P}$ et $i \in \mathcal{I}$, i.e. $e' \cdot t''' \perp i$. Il y a un morphisme h, ainsi que

$$f = h \cdot e' \cdot t''', \tag{31}$$

$$i \cdot h = v. \tag{32}$$

Plus loin, $l^Y \in \varepsilon\mathcal{L}$, donc $t' \in \varepsilon\mathcal{L}$ et $lY \in |\mathcal{L}|$. Ainsi

$$l^X \cdot h = w \cdot t' \tag{33}$$

pour un w. On a

$$l(i) \cdot w \cdot t' = l(i) \cdot l^X \cdot h = l^Y \cdot i \cdot h = l^Y \cdot v' = v \cdot t',$$

i.e.

$$l(i) \cdot w \cdot t' = v \cdot t' \tag{34}$$

ou

$$l(i) \cdot w = v. \tag{35}$$

La dernière égalité montre que $e \perp l(i)$.

$$\begin{tikzcd}
P_3 \arrow[r, dashed, "t''''"] \arrow[d, dashed, "f"'] & P_2 \arrow[r, dashed, "e'"] \arrow[dd, dashed, "t''"'] & P_1 \arrow[dll, dashed, "h"'] \arrow[dr, dashed, "v'"] \arrow[dd, dashed, "t'"] & \\
X \arrow[rrr, "i \in \mathcal{I}"] \arrow[dd, "l^X"'] & & & Y \arrow[dd, "l^Y"] \\
& P \arrow[r, "e \in \mathcal{P}"] \arrow[dl, "u"'] & T \arrow[dll, dashed, "w"] \arrow[dr, "v"] & \\
lX \arrow[rrr, "l(i)"'] & & & lY
\end{tikzcd}$$

2.8. Remarque. *La condition $\mathcal{S} \subset \mathcal{L}$ dans p.10 est nécessaire, puisque $\varepsilon\Pi = \mathcal{E}pi \cap \mathcal{M}_u$, et $\mathcal{M}_p \circ (\varepsilon\Pi) = \mathcal{M}_u$. Mais la sous-catégorie Π n'est pas c-réflective.*

2.9. Théorème. *Soit $\mathcal{R} \in \mathbb{R}$ et $\mathcal{S} \subset \mathcal{R}$. Alors il existe la plus grande sous-catégorie c-réflective qui se contient en $\mathcal{R}$.*

Démonstration. Soit $\mathcal{A}$ la classe de toutes les sous-catégories c-réflectives qui se contiennent en $\mathcal{R}$. Alors $\mathcal{S} \in \mathcal{A}$, et pour tout $\mathcal{L} \in \mathcal{A}$ on a $\varepsilon\mathcal{R} \subset \varepsilon\mathcal{L}$. La classe $\mathcal{B} = \cap\{\varepsilon\mathcal{L} | \mathcal{L} \in \mathcal{A}\}$ est une classe bicomplète, $\lambda(\mathcal{B})$ est une sous-catégorie c-réflective et $\lambda(\mathcal{B}) \subset \mathcal{R}$. On vérifie facilement que $\lambda(\mathcal{B})$ est la plus grande sous-catégorie c-réflective qui se contient en $\mathcal{R}$.

2.10. Soit $\mathcal{R} \in \mathbb{R}$, et $\mathcal{A}$ une sous-catégorie de la catégorie $\mathcal{C}_2\mathcal{V}$. Notons par $r(\mathcal{A})$ la catégorie pleine de tous les objets isomorphes à des objets avec la forme rX, quand $X \in |\mathcal{A}|$.

Lemme. *Soit $\mathcal{R}$ une sous-catégorie réflective et $\mathcal{K}$ une sous-catégorie coréflective de la catégorie $\mathcal{C}_2\mathcal{V}$. Si $\tilde{\mathcal{M}} \subset \mathcal{K}$ ou $\mathcal{S} \subset \mathcal{R}$, alors $r(\mathcal{K})$ est une sous-catégorie coréflective de la catégorie $\mathcal{R}$.*

Démonstration. Soit $A \in |\mathcal{R}|, k^A : kA \to A$ $\mathcal{K}$-coréplique de A et $r^{kA} : kA \to rkA$ $\mathcal{R}$-répliqie de kA. Alors

$$k^A = p^A \cdot r^{kA} \tag{36}$$

pour un morphisme $p^A : rkA \to A$. Montrons que p^A est une coréplique de l'objet A. Examinons la construction respective pour l'objet $B \in |\mathcal{R}|$, et soit $f : rkB \to A$. Alors

$$f \cdot r^{kB} = k^A \cdot g \tag{37}$$

pour un g et

$$r^{kA} \cdot g = h \cdot r^{kB} \tag{38}$$

pour un h. Des égalités écrites, on obtient:

$$p^A \cdot h \cdot r^{kB} = p^A \cdot r^{kA} \cdot g = k^A \cdot g = f \cdot r^{kB},$$

i.e.

$$p^A \cdot h \cdot r^{kB} = f \cdot r^{kB}, \tag{39}$$

ou

$$p^A \cdot h = f. \tag{40}$$

$$\begin{tikzcd}
kA \arrow[r, "r^{kA}"] \arrow[dr, "k^A"'] & rkA \arrow[d, dashed, "p^A"] & rkB \arrow[l, dashed, "h"'] \arrow[dl, "f"] \arrow[d, dashed, "p^B"] & kB \arrow[l, "r^{kB}"'] \arrow[dl, "k^B"] \arrow[lll, dashed, bend right, "g"'] \\
& A & B &
\end{tikzcd}$$

On a démontré ainsi que le morphisme f se factorise par p^A. L'unicité de cette factorisation a lieu dans les deux cas indiqués.

Le cas $\tilde{\mathcal{M}} \subset \mathcal{K}$. Alors $k^A \in \mathcal{E}_u \cap \mathcal{M}_u$, et $r^{kA} \in \mathcal{E}pi$. La classe $\mathcal{M}_u$ est $\mathcal{E}pi$-cohéréditaire. Donc $p^A \in \mathcal{M}_u$.

Le cas $\mathcal{S} \subset \mathcal{R}$. Puisque k^A et r^{kA} sont des applications bijectives de l'égalité (36), il résulte que p^A est aussi bijectif.

2.11. Théorème. *Soit* $\mathcal{R} \in \mathbb{R}, (\mathcal{K}, \mathcal{L}) \subset \mathbb{P}_c$, *et* $l(\mathcal{R}) \subset \mathcal{R}$. *Alors* $\mathcal{R} \cap \mu(r(\mathcal{K})) = \mathcal{R} \cap \varepsilon(r(\mathcal{L}))$. *En particulier,* $(r(\mathcal{K}), r(\mathcal{L}))$ *est un couple de sous-catégories conjuguées de la catégorie* $\mathcal{R}$.

Démonstration. Mentionnons que $\mathcal{R} \cap \mathcal{L}$ est une sous-catégorie réflective de la catégorie $\mathcal{R}$. Plus loin, de la condition $l(\mathcal{R}) \subset \mathcal{R}$, il résulte que pour tout objet $A \in |\mathcal{R}|$ $l^A : A \to lA$ est aussi $(\mathcal{R} \cap \mathcal{L})$-réplique. Observons que la condition $l(\mathcal{R}) \subset \mathcal{R}$ a lieu dans les conditions suivantes:
- $\mathcal{L} \subset \mathcal{R}$;
- $\mathcal{R}$ est fermé par rapport à $(\varepsilon\mathcal{L})$-facteur-objets, c'est-à-dire $\mathcal{R} \in \mathbb{R}_f(\varepsilon\mathcal{L})$ (voir [5]).

$\mathcal{R} \cap \mu(r(\mathcal{K})) \subset \mathcal{R} \cap \varepsilon(r\mathcal{L}))$. Soit $b : X \to Y \in \mathcal{R} \cap \mu(r(\mathcal{K}))$. Si $p^X : rkX \to X$ est $r(\mathcal{K})$-coréplique de X, alors $b \cdot p^X : rkX \to Y$ est $r(\mathcal{K})$-coréplique de Y.

De l'égalité (36), point précédent, il résulte que $b \cdot p^X \in \mu\mathcal{K}$. Comme $\mu\mathcal{K} = \varepsilon\mathcal{L}$, on a $b \cdot p^X \in \mathcal{R} \cap \varepsilon(\mathcal{L}) = \mathcal{R} \cap \varepsilon(r(\mathcal{L}))$.

$\mathcal{R} \cap \varepsilon(r(\mathcal{L})) \subset \mathcal{R} \cap \mu(r(\mathcal{K}))$. Soit $b : A \to B \in \mathcal{R} \cap \varepsilon(r(\mathcal{L}))$. Donc $b \in \mathcal{R} \cap \varepsilon(\mathcal{L}) = \mathcal{R} \cap \mu(\mathcal{K}) = \mathcal{R} \cap \mu(r(\mathcal{K}))$.

2.12. Corollaire. *1. Soit* $\mathcal{R} \in \mathbb{R}_f(\varepsilon\mathcal{S})$. *Alors pour tout* $(\mathcal{K}, \mathcal{L}) \in \mathbb{P}_c$, $(r(\mathcal{K}), r(\mathcal{L})) \in \mathbb{P}_c(\mathcal{R})$.

2. Pour tout $(\mathcal{K}, \mathcal{L}) \in \mathbb{P}_c, (r_s(\mathcal{K}), r_s(\mathcal{L})) \in \mathbb{P}_c(s\mathcal{R})$, *où* $s\mathcal{R}$ *est la sous-catégorie des espaces semi-réflexifs, et* $r_s : \mathcal{C}_2\mathcal{V} \to s\mathcal{R}$ *est le foncteur réflecteur.*

3. Soit $i\mathcal{R}$ *la sous-catégorie des espaces inductifs semi-réflexifs, et* $r_i : \mathcal{C}_2\mathcal{V} \to i\mathcal{R}$ *le foncteur réflecteur, et* $\mathcal{S}h$ *la sous-catégorie des espaces Schwartz (voir [2]). Pour* $(\mathcal{K}, \mathcal{L}) \in \mathbb{P}_c$, *si* $\mathcal{S}h \subset \mathcal{L}$, *alors* $(r_i(\mathcal{K}), r_i(\mathcal{L})) \in \mathbb{P}_c(i\mathcal{R})$.

2.13. Soit $\mathbb{R}_e(\Pi, \Gamma_0) = \{\mathcal{R} \in \mathbb{R} | \Pi \subset \mathcal{R} \subset \Gamma_0$ et $r(\mathcal{M}_f) \subset \mathcal{M}_f\}$, c'est-à-dire $\mathbb{R}_e(\Pi, \Gamma_0)$ contient les sous-catégories réflectives qui s'incluent dans la sous-catégorie Γ_0, et ont le foncteur réflecteur exactement à gauche.

Théorème. *1. Application* $\mathcal{R} \mapsto \varphi(\mathcal{R} = \mathcal{S}_{\mathcal{M}_p}(\mathcal{R})$ *pour* $\mathcal{R} \in \mathbb{R}_e(\Pi, \Gamma_0)$ *prend des valeurs dans la classe* $\mathbb{R}_c$.

2. Application $\mathcal{L} \mapsto \psi(\mathcal{L}) = \mathcal{L} \cap \Gamma$ *pour* $\mathcal{L} \in \mathbb{R}_c$ *prend des valeurs dans la classe* $\mathbb{R}_e(\Pi, \Gamma_0)$.

3. Les applications φ *et* ψ *sont réciproquement inverses*

$$\mathbb{R}_e(\Pi, \Gamma_0) \overset{\varphi}{\to} \mathbb{P}_c \overset{\psi}{\to} \mathbb{R}_e(\Pi, \Gamma_0).$$

Démonstration. 1. Soit $\mathcal{R} \in \mathbb{R}_e(\Pi, \Gamma_0)$ et $\mathcal{L} = \mathcal{S}_{\mathcal{M}_p}(\mathcal{R})$. Démontrons que le foncteur réflecteur $l : \mathcal{C}_2\mathcal{V} \to \mathcal{L}$ est exactement à gauche. Soit $m : X \to Y \in$

$\mathcal{M}_f, r^X : X \to rX$ et $r^X : Y \to rY$ $\mathcal{R}$-répliques des objets X et Y. Alors

$$r(m) \cdot l^X = l^Y \cdot m. \tag{41}$$

Soit

$$l^X = i^X \cdot p^X, \tag{42}$$

$$l^Y = i^Y \cdot p^Y, \tag{43}$$

$(\mathcal{E}_u, \mathcal{M}_p)$-fonctorisation des morphismes respectifs. Alors p^X et p^Y sont $\mathcal{L}$-répliques des objets respectifs, et

$$u \cdot p^X = p^Y \cdot m, \tag{44}$$

$$i^Y \cdot u = r(m) \cdot i^X \tag{45}$$

pour un morphisme u. Il est évident que $u = l(m)$. De l'égalité (45), puisque $r(m), i^X \in \mathcal{M}_p$, il résulte que $u \in \mathcal{M}_p$. X a l'image fermée en Y. Donc pX a l'image fermée en pY.

Donc $u \in \mathcal{M}_f$. On a démontré ainsi que $\varphi(\mathcal{R}) \in \mathbb{P}_c$.

$$\begin{tikzcd}
X \arrow[r, dashed, "p^X"] \arrow[rr, bend left, "r^X"] \arrow[d, "m"'] & pX \arrow[r, dashed, "i^X"] \arrow[d, dashed, "u=l(m)"] & rX \arrow[d, "r(m)"] \\
Y \arrow[r, dashed, "p^Y"] \arrow[rr, bend right, "r^Y"'] & pY \arrow[r, dashed, "i^Y"] & rY
\end{tikzcd}$$

Démontrons que ψ prend des valeurs dans la classe $\mathbb{R}_e(\Pi, \Gamma_0)$.

Premièrement, vérifions que $\Gamma_0 \in \mathbb{R}_e(\Pi, \Gamma_0)$. Soit $m : X \to Y \in \mathcal{M}_f, g_0^X : X \to g_0X$, et $g_0^Y : Y \to g_0Y$ Γ_0-répliques des objets respectifs. Alors

$$g_0(m) \cdot g_0^X = g_0^Y \cdot m. \tag{46}$$

Puisque $g_0^Y, m \in \mathcal{M}_p$, il résulte que $g_0(m) \cdot g_0^X \in \mathcal{M}_p$. Ayant en vue que la classe $\mathcal{M}_p$ est $\mathcal{E}pi$-cohéréditaire, décidons que $g_0(m) \in \mathcal{M}_p$. Et g_0X et g_0Y sont des espaces complets. Donc $g_0(m)$ a une image fermée. Donc $g_0(m) \in \mathcal{M}_f$.

Plus loin, tout élément de la classe $\mathbb{R}_c$ est fermé par rapport aux extensions: $(\mathcal{E}pi \cap \mathcal{M}_p)$-facteur-objets (voir [6], Théorème 3.7). Ainsi, si $\mathcal{L} \in \mathbb{R}_c$, alors

$g_0 \cdot l : \mathcal{C}_2\mathcal{V} \to \mathcal{L} \cap \Gamma_0$ est le foncteur réflecteur. Donc, si $m \in \mathcal{M}_f$, alors $l(m) \in \mathcal{M}_f$ et $g_0 l(m) \in \mathcal{M}_f$.

$\varphi \cdot \psi = 1$. Résulte du fait que $g_0 \cdot l : \mathcal{C}_2\mathcal{V} \to \mathcal{L} \cap \Gamma_0$ est le foncteur réflecteur.

$\psi \cdot \varphi = 1$. Revenons au diagramme précédent.

Une fois que $i^X : pX \to X \in \mathcal{M}_p$ et $rX \in |\Gamma_0|$, il résulte que i^X est Γ_0-réplique de pX.

2.14. Corollaire. $\mathbb{R}_c(\Pi, \Gamma_0)$ *n'est pas un ensemble.* $\mathbb{R}_e(\Pi, \Gamma_0)$ *contient une classe propre d'éléments.* (voir Théorème 3.3).

2.15. Théorème. *Soit* $(\mathcal{K}, \mathcal{L}) \in \mathbb{P}_c$. *Alors*

1. Les catégories $\mathcal{K}$ *et* $\mathcal{L}$ *sont isomorphes.*

2. Les catégories $\mathcal{K}$ *et* $\mathcal{L}$ *sont semi-abeliennes au sens de Raïcov.*

Démonstration. 1. Rezulte de la Défінion 2.6.

2. Soit $\mathcal{U}(\mathcal{L}) = \{l^X | X \in |\mathcal{C}_2\mathcal{V}|\}$. La classe des conoyaus de la catégorie $\mathcal{C}_2\mathcal{V}$(respectivement: $\mathcal{L}$) Cok (respectivement: $Cok\mathcal{L}$) vérifie les relations $Cok\mathcal{L} \subset \mathcal{U}(\mathcal{L}) \circ Cok \subset (\varepsilon\mathcal{L}) \circ Cok$. La classe $(\varepsilon\mathcal{L}) \circ Cok$ est stable tant à gauche qu'à droite. De telle manière on vérifie que $\mathcal{L}$ est semi-abelien.

3. Exemples

3.1. Théorème. *Soit* $\mathcal{A}$ *une famille d'objets* $\mathcal{M}_p$*-injectifs de la catégorie* $\mathcal{C}_2\mathcal{V}$. *Alors* $\mathcal{S}_{\mathcal{M}_p} P(\mathcal{A})$ *est une sous-categorie c-réflective.*

Démonstration. Soit $\mathcal{L} = \mathcal{S}_{\mathcal{M}_p} P(\mathcal{A}), l : \mathcal{C}_2\mathcal{V} \to \mathcal{L}$ le foncteur réflecteur, et montrons que $l(\mathcal{M}_p) \subset \mathcal{M}_p$. Vraiment, soit $m : X \to Y \in \mathcal{M}_p$. Alors

$$l(m) \cdot l^X = l^Y \cdot m. \tag{47}$$

Il existe un objet $Z \in |P(\mathcal{A})|$ et un morphisme $i : lX \to Z \in \mathcal{M}_p$. Puisque Z est $\mathcal{M}_p$-injectif et $m \in \mathcal{M}_p$, on a

$$i \cdot l^X = f \cdot m \tag{48}$$

pour un f. Et $Z \in |\mathcal{L}|$. Donc

$$f = g \cdot l^Y \tag{49}$$

pour un g. Ainsi

$$g \cdot l(m) \cdot l^X = g \cdot l^Y \cdot m = f \cdot m = i \cdot l^X,$$

i.e.

$$g \cdot l(m) \cdot l^X = i \cdot l^X, \tag{50}$$

ou

$$g \cdot l(m) = i. \tag{51}$$

De l'égalité (51), puisque $i \in \mathcal{M}_p$, il résulte que $l(m) \in \mathcal{M}_p$.

$$\begin{tikzcd}
X \arrow[r, "l^X"] \arrow[d, "m"'] & lX \arrow[r, "i"] \arrow[d, "l(m)"'] & Z \\
Y \arrow[r, "l^Y"] \arrow[urr, dashed, bend right, "f"'] & lY \arrow[ur, dashed, "g"] &
\end{tikzcd}$$

3.2. Soit X une ensemble de puissance α, et m_α l'espace Banach des fonctions bornées définies sur X. Alors m_α est un espace $\mathcal{M}_p$-injectif [12].

Corollaire [10]. *Pour tout cardinal τ, le foncteur réflecteur $r_\tau : \mathcal{C}_2\mathcal{V} \to \mathcal{S}_{\mathcal{M}_p}P(m_\tau)$ est exactment à gauche.*

3.3. Soit α et β deux cardinales, $\beta > \alpha$ et $\beta > \mathcal{X}_0$. Alors m_β n'appartient pas à la sous-catégorie $\mathcal{S}_{\mathcal{M}_p}P(m_\alpha)$ ([7], Théorème 2.14).

Théorème. *$\mathbb{P}_c, \mathbb{K}_c$ et $\mathbb{R}_c$ sont des classes propres.*

3.4. Corollaire. *$\mathcal{S}_{\mathcal{M}_p}P(K) = \mathcal{S}$, où K est le corps des nombres réels ou complexes.*

3.5. Théorème [10]. *Soit $\mathcal{L} = \mathcal{S}_{\mathcal{M}_p}P(m_\alpha)$ et $\mathcal{K}$ la conjuguée pour $\mathcal{L}$. Si $X \in |\mathcal{C}_2\mathcal{V}|$, alors la topologie sur kX est une topologie de la convergeance uniforme sur tous les ensemples absolument convexes faiblement compacts dans X' pour lequel tout sous-ensemple de puissance $\leq \alpha$ est équicontinu sur X.*

3.6. Dans les ouvrages [8, 11], on a défini et étudié les espaces ultranucléaires qui forment une sous-catégorie réflective de la catégorie $\mathcal{C}_2\mathcal{V}$ avec le foncteur réflecteur $u : \mathcal{C}_2\mathcal{V} \to u\mathcal{N}$.

Soit $t_\mathcal{N}$ la topologie nucléaire associée à la topologie normée à l'espace Hilbert ℓ^2, c'est-à-dire $(\ell^X, t_\mathcal{N})$ est $\mathcal{N}$-réplique de l'espace ℓ^2, où $\mathcal{N}$ est la sous-catégorie des espaces nucléaires. Tenant compte des notations ci-dessus, écrivons le Théorème 4 [8].

Théorème. $u\mathcal{N} = \mathcal{S}_{\mathcal{M}_p}P(\ell^2, t_\mathcal{N})$.

3.7. Là encore, on affirme que le foncteur réflecteur $u : \mathcal{C}_2\mathcal{V} \to u\mathcal{N}$ admet un adjoint à gauche. Alors du Théorème 2.7, il résulte.

Théorème. $u\mathcal{N} \in \mathbb{R}_c$.

3.8. Théorème. *$u\mathcal{N}$ est la plus grande sous-catégorie c-réfective qui se contient dans la sous-catégorie $\mathcal{N}$.*

Démonstration. Mentionnons que si $\mathcal{L}$ est une sous-catégorie c-réflective et $\mathcal{L} \subset \mathcal{S}_{\mathcal{M}_p}P(A)$ pour un objet A, alors $\mathcal{L} = \mathcal{S}_{\mathcal{M}_p}P(lA)$. Soit, maintenant, $\mathcal{L}$ une sous-catégorie c-réflective, et $u\mathcal{N} \subset \mathcal{L} \subset \mathcal{N}$. Alors $\mathcal{N} \subset \mathcal{S}_{\mathcal{M}_p}P(B)$ pour tout espace Banach B infini dimensionnel [15]. Donc $\mathcal{N} \subset \mathcal{S}_{\mathcal{M}_p}P(\ell^2)$. Pour l'objet ℓ^2 et $\mathcal{N}$-réplique, $\mathcal{L}$-réplique et $u\mathcal{N}$-réplique, on a les relations suivantes: $\ell^2 \to n\ell^2 \to l\ \ell^2 \to u\ell^2 = n\ell^2$. Donc $\mathcal{L} = u\mathcal{N}$.

3.9. Soit $\mathcal{S}h$ sous-catégorie des espaces Scwartz. $\mathcal{S}h$ est une sous-catégorie réflective, on notera le foncteur réflecteur par $s_h : \mathcal{C}_2\mathcal{V} \to \mathcal{S}h$. Si c_0 est l'espace Banach des séries qui convergent à zéro, alors $\mathcal{S}h$-réplique de c_0 s_hc_0 est un objet universel pour $\mathcal{S}h$: $\mathcal{S}h = \mathcal{S}_{\mathcal{M}_p}P(s_hc_0)$ [13], et le foncteur réflecteur s_h est exactement à gauche [2]. Soit $\mathcal{C}h$ la conjuguée de la sous-catégorie $\mathcal{S}_h$ au foncteur coréflecteur $c_h : \mathcal{C}_2\mathcal{V} \to \mathcal{C}h$. Pour tout objet $X \in |\mathcal{C}_2\mathcal{V}|$, c_hX possède la topologie de la convergeance uniforme de toutes les ensembles faiblement compactes A de X' avec la propriété: si la suite (x'_n) converge à zero dans Banach espace X'_A, alors elle est équicontinue dans X' [10].

Théorème. $(\mathcal{C}_h, \mathcal{S}_h) \in \mathbb{P}_c$.

Autres propriétés des foncteurs $s_h : \mathcal{C}_2\mathcal{V} \to \mathcal{S}h$ et $c_h : \mathcal{C}_2\mathcal{V} \to \mathcal{C}h$ (voir [5]).

References

[1] Adámek J., Herrlich H., Strecker G.E., *Abstract and Concrete Categories* - the joy of cats. Dover Publications, New York, 2004.

[2] Berezansky J.A., *Les espaces inductivement réflexifs localement convexes,* Dokl. Ak.Nauk SSSR, 182(1966), 1, 20-22 (en russe).

[3] Botnaru D., *Couples des sous-catégories conjuguées,* Uspehi Math. Nauk, XXXI (1976), 3(189), 203-204 (en russe).

[4] Botnaru D., *Structures bicatégorielles complémentaires,* ROMAI J., 5(2009), 2, 5-27.

[5] Botnaru D., *Groupoïd des sous-catégories $\mathcal{L}$-semi-réflexives,* Rev. Roumaine Math. Pures Appl., 63(2018), 1, 61-71.

[6] Botnaru D., Cerbu O., *Semireflexive product of two subcategories,* Proc. Sixth Congress of Romanian Math., Bucharest, 1 (2007), 5-19.

[7] Botnaru D., Ţurcanu A., Cerbu O., *Bicategory structures generated by injective spaces,* ROMAI J., 3(2007), 2, 31-50.

[8] Brudovsky B.S., *La topologie nucléaire associée, application du type s et les espaces ultranucléaires,* Dokl, Ak. Mauk SSSR, 178(1968), 2, 271-273 (en russe).

[9] Bucur I., Deleanu A., *Introducton to the theory of categories and functors,* London·New York· Sydney: John Wiley and Sons, Ltd, 1968.

[10] Gheyler V.A., Ghisin V.B., *Dualité généralisée pour les espaces localement convexe,* Func. an., Mejvouz.sb., Oulianovsk, 11(1978), 41-50 (en russe).

[11] Martineau A., *Sur une propriété universelle de l'espace de distributions de M.Schwartz,* C.R. Acad. Sci.Paris, 259(1964), 3162-3164.

[12] Palamodov D.P., *Homology methods in the theory of locally convex spaces,* Uspehi Mat. Nauk, 26(1971), 1(157), 3-65 (Russian).

[13] Randtke D.J., *A simple example of universal Schwartz space,* Proc. Amer. Math. Soc., 37(1973), 1, 185-188.

[14] Robertson A.P., Robertson W.J., *Topological vectors spaces,* Cambridge Tracts in Mathematics and Mathematical Phisics no 53. Cambridge University Press, 1964.

[15] Saxon S.A., *Embedding nuclear spaces in products arbitrary Banach spaces,* Proc. Amer. Math. Soc., 34(1972), 1, 138-140.

ROMAI J., v.14, no.1(2018), 43–62

ON THE RESTRICTED EIGHT BODIES PROBLEM

Elena Cebotaru
Technical University of Moldova, Chişinău, Republic of Moldova
elena.cebotaru@mate.utm.md

Abstract We consider the Newtonian restricted eight body problem. We investigate the linear stability of this configuration by some numerical methods.

Keywords: Newtonian problem; differential equation of motion; configuration; particular solutions; equilibrium points; linear stability.
2010 MSC: 70F10, 70K20.

1. INTRODUCTION

In the present article we examine the problem of the stability in the Lyapunov sense of the plane Newtonian problem formulated by the academician E.A. Grebenicov. Assume that in three-dimensional inertial space $Oxyz$ we have $(n+1)$ bodies that attract each other according to the universal attraction law. Then the vectorial form of the differential equations describing the motion of these bodies is the following (see [1]):

$$m_i\ddot{\vec{r}_i} = \overrightarrow{grad_iU}, \quad i = 0, 1, ..., n, \tag{1}$$

where $\overrightarrow{OP}_i = \overrightarrow{r}_i = (x_i, y_i, z_i,)$ are the radius of the vectors, and the scalar-potential function U is expressed by the relation:

$$U(x_0, y_0, z_0, ..., x_n, y_n, z_n) = \frac{f}{2}\sum_{k=0}^{n}\sum_{i=0}^{n}{}' \frac{m_k m_i}{\Delta_{ki}}, \tag{2}$$

$$\Delta_{ki} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2 + (z_k - z_i)^2}, \tag{3}$$

f is the gravitational constant, and the sign $'$ in sums stands for $k \neq i$. The vector $\overrightarrow{grad_iU}$ is the usual gradient one. Written in coordinate form, the equalities (1) represents a system of nonlinear differential equations of the order $(6n+6)$:

$$\begin{cases} m_i\frac{d^2x_i}{dt^2} = \frac{\partial U}{\partial x_i}, \\ m_i\frac{d^2y_i}{dt^2} = \frac{\partial U}{\partial y_i}, \\ m_i\frac{d^2z_i}{dt^2} = \frac{\partial U}{\partial z_i}, \\ i = 0, 1, 2, ..., n. \end{cases} \tag{4}$$

As a restricted of n-body problem we understand the case when one of the bodies does not fulfill the Third Law of Newtonian dynamics. That implies the existence of an infinitely small mass: one of the masses $m_0, m_1, ..., m_n$ is infinitely small and it does not attract other masses, but these attract them. Replacing the infinitely small zero mass into the system (4) will result in a similar system but of order $6n$. Thus, instead of the system of differential equations of the Newtonian $(n{+}1)$ body problem, we will obtain the differential equations of the Newtonian problem in the n bodies. Thus, after we equate with zero of one of the masses $m_0, m_1, ..., m_n$ in equations (4), the differential equations describing the motion of the infinitely small mass are not obtained. We will write the system (4) in a more detailed form, highlighting, for example, the differential equations of the mass m_0:

$$\begin{cases} m_0 \frac{d^2x_0}{dt^2} = \frac{\partial U}{\partial x_0} = f m_0 \sum\limits_{k=1}^{n} m_k \frac{x_k - x_0}{\Delta_{k0}^3}, \\ m_0 \frac{d^2y_0}{dt^2} = \frac{\partial U}{\partial y_0} = f m_0 \sum\limits_{k=1}^{n} m_k \frac{y_k - y_0}{\Delta_{k0}^3}, \\ m_0 \frac{d^2z_0}{dt^2} = \frac{\partial U}{\partial z_0} = f m_0 \sum\limits_{k=1}^{n} m_k \frac{z_k - z_0}{\Delta_{k0}^3}, \end{cases} \tag{5}$$

$$\begin{cases} m_i \frac{d^2x_i}{dt^2} = \frac{\partial U}{\partial x_i} = f m_i \left(\sum\limits_{k=1}^{n}{}' \, m_k \frac{x_k - x_i}{\Delta_{ki}^3} + m_0 \frac{x_0 - x_i}{\Delta_{0i}^3} \right), \\ m_i \frac{d^2y_i}{dt^2} = \frac{\partial U}{\partial y_i} = f m_i \left(\sum\limits_{k=1}^{n}{}' \, m_k \frac{y_k - y_i}{\Delta_{ki}^3} + m_0 \frac{y_0 - y_i}{\Delta_{0i}^3} \right), \\ m_i \frac{d^2z_i}{dt^2} = \frac{\partial U}{\partial z_i} = f m_i \left(\sum\limits_{k=1}^{n}{}' \, m_k \frac{z_k - z_i}{\Delta_{ki}^3} + m_0 \frac{z_0 - z_i}{\Delta_{0i}^3} \right), \\ i = 1, 2, ..., n, \;\; k \neq i. \end{cases} \tag{6}$$

Since the mass m_0 is small but not identical zero, we can divide both sides of system (5) by m_0. Moreover, since $m_0 = \mu$, $0 < \mu << 1$, in a first approximation we can replace in system (6) m_0 by 0. We obtain the following systems:

$$\begin{cases} \frac{d^2x_0}{dt^2} = f \sum\limits_{k=1}^{n} m_k \frac{x_k - x_0}{\Delta_{k0}^3}, \\ \frac{d^2y_0}{dt^2} = f \sum\limits_{k=1}^{n} m_k \frac{y_k - y_0}{\Delta_{k0}^3}, \\ \frac{d^2z_0}{dt^2} = f \sum\limits_{k=1}^{n} m_k \frac{z_k - z_0}{\Delta_{k0}^3}, \end{cases} \tag{7}$$

$$\begin{cases} \frac{d^2x_i}{dt^2} = f\sum\limits_{k=1}^{n}{}' \, m_k \frac{x_k - x_i}{\Delta_{ki}^3}, \\ \frac{d^2y_i}{dt^2} = f\sum\limits_{k=1}^{n}{}' \, m_k \frac{y_k - y_i}{\Delta_{ki}^3}, \\ \frac{d^2z_i}{dt^2} = f\sum\limits_{k=1}^{n}{}' \, m_k \frac{z_k - z_i}{\Delta_{ki}^3}, \\ i = 1, 2, ..., n, \;\; k \neq i. \end{cases} \tag{8}$$

Analyzing the systems (7) and (8), we notice that the second one does not contain the coordinates x_0, y_0, z_0 of the infinitely small mass. The system (8) describes the motion in the Newtonian n-body problem with the respective masses equal to m_1,..., m_n. We admit that one of the particular solutions of the system (8) has been determined:

$$\begin{cases} x_k(t) = f_k(t), \\ y_k(t) = \varphi_k(t), \\ z_k(t) = \psi_k(t), \\ k = 1, 2, ..., n. \end{cases} \tag{9}$$

By replacing these functions in the system (7) we obtain:

$$\begin{cases} \frac{d^2x_0}{dt^2} = f\sum\limits_{k=1}^{n} m_k \frac{f_k(t) - x_0}{\Delta_{k0}^3}, \\ \frac{d^2y_0}{dt^2} = f\sum\limits_{k=1}^{n} m_k \frac{\varphi_k(t) - y_0}{\Delta_{k0}^3}, \\ \frac{d^2z_0}{dt^2} = f\sum\limits_{k=1}^{n} m_k \frac{\psi_k(t) - z_0}{\Delta_{k0}^3}, \\ \Delta_{k0}^2(t) = (f_k(t) - x_0)^2 + (\varphi_k(t) - y_0)^2 + (\psi_k(t) - z_0)^2 . \end{cases} \tag{10}$$

The system (10) describes the motion of the infinitely small mass $m_0 = \mu$ in the gravitational field of the masses $m_1, m_2, ..., m_n$ and which moves following the law described by the solution $f_k(t), \varphi_k(t), \psi_k(t)$. Hence the motion of the infinitely small mass in any restricted n-body problem may be described by a non-autonomous system of the sixth differential equations.

In the case of a plane solution $z_0 = z_1 = ... = z_n \equiv 0$, it will be investigated a fourth-order system, not that of the sixth order:

$$\begin{cases} \frac{d^2x_0}{dt^2} = f\sum\limits_{k=1}^{n} m_k \frac{f_k(t) - x_0}{\Delta_{k0}^3}, \\ \frac{d^2y_0}{dt^2} = f\sum\limits_{k=1}^{n} m_k \frac{\varphi_k(t) - y_0}{\Delta_{k0}^3}, \\ \Delta_{k0}^2(t) = (f_k(t) - x_0)^2 + (\varphi_k(t) - y_0)^2 . \end{cases}$$

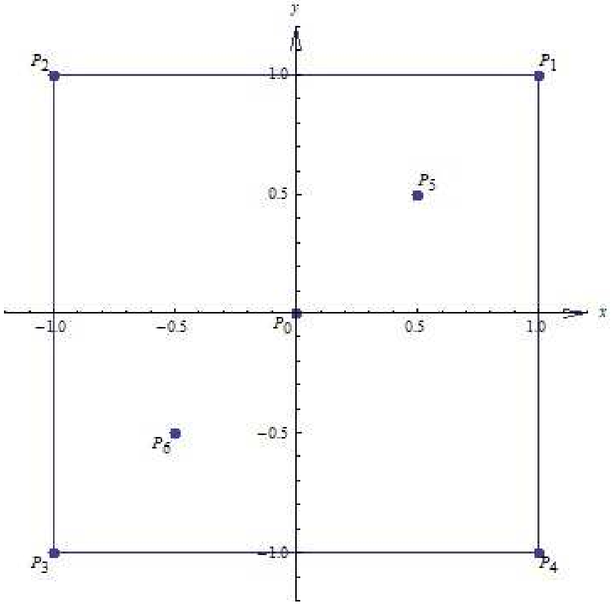


Fig. 1: Studied model

So we can state the following theorem:

Theorem 1.1. *Any known particular solution to the Newtonian problem of the n bodies serves as a generator of differential equations of motion in the restricted Newtonian problem of* $(n + 1)$ *bodies.*

2. DIFFERENTIAL EQUATIONS OF THE RESTRICTED PROBLEM OF EIGHT BODIES. DETERMINATION OF EQUILIBRIUM POSITIONS.

From the above mentioned facts, it follows that in studying of the differential equations of restricted problems, first of all it is necessary to study the existence of particular solutions of ,,equilibrium positions" in the unlimited small size problems. If the origin of the three-dimensional Euclidean space it will be chosen one of the bodies that gravitates, for example the point P_0, then, obviously, the Euclidean coordinate system P_0xyz will be non-inertial.

Assume that in a non-inertial space P_0xyz there is the motion of eight bodies P_0, P_1, P_2, P_3, P_4, P_5, P_6, P with the masses m_0, m_1, m_2, m_3, m_4, m_5, m_6, μ, which attract each other in accordance with the law of universal attraction . We will investigate the planar dynamic pattern formed by a square in the vertices of which are placed the bodies P_1, P_2, P_3, P_4. The body P_0

is the center of the square and the bodies P_5, P_6 are placed on the diagonal P_1P_3 of the square at equal distances from P_0 (see Fig.1). We consider that $m_5 = m_6$ and the configuration rotates around the body P_0 with the constant angular velocity ω, which is determined from the model parameters . It will be studied the motion of the infinitely small mass $\mu = 0$ (the so-called passive gravitational body) in the gravitational field formed by the seven bodies $P_0, P_1, P_2, P_3, P_4, P_5, P_6$, that attract each other and attract the body P.

In the case of restricted 3-body problem H. Poincaré has demonstrated that the differential equations of dynamics of the point P form a non-integrable system. The Hamiltonian equations of the restricted problem of eight bodies have an even more complicated structure, that's why they can also be considered as non-integrable in sence of Poincaré. The differential equations of the Newtonian problem of eight bodies in a non-inertial cartesian coordinate system P_0xyz have the form:

$$\begin{cases} \frac{d^2x_k}{dt^2} + \frac{f(m_0+m_k)x_k}{r_k^3} = \frac{\partial R_k^*}{\partial x_k}, \\ \frac{d^2y_k}{dt^2} + \frac{f(m_0+m_k)y_k}{r_k^3} = \frac{\partial R_k^*}{\partial y_k}, \\ \frac{d^2z_k}{dt^2} + \frac{f(m_0+m_k)z_k}{r_k^3} = \frac{\partial R_k^*}{\partial z_k}, \\ k = 1, 2, ..., 7; \end{cases} \tag{11}$$

where $R_k^*(k = 1, 2, ..., 7)$ are the perturbation functions which are expressed by the relations:

$$\begin{cases} R_k^* = f \sum\limits_{j=1}^{6} m_j \left(\frac{1}{\Delta_{kj}} - \frac{x_k x_j + y_k y_j + z_k z_j}{r_j^3} \right), \ j \neq k, \\ \Delta_{kj}^2 = (x_j - x_k)^2 + (y_j - y_k)^2 + (z_j - z_k)^2, \\ r_j^2 = x_j^2 + y_j^2 + z_j^2, \\ k = 1, 2, ..., 7. \end{cases} \tag{12}$$

To determine ω we will carry out coordinate transformation that would exclude from the right-hand sites of the equations (11) the time t:

$$\begin{cases} x_j = X_j \cos(\omega t) - Y_j \sin(\omega t), \\ y_j = X_j \sin(\omega t) + Y_j \cos(\omega t), \\ z_j = Z_j. \end{cases} \tag{13}$$

Since we study the planar configuration, we have $z_j = 0$, $j = 0, 1, ..., 7$. In the new coordinates the equations (11), written only for the bodies $P_0, P_1, P_2, P_3, P_4, P_5, P_6$, have the form:

$$\begin{cases} \frac{d^2X_k}{dt^2} = \omega^2 X_k + 2\omega \frac{dY_k}{dt} - \frac{f(m_0+m_k)X_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k}, \\ \frac{d^2Y_k}{dt^2} = \omega^2 Y_k - 2\omega \frac{dX_k}{dt} - \frac{f(m_0+m_k)Y_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k}, \end{cases} \tag{14}$$

$$\begin{cases} R_k^* = f\sum\limits_{j=1}^{6} m_j \left(\frac{1}{\Delta_{kj}} - \frac{X_k X_j + Y_k Y_j}{r_j^3} \right), \ j \neq k, \\ \Delta_{kj}^2 = (X_j - X_k)^2 + (Y_j - Y_k)^2, \\ r_j^2 = X_j^2 + Y_j^2, \\ k = 1, 2, ..., 6. \end{cases} \tag{15}$$

The differential equation system (14) describes the motion of the material system of the points in the coordinate system with rotation. For its further study, it is necessary to replace it with an equivalent system of differential equations written in coordinates of the coordinate rotation phase space and four-velocity gears. We canonize the equations of motion. Thus, we will first introduce the additional variables - the unitary impulses, applying the formulas of classical mechanics:

$$\frac{dX}{dt} = p_x, \quad \frac{dY}{dt} = p_y.$$

Then in the new phase coordinates (X, Y, p_x, p_y) the differential equations describing the motion of the seven bodies in that system will have the standard Cauchy form:

$$\begin{cases} \frac{dX_k}{dt} = p_{kx}, \quad \frac{dY_k}{dt} = p_{ky}, \\ \frac{dp_{kx}}{dt} = \omega^2 X_k + 2\omega p_{ky} - \frac{f(m_0+m_k)X_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k}, \\ \frac{dp_{ky}}{dt} = \omega^2 Y_k - 2\omega p_{kx} - \frac{f(m_0+m_k)Y_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k}, \\ k = 1, 2, ..., 6. \end{cases} \tag{16}$$

As the stationary points of the system (16) are searched, then according to their definition, they must be solutions of the system of functional equations:

$$\begin{cases} p_{kx} = 0, \ p_{ky} = 0, \\ \omega^2 X_k + 2\omega p_{ky} - \frac{f(m_0+m_k)X_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k} = 0, \\ \omega^2 Y_k - 2\omega p_{kx} - \frac{f(m_0+m_k)Y_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k} = 0, \\ k = 1, 2, ..., 6. \end{cases} \tag{17}$$

So the determination of the stationary solutions, the equilibrium points, of the Newtonian problem of the seven bodies (14) is reduced to the establishment of all real solutions of the nonlinear system of functional equations (17). This system is equivalent to the system of algebraic equations:

$$\begin{cases} \omega^2 X_k + \frac{f(m_0+m_k)X_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k} = 0, \\ \omega^2 Y_k + \frac{f(m_0+m_k)Y_k}{r_k^3} + \frac{\partial R_k^*}{\partial X_k} = 0, \end{cases} \tag{18}$$

which is made up of 12 algebraic equations and has 20 unknowns, if we consider as unknowns the coordinates X_k, Y_k of the gravitating bodies, their masses m_0, m_1, m_2, m_3, m_4, m_5, m_6, and the angular velocity ω of the coordinate system. In these conditions, the problem of determining the stationary solutions of the Newtonian problem of several bodies becomes, from one point of view, incorrect, and from the other point of view admits multiple aspects and interpretations in its study. As a most simple case we consider that the coordinates of the bodies, the geometric parameters of the configuration, are known and then the mathematical problem is essentially simplified: the system (18) becomes a linear algebraic system with respect to the masses m_0, m_1, m_2, m_3, m_4, m_5, and for solving it we can apply the methods of linear algebra. In our case, not all the coordinates of the bodies are known and the system (18) becomes non-linear with respect to the unknown coordinates.

To simplify the studied problem, we consider $P_1(1,1)$, $P_2(-1,1)$, $P_3(-1,-1)$, $P_4(1,-1)$, $P_5(\alpha,\alpha)$, $P_6(-\alpha,-\alpha)$, $f=1$, $m_0=1$, $m_5=m_6$. Then, replacing these data in the system (18) and solving it, by applying the symbolic calculus system Mathematica, we obtain:

$$\begin{gathered} m_1 = m_3,\, m_2 = m_4 = f_1(m_1,\alpha)\,, \\ m_5 = m_6 = f_2(m_1,\alpha)\,,\, \omega^2 = f_3(m_1,\alpha)\,. \end{gathered} \tag{19}$$

Theorem 2.1. *The verification of relations* (19) *represents the sufficient condition of existence of the homographic solution of the Newtonian problem of seven bodies, the its configuration of which represents a square $P_1P_2P_3P_4$ with one of the bodies (P_0) located in the origin of the coordinates, and the other two P_5, P_6 are located of the diagonal P_1P_3.*

The functional dependences $m_2 = m_4 = f_1(m_1,\alpha)$, $m_5 = m_6 = f_2(m_1,\alpha)$, $\omega^2 = f_3(m_1,\alpha)$ can be seen from graphs obtained with the graphical package of Mathematica (see Fig. 2–4).

Intervals of admissible values for the parametr α are determined by the conditions $m_2 = m_4 > 0$, $m_5 = m_6 > 0$ and $\omega^2 > 0$. In the Table 1 are displayed the admissible intervals of α according to some values of m_1, approximately calculated using the graphical tools of Mathematica.

Let's study the motion of the body $P_7(x_7, y_7, z_7)$ which gravitates passively in the field of other bodies. In the studied model $m_7 = \mu$ and we can assume

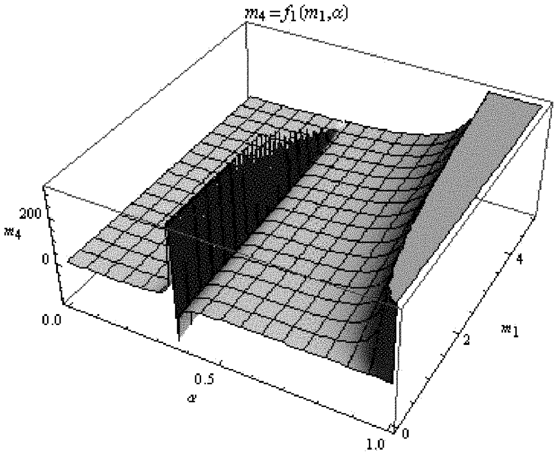


Fig. 1. Representation of the functional dependence $m_4 = f_1\,(m_1, \alpha)$

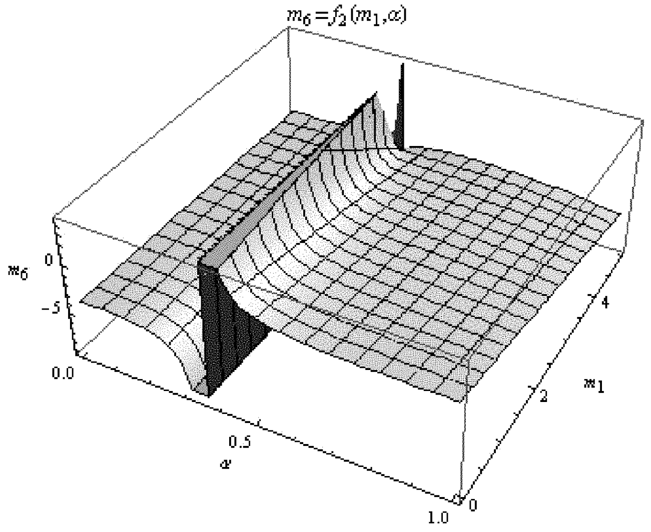


Fig. 2. Representation of the functional dependence $m_6 = f_2\,(m_1, \alpha)$

that $\mu = 0$. For simplicity it will be considered further $P_7(x_7, y_7, z_7) \equiv P(x, y, z)$. Then the equations of the point motion $P(x, y, z)$ have the form:

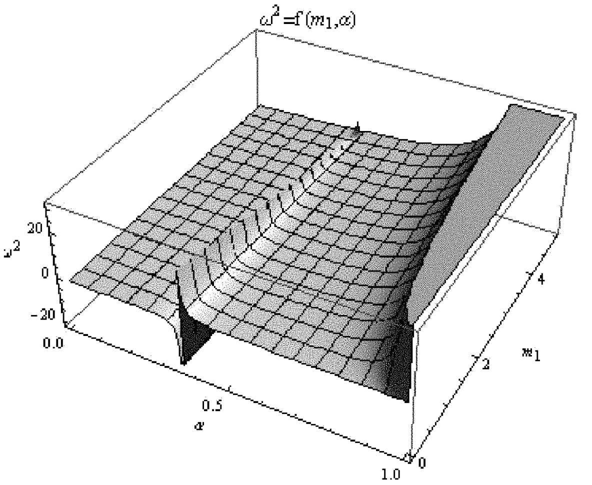


Fig. 3. Representation of the functional dependence $\omega^2 = f_3(m_1, \alpha)$

Table 1 The admissible intervals for α

m_1	Intervals allowed for α
0.0001	———————
0.001	———————
0.01	$(0.8582; 0.85857)$
0.1	$(0.715; 0.718)$
1	$(0.48965; 0.5053)$
10	$(0.291; 0.320)$
100	$(0.149; 0.2871)$
1000	$(0.050; 0.2838)$

$$\begin{cases} \dfrac{d^2x}{dt^2} + \dfrac{fm_0x}{r^3} = \dfrac{\partial R}{\partial x}, \\ \dfrac{d^2y}{dt^2} + \dfrac{fm_0y}{r^3} = \dfrac{\partial R}{\partial y}, \\ \dfrac{d^2z}{dt^2} + \dfrac{fm_0z}{r^3} = \dfrac{\partial R}{\partial z}, \end{cases} \tag{20}$$

where

$$\left\{\begin{array}{l} R = f\sum\limits_{j=1}^{6} m_j\left(\dfrac{1}{\Delta_{kj}} - \dfrac{xx_j + yy_j + zz_j}{r_j^3}\right), \\ \Delta_j^2 = \left(x_j - x\right)^2 + \left(y_j - y\right)^2 + \left(z_j - z\right)^2, \\ r_j^2 = x_j^2 + y_j^2 + z_j^2,\ r^2 = x^2 + y^2 + z^2. \end{array}\right. \tag{21}$$

The system (20) is non-autonomous. Performing the transformation

$$\left\{\begin{array}{l} x = X\cos\left(\omega t\right) - Y\sin\left(\omega t\right), \\ y = X\sin\left(\omega t\right) + Y\cos\left(\omega t\right), \\ z = Z. \end{array}\right. \tag{22}$$

we obtain an autonomous system of differential equations

$$\left\{\begin{array}{l} \dfrac{d^2X}{dt^2} = \omega^2 X + 2\omega\dfrac{dY}{dt} - \dfrac{fm_0X}{r^3} + \dfrac{\partial R}{\partial X}, \\ \dfrac{d^2Y}{dt^2} = \omega^2 Y - 2\omega\dfrac{dX}{dt} - \dfrac{fm_0Y}{r^3} + \dfrac{\partial R}{\partial Y}, \\ \dfrac{d^2Z}{dt^2} + \dfrac{fm_0Z}{r^3} = \dfrac{\partial R}{\partial Z}, \end{array}\right. \tag{23}$$

where

$$\left\{\begin{array}{l} R = f\sum\limits_{j=1}^{6} m_j\left(\dfrac{1}{\Delta_{kj}} - \dfrac{XX_j + YY_j + ZZ_j}{r_j^3}\right), \\ \Delta_j^2 = \left(X_j - X\right)^2 + \left(Y_j - Y\right)^2 + \left(Z_j - Z\right)^2, \\ r_j^2 = X_j^2 + Y_j^2 + Z_j^2,\ r^2 = X^2 + Y^2 + Z^2, \\ j = 1, 2, ..., 6, \end{array}\right. \tag{24}$$

and (X_j, Y_j, Z_j) are the previously determined coordinates of the bodies $P_1, P_2, P_3, P_4, P_5, P_6$. We have $Z_j = 0$. We will introduce the new phase coordinates x, y, z, u, v, w where

$$\left\{\begin{array}{l} x = X,\ y = Y,\ z = Z, \\ \dfrac{dx}{dt} = u,\ \dfrac{dy}{dt} = v,\ \dfrac{dz}{dt} = w, \end{array}\right. \tag{25}$$

to bring the system (23) to the normal Cauchy form. Thus in the new coordinates the system (23) has the form:

$$\begin{cases} \dfrac{dx}{dt} = u, \ \dfrac{dy}{dt} = v, \ \dfrac{dz}{dt} = w, \\ \dfrac{du}{dt} = \omega^2 x + 2\omega v - \dfrac{fm_0 x}{r^3} + \dfrac{\partial R}{\partial x}, \\ \dfrac{dv}{dt} = \omega^2 y - 2\omega u - \dfrac{fm_0 y}{r^3} + \dfrac{\partial R}{\partial y}, \\ \dfrac{dw}{dt} = \dfrac{\partial R}{\partial z}. \end{cases} \tag{26}$$

According to the definition of the stationary solutions of the differential equations, the equilibrium positions (in case when they exist) are solutions of the functional system of equations:

$$\begin{cases} u = 0, \ v = 0, \ w = 0, \\ \omega^2 x + 2\omega v - \dfrac{fm_0 x}{r^3} + \dfrac{\partial R}{\partial x} = 0, \\ \omega^2 y - 2\omega u - \dfrac{fm_0 y}{r^3} + \dfrac{\partial R}{\partial y} = \dfrac{\partial R}{\partial z} = 0, \end{cases} \tag{27}$$

or in deployed form

$$\begin{cases} u = 0, \ v = 0, \ w = 0, \\ \omega^2 x + 2\omega v - \dfrac{fm_0 x}{r^3} - f\sum\limits_{j=1}^{6} m_j \left(\dfrac{x - x_j}{\Delta_j^3} + \dfrac{x_j}{r_j^3} \right) = 0, \\ \omega^2 y - 2\omega u - \dfrac{fm_0 y}{r^3} - f\sum\limits_{j=1}^{6} m_j \left(\dfrac{y - y_j}{\Delta_j^3} + \dfrac{y_j}{r_j^3} \right) = 0, \\ -\dfrac{fm_0 z}{r^3} - f\sum\limits_{j=1}^{6} m_j \left(\dfrac{z - z_j}{\Delta_j^3} + \dfrac{z_j}{r_j^3} \right) = 0, \end{cases} \tag{28}$$

$$\begin{cases} \Delta_j^2 = (x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2, \\ r_j^2 = x_j^2 + y_j^2 + z_j^2, \ r^2 = x^2 + y^2 + z^2, \\ j = 1, 2, ..., 6. \end{cases} \tag{29}$$

For simplicity as above it has been taken $f = 1$, $m_0 = 1$. Replacing in relations (28) (X_j, Y_j, Z_j), $m_2 = m_4 = f_1(m_1, \alpha)$, $m_5 = m_6 = f_2(m_1, \alpha)$ and $\omega^2 = f_3(m_1, \alpha)$, determined above for admissible α and m_1 , we obtain the following system:

$$\left\{\begin{array}{l}
u=0,\ v=0,\ w=0,\\
f(x,y)=\omega^2x+2\omega v-\dfrac{x}{(x^2+y^2)^{\frac{3}{2}}}+\\
+m_1\left(\dfrac{-1-x}{\big((1+x)^2+(1+y)^2\big)^{3/2}}+\dfrac{1-x}{\big((1-x)^2+(1-y)^2\big)^{3/2}}\right)+\\
+m_4\left(\dfrac{1-x}{\big((1-x)^2+(1+y)^2\big)^{3/2}}+\dfrac{-1-x}{\big((1+x)^2+(1-y)^2\big)^{3/2}}\right)+\\
+m_6\left(\dfrac{-\alpha-x}{\big((\alpha+x)^2+(\alpha+y)^2\big)^{3/2}}+\dfrac{\alpha-x}{\big((\alpha-x)^2+(\alpha-y)^2\big)^{3/2}}\right)=0,\\
g(x,y)=\omega^2y-2\omega u-\dfrac{y}{(x^2+y^2)^{\frac{3}{2}}}+\\
+m_1\left(\dfrac{-1-y}{\big((1+x)^2+(1+y)^2\big)^{3/2}}+\dfrac{1-y}{\big((1-x)^2+(1-y)^2\big)^{3/2}}\right)+\\
+m_4\left(\dfrac{1-y}{\big((1-x)^2+(1+y)^2\big)^{3/2}}+\dfrac{-1-y}{\big((1+x)^2+(1-y)^2\big)^{3/2}}\right)+\\
+m_6\left(\dfrac{-\alpha-y}{\big((\alpha+x)^2+(\alpha+y)^2\big)^{3/2}}+\dfrac{\alpha-y}{\big((\alpha-x)^2+(\alpha-y)^2\big)^{3/2}}\right)=0,
\end{array}\right. \tag{30}$$

The system (30) is reduced to solving the following system consisting of two irrational algebraic equations with the unknowns x, y:

$$\begin{cases} \omega^2 x - \dfrac{x}{(x^2+y^2)^{\frac{3}{2}}} + m_1\left(\dfrac{-1-x}{\left((1+x)^2+(1+y)^2\right)^{3/2}} + \dfrac{1-x}{\left((1-x)^2+(1-y)^2\right)^{3/2}}\right) + \\ +m_4\left(\dfrac{1-x}{\left((1-x)^2+(1+y)^2\right)^{3/2}} + \dfrac{-1-x}{\left((1+x)^2+(1-y)^2\right)^{3/2}}\right) + \\ +m_6\left(\dfrac{-\alpha-x}{\left((\alpha+x)^2+(\alpha+y)^2\right)^{3/2}} + \dfrac{\alpha-x}{\left((\alpha-x)^2+(\alpha-y)^2\right)^{3/2}}\right) = 0, \\ \omega^2 y - \dfrac{y}{(x^2+y^2)^{\frac{3}{2}}} + m_1\left(\dfrac{-1-y}{\left((1+x)^2+(1+y)^2\right)^{3/2}} + \dfrac{1-y}{\left((1-x)^2+(1-y)^2\right)^{3/2}}\right) + \\ +m_4\left(\dfrac{1-y}{\left((1-x)^2+(1+y)^2\right)^{3/2}} + \dfrac{-1-y}{\left((1+x)^2+(1-y)^2\right)^{3/2}}\right) + \\ +m_6\left(\dfrac{-\alpha-y}{\left((\alpha+x)^2+(\alpha+y)^2\right)^{3/2}} + \dfrac{\alpha-y}{\left((\alpha-x)^2+(\alpha-y)^2\right)^{3/2}}\right) = 0, \end{cases} \tag{31}$$

Theorem 2.2. *Determining condition for the existence of the solutions of the system* (31) *represents the necessary and sufficient condition for the existence of the stationary solutions to the restricted eight body problem.*

The equations in the system (31) have a rather complicated structure. Its solving is quite cumbersome. If the solution of the system (31) will to be determined then by adding $u = v = w = 0$ it would be obtained the solution of the equilibrium position of differential equations describing the restricted problem of the eight bodies. In order to determine the solutions of the system (31), the graphical possibilities of Mathematica have been used. Using the graphical package of Mathematica for different parameter values α and m_1 have been constructed the graphs of the curves described by the equations in the system (31). Obviously, the points of intersection of these curves in the plan P_0xy will be the equilibrium positions of the investigated system.

For example, for $m_1 = 0.01$ and $\alpha = 0.8585$ the graphs of these curves are represented in Fig.5. In this graph the axis of the abscissae is the axis P_0x, the axis of the right-order P_0y. The system solutions (31) represent the intersection points of the curves in the drawing. We see that the coordinate axes themselves verify the equations of the given system: the points of the P_0x axis of the second equation of the totality, and the points of the axes of

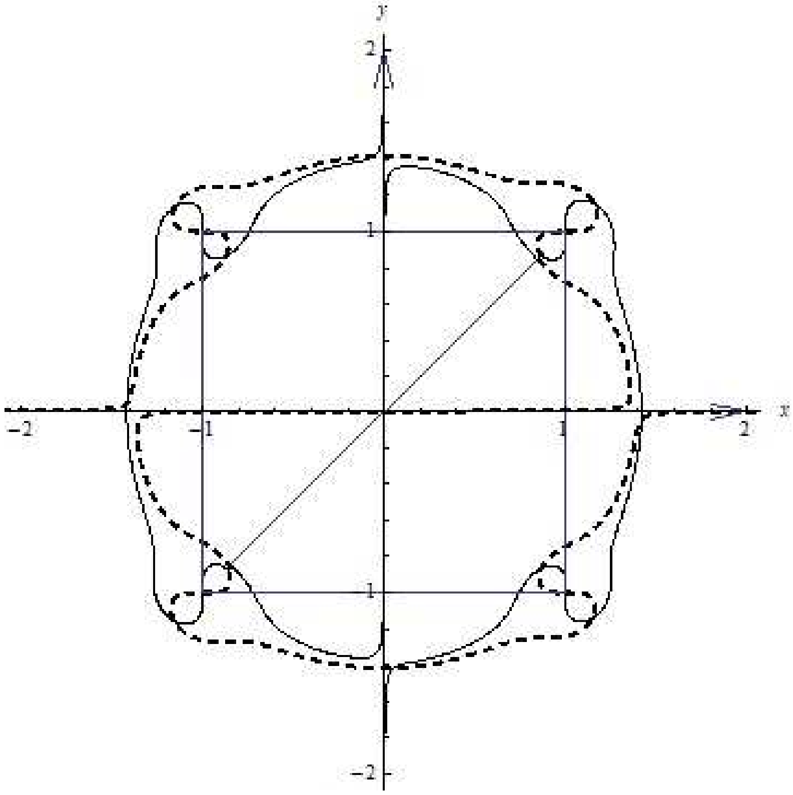


Fig. 4. Representation of the curves f, g for $m_1 = 0.01$, $\alpha = 0.8585$.

the P_0y - the first equation. In total we have 12 points of equilibrium. We will name the points that are on the straight curves passing through the center of the configuration and any peak of the square radial equilibrium position (we will note them in the future by N_i). We will name the other points as equilibrium bisectorial positions (we will note them in the future through S_i).

In figures 5-7 are represented graphical solutions of the system (31) for certain admissible values, previously determined, of the parameters m_1 and α.

The graphical method of solving the systems gives the possibility to determine the approximate values of the equilibrium positions using the *FindRoot* routine. For the aforementioned graphs the coordinates of all equilibrium points were calculated with a fairly high accuracy. The Table 2 below contains the coordinates of these points with the accuracy of up to five digits after the comma.

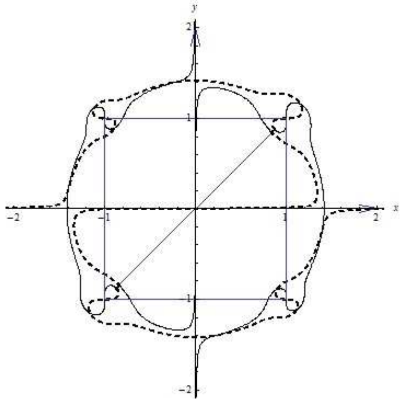


Fig. 6: Representation of the curves f, g for$m_1 = 0.01$, $\alpha = 0.8583$

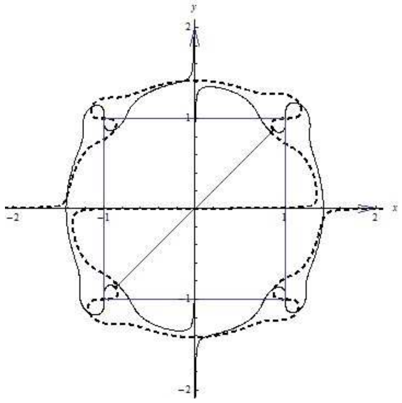


Fig. 7: Representation of the curves f, g for $m_1 = 0.01$, $\alpha = 0.8584$

Table 2 The coordinates of equilibrium points

m_1	α	N_1		S_1	
—-		x^*	y^*	x^*	y^*
0.01	0.8583	1.15589	1.15589	1.39868	−0.22286
0.01	0.8584	1.15597	1.15597	1.41168	−0.12379
0.01	0.8585	1.15604	1.15604	1.41684	−0.05223
0.01	0.85853	1.15606	1.15606	1.41760	−0.03417
0.1	0.715	1.34188	1.34188	1.34865	−0.45766
0.1	0.717	1.34324	1.34324	1.44139	−0.11335
1	0.48965	1.63351	1.63351	0.93934	−1.05917
1	0.505	1.66022	1.66022	1.82285	−0.00771
10	0.291	1.84521	1.84521	2.19692	−0.00052
100	0.2	1.82945	1.82945	0.82914	−0.02594
1000	0.2	1.81083	1.81083	2.10424	−0.05038

3. ABOUT THE STABILITY AND LINEAR INSTABILITY OF THE STATIONARY POINTS

To study the stability of the points N_i, S_i by the first method of A.M. Lyapunov it is necessary to linearize the system of the differential equation

(26) in the neighborhood of each stationary point N_i, S_i. In advance, the equations of motion of the point that passively gravitates must be written in the normal Cauchy form . For this we will introduce instead of the three-dimensional space $\{x, y, z\}$ the phase space of the sixth dimension after the formulas

$$x = X, \quad y = Y, \quad z = Z, \quad \frac{dx}{dt} = u, \quad \frac{dy}{dt} = v, \quad \frac{dz}{dt} = w.$$

In the new coordinates the system (26) takes the form:

$$\begin{cases} \frac{dx}{dt} = u, \ \frac{dy}{dt} = v, \ \frac{dz}{dt} = w, \\ \frac{du}{dt} = \omega^2 X + 2\omega v - \frac{fm_0 X}{r^3} + \frac{\partial R}{\partial X}, \\ \frac{dv}{dt} = \omega^2 Y - 2\omega u - \frac{fm_0 Y}{r^3} + \frac{\partial R}{\partial Y}, \\ \frac{dw}{dt} = \frac{\partial R}{\partial Z}, \end{cases} \tag{32}$$

where

$$\begin{cases} R = f\sum_{j=1}^{6} m_j \left(\frac{1}{\Delta_j} - \frac{XX_j + YY_j + ZZ_j}{r_j^3} \right), \\ \Delta_j^2 = (X_j - X)^2 + (Y_j - Y)^2 + (Z_j - Z)^2, \\ r_j^2 = X_j^2 + Y_j^2 + Z_j^2, \ r^2 = X^2 + Y^2 + Z^2, \\ j = 1, 2, ..., 6, \end{cases} \tag{33}$$

$(X_j, Y_j, Z_j = 0)$ and ω^2 are those previously determined.

To study the stability of the equilibrium points of the system (32), it is necessary to investigate the properties of the eigenvalues of matrix of the linearized system in the neighborhood of each point N_i, S_i. We will note, for simplicity, the coordinates of any point N_i, S_i through $x_i^*, y_i^*, z_i^* = 0$ and by x the vector (point)

$$x = (u - u^*, v - v^*, w - w^*, x - x^*, y - y^*, z - z^*). \tag{34}$$

In the notation (34), $u^* = v^* = w^* = 0$.

The six-dimensional phase space is local, therefore each of the N_i and S_i equilibrium points (taken separately) represents the point $x = 0$ of this space. Performing the linearization procedure of the right-hand sites of the system (32) in the neighborhood of the phase point $x = 0$, we obtain the following system of linear differential equations:

$$\frac{dx}{dt} = Ax, \tag{35}$$

where the matrix A of the size 6×6 has the form:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ a & b & 0 & 0 & 2\omega & 0 \\ b & c & 0 & -2\omega & 0 & 0 \\ 0 & 0 & d & 0 & 0 & 0 \end{pmatrix}. \tag{36}$$

Elements a, b, c, d of matrix A, calculated using Mathematica are expressed by the relationships

$$\begin{aligned}
a = \omega^2 + &\frac{2\,(x^*)^2 - (y^*)^2}{\left((x^*)^2+(y^*)^2\right)^{\frac{5}{2}}} + \\
+m_1\left(\frac{2(1-x^*)^2-(1-y^*)^2}{\left((1-x^*)^2+(1-y^*)^2\right)^{5/2}}\right) &+ m_1\left(\frac{3(1+x^*)^2-1}{\left((1+x^*)^2+(1+y^*)^2\right)^{5/2}}\right) + \\
+m_4\left(\frac{2(1-y^*)^2-(1+x^*)^2}{\left((1-x^*)^2+(1+y^*)^2\right)^{5/2}}\right) &+ m_4\left(\frac{2(1+y^*)^2-(1-x^*)^2}{\left((1+x^*)^2+(1-y^*)^2\right)^{5/2}}\right) + \\
+m_6\left(\frac{2(\alpha-x^*)^2-(\alpha-y^*)^2}{\left((\alpha-x^*)^2+(\alpha-y^*)^2\right)^{5/2}}\right) &+ m_6\left(\frac{2(\alpha+x^*)^2-(\alpha+y^*)^2}{\left((\alpha+x^*)^2+(\alpha+y^*)^2\right)^{5/2}}\right) +
\end{aligned} \tag{37}$$

$$\begin{aligned}
b = &\frac{3x^*y^*}{\left((x^*)^2+(y^*)^2\right)^{\frac{5}{2}}} + \\
+m_1\frac{3(1-x^*)(1-y^*)}{\left((1-x^*)^2+(1-y^*)^2\right)^{5/2}} &+ m_1\frac{3(1+x^*)(1+y^*)}{\left((1+x^*)^2+(1+y^*)^2\right)^{5/2}} + \\
-m_4\frac{3(1+x^*)(1-y^*)}{\left((1+x^*)^2+(1-y^*)^2\right)^{5/2}} &- m_4\frac{3(1-x^*)(1+y^*)}{\left((1-x^*)^2+(1+y^*)^2\right)^{5/2}} + \\
+m_6\frac{3(\alpha-x^*)(\alpha-y^*)}{\left((\alpha-x^*)^2+(\alpha-y^*)^2\right)^{5/2}} &+ m_6\frac{3(\alpha+x^*)(\alpha+y^*)}{\left((\alpha+x^*)^2+(\alpha+y^*)^2\right)^{5/2}}\,;
\end{aligned} \tag{38}$$

$$\begin{aligned}
c = \omega^2 + &\frac{2\,(y^*)^2 - (x^*)^2}{\left((x^*)^2+(y^*)^2\right)^{\frac{5}{2}}} + \\
+m_1\left(\frac{2(1-y^*)^2-(1-x^*)^2}{\left((1-x^*)^2+(1-y^*)^2\right)^{5/2}}\right) &+ m_1\left(\frac{3(1+y^*)^2-1}{\left((1+x^*)^2+(1+y^*)^2\right)^{5/2}}\right) +
\end{aligned}$$

$$+m_4\left(\frac{3(1-y^*)^2-(1+y^*)^2-(1-x^*)^2}{\left((1-x^*)^2+(1+y^*)^2\right)^{5/2}}\right)+$$

$$+m_4\left(\frac{3(1+y^*)^2-(1-y^*)^2-(1+x^*)^2}{\left((1+x^*)^2+(1-y^*)^2\right)^{5/2}}\right)+$$

$$+m_6\left(\frac{2(\alpha-y^*)^2-(\alpha-x^*)^2}{\left((\alpha-x^*)^2+(\alpha-y^*)^2\right)^{5/2}}\right)+m_6\left(\frac{2(\alpha+y^*)^2-(\alpha+x^*)^2}{\left((\alpha+x^*)^2+(\alpha+y^*)^2\right)^{5/2}}\right);\qquad(39)$$

$$d=-\frac{1}{\left((x^*)^2+(y^*)^2\right)^{3/2}}-$$

$$-m_1\left(\frac{1}{\left((1+x^*)^2+(1+y^*)^2\right)^{3/2}}+\frac{1}{\left((1-x^*)^2+(1-y^*)^2\right)^{3/2}}\right)-$$

$$-m_4\left(\frac{1}{\left((1+x^*)^2+(1-y^*)^2\right)^{3/2}}+\frac{1}{\left((1-x^*)^2+(1+y^*)^2\right)^{3/2}}\right)+$$

$$-m_6\left(\frac{1}{\left((\alpha-x^*)^2+(\alpha-y^*)^2\right)^{3/2}}+\frac{1}{\left((\alpha+x^*)^2+(\alpha+y^*)^2\right)^{3/2}}\right).\qquad(40)$$

Expressions (37)-(40) depend on the values x_i^*, y_i^*, therefore for each equilibrium position the values of the elements a, b, c, d of the matrix A will be different.

The characteristic equation from which the eigenvalues of the matrix A are determined has the form:

$$det\,(A-\lambda E)=\left(\lambda^2-d\right)\left(\lambda^4+\left(4\omega^2-a-c\right)\lambda^2+ac-b^2\right)=0.\qquad(41)$$

Let's examine the roots of the equation in more details (41). From the relation (40) we can see that always $d<0$. From the equation

$$\lambda^2-d=0\qquad(42)$$

we obtain that as two eigenvalues of the matrix A will always be imaginary. We will note them by λ_5, λ_6. Let's search the equation now

$$\lambda^4+\left(4\omega^2-a-c\right)\lambda^2+ac-b^2=0.\qquad(43)$$

Each equilibrium position is stable if all four solutions of equation (43) are imaginary. The problem at hand is then reduced to the determination of those values of the parameters m_1 and α for which the elements a, b, c, d of the matrix A receive such values that the roots of the bipatratic equation (43) are purely imaginary (in the future we will note these roots by $\lambda_1, \lambda_2, \lambda_3, \lambda_4$). The structure of the relationships (37) - (40) is rather complicated. The determination of an analytical expression that would express the dependence between of the eigenvalues of the matrix A and the parameters m_1 and α is virtually impossible.

Table 3 below contains the values $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ for stationary points N_1 and S_1. Analyzing the Table 3 we notice that for the stationary point N_1, varying the values of the parameters m_1 and α the eigenvalues of the matrix A are not purely imaginary. The same result is obtained for the other N_i points. It follows that stationary N_i points are unstable in the first approximation. We will formulate this result by the theorem:

Table 3 The eigenvalues of the matrix A

m_1	α	N_1		S_1	
—-		λ_1, λ_2	λ_3, λ_4	λ_1, λ_2	λ_3, λ_4
0.01	0.858	± 1.309	$\pm 1.123\,i$	$\pm 0.284\,i$	$\pm 0.518\,i$
0.01	0.858	± 1.307	$\pm 1.122\,i$	$\pm 0.494\,i$	$\pm 0.322\,i$
0.01	0.858	± 1.306	$\pm 1.122\,i$	$\pm 0.459\,i$	$\pm 0.369\,i$
0.01	0.858	± 1.306	$\pm 1.121\,i$	$\pm 0.004 + 0.36\,i$	$\pm 0.004 - 0.36\,i$
0.1	0.715	± 1.191	$\pm 1.067\,i$	$\pm 0.34 + 0.53\,i$	$\pm 0.34 - 0.53\,i$
0.1	0.717	± 1.178	$\pm 1.060\,i$	$\pm 0.407 + 0.56\,i$	$\pm 0.407 - 0.56\,i$
1	0.489	± 1.367	$\pm 1.306\,i$	$\pm 0.74 + 0.82\,i$	$\pm 0.74 - 0.82\,i$
1	0.505	± 1.233	$\pm 1.128\,i$	$\pm 0.75 + 0.83\,i$	$\pm 0.75 - 0.83\,i$
10	0.291	± 2.503	$\pm 2.630\,i$	$\pm 1.66 + 1.88\,i$	$\pm 1.66 - 1.88\,i$
100	0.2	± 8.226	$\pm 8.568\,i$	± 15.312	$\pm 8.390\,i$
1000	0.2	± 27.15	$\pm 28.07\,i$	$\pm 17.76 + 19.89\,i$	$\pm 17.76 - 19.89\,i$

Theorem 3.1. *The radial equilibrium points N_i of the differential equations describing the restricted eight body problem are unstable in the first approximation for any values of the parameters m_1 and α.*

From Table 3 we can see that in the equilibrium point S_1 for certain values of the parameters m_1 and α the eigenvalues of the matrix A are purely imaginary. Hence this stationary point S_1 is stable in the first approximation. Similarly, similar results are obtained for other points of type S_i.

Theorem 3.2. *There are values of the parameters m_1 and α for which the bisectorial stationary points S_i in the restricted eight body problem are stable in the first approximation.*

Concluding remarks

We have determined sufficient conditions for the existence of the linear stable configurations describing the restricted Newtonian eight body problem. We have used some built in functions of the Mathematica programming environment in order to determine the stationary points. Their linear stability has been studied. It has been demonstrated that there are values of the parameters α and m_1 for which the bisectorial stationary points are stable in the first approximation.

References

[1] E. A. Grebenikov, *Matematicheskie problemi gomograficheskoi dinamiki. (Russian) [On a mathematical problem of homographic dynamics]* , MAKS Press, Moscow, 2010.

[2] E. A. Grebenikov, D. Kozak-Skovorodkina, M. Yakubyak, *Metody komp'yuternoi algebry v probleme mnogikh tel. (Russian) [Computer algebra methods in the many-body problem]* Second edition, revised and supplemented. Izdatel'stvo Rossiiskogo Universiteta Druzhby Narodov, Moscow, 2002. 211 pp. ISBN: 5-209-01451-7

[3] V. Szebehely, *Teoria orbit. (Russian) [Theory of Orbit]*, Nauka, Moscow, 1982.

[4] V.K. Abalakin, V.P. Aksenov, E. A. Grebenikov, V.G. Demin, Iu.A. Ryabov, *Spravochnoe rukovodstvo po nebesnoi mehaniki i astrodinamiki. (Russian) [Reference manual on celestial mechanics and astrodynamics]*, Nauka, Moscow, 1976.

[5] A. Wintner, *The Analytical Foundations of Celestial Mechanics.* Princeton University Press, Priceton, 1941.

[6] S. Wolfram, *The Mathematica Book.* University Press, Cambridge, 1996.

[7] N. S. Bakhvalov, *Numerical Methodes.* Nauka, Moscow, 1973.

ROMAI J., v.14, no.1(2018), 63–72

A COMPARISON BETWEEN AKIMA AND HERMITE TYPE CUBIC SPLINE WITH MINIMAL QUADRATIC OSCILLATION IN AVERAGE

Larisa Cheregi

Technical University of Cluj-Napoca, Cluj-Napoca, Romania

cheregi.larisa@yahoo.com

Abstract A comparison between Akima and Hermite type cubic spline is presented. Some numerical examples are provided to illustrate the satisfactory shape of the interpolation curves using an example from diabetology and a design construct of aerodynamic profiles.

Keywords: curves, approximation, cubic interpolatory spline, Akima, Hermite polynomials, comparison of the interpolation methods.

2010 MSC: 65D05, 65D18, 65D07.

1. INTRODUCTION

In this paper are presented algorithms for implementation of spline functions and will be graphically illustrated by a program implemented for Hermite type cubic spline with minimal quadratic oscillation and Akima interpolation method. This numerical method is implemented here to obtain a proper soft which is tested and used on diabetology measurements.

In Section 2 we optimize Hermite type cubic spline and Akima interpolation method. The Hermite interpolation consists in determining a polynomial (as small as possible) passing through the points $(x_i, f(x_i))$ with derivatives of a given order (evaluated in nodes) one with the values of the derivatives of the same order of function $f(x)$ in these points ([2],[6]). A method of Hermite interpolation was presented in [3] by minimization of the quadratic oscillation in average. In the method, we do not need to choose parameters ([1]), but the obtained cubic interpolation curves may change from one side of the linear interpolation curve to the other side([7],[8]). Minimization of the quadratic oscillation in average [3] is considered a measure of the deviation of an interpolation function of points $(x_i, y_i), \forall i = \overline{0, n}$ by the polygonal line joining these points, and these deviations are called oscillations. These deviations can be applied to any type of spline function.

The Akima method is based on piecewise polynomials with conditions imposed at the data points([5]). The function passes through the points with specified slopes. These slopes are determined by a local procedure, at a given datum point. A method was presented in [4].

The computing method is given in Section 3. The last section contains some numerical examples which illustrate the obtained theoretical results.

2. OPTIMIZATION PROCESSES

2.1. OPTIMIZING HERMITE TYPE CUBIC SPLINE WITH MINIMAL OSCILLATION

For a partition Δ of an interval $[a, b]$,

$$\Delta : a = x_0 < x_1 < \cdots < x_n = b,$$

on which we will build a Hermite interpolation procedure based on expression

$$S(x) = (1-s)^3 y_i + 3(1-s)^2 s(y_i + \frac{h_i}{3} m_i) + 3(1-s)s^2(y_{i+1} - \frac{h_i}{3} m_{i+1}) + s^3 y_{i+1}, \tag{1}$$

$\forall i = \overline{1, n-1}$ where $h_i = x_i - x_{i-1}, s = \frac{x-x_i}{h_i}$, $\forall i = \overline{1,n}$ and $y_0, y_1, ..., y_n \in \mathbb{R}$ are given values and the derivatives m_i remain to be determined. It is known that $S(x_i) = y_i$, $S‘(x_i) = m_i$, $\forall i = \overline{1,n}$ Let us consider the functionals

$$S_k = \int_a^b \left[S^{(k)}(x) - H^{(k)}(x)\right]^2 dx = \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left[S^{(k)}(x) - H_i^{(k)}(x)\right]^2 dx, \forall k = \overline{0,2} \tag{2}$$

where $H(x) = H_i(x)$ for $x \in [x_i, x_{i+1}]$ and $H_i = (1-s)y_i + sy_{i+1}$,$s = \frac{x-x_i}{h_i}$. Different to (1), in [9], the methods of determining the derivatives m_i were presented by minimizing

$$H_k = \int_a^b \left[S^{(k)}(x)\right]^2 dx \tag{3}$$

The motive of this paper is to present a method by minimizing S_k, for $k = 1$ which is the simplest shape-preserving interpolant. The method for obtaining the values derivatives can be viewed as an optimal alternative to the natural cubic spline method and the method in [3].

As submited in [3], the method by minimizing S_0 presents the cubic interpolating spline with minimal quadratic oscillation in average. We can discuss the method of minimizing S_1. The method by minimizing S_1 presents the cubic spline interpolant so that $S`(x)$ is the optimal approximation of $H`(x)$. $S`(x)$ represents the monotonicity of the interpolant, $S_i`(x)$ represents the monotonicity of the given data for $x \in [x_i, x_{i+1}]$. We deduce that $S(x)$ has minimal derivative oscillation to $H`(x)$ by choosing all $m_i$, so we choose all $m_i$ in (1) so that $S`(x)$ optimally approximates $H`(x)$.

For $x \in [x_i, x_{i+1}]$, $\forall i = \overline{1, n-1}$, $h_i = x_i - x_{i-1}, s = \frac{x-x_i}{h_i} \in [0,1]$, $\forall i = \overline{1, n}$

$$S`(x) - \frac{y - y_i}{h_i} = (1-s)^2(m_i - \frac{y-y_i}{h_i}) + 2(1-s)s(\frac{2(y-y_i)}{h_i} - m_i - m_{i+1}) + \\ + s^2(m_{i+1} - \frac{y-y_i}{h_i}),$$

We denote

$$a = m_i - \frac{y-y_i}{h_i},$$
$$b = \frac{2(y-y_i)}{h_i} - m_i - m_{i+1},$$
$$c = m_{i+1} - \frac{y-y_i}{h_i}$$

From

$$\int_0^1 \binom{k}{i} (1-s)^{k-i} s^i ds = \frac{1}{k+1}, \forall i = \overline{0, k} \tag{4}$$

we obtain

$$S_1 = \frac{1}{5} \sum_{i=1}^{n-1} h_i [a^2 + ab + \frac{2b^2 + ac}{3} + c^2] \tag{5}$$

Hence

$$\begin{cases} \frac{\partial S_1}{\partial m_1} = \frac{h_1}{15}(4m_1 - m_2 - 3\frac{y_2 - y_1}{h_1}), \\ \frac{\partial S_1}{\partial m_i} = \frac{h_{i-1}}{15}(4m_i - m_{i-1} - 3\frac{y_i - y_{i-1}}{h_{i-1}}) + \frac{h_i}{15}(4m_i - m_{i+1} - 3\frac{y_{i+1} - y_i}{h_i}), \\ \frac{\partial S_1}{\partial m_n} = \frac{h_{n-1}}{15}(4m_n - m_{n-1} - 3\frac{y_n - y_{n-1}}{h_{n-1}}) \forall i = \overline{2, n-1} \end{cases}$$

But for minimizing we have

$$\begin{cases} \frac{\partial S_1}{\partial m_1} = 0, \\ \frac{\partial S_1}{\partial m_i} = 0, \\ \frac{\partial S_1}{\partial m_n} = 0. \end{cases}$$

Thus we obtain the system of normal equations:

$$\begin{cases} 4m_1 - m_2 = 3\frac{y_2-y_1}{h_1}, \\ m_{i-1} + \frac{h_i m_{i-1}}{4m_i+h_{i-1}+h_i} - \frac{h_i m_{i+1}}{h_{i-1}} = 3\frac{y_{i+1}-y_{i-1}}{h_{i-1}+h_i} \\ 4m_n - m_{n-1} = 3\frac{y_n-y_{n-1}}{h_{n-1}} \forall i = \overline{2, n-1} \end{cases}$$

This system is strictly diagonally dominant and therefore has unique solution. By minimizing S_2 we obtain the system of normal equations

$$\begin{cases} 2m_1 + m_2 = 3\frac{y_2-y_1}{h_1}, \\ 2m_i + \frac{h_i m_{i-1}}{4m_i+h_{i-1}+h_i} + m_{i+1} - \frac{h_i m_{i+1}}{h_{i-1}+h_i} = y_{i+1} - y_i + \frac{h_i(4y_i-3y_{i-1}-y_{i+1})}{h_{i-1}+h_i}, \\ 2m_n + m_{n-1} = 3\frac{y_n-y_{n-1}}{h_{n-1}}, \forall i = \overline{2, n-1} \end{cases}$$

On the case of parametric curves we take x_i as parametric knots and $y_i \in \mathbb{R}^2$,. We need to consider the functionals

$$S_k = \int_a^b \|S^{(k)}(x) - H_i^{(k)}(x)\|^2 dx, \forall k = \overline{0,2} \tag{6}$$

where the norm means the Euclidean norm. Let $\frac{\partial S_1}{\partial m_i}$ be the gradient of S_1 on m_i. For constructing a closed curve we consider $y_n - y_1 = 0$ and $m_n - m_1 = 0$, then by (7) we have

$$\frac{\partial S_1}{\partial m_1} + \frac{\partial S_1}{\partial m_n} = 0$$

$$\frac{h_1}{15}(4m_1 - m_2 - 3\frac{y_2 - y_1}{h_1}) + \frac{h_{n-1}}{15}(4m_n - m_{n-1} - 3\frac{y_n - y_{n-1}}{h_{n-1}}) = 0$$

$$4m_1 - m_{n-1} - \frac{h_1(m_2 + m_{n_1})}{h_1 + h_{n-1}} = 3\frac{y_2 - y_{n-1}}{h_1 + h_{n-1}}$$

Thus we obtain the we obtain the vector system of normal equations:

$$\begin{cases} 4m_1 - m_{n-1} - \frac{h_1(m_2+m_{n-1})}{h_1+h_{n-1}} = 3\frac{y_2-y_{n-1}}{h_1+h_{n-1}}, \\ 4m_i - m_{i-1} + \frac{h_i m_{i-1}}{h_{i-1}+h_i} - \frac{h_i m_{i+1}}{h_{i-1}+h_i} = 3\frac{y_{i+1}-y_{i-1}}{h_{i-1}+h_i} \\ 4m_{n-1} - m_{n-2} - \frac{h_{n-1}m_1}{h_{n-2}+h_{n-1}} + \frac{h_{n-1}m_{n-2}}{h_{n-2}+h_{n-1}} = 3\frac{y_n-y_{n-2}}{h_{n-2}+h_{n-1}} \forall i = \overline{2,n-1} \end{cases}$$

Theorem 2.1. *(See [3]) For given points $(x_i, y_i), i = \overline{0,n}$, there exists a unique cubic spline of the Hermite type having minimal quadratic oscillation in average. This cubic spline $s \in C^1[a,b]$] can be determined by using an iterative algorithm. If s interpolates a function $f \in C[a,b], f(x_i) = y_i, i = \overline{0,n}$, then its error estimation is*

$$|f(x) - s(x)| \leq \left(1 + \frac{h^3}{4\overline{h}^3}\right)\overline{\omega}(f,g), \quad \forall x \in [a,b] \tag{7}$$

where $h = max\left\{h_i : i = \overline{1,n}\right\}, \quad \overline{h} = min\left\{h_i : \overline{1,n}\right\}$ where we denote

$$\overline{\omega}(f,h) = max\left\{\overline{\omega}(f,h_i) : i = \overline{1,n}\right\}$$

and

$$\overline{\omega}(f,\delta) = sup\left\{|f(t) - f(s)| : t,s \in [a,b], |t-s| \leq \delta\right\}$$

is the uniform modulus of continuity. If $f \in C^1[a,b]$ then $|f(x) - s(x)| \leq \left(1 + \frac{h^3}{4\overline{h}^3}\right) \cdot \|f'\|_\infty h$, where $x \in [a,b]$.

For equidistant grids [4], the estimate (12) becomes

$$|f(x) - s(x)| \leq \frac{5}{4}\overline{\omega}(f,h), \quad \forall x \in [x_0, x_n]. \tag{8}$$

Optimizing at the end-points the Akima interpolation method

For a partition Δ of an interval $[a,b]$,

$$\Delta : a = x_0 < x_1 < \cdots < x_n = b,$$

$$S_k = \sum_{i=1}^{n-1} \int_{x_{i-1}}^{x_i} \left[S^{(k)}(x) - H_i^{(k)}(x)\right]^2 dx \tag{9}$$

We minimize the partial quadratic oscillation in average on the end intervals $[x_0, x_2]$ and $[x_{n-2}, x_n]$. Since the unknown derivatives m_0, m_1, m_{n-1}, m_n appear only in the intervals $[x_0, x_2]$ and $[x_{n-2}, x_n]$, respectively, we define the residual [4]

$$S_k = \sum_{i \in K} \int\limits_{x_{i-1}}^{x_i} \left[S(x) - \frac{x_i - x}{h_i} \cdot y_{i-1} - \frac{x - x_{i-1}}{h_i} \cdot y_i \right]^2 dx$$

for $K = \{1, 2, n-1, n\}$.

Theorem 2.2. *(See [4]) For given data $(x_i, y_i), i = \overline{0, n}$, and with the values $m_i, i = \overline{2, n-2}$ there are uniquely determined the values m_0, m_1, m_{n-1}, m_n.*

$$|s(x) - f(x)| \leq \left(1 + \frac{h^4}{4\underline{h}^4}\right) \cdot \overline{\omega}(f, h), \quad \forall x \in [x_0, x_2] \cup [x_{n-2}, x_n] \tag{10}$$

where $h = max\left\{h_i : i = \overline{1, n}\right\}$, $\quad \underline{h} = min\left\{h_i : \overline{1, n}\right\}$ where we denote

$$\overline{\omega}(f, \delta) = max\left\{|f(u) - f(v)| : u, v \in [a, b], |u - v| \leq h\right\}$$

is the uniform modulus of continuity. For equidistant grids the error estimate becomes

$$|s(x) - f(x)| \leq \frac{5}{4}\overline{\omega}(f, h), \quad \forall x \in [x_0, x_n]. \tag{11}$$

Since the parameters m_0, m_1, m_{n-1}, m_n appear separated in the intervals $[x_0, x_2]$ and $[x_{n-2}, x_n]$, respectively, the normal equations $\frac{\partial S}{\partial m_0} = 0$, $\frac{\partial S}{\partial m_1} = 0$, $\frac{\partial S}{\partial m_{n-1}} = 0$, $\frac{\partial S}{\partial m_n} = 0$ form two separated systems:

$$\begin{cases} \frac{\partial S}{\partial m_0} = 0 \\ \frac{\partial S}{\partial m_1} = 0 \end{cases}$$

and

$$\begin{cases} \frac{\partial S}{\partial m_{n-1}} = 0 \\ \frac{\partial S}{\partial m_n} = 0. \end{cases}$$

After elementary calculus, these systems become:

$$\begin{cases} m_0 - \frac{3}{4}m_1 = \frac{1}{4h_1}(y_1 - y_0) \\ -\frac{3h_i^3}{4(h_1^3 + h_2^3)}m_0 + m_1 = \frac{3h_2^3}{4(h_1^3 + h_2^3)}m_2 + \frac{h_1^2(y_1 - y_0)}{4(h_1^3 + h_2^3)} + \frac{h_2^2(y_2 - y_1)}{4(h_1^3 + h_2^3)} \end{cases}$$

and

$$\begin{cases} m_{n-1} - \frac{3h_n^3}{4(h_{n-1}^3+h_n^3)} m_n = \frac{3h_{n-1}^3}{4(h_{n-1}^3+h_n^3)} m_{n-2} + \\ + \frac{h_{n-1}^2}{4(h_{n-1}^3+h_n^3)} (y_{n-1} - y_{n-2}) + \frac{h_n^2}{4(h_{n-1}^3+h_n^3)} (y_n - y_{n-1}) \\ -\frac{3}{4} m_{n-1} + m_n = \frac{1}{4h_n} (y_n - y_{n-1}) \end{cases}$$

respectively. The systems (19) and (20) have unique solution, and therefore the values m_0, m_1, m_{n-1}, m_n are uniquely obtained. The Hesse matrix ([4]) of S_k is $H = \left(\frac{\partial^2 S_k}{\partial m_i^2} \right)$, $\forall i = \overline{0, n}$ having the form:

$$H = 2 \cdot \begin{pmatrix} \frac{h_1^3}{105} & \frac{h_1^3}{140} & 0 & 0\frac{3}{4} \\ \frac{h_1^3}{105} & \frac{h_1^3}{105} + \frac{h_2^3}{105} & 0 & 0 \\ 0 & 0 & \frac{h_{n-1}^3}{105} + \frac{h_n^3}{105} & -\frac{h_n^3}{140} \\ 0 & 0 & -\frac{h_n^3}{140} & -\frac{h_n^3}{105} \end{pmatrix} \qquad (12)$$

Since all the diagonal minors of H are strictly positive, we infer that the joined solution (m_0, m_1, m_{n-1}, m_n) of the systems (19) and (20) is the unique critical point of the residual S_k and minimize it. So, (m_0, m_1, m_{n-1}, m_n) is the unique minimal point of S_k .

3. ILLUSTRATION OF INTERPOLATION METHODS

We first present the performance of the proposed methods by using the experimental data in [3] and then a design construct of aerodynamic profiles to determine which function is the most efficient for increasing speed.

This numerical method is used to obtain a soft applicable in diabetology at the fitting of glycemic profile experimental data. The soft was created in Visual C#.

The usual measurements of blood-glucose levels (represented in mg/dl) were realized at seven moments: at 07:00 AM, 9:00 AM, 11:30 AM, 13:30 PM, 17:45 PM, 20:30 PM, and 23:00 PM on the first day. On the time scale, half an hour was rendered as 0.5. A translation was realized with start step at 7:00 AM, therefore the moments of measurement were translated to 0, 2, 4.5, 6.5, 10.75, 13.5 and 16. The blood-glucose levels of the subject were 93, 98, 107, 98, 97, 85, 92.

For this patient, interpolations of the quadratic spline type, cubic spline, Akima and minimal quadratic oscillation of Hermite type were made on the recorded data. In the illustration, the blue color represents the minimal square oscillation of the Hermite type, the cubic spline is represented by the red color, Akima with green, and the gray colored quadratic spline. The polygon line is black.

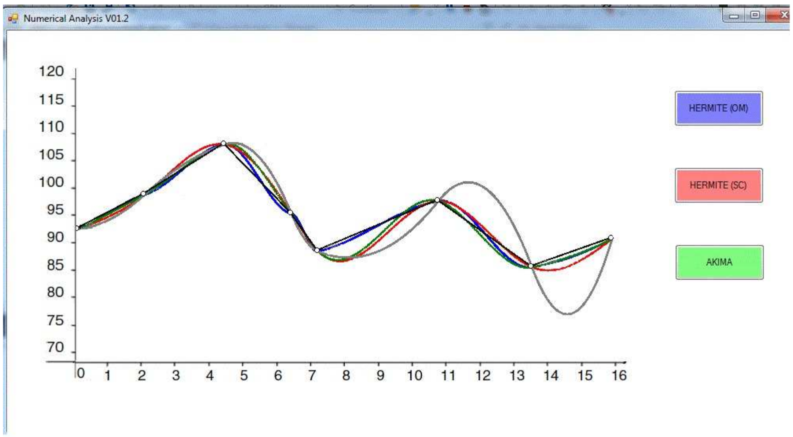


Fig. 1. Hermite (CS), Hermite (OM), Akima, Quadratic spline

Therefore, we can see that the minimal quadratic oscillation of Hermite type shows the best approximation.

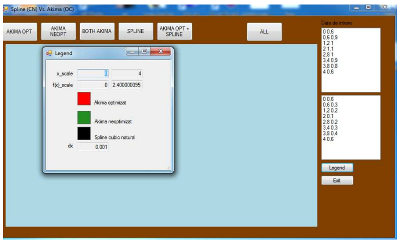
INDUSTRIAL APPLICATIONS OF SPLINE INTERFACE

We have set millimeter points to get the longitudinal section of a boat. The points are:

$x \rightarrow 0; 0.6; 1.2; 2; 2.8; 3.4; 3.8; 4,$

$y_{superior} \rightarrow 0.6; 0.9; 1; 1.1; 1; 0.9; 0.8; 0.6,$

$y_{inferior} \rightarrow 0.6; 0.3; 0.2; 0.1; 0.2; 0.3; 0.4; 0.6.$



Date de intrare

0 0,6
0,6 0,9
1,2 1
2 1,1
2,8 1
3,4 0,9
3,8 0,8
4 0,6

0 0,6
0,6 0,3
1,2 0,2
2 0,1
2,8 0,2
3,4 0,3
3,8 0,4
4 0,6

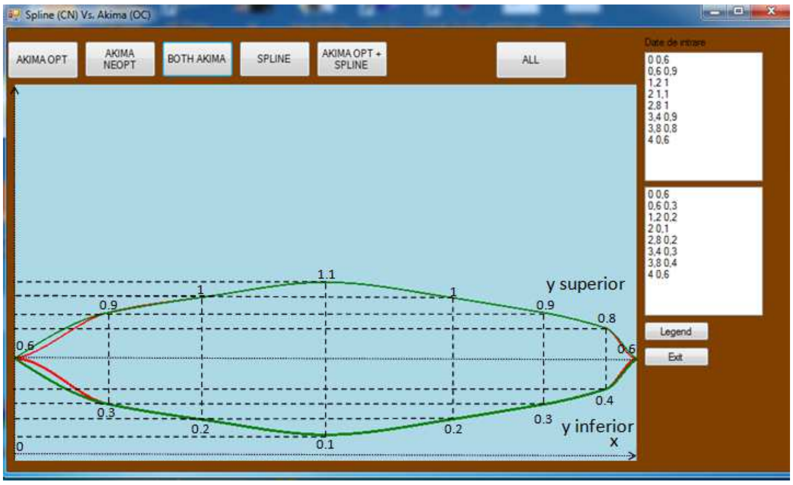
x_scale 0 4
f(x)_scale 0 2,400000095:

🟥 Akima optimizat
🟩 Akima neoptimizat
⬛ Spline cubic natural

dx 0,001

Fig. 2. Legend - Akima‘s not optimized method and Akima’s optimized method

We can see that the optimized end-points method of Akima shows the best aerodynamic properties of the profiles due to attack angle formed by the tangents at this peak. This is possible due to infiltration point in the last subin-



Date de intrare

0 0,6
0,6 0,9
1,2 1
2 1,1
2,8 1
3,4 0,9
3,8 0,8
4 0,6

0 0,6
0,6 0,3
1,2 0,2
2 0,1
2,8 0,2
3,4 0,3
3,8 0,4
4 0,6

Fig. 3. Akima‘s not optimized method (green) , Akima’s optimized method (red)

terval of the optimized method of Akima. As the forward speed is routed along the longitudinal axis of the boat, this speed becomes higher at a lower angle. Therefore the aerodynamic properties of this profile are improved at the same engine speed resulting in a higher displacement speed for the optimized Akima method at the extremities.

References

[1] H. Akima, *A NEW METHOD FOR INTERPOLATION AND SMOOTH CURVE FITTING BASED ON LOCAL PROCEDURES*, J. Asocc. Comp. Machinery, 4(1970), 589–602.

[2] J. H. Ahlberg, E. N. Nilson, J. L. Walsh, *The Theory of Splines and Their Applications*, Academic Press, New York, London, 1967.

[3] A. M. Bica, *FITTING DATA USING OPTIMAL HERMITE TYPE CUBIC INTERPOLATING SPLINES*, Applied Mathematics Letters, Elsevier, Volume 25(2012), 2047–2051.

[4] A. M. Bica, *OPTIMIZING AT THE END-POINTS THE AKIMAS INTERPOLATION METHOD OF SMOOTH CURVE FITTING*, Computer Aided Geometric Design, Elsevier, Volume 31, 2014, 245–257.

[5] C. Iacob, D. Homentcovschi, N. Marcov, A. Nicolau, *MATEMATICI CLASICE SI MODERNE*, vol.IV, Ed. Tehnica, Bucuresti, 1983.

[6] D. Kincaid, W. Cheney: *NUMERICAL ANALYSIS MATHEMATICS OF SCIENTIFIC COMPUTING*, Third Edition, American Mathematical Society, Providence, Rhode Island, 2002 .

[7] D. Kincaid, W. Cheney, *NUMERICAL ANALYSIS MATHEMATICS OF SCIENTIFIC COMPUTING*, Sixth Edition, American Mathematical Society, Providence, Rhode Island, 2008.

[8] D. D. Stancu, Gh. Coman, O. Agratini, R. Trâmiţaş, *ANALIZĂ NUMERICĂ ŞI TEORIA APROXIMĂRII*, vol.I, Presa Universitară Clujeană, Cluj Napoca, 2001.

[9] J. Kobza, *Cubic splines with minimal norm*, Appl. Math. 47(2002), 285–295.

ROMAI J., v.14, no.1(2018), 73–87

ABOUT THE CONSTRUCTION OF THE WEIGHTED MEANS OF A PAIR OF STRINGS

Mitrofan M. Choban, Ivan A. Budanaev

Tiraspol State University, Republic of Moldova,

Institute of Mathematics and Computer Sciences of ASM,

Chişinău, Republic of Moldova

mmchoban@gmail.com, ivan.budanaev@gmail.com

Abstract In information theory, linguistics and computer science are important distinct string metrics for measuring the difference between two given strings (sequences). In distinct domains of research are well known the distances of Hamming and Graev-Levenshtein between two strings. For any string distance there are distinct geometrical compute problems. Some of them are as follows:

- the calculation of the median of two strings;
- the calculation of the weighted means of two strings;
- the problem of the convexity of the weighted mean set.

In the present article the above geometrical problems are examined for the Hamming and Graev-Levenshtein string distances.

Keywords: quasimetric, Hamming distance, Graev distance, Levenshtein distance, weighted means of two strings.

2010 MSC: 20M35, 32F45, 54A20, 54E35, 68W32, 94A24, 94A55.

1. INTRODUCTION

Let X be a non-empty set and $d : X \times X \to \mathbb{R}$ be a mapping such that for all $x, y \in X$ we have:

(i_m) $d(x, y) \geq 0$;

(ii_m) $d(x, y) + d(y, x) = 0$ if and only if $x = y$,

(iii_m) $d(x, z) \leq d(x, y) + d(y, z)$.

Then (X, d) is called a *quasimetric space* and d is called a *quasimetric* on X. A function d with properties (i_m) and (ii_m) is called a distance on X.

If d is a quasimetric on X with property

(iv_m) $d(x, y) = d(y, x)$,

then (X, d) is called a *metric space* and d is called a *metric.*

If

(v_m) $d(x, y) = 0$ if and only if $x = y$,

then (X, d) is called a *strong quasimetric space* and d is called a *strong quasimetric* on X.

General problems in distance spaces were studied by distinct authors (see [1, 2, 5, 6, 17, 18, 10, 11, 14]).

Let d be a quasimetric on the non-empty set X. Fix an oriented pair $a, b \in X$. We put:

- $\alpha_d(a; a, b) = \{x \in X : d(a, x) = 0, d(x, b) = d(a, b)\}$ is the left annihilator of the oriented pair of points a, b;

- $\alpha_d(b; a, b) = \{x \in X : d(a, x) = d(a, b), d(x, b) = 0\}$ is the right annihilator of the oriented pair of points a, b;

- $M_d(a, b) = \{x \in X : d(a, x) + d(x, b) = d(a, b)\}$ is the weighted mean of the oriented pair of points a, b;

By definition, $\{a, b\} \subset \alpha_d(a; a, b) \cup \alpha_d(b; a, b) \subset M_d(a, b)$.

We assume that for a given finite space X with a given quasimetric d there exist "effective algorithms" of the calculation of the weighted mean $M_d(a, b)$ for any given oriented pair of points $a, b \in X$ [3, 4].

Let G be a semigroup and d be a pseudo-distance on G. The pseudo-distance d is called:

- Left (respectively, right) invariant if $d(xa, xb) \leq d(a, b)$ (respectively, $d(ax, bx) \leq d(a, b)$) for all $x, a, b \in G$;

- Invariant if it is both left and right invariant.

A distance d on a semigroup G is called *stable* if $d(xy, uv) \leq d(x, u) + d(y, v)$ for all $x, y, u, v \in G$.

Proposition 1.1. *Let d be a quasimetric on a semigroup G. The next assertions are equivalent:*

1 d is invariant.

2 d is stable.

A monoid is a semigroup with an identity element.

Fix a non-empty set A. The set A is called an alphabet. Let $L(A)$ be the set of all finite strings $a_1a_2 \dots a_n$ with $a_1, a_2, \dots, a_n \in A$ and $n \in \mathbb{N} = \{0, 1, 2, ...\}$. Let ε be the empty string. If $a = a_1a_2...a_n$ and $n = 0$, then $a = \varepsilon$. Consider the strings $a_1a_2 \dots a_n$ for which $a_i = \varepsilon$ for some $i \leq n$. If $a_i \neq \varepsilon$, for any $i \leq n$, or $n = 1$ and $a_1 = \varepsilon$, the string $a_1a_2 \dots a_n$ is called a *canonical string* or an *irreducible string*. The following set:

$$Sup(a_1a_2 \dots a_n) = \{a_1, a_2, \dots, a_n\} \cap A$$

is called the *support* of the string $a_1a_2 \dots a_n$ and the function:

$$l(a_1a_2 \dots a_n) = |\{i \leq n : a_i \neq \varepsilon\}|$$

represents the *length* of the string $a_1a_2\dots a_n$.

The number of elements $l^*(a_1a_2...a_n) = n$ of the string $a_1a_2...a_n$ is called the general length of the string $a_1a_2...a_n$. For two strings $a_1\dots a_n$ and $b_1\dots b_m$, their product (concatenation) is $a_1\dots a_nb_1\dots b_m$. If $n \geq 2$, $i < n$ and $a_i = \varepsilon$, then the strings $a_1\dots a_n$ and $a_1\dots a_{i-1}a_{i+1}\dots a_n$ are considered *equivalent*. In this case any string is equivalent to one unique canonical string.

Denote by $L^*(A)$ the family of all strings and by $L(A)$ the family of all canonical strings. In this case $L^*(A)$ became a semigroup and $L(A)$ becomes a monoid with identity ε. The monoid $L(A)$ is not a subsemigroup of the semigroup $L^*(A)$, but $L(A)\setminus\{\varepsilon\}$ is a subsemigroup of the semigroups $L^*(A)$ and $L(A)$! We identify the equivalent strings and $\kappa : L^*(A) \longrightarrow L(A)$ is the operation of the identification. By construction, $L(A) \subset L^*(A)$ and $\kappa(a) = a$ for any $a \in L(A)$. Hence, the mapping κ is a homomorphism and a retract.

If the strings a and b are equivalent, then we denote $a \sim b$ and say that the string b is a representation of the string a.

Let $a, b \in L(A)$. We consider the oriented pairs of strings. If a and b is a given pair of strings, then a is considered the first element of the pair a and b. The representations $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_m$ of the strings a and b are called *parallel decompositions* if $n = m$. The parallel representations $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ of the strings a and b are *irreducible* if $A \cap \{a_i, b_i\} \neq \emptyset$ for each $i \leq n$, or $n = 1$ and $a_1 = b_1 = \varepsilon$. The family of all parallel decompositions is infinite and the family of all irreducible parallel decompositions is finite. Moreover, $n \leq l(a) + l(b)$ for any irreducible parallel representations $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ of the strings a and b.

Let d be a quasimetric on $\bar{A} = A \cup \{\varepsilon\}$.

For any two strings $a = a_1a_2...a_n$ and $b = b_1b_2...b_m$ from $L^*(A)$ we put:

$$\begin{aligned} d_H(a,b) = {} & \Sigma\{d(a_i,b_i) : i \leq min\{n,m\}\} \\ & + \Sigma\{d(\varepsilon,b_j) : n+1 \leq j \leq m\} \\ & + \Sigma\{d(a_j,\varepsilon) : m+1 \leq j \leq n\}. \end{aligned}$$

The function d_H is a quasimetric on $L^*(A)$ and is called the *Hamming extension* of the distance d on $L^*(A)$. If d is a metric, then d_H is a metric as well [13, 7].

For any two strings $a, b \in L(A)$ we define:

$$d_G(a,b) = min\{d_H(a',b') : a', b' \in L^*(A), a = \kappa(a'), b = \kappa(b')\}.$$

The distance d_G is called the *Graev extension* or the *Graev-Markov-Livenstein extension* of the distance d [12, 7, 8, 16]. Obviously, we have that $d_G(a,b) \leq d_H(a,b)$ for all $a, b \in L(A)$.

The representations $a' = a_1 a_2 ... a_n$ and $b' = b_1 b_2 ... b_m$ of the $a, b \in L(A)$ are called *parallel d-optimal decompositions* of the pair of strings a, b if $n = m$ and $d_G(a,b) = \Sigma\{d(a_i, b_i) : i \leq n\}$.

The following assertions were proved in [7, 8].

Theorem 1.1. *For any quasimetric d on the alphabet $\bar{A}$ there exists a unique invariant quasimetric d^* on the monoid $L(A)$ with following properties:*

1. $d^*(x,y) = d(x,y)$ for any $x, y \in \bar{A}$.

2. If ρ is an invariant quasimetric on $L(A)$ and $\rho(x,y) \leq d(x,y)$ for all $x, y \in \bar{A}$, then $\rho(x,y) \leq d^*(x,y)$ for all $x, y \in L(A)$.

3. If $a' = a_1 a_2 ... a_n$ and $b' = b_1 b_2 ... b_n$ are parallel representations of the two strings $a, b \in L(A)$, then $d^*(a,b) \leq \Sigma\{d(a_i, b_i) : i \leq n\}$.

4. For any two strings $a, b \in L(A)$ there exist $n \geq 1$ and the parallel representations $a' = a_1 a_2 ... a_n$ and $b' = b_1 b_2 ... b_n$, such that $d^*(a,b) = \Sigma\{d(a_i, b_i) : i \leq n\}$. Therefore, for any oriented pair of strings $a, b \in L(A)$ there exist parallel d-optimal decompositions.

5. $d^* = d_G$.

6. If d is a metric, then d^* is a metric as well.

There exist algorithms of construction of parallel d-optimal decompositions of the given pair of strings $a, b \in L(A)$ [7, 8]. Our aim is to propose some algorithms of the calculation of the weighted means $M_{d_G}(a,b)$ for any given oriented pair of strings $a, b \in L(A)$. Some cases were presented in [3, 4].

We mention that the Graev extension d_G [12, 19, 9, 7, 8] coincides with the Levenshtein [15] extension d_L of the quasimetric d on $\bar{A}$. Levenshtein distance between two strings $a = a_1 a_2 \cdots a_n$ and $b = b_1 b_2 \cdots b_m$ is defined as the minimum number of insertions, deletions, and substitutions required to transform one string to the other.

We state the following problems and present algorithms to solve them for discrete distances.

Problem 1.1 *Find the methods (algorithms) for calculation of the distances $d_G(x,y)$ for the given $x, y \in L(A)$.*

Problem 1.2 *Find the applications of the quasi-metrics d_H and d_G in the fields of the information theory.*

For the case of discrete distance d, i.e. $d(x,x) = 0$ and $d(x,y) = 1$ for any distinct $x, y \in \bar{A}$, is well known the following algorithm for computing the distance $d^* = d_G = d_L$:

Algorithm 1 Graev-Markov-Levenshtein Distance:
Given $x, y \in L(\bar{A})$ compute $d_L(x, y)$, for the case of a discrete metric.

```
procedure COMPUTE_D($x, y$)                               ▷ The $d_L$ of $x$ and $y$
    for $i \leftarrow 1, m$ & $j \leftarrow 1, n$ do
        $d[i, 0] \leftarrow i, d[0, j] \leftarrow j$          ▷ initializing the memorization matrix
    end for
    $i \leftarrow 1, j \leftarrow 1$                          ▷ initializing loop variables
    for $j \leftarrow 1, n$ do
        for $i \leftarrow 1, m$ do
            if $x_i = y_j$ then
                $d[i, j] \leftarrow d[i - 1, j - 1]$
            else
                $d[i, j] := min(d[i - 1, j] + 1, min(d[i, j - 1] + 1, d[i - 1, j - 1] + 1))$
            end if
        end for
    end for
    return $d[m, n], d$                                   ▷ value of $d_L$ and matrix $d$
end procedure
```

For computation of the distance $d^* = d_G = d_L$ for any quasimetric d we propose the following algorithm:

Algorithm 2 QuasiMetric:
Given $x, y \in L(\bar{A})$ compute $d^*(x, y)$, for the case of quasimetric.

```
procedure COMPUTE_QUASI_D($x, y, dist$)                   ▷ The $d_L$ of $x$ and $y$
    for $i \leftarrow 1, m$ & $j \leftarrow 1, n$ do
        $d[i, 0] \leftarrow i, d[0, j] \leftarrow j$          ▷ initializing the memorization matrix
    end for
    $i \leftarrow 1, j \leftarrow 1$                          ▷ initializing loop variables
    for $j \leftarrow 1, n$ do
        for $i \leftarrow 1, m$ do
            if $dist(x_i, y_j) = 0$ then
                $d[i, j] \leftarrow d[i - 1, j - 1]$
            else
                $d[i, j] := min(d[i - 1, j] + cost_{remove},$
                $min(d[i, j - 1] + cost_{insert}, d[i - 1, j - 1] + dist(x_i, y_i)))$
            end if
        end for
    end for
    return $d[m, n], d$                                   ▷ value of $d_L$ and matrix $d$
end procedure
```

2. WEIGHTED MEANS OF THE PAIR OF STRINGS

On the given alphabet $\bar{A} = A \cup \{\varepsilon\}$ fix a quasimetric d with the Graev extension d_G.

Lemma 2.1. *Let $a, b, c \in L(A)$, $n \geq 1$ and $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$, $c' = c_1c_2...c_n$ be representations of the strings a, b, c respectively. If*

$$d_G(a,b) = \Sigma\{d(a_i,b_i) : i \leq n\} = \Sigma\{d(a_i,c_i) + d(c_i,b_i) : i \leq n\},$$

then the following assertions hold:

1. The strings $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ form the parallel d-optimal representations of the pair of strings a and b.

2. The strings $a' = a_1a_2...a_n$ and $c' = c_1c_2...c_n$ form the parallel d-optimal representations of the pair of strings a and c.

3. The strings $c' = c_1c_2...c_n$ and $b' = b_1b_2...b_n$ form the parallel d-optimal representations of the pair of strings c and b.

Proof. Follows from the inequality $d_G(x,y) \leq d_G(x,z) + d_G(z,y)$, for any strings $x, y, z \in L(A)$. ▮

We define the following sets:

$$M_{d_G}(a,b) = \{x \in L(A) : d_G(a,b) = d_G(a,x) + d_G(x,b)\}$$

and

$$M^*_{d_G}(a,b) = \{x \in L^*(A) : d_G(a,b) = d_G(a,\kappa(x)) + d_G(\kappa(x),b)\}$$

as the sets of weighted d-means of the oriented pair of strings $a, b \in L(A)$.

Assume that

$$M_{d_H}(a,b) = \{x \in L^*(A) : d_H(a,b) = d_H(a,x) + d_H(x,b)\}$$

is the set of H-weighted d-means of the oriented pair of strings $a, b \in L^*(A)$.

First, we construct equivalent representations of strings from $M_{d_G}(a,b)$ with respect to given parallel d-optimal decompositions of a and b.

Theorem 2.1. *Any fixed parallel d-optimal decompositions of a pair given strings $a, b \in L(A)$ generate weighted means, simultaneously with their equivalent representations, which form parallel d-optimal decompositions with the fixed representations of the given strings.*

Proof. We present the proof by construction. Let $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ be the fixed parallel d-optimal decompositions of the strings

a and b. Denote $\bar{M}^*_{d_G}(a', b') = \{c = c_1c_2...c_n \in L^*(A) : d_H(a', c) + d_H(c, b') = d_G(a, b)\}$ and $\bar{M}_{d_G}(a', b') = \{\kappa(c) : c \in \bar{M}^*_{d_G}(a', b')\}$.

We aim to construct strings of form $c' = c_1c_2...c_n$ such that for $c = \kappa(c')$ we have $d_G(a, b) = d_G(a, c) + d_G(c, b) = \Sigma\{d(a_i, c_i) + d(c_i, b_i) : i \leq n\}$.

For each $i \leq n$ we fix $c_i \in M_d(a_i, b_i) = \{x \in \bar{A} : d(a_i, x) + d(x, b_i) = d(a_i, b_i)\}$ and put $c' = c_1c_2...c_n$. Let $c = \kappa(c')$. From Lemma 2.1 it follows:

- the strings $a' = a_1a_2...a_n$ and $c' = c_1c_2...c_n$ form the parallel d-optimal representations of the pair of strings a and c;
- the strings $c' = c_1c_2...c_n$ and $b' = b_1b_2...b_n$ form the parallel d-optimal representations of the pair of strings c and b;
- $d_G(a, b) = d_G(a, c) + d_G(c, b)$;
- $c \in M_{d_G}(a, b)$.

The numbers $n(a', b') = |\bar{M}_{d_G}(a', b')|$ and $n^*(a', b') = |\bar{M}^*_{d_G}(a', b')|$ are estimated by the following relations:

$$n(a', b') \leq n^*(a', b') = \Pi\{|M_d(a_i, b_i)| i \leq n\},$$
$$\Pi\{|M_d(a_i, b_i)| i \leq n\} \geq 2^{|\{i \leq n : a_i \neq b_i\}|}.$$

This completes the proof of the theorem. ▮

In the case of discrete metric when $d_G(a, b)$ is an even number, we have the following algorithm for constructing the medians of a pair strings:

Algorithm 3 Medians of OPD of x and y:
Given $x, y \in L(\bar{A})$ construct $m \in L(\bar{A})$, s.t. $d^*(x, m) = d^*(m, y)$.

```
procedure MEDIANS($x, y$)                              ▷ Generates medians of $x$ and $y$
    $d^*$ ← compute_d(x,y)                            ▷ calculates distance between $x$ and $y$
    if $d^*$ is odd then
        return "distance $d^*(x, y)$ is odd, set $M$ is an empty set."
    end if
    OPD ← generate_OPD(x,y)                            ▷ generates optimal parallel decomp.
    $I = \{i : 1 \leq i \leq l^*(x')\}$
    for all $(x', y')$ in OPD do
        $I_1 = \{i : x'_i = y'_i\}$
        $I_2 = I \setminus I_1$
        for all ($I_3$ = Choose $(|I| - d)/2$ elements from $I_2$) do
            $m := m_1 m_2 \dots m_{|I|}$, where $m_i = x'_i$ if $i \in I_1 \cup I_3$, else $m_i = y'_i$
            $M := M \cup \{m\}$;
        end for
    end for
    return $M$                                         ▷ Median set of $x$ and $y$
end procedure
```

Remark 2.1. *One can notice that the median of a pair of strings is a special case of the above theorem. In particular, if $C \subset \{1, 2, ..., n\}$, and*

$$\Sigma\{d(a_i, b_i) : i \in C\} = \Sigma\{d(a_i, b_i) : i \notin C\},$$

putting $c_i = a_i$ for $i \in C$ and $c_i = b_i$ for $i \notin C$, for $c = \kappa(c_1c_2...c_n)$ we get $d_G(a, b) = 2d_g(a, c) = 2d_G(c, b)$ and c is an element of the median of a pair of strings a, b.

Further we present an important result which will be used to prove the converse of Theorem 2.1.

Lemma 2.2. *Let a, b and c be three strings for which $d_G(a, b) = d_G(a, c) + d_G(c, b)$. Then there exist $n \geq 1$ and the strings $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$ and $c' = c_1c_2...c_n$ such that:*

1. The strings $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ form the parallel d-optimal representations of the pair of strings a and b.

2. The strings $a' = a_1a_2...a_n$ and $c' = c_1c_2...c_n$ form the parallel d-optimal representations of the pair of strings a and c.

3. The strings $c' = c_1c_2...c_n$ and $b' = b_1b_2...b_n$ form the parallel d-optimal representations of the pair of strings c and b.

4. The representation $c' = c_1c_2...c_n$ of the string $c \in M_{d_G}(a, b)$ is generated by the parallel d-optimal representations $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$ of the pair of strings a and b.

Proof. First we examine the case when $c \sim e$. We fix the parallel d-optimal representations $a' = a_1a_2...a_n$ and $b' = b_1b_2...b_n$ of the pair of strings a and b. Then we put $c' = c_1c_2...c_n$, where $c_i = \varepsilon$ for each $i \leq n$. In this case the assertions of Lemma are proved.

Assume now that the $\kappa(c) \neq \varepsilon$. Then $l(c) = k \geq 1$. In this case we use the following algorithm:

1. Fix the parallel d-optimal representations $a^1 = u_1u_2...u_p$ and $c^1 = v_1v_2...v_p$ of the pair of strings a and c and the parallel d-optimal representations $c^2 = w_1w_2...w_m$ and $b^2 = z_1z_2...z_m$ of the pair of strings c and b.

2. We determine the sets $\{i \leq p : v_i \neq \varepsilon\} = \{i_j : j \leq k\}$ and $\{i \leq m : s_i \neq \varepsilon\} = \{s_j : j \leq k\}$, where $1 \leq i_1 < i_2 < ... < i_k \leq p$ and $1 \leq s_1 < s_2 < ... < s_k \leq m$.

3. We calculate $n_1 = max\{i_1, s_1\}$, $n_2 = max\{i_2 - i_1, s_2 - s_1\} + n_1$, ..., $n_k = max\{i_k - i_{k-1}, s_k - s_{k-1}\} + n_{k-1}$, $n = n_{k+1} = max\{p - i_k, m - s_k\} + n_k$.

4. Fix two monotone injection mappings $f : \{1, 2, ..., p\} \to \{1, 2, ..., n\}$ and $g : \{1, 2, ..., m\} \to \{1, 2, ..., n\}$ such that $f(i_1) = g(s_1) = n_1$ and $f(i_j) = g(s_j) = n_j$ for each $j \leq k$.

5. We construct the string $c' = c_1c_2...c_n$, where $c_{n_j} = v_{i_j} = w_{s_j}$ for each $j \leq k$ and $c_i = \varepsilon$ if $i \notin \{n_1, n_2, ..., n_k\}$.

6. Fix the representation $a' = a_1a_2...a_n$ of the string a such that $a_{n_j} = u_{i_j}$ for each $j \leq k$. We can assume that $a_{f(i)} = u_i$ for each $i \leq p$ and $a_i = \varepsilon$ for $i \notin f(\{1, 2, ..., p\})$.

7. Fix the representation $b' = b_1b_2...b_n$ of the string a such that $b_{n_j} = z_{s_j}$ for each $j \leq k$. We can assume that $b_{g(i)} = z_i$ for each $i \leq m$ and $b_i = \varepsilon$ for $i \notin g(\{1, 2, ..., m\})$.

8. The representations $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$ and $c' = c_1c_2...c_n$ are constructed.

From the above, by construction, we obtain the following:

$$d_H(a_1a_2...a_n, c_1c_2...c_n) = d_H(u_1u_2...u_p, v_1v_2...v_p) = d_G(a, c),$$
$$d_H(c_1c_2...c_n, b_1b_2...b_n) = d_H(w_1w_2...w_m, z_1z_2...z_m) = d_G(c, b),$$
$$d_G(a, b) \leq d_H(a_1a_2...a_n, b_1b_2...b_n) \leq \Sigma\{d(a_i, b_i) : i \leq n\}.$$

Also, the following equalities hold:

$$\begin{aligned}\Sigma\{d(a_i, b_i) : i \leq n\} &= \Sigma\{d(a_i, c_i) + d(c_i, b_i) : i \leq n\} \\ &= d_G(a, c) + d_G(c, b) = d_G(a, b).\end{aligned}$$

Hence $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$ and $c' = c_1c_2...c_n$ are the desired representations. The proof is complete. ∎

We are now ready to state the converse of Theorem 2.1.

Corollary 2.1. *Any weighted mean of a fixed pair of strings is generated by some of their optimal parallel decompositions.*

Remark 2.2. *Let $a, b \in L(A)$. Then from Lemma 2.2 it follows:*

1. Any weighted mean of a fixed pair of strings is generated by some of their optimal irreducible parallel decompositions.

2. If for any $x, y \in \bar{A}$ the set $M_d(x, y)$ of all weighted means is finite, then of the oriented pair of points $a, b \in L(A)$ the set $M_d(a, b)$ of all weighted means is finite too.

In the case of the discrete metric there exist algorithms of construction of all parallel d-optimal decompositions of the given pair of strings $a, b \in L(A)$ [7, 8]. The pseudo-code of such algorithm is presented below:

Algorithm 4 Optimal Parallel Decompositions (OPD):
Generate all optimal parallel decompositions of given $x, y \in L(\bar{A})$.

```
procedure GENERATE_OPD($x, y$)
    parameters costs of insertion and
    removal operations - $cost_{insert}$ and $cost_{remove}$ respectively.
    d, D := compute_D(x, y);
    return build_OPD(n,m,x,y,"","",D)                    ▷ n,m lengths of $x$ and $y$
end procedure
Backtracking function to incrementally generate OPD
procedure BUILD_OPD(n,m,x,y,a,b,D)
    if (n=0) and (m=0) then
        return ($reverse(a)$, $reverse(b)$)
    end if
    if $((n > 0) and (m > 0))$ and
    $((D[n, m] = D[n-1, m-1] + dist(x_n, y_m))$ or
    $((D[n, m] = D[n-1, m-1])$ and $(dist(x_n, y_m) = 0)))$ then
        Build_OPD(n-1,m-1,a+$x_n$,b+$y_m$)
    else if $(n > 0)$ and $(D[n, m] = D[n-1, m] + cost_{remove})$ then
        Build_OPD(n-1,m,a+$x_n$,b+$\varepsilon$)
    else if $(m > 0)$ and $(D[n, m] = D[n, m-1] + cost_{insert})$ then
        BuildOPD(n,m-1,a+$\varepsilon$,b+$y_m$)
    end if
end procedure
```

Lemma 2.2 is not true for arbitrary strings.

Example 2.1. *Let* $\{0, 1\} \subset A$, *where* $0 \neq 1$. *Consider that* $d(x, x) = 0$ *for any* $x \in \bar{A}$ *and* $d(x, y) = 1$ *for any distinct elements* $x, y \in \bar{A}$. *We say that* d *is the discrete metric on* $\bar{A}$. *Then* d, d_H *and* d_G *are metrics.*

Consider the canonical strings $a = 01$, $b = 0$ *and* $c = 1$. *We have* $d_G(a, b) = d_H(a, b) = d_G(a, c) = d_H(a, c) = d_G(c, b) = d_H(c, b) = 1$.

Fix the representations $a' = a_1a_2...a_n$, $b' = b_1b_2...b_n$ *and* $c' = c_1c_2...c_n$ *of the strings* a, b *and* c *respectively. Assume that:*

- the strings $a' = a_1a_2...a_n$ *and* $b' = b_1b_2...b_n$ *form the parallel d-optimal representations of the pair of strings* a *and* b.

- the strings $a' = a_1a_2...a_n$ *and* $c' = c_1c_2...c_n$ *form the parallel d-optimal representations of the pair of strings* a *and* c.

There exist $1 \leq i < j \leq n$ *such that* $a_i = 0$, $a_j = 1$ *and* $a_s = \varepsilon$ *for* $s \notin \{i, j\}$. *Since* $d_G(a, b) = \Sigma\{d(a_s, b_s) : s \leq n\} = 1$, *we have* $b_i = 0$ *and* $b_s = \varepsilon$ *for* $s \neq i$. *Since* $d_G(a, c) = \Sigma\{d(a_s, c_s) : s \leq n\} = 1$, *we have* $c_j = 1$ *and*

and $c_s = \varepsilon$ for $s \neq j$. Thus $d_H(b_1b_2...b_n, c_1c_2...c_n)$ $=2 > 1$ $=d_G(b,c)$ and the strings $c' = c_1c_2...c_n$ and $b' = b_1b_2...b_n$ does not form the parallel d-optimal representations of the pair of strings c and b. Thus the requirement $d_G(a,b)$ $= d_G(a,c) + d_G(c,b)$ is essential in the conditions of Lemma 2.2.

Example 2.2. *Let $\{0,1\} \subset A$, where $0 \neq 1$. Consider that $d(x,x) = 0$ for any $x \in \bar{A}$ and $d(x,y) = 1$ for any distinct elements $x, y \in \bar{A}$. Then d, d_H and d_G are metrics.*

Let $a' = a_1a_2...a_n$ and $b' = b_1b_2w_2 \ldots b_nu_n$ be the fixed parallel d-optimal decompositions of the strings a and b. Let $N = \{i \leq n : a_i \neq b_i\}$. For any proper subset M of N we put $c_M = c_1c_2...c_n$, where $c_i = a_i$ for $i \notin M$ and c_i $= b_i$ for $i \in M$. For the improper subsets we have $c_\emptyset$ $=a$ and $c_N = b$. As was proved in Theorem 2.1, $c = \kappa(c_M) \in M_{d_G}(a,b)$. We observe that $\bar{M}^*_{d_G}(a',b')$ is the set of all strings c_M, $M \subset N$, and $\bar{M}_{d_G}(a',b') = \kappa(\bar{M}^*_{d_G}(a',b'))$.

The number $n^*(a',b')$ of such strings from the set $\bar{M}^*_{d_G}(a',b')$, generated by the above method, is equal to $2^{|N|}$. We mention that the number $n(a',b')$ of the canonical strings $\bar{M}_{d_G}(a',b')$ may be $< 2^{|N|}$.

Let $a = 1$ and $b = 0000$ be the canonical representation of the given strings. We have $d_G(a,b) = d_H(a,b) = 4$. For a and b we have the following parallel d-optimal decompositions $a' = 1\varepsilon\varepsilon\varepsilon$, $b' = 0000$. These parallel decompositions generate the following eight canonical strings 1, 0, 00, 10, 000, 100, 0000, 1000. We have $\bar{M}_{d_G}(a',b') = \{1, 0, 00, 10, 000, 100, 0000, 1000\}$, $N = \{1,2,3,4\}$ and $8 = |C_{d_G}(a',b')| < 2^{|N|} = 2^4 = 16$. The other parallel d-optimal decompositions $a'' = \varepsilon\varepsilon\varepsilon 1$, $b'' = 0000$ of a, b present the following set of canonical strings $\bar{M}_{d_G}(a'',b'') = \{1, 0, 00, 01, 000, 001, 0000, 0001\}$ with $N = \{1,2,3,4\}$. We have that $\bar{M}_{d_G}(a',b') \cap \bar{M}_{d_G}(a'',b'') = \{1, 0, 00, 000, 0000\}$.

Let $a, b \in L^*(A)$. The following remarks shows that the construction of the H-weighted d-means $c \in M_{d_H}(a,b)$ is more simple than the construction of the weighted d-means $c \in M_{d_G}(a,b)$.

Remark 2.3. *Let $a, b \in L(A)$. Then:*

1. If $x, y \in L^*(A)$, $x \sim y$ and $x \in M^*_{d_G}(a,b)$, then $y \in M^*_{d_G}(a,b)$.

2. If $x \in L^*(A)$, $x \sim y$, then $x \in M^*_{d_G}(a,b)$ if and only if $\kappa(x) \in M_{d_G}(a,b)$.

If $a \in L^*(A)$, $c = c_1c_2...c_n$, $n \geq 1$ and $c_i = a$ for any $i \leq n$, then we put $c = a^n$.

Remark 2.4. *Let $a, b, c \in L^*(A)$ and $n \geq 1$. Then $c \in M_{d_H}(a,b)$ if and only if $c \cdot \varepsilon^n \in M_{d_H}(a,b)$. The string $c = c_1c_2...c_n$ is called H-irreducible if $n = 1$ or $c_n \neq \varepsilon$. Hence are true the following two assertions:*

1. $l^*(c) \leq max\{l^*(a), l^*(b)\}$ for any H-irreducible element $c \in M_{d_H}(a,b)$.

2. If the string $c \in M_{d_H}(a,b)$ is not H-irreducible, then there exist a unique H-irreducible string $c' \in M_{d_H}(a,b)$ and a number $n = l^*(c) - l^*(c')$ such that $c = c' \cdot \varepsilon^n$.

Assume that

$$\bar{M}_{d_H}(a,b) = \{x \in M_{d_H}(a,b) : l^*(c) = max\{l^*(a), l^*(b)\}\}$$

is the set of H-weighted d-means c of the oriented pair of strings $a, b \in L^*(A)$ with $l^*(c) = max\{l^*(a), l^*(b)\}$.

Remark 2.5. *We present below the algorithm of construction of elements from $M_{d_H}(a,b)$. From the above remark it follows that is sufficient to construct the strings $c \in M_{d_H}(a,b)$ for which $l^*(c) = max\{l^*(a), l^*(b)\}$. Fix two strings $a, b \in L^*(A)$ with $p = l^*(a)$ and $l^*(b) = q$.*

1. We put $n = max\{p, q\}$.

2. We construct:

- $a' = a = a_1a_2...a_n$ and $b' = b = b_1b_2...b_n$ if $p = q$;

- $a' = a \cdot \varepsilon^{q-p} = a_1a_2...a_n$ and $b' = b$ $b_1b_2...b_n$ if $p < q$;

- $a' = a = a_1a_2...a_n$ and $b' = b \cdot \varepsilon p - q = b_1b_2...b_n$ if $q < p$.

3. For each $i \leq n$ we fix $c_i \in M_d(a_i, b_i) = \{x \in \bar{A} : d(a_i, x) + d(x, b_i) = d(a_i, b_i)\}$.

4. Put $c = c_1c_2...c_n$.

5. Have $c \in \bar{M}_{d_H}(a,b)$.

By construction, we have $d_H(a,b) = d_H(a',b') = d_H(a',c) + d_H(c,b') = d_H(a,c) + d_H(c,b)$ and $c \in M_{d_H}(a,b)$.

One can observe that from $d_H(a,c) + d_H(c,b) = d_H(a,b)$ it follows that $c_i \in M_d(a_i, b_i)$ for each $i \leq n$. Therefore, any string $c \in \bar{M}_{d_H}(a,b)$ with $l^*(c) = n$ can be constructed by the above algorithm. Hence that algorithm permit to construct all strings $c \in M_{d_H}(a,b)$

The number $m^*(a,b) = |\bar{M}_{d_H}(a,b)|$ is estimated by the following relations:

$$m^*(a,b) = \Pi\{|M_d(a_i,b_i)| i \leq n\} \geq 2^{|\{i \leq n : a_i \neq b_i\}|}.$$

If d is discrete metric on $\bar{A}$ with $d(x,y) = 1$ for any pair of distinct elements $x, y \in \bar{A}$, then $M_d(x,y) = \{x, y\}$ for any $x, y \in \bar{A}$ and

$$m^*(a,b) = 2^{|\{i \leq n : a_i \neq b_i\}|} \text{ for any } a, b \in L^*(A).$$

3. PROBLEM OF CONVEXITY

Let (X, d) be a metric space. A subset $L \subseteq X$ is called d-convex if $M_d(a,b) \subseteq L$ for any $a, b \in L$.

On the alphabet $\bar{A} = A \cup \{\varepsilon\}$ consider the distance metric $d : d(x,x) = 0$ and $d(x,y) = 1$ for distinct $x, y \in \bar{A}$. Any subset of (A, d) is d-convex. In 2016 Professor Gh. Zbăganu informed us about the following questions: **Question 1.** *Is it true that the set $M_{d_H}(a,b)$ is d_H-convex in $(L^*(A), d_H)$ for any $a, b \in L^*(A)$?*

Question 2. *Is it true that the set $M_{d_G}(a,b)$ is d_G-convex in $(L^*(A),d_G)$ for any $a,b \in L^*(A)$?*

Theorem 3.1. *The set $M_{d_H}(a,b)$ is d_H-convex in $(L^*(A),d_H)$ for any $a,b \in L^*(A)$.*

Proof. We can assume that $a = a_1a_2...a_n$ and $b = b_1b_2...b_n$. Then $x = x_1x_2...x_n \in M_{d_H}(a,b)$ if and only if $x_i \in \{a_i,b_i\}$ for any $i \le n$. If $c = c_1c_2...c_n$, $f = f_1f_2...f_n$ are two strings from $M_{d_H}(a,b)$ and $x = x_1x_2...x_n \in M_{d_H}(c,f)$, then $x_i \in \{c_i,f_i\} \subseteq \{a_i,b_i\} \cup \{a_i,b_i\} = \{a_i,b_i\}$ and $x \in M_{d_H}(a,b)$. The proof is complete. ▮

Theorem 3.2. *There exists a finite alphabet A and two strings $a,b \in L(A)$ for which the set $M_{d_G}(a,b)$ is not d_G-convex.*

Proof. The proof follows from the following examples. ▮

Example 3.1. *Let $A = \{B,C,D,J,K,L,M,N,O,P,Q,R\}$,*

$$a = DJCJNRCKCRBP, b = DBCNJROCLCRPM,$$

$$a' = DJCNJNRCKCRBP, b' = DBCJNJROCLCRBPM,$$

$$c = DJCJNJNROCKCRBPM.$$

For the above strings, we have that:

$$d_G(a,b) = 7, d_G(a,a') = 1, d_G(a',b) = 6, d_G(a,b') = 5,$$
$$d_G(b',b) = 2, d_G(a',b') = 6, d_G(a',c) = d_G(c,b') = 3,$$
$$d_G(a,c) = 4, d_G(c,b) = 5.$$

Hence $a',b' \in M_{d_G}(a,b)$, $c \in M_{d_G}(a',b')$, but $c \notin M_{d_G}(a,b)$. Therefore, it follows that the set $M_{d_G}(a,b)$ is not convex.

In construction of strings a',b' and c we used the d_G-optimal parallel representations of pairs of strings a,b and a',b' respectively. The string a' is constructed using the following d_G-optimal parallel representations:

$$\begin{pmatrix} D \\ D \end{pmatrix} \begin{matrix} J \\ B \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} \varepsilon \\ N \end{matrix} \begin{pmatrix} J \\ J \end{pmatrix} \begin{matrix} N & R \\ R & O \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} K \\ L \end{matrix} \begin{pmatrix} C & R \\ C & R \end{pmatrix} \begin{matrix} B & P \\ P & M \end{matrix}$$

The string b' is constructed using the following d_G-optimal parallel representations:

$$\begin{pmatrix} D \\ D \end{pmatrix} \begin{matrix} J \\ B \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} J \\ \varepsilon \end{matrix} \begin{pmatrix} N \\ N \end{pmatrix} \begin{matrix} \varepsilon \\ J \end{matrix} \begin{pmatrix} R \\ R \end{pmatrix} \begin{matrix} \varepsilon \\ O \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} K \\ L \end{matrix} \begin{pmatrix} C & R \\ C & R \end{pmatrix} \begin{matrix} B \\ \varepsilon \end{matrix} \begin{pmatrix} P \\ P \end{pmatrix} \begin{matrix} \varepsilon \\ M \end{matrix}$$

The string c is constructed using the following d_G-optimal parallel representations:

$$\begin{pmatrix} D \\ D \end{pmatrix} \begin{matrix} J \\ B \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} \varepsilon \\ J \end{matrix} \begin{pmatrix} N & J \\ N & J \end{pmatrix} \begin{matrix} N \\ \varepsilon \end{matrix} \begin{pmatrix} R \\ R \end{pmatrix} \begin{matrix} \varepsilon \\ O \end{matrix} \begin{pmatrix} C \\ C \end{pmatrix} \begin{matrix} K \\ L \end{matrix} \begin{pmatrix} C & R & B & P \\ C & R & B & P \end{pmatrix} \begin{matrix} \varepsilon \\ M \end{matrix}$$

Example 3.2. *Let alphabet A and strings a, b, a', b', c be as in the previous example. We put $m = QQQQQQQ$. Consider the strings amb, bma, $a'ma'$, $b'mb'$ and cmc. We obtain the following:*

$$d_G(amb, bma) = 14,$$
$$d_G(amb, a'ma') = d_G(a'ma', bma) = 7,$$
$$d_G(amb, b'mb') = d_G(b'mb', bma) = 7,$$
$$d_G(a'ma', b'mb') = 12,$$
$$d_G(a'ma', cmc) = d_G(cmc, b'mb') = 6,$$
$$d_G(amb, cmc) = d_G(cmc, bma) = 9.$$

Hence $a'ma', b'mb'$ are from the middle of the segment $M_{d_G}(amb, bma)$, the string cmc is from the middle of the segment $M_{d_G}(a'ma', b'mb')$, but $cmc \notin M_{d_G}(amb, bma)$.

4. CONCLUSIONS

The optimal decompositions:

- permit the calculation of the median of two strings;
- permit the calculation of the weighted means of two strings;
- describe the proper similarity of two strings;
- permit to obtain long common sub-sequences;
- permit to calculate the distance between strings;
- permit to solve the problem of text editing and correction;
- permit to appreciate changeability of information over time;
- permit to solve the problem of convexity of the weighted means of two strings.

Acknowledgments. The second author's research was supported by the World Federation of Scientists.

References

[1] A. V. Arkhangel'skii, *Mappings and spaces*, Uspekhi Mat. Nauk 21 (1966), vyp. 4, 133–184. (in Russian) (English translation: Russian Math. Surveys **21** (1966), no. 4, 115–162).

[2] L. M. Blumenthal, *Distance Geometry*, Clarendon Press, Oxford, 1953.

[3] I. Budanaev, *On Hamming type Distance Functions*, ROMAI J., 12 (2016), no. 2, 25-32.

[4] I. Budanaev, *Parallel Decompositions and The Weighted Mean of a Pair of Strings*, Proceedings of International Conference "Contemporary Trends in Science Development: Visions of Young Researchers", 6th Edition, Chisinau, 2017, 7–11.

[5] D. Burago, Yu. Burago, S. Ivanov, *A Course in Metric Geometry*, Graduate Studies in Mathematics. Vol 33. American Mathematical Society, 2001.

[6] M. M. Choban, *The theory of stable metrics*, Math. Balkanica, 2 (1988), 357–373.

[7] M. M. Choban, I. A. Budanaev, *Distances on Monoids of Strings and Their Applications*, Proceedings of the Conference on Mathematical Foundations of Informatics MFOI2016, July 25-29, 2016, Chisinau, Republic of Moldova, 2016, 144–159.

[8] M.M. Choban, I.A. Budanaev, *About Applications of Distances on Monoids of Strings.* Computer Science Journal of Moldova, 24 (2016), no. 3, 335-356.

[9] M. M. Choban, L. L. Chiriac, *On free groups in classes of groups with topologies*, Bul. Acad. Ştiinţe Repub. Moldova, Matematica (2013), no. 2-3, pp. 61–79.

[10] M. M. Deza, E. Deza, *Dictionary of Distances*, Elsevier, Amsterdam, 2006.

[11] M. M. Deza, E. Deza, *Encyclopedia of Distances*, Springer, Berlin, 2014.

[12] M. I. Graev, *Free topological groups*, Izv. Akad. Nauk SSSR, Ser. Mat., 12 (1948), Issue 3, 279-324 (In Russian). (English translation: Amer. Math. Soc. Transl. 35 (1951). Beprint Amer. Math. Soc. Trarasl. (1) 8 (1962) 305-364).

[13] R. W. Hamming, *Error Detecting and Error Correcting Codes,* Bell System Technical Journal 29 (1952), no 2, pp. 147–160.

[14] H.-P. A. Künzi, *Nonsymmetric distances and their associated topologies: about the origins of basic ideas in the area of asymmetric topology*, In: *Handbook of the history of general topology*, Vol. 3, Hist. Topol., vol. 3, Kluwer Acad. Publ., Dordrecht, 2001, 853–968.

[15] V. I. Levenshtein, *Binary codes capable of correcting deletions, insertions, and reversals*, Dokl. AN SSSR 163 (1965), no 4, pp. 845–848 (in Russian). (English translation: Soviet Physics – Doklady 10 (1966), no. 8, 707–710).

[16] A. A. Markov, *On free topological groups*, Izv. Akad. Nauk SSSR, Ser. Mat., 8 (1945), 225–232 (In Russian). (English translation: Trans. Moscow Math. Soc. 8 (1962), pp. 195–272).

[17] K. Menger, *Untersuchungen über allegemeine Metrik*, Math. Ann. 100 (1928), 75–63.

[18] S. I. Nedev, *o-metrizable spaces*, Trudy Moskov. Mat.Ob-va 24 (1974), 213–247 (in Russian). (English translation: Trans. Moscow Math. Soc. 24 (1974), 213–247).

[19] S. Romaguera, M. Sanchis, M. Tkachenko, *Free paratopological groups*, Topology Proceed. 27 (2003), no 2, 613–640.

ROMAI J., v.14, no.1(2018), 89–100

ON INEQUALITIES INVOLVING CONVEX FUNCTIONS AND INTEGRAL CONDITIONS

Zoubir Dahmani, Mohamed Doubbi Bounoua

Laboratory LPAM, UMAB, University of Mostaganem, Algeria

zzdahmani@yahoo.fr

Abstract In this paper, using the Riemann-Liouville fractional integral operator, we establish new results that generalize some theorems of the work: [A note on some new fractional results involving convex functions. Acta Math. Univ. Comenianae, Vol. LXXXI, 2, 2012]. We also discuss other integral inequalities generalizing some theorems in the paper: [Some new results of two open problems related to integral inequalities, Journal of Mathematical Inequalities, 10(3), 2016].

Keywords: integral inequality, Riemann-Liouville integral, convex function .

2010 MSC: 26D10, 26A33.

1. INTRODUCTION

In [5] W.J. Liu et al. studied some interesting inequalities for a convex function $(x-a)^\delta$ for $\delta \geq 1$ and established the following result:

Theorem 1.1. *Let $\beta > 0$ and $f \geq 0$ be a continuous function on $[a,b]$ with*

$$\int_x^b f^{\min\{1,\beta\}}(t)dt \geq \int_x^b (t-a)^{\min\{1,\beta\}}dt, x \in [a,b]. \tag{1}$$

Then,

$$\int_a^b f^{\alpha+\beta}(x)dx \geq \int_a^b (x-a)^\alpha f^\beta(x)dx \tag{2}$$

is valid for all $\alpha > 0$.

Then, in 2009, W.J. Liu, Q. Ngo and V.N. Huy [6] proved the following important result includes more general convex function :

Theorem 1.2. *Let f, g, h be positive three continuous functions on $[a,b]$, with $f \leq h$ on $[a,b]$ and such that $\frac{f}{h}$ is decreasing and f, g are increasing. If φ is a* *convex function with* $\varphi(0)=0$*, then*

$$\frac{\int\limits_a^b f(x)dx}{\int\limits_a^b h(x)dx} \geq \frac{\int\limits_a^b \varphi(f(x))g(x)dx}{\int\limits_a^b \varphi(h(x))g(x)dx}.$$

Recently, Z. Dahmani [1] established generalization for the above theorem, he proved that for any three positive continuous functions f,g and h defined on $[a,b]$, with $f\leq h$, f and g are increasing and $\frac{f}{h}$ is decreasing, then for any $x\in]a,b]$, we have:

$$\frac{J_a^\alpha f(x)}{J_a^\alpha h(x)} \geq \frac{J_a^\alpha[\varphi(f)g](x)}{J_a^\alpha[\varphi(h)g](x)},$$

where $\alpha>0$ and φ is a positive and convex function, with $\varphi(0)=0$.
Very recently, A. Kashuri and R. Liko [4] proposed another result, as a response to an open problem posed by Liu et al. in [6]. In fact, for three positive continuous functions f,g and h defined on $[a,b]$, such that $f\leq h$ on $[a,b]$, f,g are increasing and $\frac{f}{h}$ is decreasing, if φ is a positive and convex function, with $\varphi(0)=0$, the authors of [4] proved that the inequality

$$\frac{\int\limits_a^b f(x)dx}{\int\limits_a^b h(x)dx} \geq \frac{(\int\limits_a^b \varphi(f(x))g(x)dx)^\delta}{(\int\limits_a^b \varphi(h(x))g(x)dx)^\lambda}$$

is valid, under some conditions on $\lambda,\delta,\varphi f(a),\varphi f(b),g(a),g(b)$. Other important results introducing a parameter λ and generalizing Theorem 1.1 are also discussed by the authors of [4].
In this paper, we prove new classical and fractional integral inequalities that generalise some integral results of the papers [1, 4].

2. RIEMANN-LIOUVILLE INTEGRATION

We recall the following definition and some properties.

Definition 2.1. *[3] The Riemann-Liouville fractional integral operator of order* $\delta\geq 0$*, for a continuous function* f *on* $[a,b]$ *is defined as*

$$\begin{gathered} J_a^\delta f(x)=\tfrac{1}{\Gamma(\delta)}\int_a^x (x-u)^{\delta-1}f(u)du;\quad \delta>0, a<x\leq b, \\ J_a^0 f(x)=f(x). \end{gathered} \tag{3}$$

We give the semigroup property:

$$J_a^{\alpha} J_a^{\delta} f(x) = J_a^{\alpha+\delta} f(x), \alpha \geq 0, \delta \geq 0, \tag{4}$$

In the particular case where $f(x) = (x-a)^{\beta}$ on $[a,b]$, we have

$$J_a^{\delta}(x-a)^{\beta} = \frac{\Gamma(\beta+1)}{\Gamma(\delta+\beta+1)}(x-a)^{\delta+\beta}. \tag{5}$$

3. MAIN RESULTS

Theorem 3.1. *Let f, g and h be three positive continuous functions on $[a,b]$ with $f \leq h$. Suppose that f and g are increasing and $\frac{f}{h}$ is a decreasing function, and assume that φ is a positive and convex function, with $\varphi(0) = 0$. Then for any $x \in]a,b]$, we have:*

$$\frac{J_a^{\alpha} f(x)}{J_a^{\alpha} h(x)} \geq \left(\frac{J_a^{\alpha}[\varphi(f)g](x)}{J_a^{\alpha}[\varphi(h)g](x)} \right)^{\lambda}, \tag{6}$$

where $\lambda \geq 1, \alpha > 0$.

Proof. Since $f \leq h$, then for all $\lambda \geq 1$, we can write:

$$\frac{J_a^{\alpha} f(x)}{J_a^{\alpha} h(x)} \geq \left(\frac{J_a^{\alpha} f(x)}{J_a^{\alpha} h(x)} \right)^{\lambda}, x \in]a,b]. \tag{7}$$

On the other hand, for any $x \in]a,b]$, we have (see [1]):

$$\frac{J_a^{\alpha} f(x)}{J_a^{\alpha} h(x)} \geq \frac{J_a^{\alpha}[\varphi(f)g](x)}{J_a^{\alpha}[\varphi(h)g](x)}. \tag{8}$$

Therefore, it yields that

$$\left(\frac{J_a^{\alpha} f(x)}{J_a^{\alpha} h(x)} \right)^{\lambda} \geq \left(\frac{J_a^{\alpha}[\varphi(f)g](x)}{J_a^{\alpha}[\varphi(h)g](x)} \right)^{\lambda}, x \in]a,b]. \tag{9}$$

Using (7) and (9) we obtain (6). ∎

Remark 3.1. *Taking $\lambda = 1$ in Theorem 3.1, we obtain Theorem 3.5 proved in [1].*

Another main result is the following theorem, in which we will generalize a theorem in the paper [4]. We prove:

Theorem 3.2. *Let f, g and h be three positive continuous functions on $[a,b]$, such that $f \leq h$ on $[a,b]$, f and g are increasing and $\frac{f}{h}$ is decreasing. Assume*

that φ is a positive and convex function, with $\varphi(0)=0$.
In the case where $1\leq\theta<\lambda$, if $\varphi[f(a)]g(a)J_a^\alpha(1)\geq 1$, then, we have:

$$\frac{J_a^\alpha f(x)}{J_a^\alpha h(x)}\geq\frac{\left(J_a^\alpha[\varphi(f)g](x)\right)^\theta}{\left(J_a^\alpha[\varphi(h)g](x)\right)^\lambda}, x\in]a,b],\alpha>0. \qquad (10)$$

The same inequality is valid in the case: $1\leq\lambda<\theta$, under the condition: $\varphi[f(b)]g(b)J_a^\alpha(1)\leq 1$.

Proof. We prove the theorem in two steps:
Case 1: For $1\leq\theta<\lambda$, there exists $s>0$, such that $\lambda=\theta+s$.
So, we have:

$$\frac{\left(J_a^\alpha[\varphi(f)g](x)\right)^\theta}{\left(J_a^\beta[\varphi(h)g](x)\right)^\lambda}=\left(\frac{J_a^\alpha[\varphi(f)g](x)}{J_a^\alpha[\varphi(h)g](x)}\right)^\theta\times\frac{1}{\left(J_a^\alpha[\varphi(h)g](x)\right)^s}$$

Thanks to Theorem 3.1, we obtain

$$\frac{\left(J_a^\alpha[\varphi(f)g](x)\right)^\theta}{\left(J_a^\alpha[\varphi(h)g](x)\right)^\lambda}\leq\frac{J_a^\alpha f(x)}{J_a^\alpha h(x)}\times\frac{1}{\left(J_a^\alpha[\varphi(h)g](x)\right)^s}.$$

Now, we shall prove that $\left(J_a^\alpha[\varphi(h)g](x)\right)^s\geq 1$.

We have:

$$\begin{aligned}J_a^\alpha[\varphi(h)g](x) &= J_a^\alpha[\frac{\varphi(h)}{h}hg](x)\\ &\geq J_a^\alpha[\frac{\varphi(h)}{h}fg](x).\end{aligned}$$

Since φ is a convex function, then, for all x,y, we can write

$$(y-x)\varphi'(x)\leq\varphi(y)-\varphi(x).$$

Hence for $y=0$, we obtain $x\varphi'(x)-\varphi(x)\geq 0$. Therefore, we get $\left(\frac{\varphi(x)}{x}\right)'=\frac{x\varphi'(x)-\varphi(x)}{x^2}\geq 0$, which implies that $\frac{\varphi(x)}{x}$ is an increasing function and by the hypothesis of $f\leq h$, we conclude that $\frac{\varphi[f]}{f}\leq\frac{\varphi[h]}{h}$.
Consequently, we obtain

$$J_a^\alpha[\frac{\varphi(h)}{h}fg](x)\geq J_a^\alpha[\frac{\varphi(f)}{f}fg](x).$$

On the other hand, since f, g and $\frac{\varphi(t)}{t}$ are increasing, then $[\frac{\varphi(f)}{f}fg](x)$ is increasing. So, we have $\forall x \in [a,b]$, $[\frac{\varphi(f)}{f}fg](x) \geq \varphi[f(a)]g(a)$.
Finally,

$$J_a^\alpha[\frac{\varphi(f)}{f}fg](x) \quad \geq \quad \varphi[f(a)]g(a)J_a^\alpha(1) \geq 1.$$

Case 2: For $1 \leq \lambda < \theta$, there exists $s > 0$, such that $\theta = \lambda + s$.
We have

$$\begin{aligned} \frac{(J_a^\alpha[\varphi(f)g](x))^\theta}{(J_a^\alpha[\varphi(h)g](x))^\lambda} &= \left(\frac{J_a^\alpha[\varphi(f)g](x)}{J_a^\alpha[\varphi(h)g](x)}\right)^\lambda \times \left(J_a^\beta[\varphi(f)g](x)\right)^s. \\ &\leq \frac{J_a^\alpha f(x)}{J_a^\alpha h(x)} \times \left(J_a^\beta[\varphi(f)g](x)\right)^s. \end{aligned}$$

Now, we need to prove that $\left(J_a^\beta[\varphi(f)g](x)\right)^s \leq 1$.

Since $\varphi(f)g$ is increasing on $[a,b]$, we have $[\varphi(f)g](x) \leq \varphi(f(b))g(b)$, $\forall x \in [a,b]$, which implies

$$(J_a^\alpha[\varphi(f)g](x))^s \leq (\varphi(f(b))g(b)J_a^\alpha(1))^s \leq 1.$$

The proof of Theorem 3.2 is thus achieved. ▮

Remark 3.2. *In Theorem 3.2, if we take $\alpha = 1$, we obtain Theorem 2.2 of [4].*

Changing the hypotheses of Theorem 2.1 in [4] by considering two integral conditions on f, we obtain the following result:

Theorem 3.3. *Let $f : [a,b] \to \mathbb{R}^+$ be a continuous function, such that:*

$$\int_x^b (u-a)^{\min(1,\beta)}du \leq \int_x^b f^{\min(1,\beta)}(u)du, x \in [a,b], \beta > 0 \tag{11}$$

and

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha - n + 1)}(b-a)^{n+1}\int_a^b f^\beta(t)dt \leq 1, n = [\alpha], \alpha > 0. \tag{12}$$

Then for any $\lambda \geq 1$, $b - a \geq 1$, we have

$$\int_a^b f^{\alpha+\beta}(u)du \geq \left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)^\lambda.$$

Proof. For $\lambda \geq 1$, we have

$$\left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)^\lambda = \left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)\left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)^{\lambda-1}.$$

By Theorem 2.1 of [5], we can write $\int_a^b (u-a)^\alpha f^\beta(u)du \leq \int_a^b f^{\alpha+\beta}(u)du$. Therefore,

$$\left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)^\lambda \leq \int_a^b f^{\alpha+\beta}(u)du \left(\int_a^b (u-a)^\alpha f^\beta(u)du\right)^{\lambda-1}.$$

Now we need to prove that $\int_a^b (u-a)^\alpha f^\beta(u)du \leq 1$. We have

$$\begin{aligned}\int_a^b (u-a)^\alpha f^\beta(u)du &= -(u-a)^\alpha \int_u^b f^\beta(t)du|_{u=a}^{u=b} \\ &\quad +\alpha \int_a^b (u-a)^{\alpha-1} \int_u^b f^\beta(t)dtdu \\ &= \alpha \int_a^b (u-a)^{\alpha-1} \int_u^b f^\beta(t)dtdu.\end{aligned}$$

Hence, we can write

$$\begin{aligned}\int_a^b (u-a)^\alpha f^\beta(u)du &= \alpha \int_a^b (u-a)^{\alpha-1} \int_u^b f^\beta(t)dtdu \\ &= \alpha(\alpha-1) \int_a^b (u-a)^{\alpha-2} \int_u^b \int_{u_1}^b f^\beta(t)dtdu_1du \\ &\quad \ldots\end{aligned}$$

$$= \alpha(\alpha-1)...(\alpha-n+1)\int\limits_a^b (u-a)^{\alpha-n}$$

$$\times \int\limits_u^b \int\limits_{u_1}^b ... \int\limits_{u_{n-1}}^b f^{\beta}(t)dtdu_{n-1}...du_1du.$$

On the other hand, since $(u-a)^{\alpha-n} \leq (b-a)$, it follows that

$$\begin{aligned}
\int\limits_a^b (u-a)^{\alpha} f^{\beta}(u)du &\leq \alpha(\alpha-1)...(\alpha-n+1)\int\limits_a^b (b-a) \\
&\quad \times \int\limits_u^b \int\limits_{u_1}^b ... \int\limits_{u_{n-1}}^b f^{\beta}(t)dtdu_{n-1}...du_1du \\
&= \alpha(\alpha-1)...(\alpha-n+1)\int\limits_a^b \int\limits_a^b \int\limits_a^b ... \int\limits_a^b f^{\beta}(t)dtdu_{n-1}...du_1du_0du \\
&= \alpha(\alpha-1)...(\alpha-n+1)\int\limits_a^b f^{\beta}(t)dt \\
&\quad \times \int\limits_a^b \int\limits_a^b ... \int\limits_a^b du_{n-1}...du_1du_0du \\
&= \alpha(\alpha-1)...(\alpha-n+1)(b-a)^{n+1}\int\limits_a^b f^{\beta}(t)dt \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha-n+1)}(b-a)^{n+1}\int\limits_a^b f^{\beta}(t)dt \leq 1.
\end{aligned}$$

▮

We prove also the following theorem:

Theorem 3.4. *Let $f : [a,b] \longrightarrow \mathbb{R}^+$ be a continuous function, such that:*

$$\int_x^b (u-a)^{\min\{1,\beta\}}du \leq \int_x^b f^{\min\{1,\beta\}}(u)du, x \in [a,b], \beta > 0$$

and

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha-n+1)}(b-a)^{n+1}J_a^{\delta}f^{\beta}(b)\leq 1; n=[\alpha],\alpha>0.$$

Then, for all $\lambda\geq 1,\delta\geq 1,b-a\geq 1$, *we have:*

$$J_a^{\delta}f^{\alpha+\beta}(b)\geq\left(J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\right)^{\lambda}.$$

Proof. For $\lambda\geq 1$, we have

$$\left(J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\right)^{\lambda}=\left(J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\right)\left(J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\right)^{\lambda-1}.$$

Now, we begin by proving that $J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\leq J_a^{\delta}f^{\alpha+\beta}(b)$.
To do this, we need to prove that

$$J_a^{\delta}(x-a)^{\alpha}f^{\beta}(x)|_{x=b}\geq\frac{\Gamma(\alpha+\beta+1)(b-a)^{\alpha+\beta+\delta}}{\Gamma(\alpha+\beta+\delta+1)}. \tag{13}$$

For $\beta\in]0,1]$, we have

$$\begin{aligned}J_a^{\delta}(t-a)^{\alpha}f^{\beta}(t) \quad | \quad {}_{t=b} &= \frac{1}{\Gamma(\delta)}\int_a^b(b-x)^{\delta-1}(x-a)^{\alpha}f^{\beta}(x)dx\\ &= \frac{1}{\Gamma(\delta)}\Big[(b-x)^{\delta-1}(x-a)^{\alpha}\int_x^b f^{\beta}(u)du\ |_{x=a}^{x=b}\\ &\quad +\frac{1}{\Gamma(\delta)}\int_a^b g(x)\left(\int_x^b f^{\beta}(u)du\right)dx\Big]\\ &= \frac{1}{\Gamma(\delta)}\int_a^b g(x)\left(\int_x^b f^{\beta}(u)du\right)dx,\end{aligned}$$

where $g(x)=(\delta-1)(b-x)^{\delta-2}(x-a)^{\alpha}+\alpha(b-x)^{\delta-1}(x-a)^{\alpha-1}$.
Thanks to the imposed condition, we observe that

$$\begin{aligned}\tfrac{1}{\Gamma(\delta)}\textstyle\int_a^b g(x)\left(\int_x^b f^{\beta}(u)du\right)dx &\geq \frac{1}{\Gamma(\delta)}\int_a^b g(x)\left(\int_x^b(u-a)^{\beta}du\right)dx\\ &= \frac{1}{(\beta+1)\Gamma(\delta)}\int_a^b g(x)\left[(b-a)^{\beta+1}-(x-a)^{\beta+1}\right]dx\\ &= \frac{\Gamma(\alpha+\beta+1)(b-a)^{\alpha+\beta+\delta}}{\Gamma(\alpha+\beta+\delta+1)}.\end{aligned} \tag{14}$$

If we take $\beta = 1$ in (14), then we get

$$J_a^\delta (t-a)^\alpha f(t)|_{t=b} \geq \frac{\Gamma(\alpha+2)}{\Gamma(\alpha+\delta+2)}(b-a)^{\alpha+\delta+1}. \tag{15}$$

Using the following inequality (see Lemma 2.2 of [7]):

$$ps + qr \geq s^p r^q, \forall p, q, s, r > 0, p + q = 1, \tag{16}$$

with $p = \frac{1}{\beta}$, $q = \frac{\beta-1}{\beta}$, $s = f^\beta(x)$ and $r = (x-a)^{\beta-1}$, we obtain

$$\frac{1}{\beta} f^\beta(x) + \frac{\beta-1}{\beta}(x-a)^\beta \geq f(x)(x-a)^{\beta-1}.$$

Consequently,

$$f^\beta(x) + (\beta-1)(x-a)^\beta \geq \beta f(x)(x-a)^{\beta-1}. \tag{17}$$

Multiplying both sides of (17) by $\frac{1}{\Gamma(\delta)}(b-x)^{\delta-1}(x-a)^\alpha$ and integrating the resulting inequality with respect to x over $[a,b]$, yields

$$J_a^\delta (t-a)^\alpha f^\beta(t)|_{t=b} + (\beta-1) J_a^\delta (t-a)^{\alpha+\beta}|_{t=b} \geq \beta J_a^\delta (t-a)^{\alpha+\beta-1} f(t)|_{t=b}.$$

Then, thanks to (15), we obtain

$$J_a^\delta (t-a)^\alpha f^\beta(t)|_{t=b} + (\beta-1) J_a^\delta (t-a)^{\alpha+\beta}|_{t=b} \geq \beta \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta+\delta+1)}(b-a)^{\alpha+\beta+\delta}.$$

Hence,

$$J_a^\delta (t-a)^\alpha f^\beta(t)|_{t=b} \geq \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta+\delta+1)}(b-a)^{\alpha+\beta+\delta}. \tag{18}$$

Now, we prove that $J_a^\delta (x-a)^\alpha f^\beta(x)|_{x=b} \leq J_a^\delta f^{\alpha+\beta}(b)$.
As before, by Lemma 2.2 of [7]), we get

$$\frac{\beta}{\alpha+\beta} f^{\alpha+\beta}(x) + \frac{\alpha}{\alpha+\beta}(x-a)^{\alpha+\beta} \geq (x-a)^\alpha f^\beta(x). \tag{19}$$

Multiplying both sides of (19) by $\frac{1}{\Gamma(\delta)}(b-x)^{\delta-1}$ and integrating the resulting inequality with respect to x over $[a,b]$, we obtain

$$\frac{\beta}{\alpha+\beta} J_a^\delta f^{\alpha+\beta}(x)|_{x=b} + \frac{\alpha}{\alpha+\beta} J_a^\delta (x-a)^{\alpha+\beta}|_{x=b} \geq J_a^\delta (x-a)^\alpha f^\beta(x)|_{x=b}.$$

Therefore,

$$\beta J_a^\alpha f^{\alpha+\beta}(x)|_{x=b}+\alpha J_a^\delta(x-a)^{\alpha+\beta}|_{x=b} \geq \alpha J_a^\delta(x-a)^\alpha f^\beta(x)|_{x=b}+\beta J_a^\delta(x-a)^\alpha f^\beta(x)|_{x=b}.$$

Using (18), we obtain

$$\begin{aligned}\beta J_a^\delta f^{\alpha+\beta}(x)|_{x=b}+\alpha\frac{\Gamma(\alpha+\beta+1)(b-a)^{\alpha+\beta+\delta}}{\Gamma(\alpha+\beta+\delta+1)} &\geq \alpha\frac{\Gamma(\alpha+\beta+1)(b-a)^{\alpha+\beta+\delta}}{\Gamma(\alpha+\beta+\delta+1)}\\ &\quad +\beta J_a^\delta(t-a)^\alpha f^\beta(t)|_{t=b}.\end{aligned}$$

Hence

$$J_a^\delta f^{\alpha+\beta}(b) \geq J_a^\delta(t-a)^\alpha f^\beta(t)|_{t=b}. \tag{20}$$

Now, we need to show that

$$\left(J_a^\delta(x-a)^\alpha f^\beta(x)|_{x=b}\right)^{\lambda-1} \leq 1,$$

which is equivalent to

$$J_a^\delta(x-a)^\alpha f^\beta(x)|_{x=b} \leq 1.$$

An integration by parts allows us to obtain:

$$\begin{aligned}J^\delta(x-a)^\alpha f^\beta(x)|_{x=b} &= \frac{1}{\Gamma(\delta)}\int_a^b(b-u)^{\delta-1}(u-a)^\alpha f^\beta(u)du\\ &= \frac{1}{\Gamma(\delta)}\Big[-(u-a)^\alpha\int_u^b(b-t)^{\delta-1}f^\beta(t)du|_{u=a}^{u=b}\\ &\quad +\alpha\int_a^b(u-a)^{\alpha-1}\int_u^b(b-t)^{\delta-1}f^\beta(t)dtdu\Big]\\ &= \frac{1}{\Gamma(\delta)}\left[\alpha\int_a^b(u-a)^{\alpha-1}\int_u^b(b-t)^{\delta-1}f^\beta(t)dtdu\right]\\ &= \frac{1}{\Gamma(\delta)}\left[\alpha(\alpha-1)\int_a^b(u-a)^{\alpha-2}\int_u^b\int_{u_1}^b(b-t)^{\delta-1}f^\beta(t)dtdu_1du\right]\\ &\quad \dots\end{aligned}$$

$$
\begin{aligned}
= \ & \frac{1}{\Gamma(\delta)}\Big[\alpha(\alpha-1)...(\alpha-n+1)\int_a^b (u-a)^{\alpha-n} \\
& \times \int_u^b \int_{u_1}^b ... \int_{u_{n-1}}^b (b-t)^{\delta-1} f^{\beta}(t)dtdu_{n-1}...du_1du\Big].
\end{aligned}
$$

On the other hand, using the fact that $(u-a)^{\alpha-n} \leq (b-a)$, we can write

$$
\begin{aligned}
J^{\delta}(x-a)^{\alpha} f^{\beta}(x)|_{x=b} \ \leq \ & \frac{1}{\Gamma(\delta)}\Big[\alpha(\alpha-1)...(\alpha-n+1)\int_a^b (b-a) \\
& \times \int_u^b \int_{u_1}^b ... \int_{u_{n-1}}^b (b-t)^{\delta-1} f^{\beta}(t)dtdu_{n-1}...du_1du\Big] \\
= \ & \frac{1}{\Gamma(\delta)}\Big[\alpha(\alpha-1)...(\alpha-n+1) \\
& \times \int_a^b \int_a^b \int_u^b \int_{u_1}^b ... \int_{u_{n-1}}^b (b-t)^{\delta-1} f^{\beta}(t)dtdu_{n-1}...du_1du_0du\Big] \\
\leq \ & \frac{1}{\Gamma(\delta)}\Big[\alpha(\alpha-1)...(\alpha-n+1) \\
& \times \int_a^b \int_a^b \int_a^b ... \int_a^b (b-t)^{\delta-1} f^{\beta}(t)dtdu_{n-1}...du_1du_0du\Big] \\
= \ & \frac{1}{\Gamma(\delta)}\Big[\alpha(\alpha-1)...(\alpha-n+1)\int_a^b (b-t)^{\delta-1} f^{\beta}(t)dt \\
& \times \int_a^b \int_a^b ... \int_a^b du_{n-1}...du_1du_0du\Big] \\
= \ & \alpha(\alpha-1)...(\alpha-n+1)(b-a)^{n+1} J_a^{\delta} f^{\beta}(b). \\
= \ & \frac{\Gamma(\alpha)}{\Gamma(\alpha-n+1)}(b-a)^{n+1} J_a^{\delta} f^{\beta}(b). \\
\leq \ & 1.
\end{aligned}
$$

Theorem 3.4 is thus proved. ▮

Remark 3.3. *If we take $\delta = 1$ in Theorem 3.4, we obtain Theorem 3.3.*

To finish, we present to the reader the following corollary:

Corollary 3.1. *Let f and h be two positive continuous functions on $[a, b]$, with $f \leq h$, f is increasing and $\frac{f}{h}$ is decreasing. Then, for any $x \in]a, b]$, we have:*

$$\frac{J_a^\delta f(x)}{J_a^\delta h(x)} \geq \left(\frac{J_a^\delta (x-a)^\alpha f^\beta(x)}{J_a^\delta (x-a)^\alpha h^\beta(x)} \right)^\lambda,$$

where $\delta, \alpha, \beta > 0$, $\lambda \geq 1$.

Proof. We take $\varphi(x) = x^\beta$ and $g(x) = (x-a)^\alpha$ in Theorem 3.1. ∎

References

[1] Z. Dahmani, *A note on some new fractional results involving convex functions*, Acta Math. Univ. Comenianae, Vol. LXXXI, 2 (2012), 241-246.

[2] Z. Dahmani, N. Bedjaoui, Some generalized integral inequalities. Journal of Advanced Research in Applied Mathematics, Vol. 3, Iss. 2, (2011), 1-9.

[3] R. Gorenflo, F. Mainardi, *Fractional calculus: integral and differential equations of fractional order*, Springer Verlag, Wien, (1997), 223-276.

[4] A. Kashuri, R. Liko, *Some new results of two open problems related to integral inequalities*, Journal of Mathematical Inequalities, 10(3), (2016), 877-883.

[5] W.J. Liu, G. Cheng, C.C. Li, *Further development of an open problem*, J. Inequal. Pure Appl. Math., 9(1), (2008), Art. 14.

[6] W.J. Liu, Q. Ngo, V.N. Huy, *Several interesting integral inequalities*, Journal of Mathematical Inequalities, 3(2), (2009), 201-212.

[7] Q.A. Ngo, D.D. Thang, T.T. Dat, D.A. Tuan, *Note on an integral inequality*, J. Inequal. Pure Appl. Math., 7(4), (2006), Art. 120.

A CASE-STUDY OF OPTIMAL PORTFOLIO COMPOSITION WITH INVESTMENTS LIMITS

Fitim Deari[1], Carmen Rocşoreanu[2]

[1] *South East European University, Tetovo, Macedonia*

[2] *University of Craiova, Craiova, Romania*

f.deari@seeu.edu.mk, rocsoreanu@yahoo.com

Abstract A portfolio composition, where the weights are supposed to depend on the investor's risk tolerance, is considered. The portfolio's certainty equivalent return has to be maximized. The mathematical model leads to a convex problem. The attached Kuhn-Tucker system is solved using the critical lines method.

Keywords: Kuhn-Tucker system, portfolio composition, critical lines.

2010 MSC: 90C25, 90C90, 91G10.

1. INTRODUCTION

Starting with the pioneering work of Markowitz from 1952 [9], the portfolio theory was studied by many researchers. Among those with recent contributions in this field we quote Jacobs, Levy and Markowitz [6], Bailey and López de Prado [1], Kwan [8], Norstad [11], Cumova, Moreno and Nawrocki [3], Marling and Emanuelsson [10], Calvo, Ivorra and Liern [2], Kan and Zhou [7].

The aim of our study is to illustrate how to compose an efficient portfolio with positive weights and investments limits for five companies.

Consider P_t the closing price of a stock at the end of time t and P_{t-1} the closing price of a stock at the end of earlier time $t-1$. It follows that $P_t - P_{t-1}$ is the price return of the stock at time t

The *discretely compounded rate of return* in the period $(t-1, t)$ is:

- $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$, if the stock has not paid dividend;
- $R_t = \frac{P_t + D_t}{P_{t-1}} - 1$, if the stock paid dividend D_t

The *continuously compounded rate of return* in the period $(t-1, t)$ is defined as:

- $R_t = \ln \frac{P_t}{P_{t-1}}$, if the stock has not paid dividend;
- $R_t = \ln \frac{P_t + D_t}{P_{t-1}}$, if the stock paid dividend D_t.

The discretely compounded rate of return is slightly larger than the continuously compounded return rate. In the case when it is assumed that historical returns denote the distribution of the returns for the coming period, continuously compounded return is more appropriate.

In our study for experimental purpose five firms from New York Stock Exchange (NYSE) were selected, namely, Applied Materials (AMAT), Amazon (AMZN), Alibaba Group Holding Limited (BABA), Advanced Micro Devices, Inc. (AMD) and AT&T (T).

Data were downloaded from http://finance.yahoo.com searching the historical data for every company. They covered the period between February 2015 and April 2016. Monthly adjusted price are used. The observations were examined using Excel and Stata. The continuously compounded rates of return for the five firms were computed. They are given in Table 1.

Date	AMAT	AMZN	BABA	AMD	T
02.05.2016	0.1016	0.0635	0.0238	0.0863	-0.0096
01.04.2016	-0.0341	0.1053	-0.0268	0.2196	0.0033
01.03.2016	0.1155	0.0718	0.1385	0.2865	0.0583
01.02.2016	0.0722	-0.0605	0.0262	-0.0277	0.0244
04.01.2016	-0.0562	-0.1410	-0.1926	-0.2659	0.0608
01.12.2015	-0.0053	0.0165	-0.0340	0.1957	0.0217
02.11.2015	0.1183	0.0603	0.0030	0.1072	0.0048
01.10.2015	0.1324	0.2011	0.3518	0.2091	0.0424
01.09.2015	-0.0910	-0.0020	-0.1144	-0.0510	-0.0189
03.08.2015	-0.0701	-0.0444	-0.1696	-0.0642	-0.0453
01.07.2015	-0.1018	0.2112	-0.0489	-0.2179	-0.0090
01.06.2015	-0.0463	0.0113	-0.0822	0.0513	0.0280
01.05.2015	0.0220	0.0175	0.0942	0.0088	-0.0029
01.04.2015	-0.1310	0.1253	-0.0237	-0.1705	0.0734
02.03.2015	-0.1047	-0.0214	-0.0223	-0.1488	-0.0569
02.02.2015	0.0965	0.0698	-0.0455	0.1907	0.0486

Table 1. Continuously compounded rates of return.

Consider N stocks and let $R_{i,t}$ be the stock i's return rate at time $t, t = 1, \ldots, T$ and $i = 1, \ldots, N$. We recall some well-known formulae from statistics:

- The *mean* of periodic returns percentages for the stock i is

$$\overline{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t}.$$

- The *variance* of the return rates for the company i is

$$\sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(R_{i,t} - \overline{R}_i)^2,$$

while the s*ample variance*, used in our model (and denoted in the same way), is

$$\sigma_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}(R_{i,t} - \overline{R}_i)^2.$$

- The *standard deviation* is $\sigma_i = \sqrt{\sigma_i^2}$.
- The *covariance* of the stocks i and j return rates is

$$cov_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(R_{i,t} - \overline{R}_i)(R_{j,t} - \overline{R}_j).$$

- The *correlation coefficient* of the stocks i and j return rates is $\rho_{ij} = \frac{cov_{ij}}{\sigma_i \sigma_j}$.

Some numerical characteristics of the return rates for the five firms considered by us are given in Table 2.

	AMAT	AMZN	BABA	AMD	T
Mean	0.0011	0.0428	-0.0077	0.0256	0.0140
Standard Error	0.0231	0.0230	0.0320	0.0427	0.0094
Median	-0.0197	0.0389	-0.0253	0.0301	0.0133
Standard Deviation	0.0925	0.0920	0.1279	0.1706	0.0376
Sample Variance	0.0086	0.0085	0.0163	0.0291	0.0014
Kurtosis	-1.6342	0.2295	3.4787	-1.1154	-0.5939
Skewness	0.1495	0.1194	1.3969	-0.1826	-0.2301
Range	0.2634	0.3522	0.5444	0.5524	0.1303
Minimum	-0.1310	-0.1410	-0.1926	-0.2659	-0.0569
Maximum	0.1324	0.2112	0.3518	0.2865	0.0734
Sum	0.0181	0.6843	-0.1227	0.4093	0.2233
Count	16	16	16	16	16

Table 2. Summary statistics.

Using the above formula for the covariance of the stocks i and j return rates, the covariance matrix of the stocks return rates for our case-study is obtained as follows:

	AMAT	AMZN	BABA	AMD	T
AMAT	0.008	0.0014	0.0071	0.0107	0.001
AMZN	0.0014	0.0079	0.0064	0.0047	0.0005
BABA	0.0071	0.0064	0.0153	0.0112	0.0011
AMD	0.0107	0.0047	0.0112	0.0273	0.0013
T	0.001	0.0005	0.0011	0.0013	0.0013

Let x_i be the weight of security i included in the portfolio, $i = 1, \ldots, N$. Thus, if S is the total amount of money which will be invested, then $x_i S$ will be invested in firm i. Obviously, $\sum\limits_{i=1}^{N} x_i = 1$.

The *portfolio expected return* is defined as:

$$E\left(R_p\right) = \sum_{i=1}^{N} x_i E\left(R_i\right),$$

where $E\left(R_i\right)$ is the expected return value of security i.

The *portfolio variance* is defined as:

$$\sigma_p^2\left(x\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j cov_{ij}.$$

In our case-study we consider that the expected return of security i is the mean of periodic returns percentages, thus the portfolio expected return reads:

$$E\left(R_p\right) = 0.0011x_1 + 0.0428x_2 - 0.0077x_3 + 0.0256x_4 + 0.014x_5.$$

The portfolio variance reads:

$$\begin{aligned} \sigma_p^2\left(x\right) &= 0.008x_1^2 + 0.0079x_2^2 + 0.0153x_3^2 + 0.0273x_4^2 + 0.0013x_5^2 \\ &\quad +0.0028x_1x_2 + 0.0142x_1x_3 + 0.0214x_1x_4 + 0.002x_1x_5 + 0.0128x_2x_3 \\ &\quad +0.0094x_2x_4 + 0.001x_2x_5 + 0.0224x_3x_4 + 0.0022x_3x_5 + 0.0026x_4x_5. \end{aligned}$$

A simple computation shows that $\sigma_p^2\left(x\right)$ is a positive definite quadratic form, so it is a convex function.

2. THE MATHEMATICAL MODEL FOR OPTIMAL PORTFOLIO SELECTION WITH POSITIVE WEIGHTS AND INVESTMENTS LIMITS

Consider $r > 0$ the parameter that quantifies the investor's risk tolerance.

As the portfolio's certainly equivalent return is $E\left(R_p\right) - \frac{\sigma_p^2(x)}{r}$, we intend to minimize $\sigma_p^2\left(x\right) - rE\left(R_p\right)$. In addition, we suppose that the weight x_i is limited by c_i. Thus, the problem can be written as:

$$\begin{aligned}(\min) f\left(x\right) &= \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j cov_{ij} - r\sum_{i=1}^{N} x_i E\left(R_i\right)\\ \sum_{i=1}^{N} x_i &= 1\\ 0 &\leq x_i \leq c_i,\ i = 1, ..., N.\end{aligned}$$

As the function f is the sum between a positive definite quadratic form and a linear function, it follows that f is a convex function. The constraints are linear, so we have a convex differentiable programming problem, which can be solved using the Kuhn-Tucker theory.

The associated Lagrange-type function L reads:

$$L = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j cov_{ij} - r\sum_{i=1}^{N} x_i E\left(R_i\right) - \theta\left(\sum_{i=1}^{N} x_i - 1\right) - \sum_{i=1}^{N}\delta_i x_i + \sum_{i=1}^{N}\eta_i (x_i - c_i),$$

where $\theta \in R$ and $\delta_i \geq 0$, $\eta_i \geq 0$, $i = 1, ..., N$.

Thus, the Kuhn-Tucker system reads:

$$\begin{cases} \frac{\partial L}{\partial x_i} = 0,\ i = 1, ..., N\\ \sum\limits_{i=1}^{N} x_i = 1\\ x_i \geq 0,\ i = 1, ..., N\\ x_i \leq c_i,\ i = 1, ..., N\\ \delta_i x_i = 0,\ i = 1, ..., N\\ \eta_i (x_i - c_i) = 0,\ i = 1, ..., N.\end{cases}$$

For our study $N = 5$ and we consider that all investments limits are $ci = 0.6$, $i = 1, \ldots, 5$. Thus, the Lagrange function is:

$$\begin{aligned} L \; = \; & 0.008x_1^2 + 0.0079x_2^2 + 0.0153x_3^2 + 0.0273x_4^2 + 0.0013x_5^2 + 0.0028x_1x_2 \\ & + 0.0142x_1x_3 + 0.0214x_1x_4 + 0.002x_1x_5 + 0.0128x_2x_3 + 0.0094x_2x_4 \\ & + 0.001x_2x_5 + 0.0224x_3x_4 + 0.0022x_3x_5 + 0.0026x_4x_5 \\ & - r\,(0.0011x_1 + 0.0428x_2 - 0.0077x_3 + 0.0256x_4 + 0.014x_5) \\ & - \theta\,(x_1 + x_2 + x_3 + x_4 + x_5 - 1) - \sum_{i=1}^{5} \delta_i x_i + \sum_{i=1}^{5} \eta_i (x_i - 0.6). \end{aligned}$$

The Kuhn-Tucker system reads:

$$\left\{ \begin{array}{l} 0.016x_1 + 0.0028x_2 + 0.0142x_3 + 0.0214x_4 \\ \qquad +0.002x_5 - 0.0011r - \theta - \delta_1 + \eta_1 = 0, \\ 0.0028x_1 + 0.0158x_2 + 0.0128x_3 + 0.0094x_4 \\ \qquad +0.001x_5 - 0.0428r - \theta - \delta_2 + \eta_2 = 0, \\ 0.0142x_1 + 0.0128x_2 + 0.0306x_3 + 0.0224x_4 \\ \qquad +0.0022x_5 + 0.0077r - \theta - \delta_3 + \eta_3 = 0, \\ 0.0214x_1 + 0.0094x_2 + 0.0224x_3 + 0.0546x_4 \\ \qquad +0.0026x_5 - 0.0256r - \theta - \delta_4 + \eta_4 = 0, \\ 0.002x_1 + 0.001x_2 + 0.0022x_3 + 0.0026x_4 \\ \qquad +0.0026x_5 - 0.014r - \theta - \delta_5 + \eta_5 = 0, \\ x_1 + x_2 + x_3 + x_4 + x_5 = 1, \\ 0 \leq x_i \leq 0.6, \; i = 1, ..., 5 \\ \delta_i x_i = 0, \; i = 1, ..., 5 \\ \eta_i (x_i - 0.6) = 0, \; i = 1, ..., 5. \end{array} \right.$$

with $\theta \in R$ and $\delta_i \geq 0$, $\eta_i \geq 0$, $i = 1, ..., 5$.

This system consists both of equations and inequations and have 16 unknown variables, namely x_i, δ_i, η_i, $i = 1, \ldots, 5$ and θ. Obviously, the solutions depend on the values of the parameter r.

Remark that the status of any security can be *in*, *out* or *up*. More precisely:

- The status is *in*, when $0 < x_i < c_i = 0.6$. In this case, $\delta_i = 0$ and $\eta_i = 0$.
- The status is *out*, when $x_i = 0$. In this case $\eta_i = 0$.
- The status is *up*, when $x_i = c_i = 0.6$. In this case, $\delta_i = 0$.

In order to solve the Kuhn-Tucker system for different values of the investor's risk tolerance r, we follow the lines in Kwan [8]. Thus, the initial portfolio consists of those securities which have the highest expected return and we follow the steps 1-6 below.

Step 1. The initial portfolio consists of AMZN and AMD, which have the highest expected return. Thus $x_2 = 0.6$, $x_4 = 0.4$, so AMZN is *up*, while AMD is *in*. It follows $\delta_2 = \delta_4 = 0$ and $\eta_4 = 0$. As AMAT, BABA and T are *out*, we get $x_1 = x_3 = x_5 = 0$, so $\eta_1 = \eta_3 = \eta_5 = 0$. Thus, the first 6 equations of the Kuhn-Tucker system have the solution:

$$\begin{aligned}
&\delta_1 = -0.01724 + 0.0245r, \quad \delta_3 = -0.01084 + 0.0333r,\\
&\delta_5 = -0.02584 + 0.0116r, \quad \eta_2 = 0.01424 + 0.0172r,\\
&\theta = 0.02748 - 0.0256r.
\end{aligned}$$

From the conditions $\delta_i \geq 0$, $i = 1, 3, 5$, it follows $r \geq 2.224$. The critical value $r = 2.224$ is obtained when $\delta_5 = 0$. As $r < 2.224$, δ_5 would become negative. But $\delta_5 \geq 0$, so δ_5 must be chosen zero at the next step.

Step 2. As $r < 2.224$, $\delta_5 = 0$, so $x_5 > 0$ and T must changes its status from *out* to *in*. In addition, it follows $\eta_5 = 0$. AMZN remains *up*, thus, $x_2 = 0.6$ and $\delta_2 = 0$. AMD remains *in*, so $\delta_4 = \eta_4 = 0$. AMAT and BABA are *out*, thus $x_1 = x_3 = 0$ and $\eta_1 = \eta_3 = 0$. The Kuhn-Tucker system has the solution:

$$\begin{aligned}
&\delta_1 = -0.001 + 0.017r, \quad \delta_3 = 0.005 + 0.0262r,\\
&\eta_2 = -0.0074 + 0.0269r, \quad x_4 = -0.0969 + 0.223r,\\
&x_5 = 0.4969 - 0.223r, \quad \theta = 0.00164 - 0.014r.
\end{aligned}$$

From the conditions $\delta_1 \geq 0, \eta_2 \geq 0$ and $x_4, x_5 \in [0, 0.6]$, it follows $r \geq 0.43$. Thus, the solutions from Step 2 are valid for $r \in [0.43, 2.224]$. The critical value $r = 0.43$ is obtained for $x_4 = 0$, so at the next step AMD must change the status.

Step 3. As $r < 0.43, x_4 = 0$, so AMD is now *out* and $\eta_4 = 0$. AMZN remains *up*, so $x_2 = 0.6$ and and $\delta_2 = 0$. T is *in*, while AMAT and BABA remain *out*. Thus, $x_1 = x_3 = 0$, $\delta_5 = 0$, $\eta_1 = \eta_3 = \eta_5 = 0$, and the solution is:

$$\begin{aligned}
&\delta_1 = 0.00084 + 0.0129r, \quad \delta_3 = 0.00692 + 0.0217r,\\
&\delta_4 = 0.00504 - 0.0116r, \quad \eta_2 = -0.00824 + 0.0288r,\\
&x_5 = 0.4, \quad \theta = 0.00164 - 0.014r.
\end{aligned}$$

From the conditions $\delta_4 \geq 0, \eta_2 \geq 0$, it follows $r \geq 0.286$. Thus, the solutions from Step 3 are valid for $r \in [0.286, 0.43]$.

Step 4. As $r < 0.286$, $\eta_2 = 0$, thus AMZN must change its status from *up* to *in*, so $0 < x_2 < 0.6$. It follows $\delta_2 = \eta_2 = 0$. As T remains *in*, $\delta_5 = \eta_5 = 0$. AMAT, BABA and AMD remain *out*, so it follows $x_1 = x_3 = x_4 = 0$ and

$\eta_1 = \eta_3 = \eta_4 = 0$. The Kuhn-Tucker system has the solution:

$$\delta_1 = -0.00036 + 0.0171r, \quad \delta_3 = 0. + 0.04312r,$$
$$\delta_4 = 0.0008 \ + 0.0031r, \quad x_2 = 0.0975 \ + 1.756r,$$
$$x_5 = 0.9024 \ - 1.7561r, \quad \theta = 0.0024 \ - 0.0168r.$$

From the conditions $\delta_1 \geq 0, x_2, x_5 \in [0, 0.6]$, it follows $r \geq 0.171$. As the critical value 0.171 is obtained when $x_5 = 0.6$, at the next step T must be *up*. The solution at Step 4 holds for $r \in [0.171, 0.286]$.

Step 5. As $r < 0.171$, T change its status from *in* to *up*, so $x_5 = 0.6$. Thus, $\delta_5 = 0$. AMZN remains *in*, so $\delta_2 = \eta_2 = 0$. AMAT, BABA and AMD remain *out*, so $x_1 = x_3 = x_4 = 0$ and $\eta_1 = \eta_3 = \eta_4 = 0$. The Kuhn-Tucker system has the solution:

$$\delta_1 = -0.0046 + 0.0417r, \quad \delta_3 = -0.00048 + 0.0505r,$$
$$\delta_4 = -0.0016 + 0.0172r, \quad \eta_5 = 0.00496 \ - 0.0288r,$$
$$x_2 = 0.4, \quad \theta = 0.00692 \ - 0.0428r.$$

From the constraints it follows that this solution holds for $r \geq 0.11$. As $r = 0.11$, the conditin $\delta_1 \geq 0$ is violated, so at the next step δ_1 must be zero, thus $x_1 > 0$.

Step 6. As $r < 0.11$, AMAT change its status from *out* to *in*. Thus, $\delta_1 = \eta_1 = 0$. T remains *up*, so $x_5 = 0.6$ and $\delta_5 = 0$. AMZN remains *in*, so $\delta_2 = \eta_2 = 0$. BABA and AMD remain *out*, so $x_3 = x_4 = 0$ and $\eta_3 = \eta_4 = 0$. The Kuhn-Tucker system has the solution:

$$\delta_3 = 0.002 \ + 0.0275r, \quad \delta_4 = 0.0027 \ - 0.0226r,$$
$$\eta_5 = 0.0025 - 0.0065r, \quad x_1 = 0.1755 \ - 1.5916r,$$
$$x_2 = 0.2244 \ + 1.5916r, \quad \theta = 0.0046 \ - 0.0221r.$$

This solution holds for $0 < r < 0.11$.

3. RESULTS AND COMMENTS

Synthesizing the results from the above section, the optimal portfolio composition for different values of the risk tolerance r is given in Table 3.

Risk tolerance	AMAT	AMZN	BABA	AMD	T
$r \geq 2.224$	0	0.6	0	0.4	0
$0.43 \leq r < 2.224$	0	0.6	0	$-0.0969 + 0.223r$	$0.4969 - 0.223r$
$0.286 \leq r < 0.43$	0	0.6	0	0	0.4
$0.171 \leq r < 0.286$	0	$0.0975 + 1.756r$	0	0	$0.9024 - 1.7561r$
$0.11 \leq r < 0.171$	0	0.4	0	0	0.6
$r < 0.11$	$0.1755 - 1.5916r$	$0.2244 + 1.5916r$	0	0	0.6

Table 3. Optimal weights depending on risk tolerance.

As the investor's risk tolerance r takes its critical values, the corner portfolios are obtained.

They are presented in Table 4, together with the corresponding portfolio expected return. As expected, for big risk tolerance, the expected return is also big.

Critical values of risk tolerance	AMAT	AMZN	BABA	AMD	T	Portfolio expected return
$r = 2.224$	0	0.6	0	0.4	0	0.03592
$r = 0.43$	0	0.6	0	0	0.4	0.03128
$r = 0.286$	0	0.6	0	0	0.4	0.03128
$r = 0.171$	0	0.4	0	0	0.6	0.02552
$r = 0.11$	0	0.4	0	0	0.6	0.02552
$r \to 0$	0.175	0.225	0	0	0.6	0.01822

Table 4. Corner portfolios.

In Figures 1-4, the optimal investment weights for AMAT, AMZN, AMD and T are represented as functions of the risk tolerance using [14]. In these figures, the dots correspond to the corner portfolios.

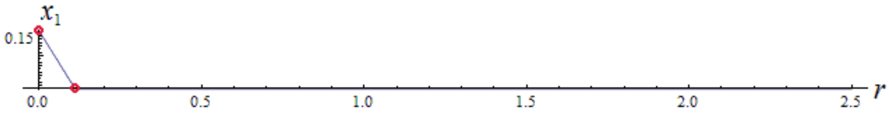


Fig. 1. Optimal AMAT investment weight as function of risk tolerance.

If the parameter r is eliminated between different optimal weights, the critical lines are obtained. Some of these critical lines are represented in Figures 5 and 6, namely those from the (x_5, x_4) and (x_5, x_2) planes. As in the previous figures, the dots correspond to the corner portfolios.

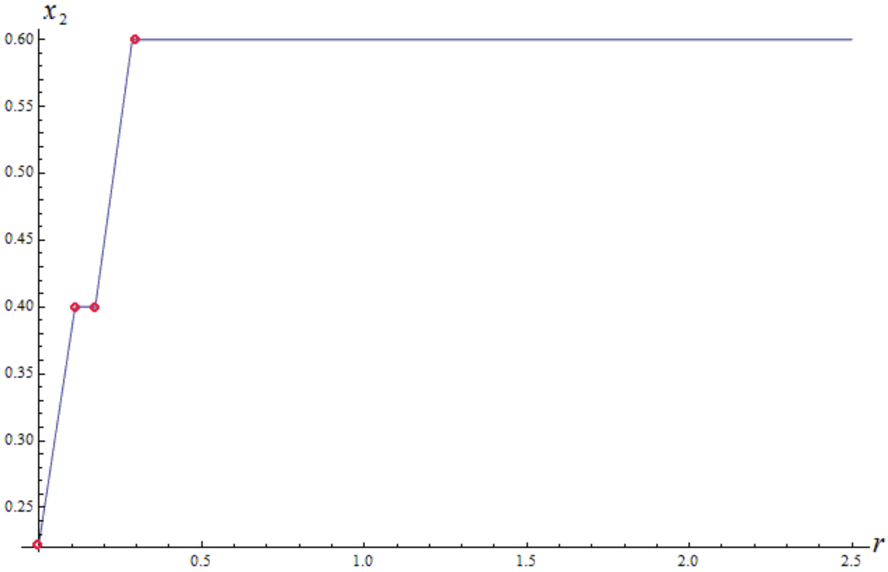


Fig. 2. Optimal AMZN investment weight as function of risk tolerance.

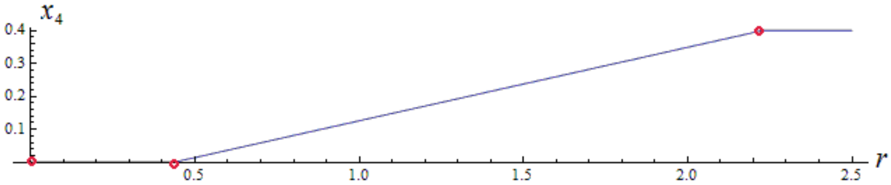


Fig. 3. Optimal AMD investment weight as function of risk tolerance.

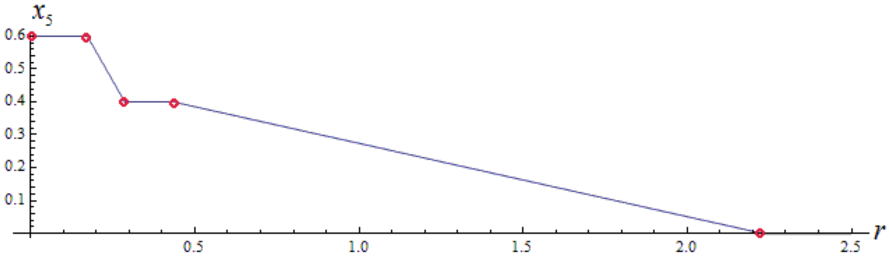


Fig. 4. Optimal T investment weight as function of risk tolerance.

Two diagrams illustrating the optimal portfolio composition for big and small risk tolerance are represented in Figures 7 and 8 respectively.

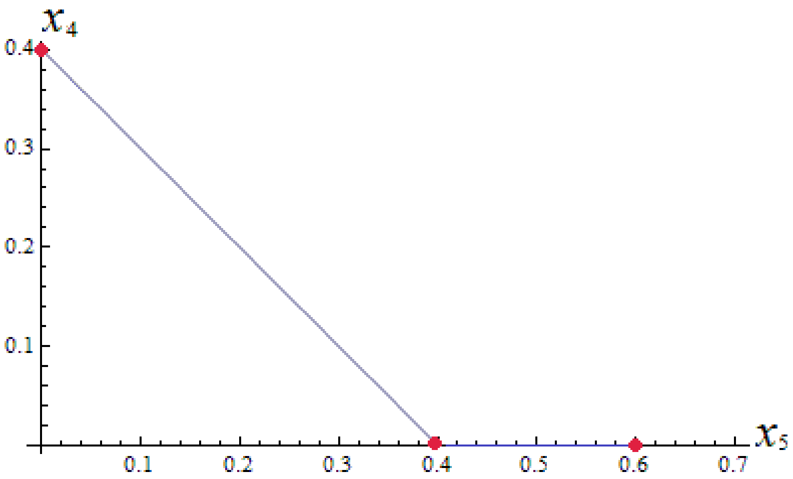


Fig. 5. Critical lines in the (x_5, x_4) -plane.

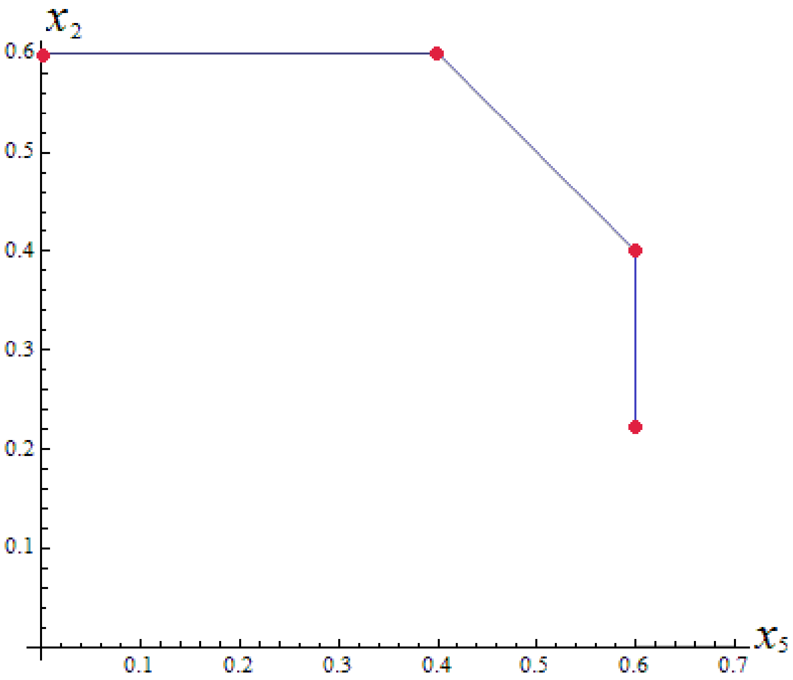


Fig. 6. Critical lines in the (x_5, x_2)-plane.

Finally, the variation of the portfolio expected return with respect to the risk tolerance is given in Table 5.

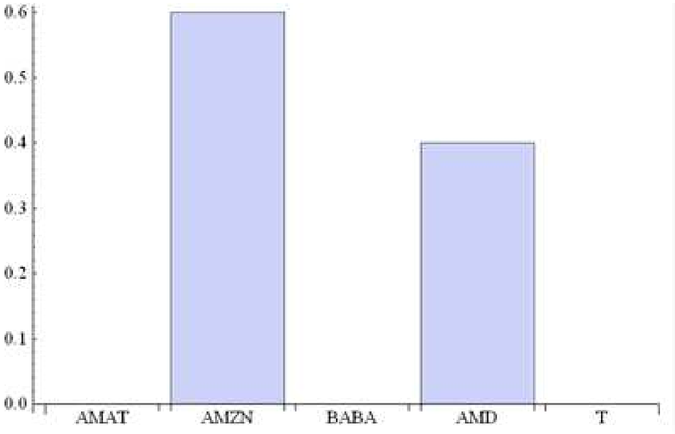


Fig. 7. Optimal portfolio composition for big risk tolerance.

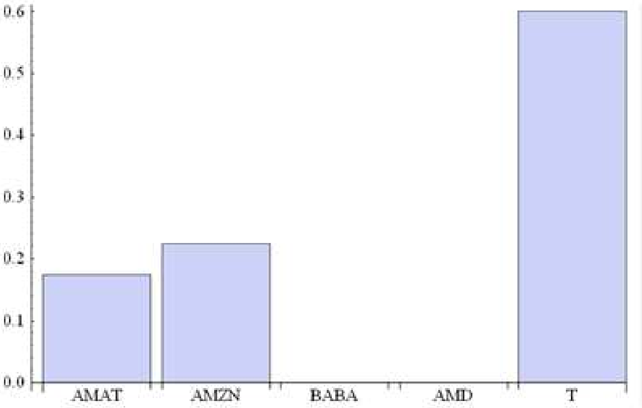


Fig. 8. Optimal portfolio composition for small risk tolerance ($r \to 0$).

Risk tolerance	Portfolio expected return
$0 < r < 0.11$	$0.01819 + 0.06636r$
$0.11 \leq r < 0.171$	0.02552
$0.171 \leq r < 0.286$	$0.01680 + 0.05057r$
$0.286 \leq r < 0.43$	0.03128
$0.43 \leq r < 2.224$	$0.03015 + 0.00258r$
$r \geq 2.224$	0.03592

Table 5. Portfolio expected return depending on risk tolerance.

The portfolio expected return as function of the risk tolerance is represented in Figure 9 as a polygonal line, where the dots correspond to the corner portfolios.

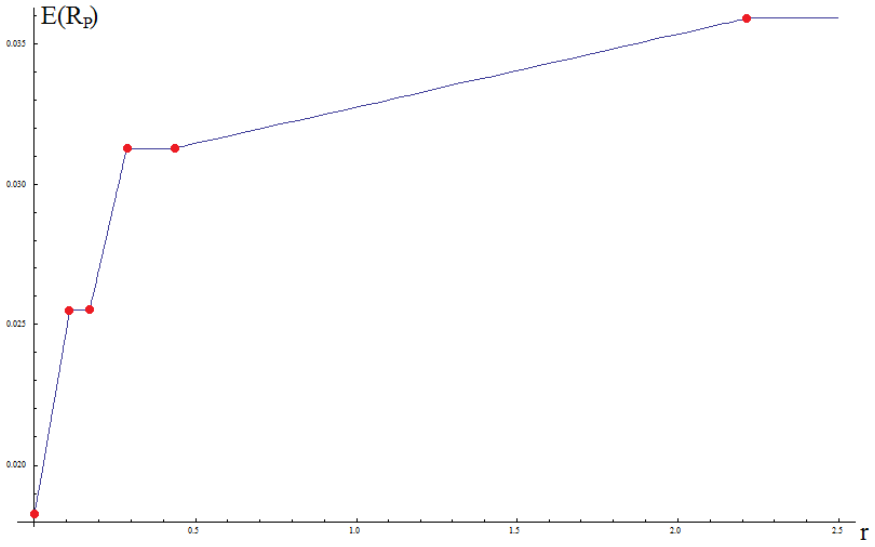


Fig. 9. The portfolio expected return as function of the risk tolerance.

4. CONCLUSIONS

For our case-study, 5 companies were chosen, namely AMAT, AMZN, BABA, AMD and T. We considered that the expected return of every security is the mean of periodic returns percentages obtained between February 2015 and April 2016. Of course, in real life, many other factors, such as the news concerning the firms or the changes of their management strategy influence these expected returns. Thus, our study is just a theoretical one. The aim was to find the optimal portfolio composition with positive weights and investments for every company limited at 60%, for different values of the investor's risk tolerance. We obtained a convex problem, which was solved using the Kuhn-Tucker theory. We found 5 critical values of the risk tolerance and the optimal weights as functions of this tolerance. In our conditions, no investments have to be made in BABA. For big risk tolerance, the optimal portfolio contains AMZN and AMD, while for small risk tolerance, T, AMZN and AMAT have to be selected.

References

[1] H.D. Bailey, M. López de Prado, *An Open-Source Implementation of the Critical-Line Algorithm for Portfolio Optimization*, Algorithms, **6**(2013), 169–196.

[2] C. Calvo, C. Ivorra, V Liern, *On the Computation of the Efficient Frontier of the Portfolio Selection Problem*, Journal of Applied Mathematics, **2012**(2012), Article ID 105616, 25 p.

[3] D. Cumova, D. Moreno, D. Nawrocki, *The Critical Line Algorithm for UPM-LPM Parametric General Asset Allocation Problem with Allocation Boundaries and Linear Constraints*, 2004. Retrieved from http://www90.homepage.villanova.edu/michael.pagano/DN

[4] F. Deari, P.V. Hudym, *Construction of Dynamic Investment Portfolio Management Model*, International Scientific Journal, Economic Sciences, Kiev, **4** (2015), 5–14.

[5] F. Deari, *Portfolio Composition: A Methodological Solution Using Lagrange Multiplier*, ICESOS'15, Sarajevo, 2015.

[6] B. Jacobs, K. Levy, H. Markowitz, *Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions*, Operations Research, **53**(2005), No. 4, 586–599.

[7] R. Kan, GF Zhou, *Tests of Mean-Variance Spanning*, Annals of Economics and Finance, **13**-1(2012), 145–193.

[8] C.C. Kwan, *A Simple Spreadsheet-Based Exposition of the Markowitz Critical Line Method for Portfolio Selection*, Spreadsheets in Education (eJSiE), **2**(2007), No. 3, 30 p.

[9] H. Markowitz, *Portfolio Selection*, Journal of Finance, **7**(1952), No. 1, 77–91.

[10] H. Marling, S. Emanuelsson, *The Markowitz Portfolio Theory*, 2012. Retrieved from http://www.math.chalmers.se/ rootzen/finrisk/gr1_HannesMarling_Sara Emanuelsson_MPT.pdf.

[11] J. Norstad, *Portfolio Optimization, Part 2 Constrained Portfolios*, 2011. Retrieved from http://www.norstad.org/finance/portopt2.pdf

[12] M. Rubinstein, *Markowitz's Portfolio Selection: A Fifty-Year Retrospective*, Journal of Finance, **LVII**(2002), No. 3, 1041–1045.

[13] F.W. Sharpe, P. Chen, E.J. Pinto, W.D. McLeavey, *Asset Allocation*, In: Maginn, L. J., Tuttle, L. D., McLeavey, W. D., Pinto, E. J., Managing Investment Portfolios: A Dynamic Process, CFA Institute, USA, John Wiley & Sons Inc., 2007.

[14] Wolfram Research Inc. *Wolfram Mathematica 9*, 2012.

ROMAI J., v.14, no.1(2018), 115–128

ESTIMATION OF HYPER-ORDER OF SOLUTIONS TO HIGHER ORDER COMPLEX LINEAR DIFFERENTIAL EQUATIONS WITH ENTIRE COEFFICIENTS OF SLOW GROWTH

Amina Ferraoun, Benharrat Belaïdi

Department of Mathematics, Laboratory of Pure and Applied Mathematics,

University of Mostaganem (UMAB), Mostaganem, Algeria

aferraoun@yahoo.fr, benharrat.belaidi@univ-mosta.dz

Abstract In this paper, we study the growth of meromorphic solutions of higher order linear differential equations with entire coefficients and we obtain some estimations on the hyper-order and hyper convergence exponent of zeros of these solutions. We extend some results due to C. Y. Zhang, J. Tu [16]; L. Wang, H. Liu [14].

Keywords: entire functions, meromorphic functions, differential equations, growth, order.
2010 MSC: 34M10, 30D35.

1. INTRODUCTION

We assume that the reader is familiar with the fundamental results and the standard notations of Nevanlinna's theory (see e.g. [10, 12, 15]).

Definition 1.1. *The order of a meromorphic function f is defined as*

$$\sigma(f) = \limsup_{r\to+\infty} \frac{\log T(r,f)}{\log r},$$

here $T(r,f)$ is the Nevanlinna characteristic function of f which is defined by

$$T(r,f) = N(r,f) + m(r,f)\,, \quad (r>0)\,,$$

where

$$N(r,f) = \int_0^r \frac{[n(t,\infty,f) - n(0,\infty,f)]}{t} dt + n(0,\infty,f)\log r,$$

$$m(r,f) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ \left| f\left(re^{i\theta}\right)\right| d\theta$$

and $n(t,\infty,f)$ denote the number of poles of f in the disc $|z|\leq t$. If f is an entire function, then

$$\sigma(f)=\limsup_{r\to+\infty}\frac{\log\log M(r,f)}{\log r},$$

where $M(r,f)=\max_{|z|=r}|f(z)|$.

Definition 1.2. *The hyper-order of a meromorphic function f is defined as*

$$\sigma_2(f)=\limsup_{r\to+\infty}\frac{\log\log T(r,f)}{\log r}.$$

If $f(z)$ is an entire function, then

$$\sigma_2(f)=\limsup_{r\to+\infty}\frac{\log\log\log M(r,f)}{\log r}.$$

Definition 1.3. *The lower order of a meromorphic function f is defined as*

$$\mu(f)=\liminf_{r\to+\infty}\frac{\log T(r,f)}{\log r}.$$

If f is an entire function, then

$$\mu(f)=\liminf_{r\to+\infty}\frac{\log\log M(r,f)}{\log r}.$$

Definition 1.4. *The convergence exponent of zeros and distinct zeros of a meromorphic function f are respectively defined by*

$$\lambda(f)=\limsup_{r\to+\infty}\frac{\log N(r,\frac{1}{f})}{\log r},\qquad \overline{\lambda}(f)=\limsup_{r\to+\infty}\frac{\log \overline{N}(r,\frac{1}{f})}{\log r},$$

where $N\left(r,\frac{1}{f}\right)$ $\left(\overline{N}\left(r,\frac{1}{f}\right)\right)$ is the integrated counting function of zeros (distinct zeros) of f in $\{z:|z|\leq r\}$.

Definition 1.5. *The hyper convergence exponent of zeros and distinct zeros of a meromorphic function f are respectively defined by*

$$\lambda_2(f)=\limsup_{r\to+\infty}\frac{\log\log N(r,\frac{1}{f})}{\log r},\qquad \overline{\lambda}_2(f)=\limsup_{r\to+\infty}\frac{\log\log \overline{N}(r,\frac{1}{f})}{\log r}.$$

Definition 1.6. *Let $f(z)=\sum\limits_{n=0}^{+\infty}a_nz^n$ be an entire function. We denote by $\mu(r)=\max\{|a_n|r^n:\ n=0,1...\}$ the maximal term of f. Then the central* *index of f is defined by*

$$\nu_f(r) = \max\{m; \mu(r) = |a_m| r^m\}.$$

In the past years, many authors investigated the growth of solutions of the higher order linear differential equation

$$f^{(k)} + A_{k-1}(z)f^{(k-1)} + \cdots + A_1(z)f' + A_0(z)f = F(z), \tag{1.1}$$

when $A_j(z)$ $(j = 0, 1, \cdots, k-1)$, $F(z)(\not\equiv 0)$ are entire (or meromorphic) functions and obtained some valuable results, (see e.g. [2, 3, 4, 11, 12, 13, 14, 16]). In 2014, Wang and Liu investigated the properties of solutions of equation (1.1) when there exists some coefficient $A_s(z)$ $(0 \leq s \leq k-1)$ verifying the condition $\mu(A_s) < \frac{1}{2}$ and obtained the following result.

Theorem A [14] *Suppose that $A_0(z), \cdots, A_{k-1}(z)$, $F(z)$ are meromorphic functions of finite order. If there exists some $s \in \{0, 1, \cdots, k-1\}$ such that*

$$b = \max\left\{\sigma(A_j), (j \neq s), \sigma(F), \lambda\left(\frac{1}{A_s}\right)\right\} < \mu(A_s) < \frac{1}{2},$$

then
(i) Every transcendental meromorphic solution f of (1.1) whose poles are of uniformly bounded multiplicities, satisfies $\mu(A_s) \leq \sigma_2(f) \leq \sigma(A_s)$. Furthermore, if $F \not\equiv 0$, then we have $\mu(A_s) \leq \bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_s)$.
(ii) If $s \geq 2$, then every non-transcendental meromorphic solution f of (1.1) is a polynomial with $\deg f \leq s-1$. If $s = 0$ or 1, then every nonconstant solution f of (1.1) is transcendental.

When $F(z)$ is of infinite order, Wang and Liu considered the linear differential equation

$$f^{(k)} + A_{k-1}(z)f^{(k-1)} + \cdots + A_1(z)f' + A_0(z)f = Qe^P, \tag{1.2}$$

when $A_j(z)$ $(j = 0, 1, \cdots, k-1)$, $Q(z)(\not\equiv 0)$ are meromorphic functions and P is a transcendental entire function and obtained the following result.

Theorem B [14] *Suppose that $A_0(z), \cdots, A_{k-1}(z)$, $Q(z)(\not\equiv 0)$ are meromorphic functions of finite order, P is a transcendental entire function such that*

$$\max\left\{\sigma(P), \sigma(Q), \sigma(A_j), (1 \leq j \leq k-1), \lambda\left(\frac{1}{A_0}\right)\right\} < \mu(A_0) < \frac{1}{2}.$$

Then every solution f of (1.2) *is transcendental, and every transcendental meromorphic solution f of* (1.2) *whose poles are of uniformly bounded multiplicities satisfies $\mu(A_0) \leq \overline{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_0)$.*

For $k \geq 2$, we consider the linear differential equation

$$A_k(z)f^{(k)} + A_{k-1}(z)f^{(k-1)} + \cdots + A_1(z)f' + A_0(z)f = F(z), \tag{1.3}$$

when $A_j(z)$ $(j = 0, 1, \cdots, k)$, $F(z)$ are entire functions such that $A_0A_kF \not\equiv 0$. It well-known that if $A_k(z) \equiv 1$, then all solutions of (1.3) are entire functions, but when $A_k(z)$ is a nonconstant entire function, then equation (1.3) can possess meromorphic solutions. For instance the equation

$$zf''' + 4f'' + \left(-1 - \frac{1}{2}z^2 - z\right)e^{-z}f' + \left(\left(1 - \frac{1}{2}z^2 + 2z\right)e^{-2z} + ze^{-3z}\right)f$$
$$= \left(-1 - \frac{1}{2}z^2 - z\right)e^{-z} + \left(z - \frac{1}{2}z^3 + 2z^2\right)e^{-2z} + z^2e^{-3z}$$

has a meromorphic solution $f(z) = \frac{1}{z^2}e^{e^{-z}} + z$. Thus, there exist two questions. Firstly, can we have the same properties as in Theorem A for the linear differential equation (1.3), when there exists some coefficient $A_s(z)$ $(0 \leq s \leq k)$ verifying the condition $\mu(A_s) < \frac{1}{2}$? Secondly, how about the growth of meromorphic solutions of the linear differential equation

$$A_k(z)f^{(k)} + A_{k-1}(z)f^{(k-1)} + \cdots + A_1(z)f' + A_0(z)f = Qe^P, \tag{1.4}$$

when $A_j(z)$ $(j = 0, 1, \cdots, k)$, $Q(z)(\not\equiv 0)$ are entire functions and P is a transcendental entire function? In this paper, we proceed this way and we obtain the following results.

Theorem 1.1. *Suppose that $A_0(z), \cdots, A_k(z)$, $F(z)$ are entire functions of finite order. If there exists some $s \in \{0, 1, \cdots, k\}$ such that*

$$\alpha = \max\{\sigma(A_j), (j \neq s), \sigma(F)\} < \mu(A_s) < \frac{1}{2}, \tag{1.5}$$

then
(i) Every transcendental meromorphic solution f of (1.3) *such that $\lambda\left(\frac{1}{f}\right) < \mu(f)$ satisfies $\mu(A_s) \leq \sigma_2(f) \leq \sigma(A_s)$. Furthermore, if $F \not\equiv 0$, then we have $\mu(A_s) \leq \overline{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_s)$.*
(ii) If $s \geq 2$, then every rational solution f of (1.3) *is a polynomial with* $\deg f \leq s - 1$. *If $s = 0$ or 1, then every nonconstant solution f of* (1.3) *is transcendental.*

Remark 1.1. Setting $A_k(z) \equiv 1$ in Theorem 1.1 we obtain the result of Zhang and Tu ([16], Theorem 1.8).

Corollary 1.1 *Suppose that* $A_0(z), \cdots, A_k(z)$, $F(z)(\not\equiv 0)$ *are entire functions. If there exists some* $s \in \{0, 1, \cdots, k\}$ *such that*

$$\alpha = \max\{\sigma(A_j), (j \neq s), \sigma(F)\} < \mu(A_s) = \sigma(A_s) < \frac{1}{2},$$

then every transcendental meromorphic solution f *of* (1.3) *such that* $\lambda\left(\frac{1}{f}\right) < \mu(f)$ *satisfies* $\bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) = \sigma(A_s)$, *and every rational solution* f *of* (1.3) *is a polynomial with* $\deg f \leq s - 1$.

Theorem 1.2. *Suppose that* $A_0(z), \cdots, A_k(z)$, $Q(z)(\not\equiv 0)$ *are entire functions of finite order,* P *is a transcendental entire function such that*

$$\max\{\sigma(P), \sigma(Q), \sigma(A_j), (1 \leq j \leq k)\} < \mu(A_0) < \frac{1}{2}. \tag{1.6}$$

Then every solution f *of* (1.4) *is transcendental, and every transcendental meromorphic solution* f *of* (1.4) *such that* $\lambda\left(\frac{1}{f}\right) < \mu(f)$ *satisfies* $\mu(A_0) \leq \bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_0)$.

Remark 1.2. In Theorems 1.1 and 1.2, we remove the restriction $\lambda\left(\frac{1}{A_s}\right) < \mu(A_s)$.

Corollary 1.2 *Suppose that* $A_0(z), \cdots, A_k(z)$, $Q(z)(\not\equiv 0)$ *are entire functions of finite order,* P *is a transcendental entire function such that*

$$\max\{\sigma(P), \sigma(Q), \sigma(A_j), (1 \leq j \leq k)\} < \mu(A_0) = \sigma(A_0) < \frac{1}{2}.$$

Then every solution f *of* (1.4) *is transcendental, and every transcendental meromorphic solution* f *of* (1.4) *such that* $\lambda\left(\frac{1}{f}\right) < \mu(f)$ *satisfies* $\bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) = \sigma(A_0)$.

Remark 1.3. Obviously, Theorem 1.1 and Theorem 1.2 are generalization of Theorems A, B of Wang and Liu [14] and Theorem 1.8 of Zhang and Tu [16].

2. PRELIMINARY LEMMAS

Lemma 2.1 [8] *Let* f *be a transcendental meromorphic function in the plane, and let* $\alpha > 1$ *be a given constant. Then there exist a set* $E_1 \subset (1, +\infty)$ *that* *has a finite logarithmic measure, and a constant $B > 0$ depending only on α and (m, n) $(m, n \in \{0, 1, \cdots, k\})$ $m < n$ such that for all z with $|z| = r \notin [0, 1] \cup E_1$, we have*

$$\left|\frac{f^{(n)}(z)}{f^{(m)}(z)}\right| \leq B\left(\frac{T(\alpha r, f)}{r}(\log^{\alpha} r)\log T(\alpha r, f)\right)^{n-m}.$$

Lemma 2.2 [6] *Let $f(z) = \frac{g(z)}{d(z)}$ be a meromorphic function, where $g(z)$ and $d(z)$ are entire functions satisfying $\mu(g) = \mu(f) = \mu \leq \sigma(g) = \sigma(f) \leq +\infty$ and $\lambda(d) = \sigma(d) = \lambda(\frac{1}{f}) < \mu$. Then there exists a set $E_2 \subset (1, +\infty)$ of finite logarithmic measure, such that for all z satisfying $|z| = r \notin [0, 1] \cup E_2$ and $|g(z)| = M(r, g)$ we have*

$$\left|\frac{f(z)}{f^{(s)}(z)}\right| \leq r^{2s}, \ (s \in \mathbb{N}).$$

Lemma 2.3 [9] *Let $g : [0, +\infty) \to \mathbb{R}$ and $h : [0, +\infty) \to \mathbb{R}$ be monotone nondecreasing functions such that $g(r) \leq h(r)$ for all $r \notin E_3 \cup [0, 1]$, where $E_3 \subset (1, +\infty)$ is a set of finite logarithmic measure. Then for any $\alpha > 1$, there exists an $r_0 = r_0(\alpha) > 0$ such that $g(r) \leq h(\alpha r)$ for all $r > r_0$.*

Lemma 2.4 [6] *Let $f(z) = \frac{g(z)}{d(z)}$ be a meromorphic function, where $g(z)$ and $d(z)$ are entire functions satisfying $\mu(g) = \mu(f) = \mu \leq \sigma(g) = \sigma(f) \leq +\infty$ and $\lambda(d) = \sigma(d) = \lambda(\frac{1}{f}) < \mu$. Then there exists a set $E_4 \subset (1, +\infty)$ of finite logarithmic measure, such that for all z satisfying $|z| = r \notin [0, 1] \cup E_4$ and $|g(z)| = M(r, g)$, we have*

$$\frac{f^{(n)}(z)}{f(z)} = \left(\frac{\nu_g(r)}{z}\right)^n (1 + o(1)), \quad (n \geq 1),$$

where $\nu_g(r)$ denote the central index of $g(z)$.

Lemma 2.5 [5] *Let $g(z)$ be an entire function of order $\sigma(g) = \alpha < \infty$. Then for any $\varepsilon > 0$, there exist a set $E_5 \subset [1, +\infty)$ that has a finite linear measure and finite logarithmic measure, such that for all z satisfying $|z| = r \notin [0, 1] \cup E_5$, we have*

$$\exp\{-r^{\alpha+\varepsilon}\} \leq |g(z)| \leq \exp\{r^{\alpha+\varepsilon}\}.$$

Lemma 2.6 [7] *Let $g(z)$ be an entire function of infinite order, with the hyper-order $\sigma_2(g) = \sigma$, and $\nu_g(r)$ denote the central index of $g(z)$. Then*

$$\limsup_{r\to+\infty} \frac{\log\log \nu_g(r)}{\log r} = \sigma.$$

Lemma 2.7 [1] *Let $g(z)$ be an entire function with $0 \le \mu(g) < 1$. Then for every $\alpha \in (\mu(g), 1)$, there exists a set $E_6 \subset [0, \infty)$ such that*

$$\overline{\log dens} E_6 \ge 1 - \frac{\mu(g)}{\alpha},$$

where $E_6 = \{r \in [0, \infty) : m(r) > M(r) \cos \pi\alpha\}$, $m(r) = \inf\limits_{|z|=r} \log |g(z)|$, $M(r) = \sup\limits_{|z|=r} \log |g(z)|$.

Lemma 2.8 *Let $f(z)$ be an entire function such that $\mu(f) < \frac{1}{2}$. Then for any given $\varepsilon > 0$, there exists a set $E_7 \subset (1, +\infty)$ with $\overline{\log dens} E_7 > 0$, such that for all z satisfying $|z| = r \in E_7$, we have*

$$|f(z)| \ge \exp\{r^{\mu(f)-\varepsilon}\}.$$

Proof. Set $\alpha_0 = \frac{\frac{1}{2}+\mu(f)}{2}$. Then, by Lemma 2.7, there exists a set H with $\overline{\log dens} H \ge 1 - \frac{\mu(f)}{\alpha_0}$, such that for all z satisfying $|z| = r \in H$, we have

$$\log |f(z)| \ge \cos(\pi\alpha_0) \log M(r, f). \tag{2.1}$$

By the definition of the lower order, for any given $\varepsilon > 0$, there exists $r_1 > 0$ such that

$$\log M(r, f) \ge r^{\mu(f)-\frac{\varepsilon}{2}}, \tag{2.2}$$

holds for $r > r_1$. Since

$$\frac{\cos(\pi\alpha_0) r^{\mu(f)-\frac{\varepsilon}{2}}}{r^{\mu(f)-\varepsilon}} \to +\infty, \ (r \to +\infty), \tag{2.3}$$

then, by (2.1) – (2.3), there exists $r_2 (\ge r_1)$, such that for all z satisfying $|z| = r \in H \setminus [0, r_2]$, we have

$$|f(z)| \ge \exp\left\{\cos(\pi\alpha_0) r^{\mu(f)-\frac{\varepsilon}{2}}\right\} \ge \exp\{r^{\mu(f)-\varepsilon}\}.$$

Setting $E_7 = H \cap [r_2, +\infty]$, then $\overline{\log dens} E_7 > 0$. ∎

Lemma 2.9 [15] *Let f, g be nonconstant meromorphic functions with $\sigma(f)$ as order and $\mu(g)$ as lower order. Then we have*

$$\mu(f+g) \leq \max\{\sigma(f), \mu(g)\}$$

and

$$\mu(fg) \leq \max\{\sigma(f), \mu(g)\}.$$

Furthermore, if $\mu(g) > \sigma(f)$, then we obtain

$$\mu(f+g) = \mu(fg) = \mu(g).$$

3. PROOF OF THEOREM 1.1

(i) Assume that f is a transcendental meromorphic solution of (1.3) such that $\lambda\left(\frac{1}{f}\right) < \mu(f)$. From (1.3), we obtain

$$|A_s(z)| \leq \left|\frac{f}{f^{(s)}}\right| \left[|A_k(z)|\left|\frac{f^{(k)}}{f}\right| + |A_{k-1}(z)|\left|\frac{f^{(k-1)}}{f}\right| + \cdots + |A_{s+1}(z)|\left|\frac{f^{(s+1)}}{f}\right|\right.$$

$$\left. + |A_{s-1}(z)|\left|\frac{f^{(s-1)}}{f}\right| + \cdots + |A_1(z)|\left|\frac{f'}{f}\right| + |A_0(z)| + \left|\frac{F}{f}\right|\right]. \quad (3.1)$$

By Lemma 2.1, there exists a constant $B > 0$ and a set $E_1 \subset (1, +\infty)$ of finite logarithmic measure such that for all z satisfying $|z| = r \notin [0,1] \cup E_1$, we have

$$\left|\frac{f^{(j)}(z)}{f(z)}\right| \leq B\left(T(2r, f)\right)^{k+1}, \quad 1 \leq j \leq k. \quad (3.2)$$

Since $\lambda\left(\frac{1}{f}\right) < \mu(f)$, then by Hadamard's factorization theorem, we can write f as $f(z) = \frac{g(z)}{d(z)}$, where $g(z)$ and $d(z)$ are entire functions satisfying

$$\mu(g) = \mu(f) = \mu \leq \sigma(g) = \sigma(f), \ \sigma(d) = \lambda\left(\frac{1}{f}\right) < \mu.$$

Then by Lemma 2.2, there exists a set E_2 of finite logarithmic measure such that for all $|z| = r \notin [0,1] \cup E_2$ and $|g(z)| = M(r, g)$ and for r sufficiently large, we have

$$\left|\frac{f(z)}{f^{(s)}(z)}\right| \leq r^{2s}. \quad (3.3)$$

By (1.5), for any given ε with $0 < 2\varepsilon < \mu(A_s) - \alpha$, we have for sufficiently large r

$$|A_j(z)| \leq \exp\{r^{\alpha+\varepsilon}\}, \ (j \neq s), \ \ |F(z)| \leq \exp\{r^{\alpha+\varepsilon}\}. \tag{3.4}$$

By Lemma 2.8, for any given $\varepsilon > 0$, there exists a set $E_7 \subset (1, +\infty)$ with $\overline{\log dens} E_7 > 0$, such that for all z satisfying $|z| = r \in E_7$, we have

$$|A_s(z)| \geq \exp\{r^{\mu(A_s)-\varepsilon}\}. \tag{3.5}$$

Since $\sigma(d) = \lambda\left(\frac{1}{f}\right) < \mu(f) = \mu(g)$, then for any ε with $0 < 2\varepsilon < \mu(f) - \lambda\left(\frac{1}{f}\right)$ and for sufficiently large r we have

$$\left|\frac{F\,(z)}{f\,(z)}\right| = \left|\frac{d(z)}{g(z)}\right| |F\,(z)| = \left|\frac{d(z)}{M(r,g)}\right| |F\,(z)|$$

$$\leq \frac{\exp\{r^{\lambda\left(\frac{1}{f}\right)+\varepsilon}\}}{\exp\{r^{\mu(f)-\varepsilon}\}} \exp\{r^{\alpha+\varepsilon}\} \leq \exp\{r^{\alpha+\varepsilon}\}. \tag{3.6}$$

Let $E_8 = E_7 \diagdown ([0,1] \cup E_1 \cup E_2)$, then we have $\overline{\log dens} E_8 > 0$. Then, by substituting (3.2) – (3.6) into (3.1), for all z satisfying $|z| = r \in E_8$ and $|g(z)| = M(r,g)$, we obtain

$$\exp\{r^{\mu(A_s)-\varepsilon}\} \leq B(k+1) r^{2s} (T(2r,f))^{k+1} \exp\{r^{\alpha+\varepsilon}\}. \tag{3.7}$$

By (3.7) and Lemma 2.3, we get $\mu(A_s) - \varepsilon \leq \sigma_2(f)$. Since $\varepsilon > 0$ is arbitrary, we have $\mu(A_s) \leq \sigma_2(f)$. Now, we prove that $\sigma_2(f) \leq \sigma(A_s)$. We can write (1.3) as

$$-A_k(z)\frac{f^{(k)}}{f} = A_{k-1}(z)\frac{f^{(k-1)}}{f} + \cdots + A_{s+1}(z)\frac{f^{(s+1)}}{f}$$

$$+A_s(z)\frac{f^{(s)}}{f} + A_{s-1}(z)\frac{f^{(s-1)}}{f} + \cdots + A_1(z)\frac{f'}{f} + A_0(z) - \frac{F(z)}{f(z)}. \tag{3.8}$$

By Lemma 2.4, there exists a set $E_4 \subset (1, +\infty)$ of finite logarithmic measure such that for all $|z| = r \notin [0,1] \cup E_4$ and $|g(z)| = M(r,g)$, we have

$$\frac{f^{(j)}(z)}{f(z)} = \left(\frac{\nu_g(r)}{z}\right)^j (1 + o(1)), \quad (j = 1, \cdots, k)\,. \tag{3.9}$$

For any given $\varepsilon > 0$, for sufficiently large r we have

$$|A_j(z)| \leq \exp\{r^{\sigma(A_s)+\varepsilon}\}, \ j = 0, \cdots, k-1. \tag{3.10}$$

By Lemma 2.5, for any given $\varepsilon > 0$, there exists a set $E_5 \subset (1, +\infty)$ of finite logarithmic measure such that for all z satisfying $|z| = r \notin [0, 1] \cup E_5$, we have

$$|A_k(z)| \geq \exp\{-r^{\sigma(A_k)+\varepsilon}\} \geq \exp\{-r^{\sigma(A_s)+\varepsilon}\}. \tag{3.11}$$

From (3.8) and (3.9), we have

$$-\left(\frac{\nu_g(r)}{z}\right)^k (1 + o(1)) = \frac{1}{A_k(z)} \left[\sum_{j=1}^{k-1} A_j(z) \left(\frac{\nu_g(r)}{z}\right)^j (1 + o(1)) \right.$$
$$\left. + A_0(z) - \frac{F(z)}{f(z)}\right],$$

it follows

$$\left|\left(\frac{\nu_g(r)}{z}\right)^k\right| |1 + o(1)| \leq \frac{1}{|A_k(z)|} \left[\sum_{j=1}^{k-1} |A_j(z)| \left|\left(\frac{\nu_g(r)}{z}\right)^j\right| |1 + o(1)| \right.$$
$$\left. + |A_0(z)| + \left|\frac{F(z)}{f(z)}\right|\right]. \tag{3.12}$$

By (3.6) and (3.10) – (3.12) for all z satisfying $|z| = r \notin [0, 1] \cup E_4 \cup E_5$ and $|g(z)| = M(r, g)$, we have

$$\left(\frac{\nu_g(r)}{r}\right) |1 + o(1)| \leq (k + 1) |1 + o(1)| \exp\{r^{\sigma(A_s)+\varepsilon}\},$$

so,

$$\limsup_{r\to+\infty} \frac{\log\log \nu_g(r)}{\log r} \leq \sigma(A_s) + \varepsilon. \tag{3.13}$$

Since $\varepsilon > 0$ is arbitrary, then by (3.13), Lemma 2.3 and Lemma 2.6, we have $\sigma_2(g) \leq \sigma(A_s)$, that is $\sigma_2(f) \leq \sigma(A_s)$. Therefore, we get

$$\mu(A_s) \leq \sigma_2(f) \leq \sigma(A_s).$$

Let $F \not\equiv 0$. Now, we prove $\bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f)$. By (1.3), we have

$$\frac{1}{f} = \frac{1}{F}\left(A_k\frac{f^{(k)}}{f} + A_{k-1}\frac{f^{(k-1)}}{f} + \cdots + A_1\frac{f'}{f} + A_0\right). \tag{3.14}$$

If f has a zero at z_0 of order $\gamma > k$, then F has a zero at z_0 of order $\gamma - k$. Hence we have

$$n(r, \frac{1}{f}) \leq k\overline{n}(r, \frac{1}{f}) + n(r, \frac{1}{F}),$$

$$N(r, \frac{1}{f}) \leq k\overline{N}(r, \frac{1}{f}) + N(r, \frac{1}{F}). \tag{3.15}$$

By (3.14), we have by the lemma of logarithmic derivative [10]

$$m(r, \frac{1}{f}) \leq m(r, \frac{1}{F}) + \sum_{j=0}^{k} m(r, A_j) + O(\log rT(r, f)), \ (r \notin E), \tag{3.16}$$

where E is a set of a finite linear measure. By (3.15) and (3.16), we get

$$T(r, f) \leq k\overline{N}(r, \frac{1}{f}) + T(r, F) + \sum_{j=0}^{k} T(r, A_j) + O(\log rT(r, f)), \ (r \notin E). \tag{3.17}$$

For sufficiently large r and any given $\varepsilon > 0$, we have

$$O(\log rT(r, f)) = o\left(T\left(r, f\right)\right), \tag{3.18}$$

$$T(r, F) + \sum_{j=0}^{k} T(r, A_j) \leq (k + 2)\, r^{\sigma(A_s)+\varepsilon}. \tag{3.19}$$

Hence, from (3.17), (3.18) and (3.19) , for sufficiently large $r \notin E$, we get that

$$(1 - o(1))\, T(r, f) \leq k\overline{N}(r, \frac{1}{f}) + (k + 2)\, r^{\sigma(A_s)+\varepsilon},$$

so $\sigma_2(f) \leq \bar{\lambda}_2(f)$. Since $\bar{\lambda}_2(f) \leq \sigma_2(f)$, we get

$$\mu(A_s) \leq \bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_s).$$

(ii) Assume that f is a rational solution of (1.3). If either f is a rational function, which has a pole at z_0 of degree $m \geq 1$, or f is a polynomial with $\deg f \geq s$, then $f^{(s)}(z) \not\equiv 0$. From (1.3), we obtain

$$|A_s(z)| \leq \Bigg[|A_k(z)| \left| \frac{f^{(k)}}{f^{(s)}} \right| + |A_{k-1}(z)| \left| \frac{f^{(k-1)}}{f^{(s)}} \right| + \cdots + |A_{s+1}(z)| \left| \frac{f^{(s+1)}}{f^{(s)}} \right|$$

$$+ |A_{s-1}(z)| \left| \frac{f^{(s-1)}}{f^{(s)}} \right| + \cdots + |A_1(z)| \left| \frac{f'}{f^{(s)}} \right| + |A_0(z)| + \left| \frac{1}{f^{(s)}} \right| |F| \Bigg]. \tag{3.20}$$

Then, by substituting (3.4) and (3.5) into (3.20), we obtain

$$\exp\{r^{\mu(A_s)-\varepsilon}\} \leq (k+1)r^M \exp\{r^{\alpha+\varepsilon}\},$$

where M is a constant. This is a contradiction. Therefore, f must be a polynomial with $\deg f \leq s-1$.

If $s=0$ or 1 and f is a polynomial solution of (1.3), then we get that $\deg f \leq s-1$. Thus, f must be a constant. Therefore, every nonconstant solution $f(z)$ of (1.3) is transcendental.

4. PROOF OF THEOREM 1.2

By hypothesis, it is known that every meromorphic solution of (1.4) is of infinite order. Then, every meromorphic solution of (1.4) is transcendental. Assume that f is a transcendental meromorphic solution, such that $\lambda\left(\frac{1}{f}\right) < \mu(f)$. Set $f = ge^P$. Then, we get that

$$\bar{\lambda}_2(g) = \bar{\lambda}_2(f), \quad \lambda_2(g) = \lambda_2(f). \tag{4.1}$$

By substituting $f = ge^P$ into (1.4), we have

$$g^{(k)} + B_{k-1}(z)g^{(k-1)} + \cdots + B_1(z)g' + B_0(z)g = \frac{Q}{A_k(z)}, \tag{4.2}$$

where

$$B_{k-1} = \frac{A_{k-1}}{A_k} + kP', \tag{4.3}$$

$$B_{k-j} = \frac{A_{k-j}}{A_k} + (k-j+1)\frac{A_{k-j+1}}{A_k}P'$$
$$+\sum_{m=2}^{j} \frac{A_{k-j+m}}{A_k}\left[\binom{k-j+m}{m}(P')^m + D_{m-1}(P')\right],\ j=2,...,k \tag{4.4}$$

and $D_{m-1}(P')$ is a differential polynomial in P' of degree $m-1$, its coefficients are constants. By (4.4) , we get

$$B_0 = \frac{A_0}{A_k} + \frac{A_1}{A_k}P' + \sum_{m=2}^{k} \frac{A_m}{A_k}\left[(P')^m + D_{m-1}(P')\right]$$
$$= \frac{1}{A_k}\left[A_0 + A_1P' + \sum_{m=2}^{k} A_m\left[(P')^m + D_{m-1}(P')\right]\right]. \tag{4.5}$$

Using (1.6) , (4.3) , (4.4) , (4.5) and Lemma 2.9, we obtain

$$\mu(B_0) = \max\left\{\mu(A_0), \sigma(A_j) \ \ (1 \leq j \leq k)\right\} = \mu(A_0) \tag{4.6}$$

and

$$\sigma\left(\frac{Q}{A_k}\right) \leq \max\left\{\sigma(A_k), \sigma(Q)\right\} < \mu(A_0),\ \sigma(B_j) < \mu(A_0),\ j = 1, \cdots, k-1. \tag{4.7}$$

By (4.2) , (4.6) , (4.7) and applying Theorem 1.1 for $A_k(z) \equiv 1$ and $s = 0$, we get

$$\mu(A_0) \leq \bar{\lambda}_2(g) = \lambda_2(g) = \sigma_2(g) \leq \sigma(A_0). \tag{4.8}$$

Since $\sigma_2(e^P) = \sigma(P) < \mu(A_0) \leq \sigma_2(g)$, then we obtain $\sigma_2(f) = \sigma_2(g)$. Hence, by (4.1) and (4.8) , we have

$$\mu(A_0) \leq \bar{\lambda}_2(f) = \lambda_2(f) = \sigma_2(f) \leq \sigma(A_0).$$

Acknowledgement. The authors would like to thank the anonymous referees for their valuable comments and suggestions to improve the quality of the paper.

References

[1] P. D. Barry, *Some theorems related to the* $\cos\pi\rho$ *theorem*, Proc. London Math. Soc., (3) 21(1970), 334–360.

[2] B. Belaïdi, H. Habib, *On the growth of solutions to non-homogeneous linear differential equations with entire coefficients having the same order*, Facta Univ. Ser. Math. Inform., **28**, 1(2013), 17–26.

[3] B. Belaïdi, H. Habib, *On the growth of solutions of some non-homogeneous linear differential equations*, Acta Math. Acad. Paedagog. Nyházi, (N.S.), **32**, 1(2016), 101-111.

[4] T. B. Cao, J. F. Xu, Z. X. Chen, *On the meromorphic solutions of linear differential equations on the complex plane*, J. Math. Anal. Appl., **364**, 1(2010), 130–142.

[5] Z. X. Chen, *On the hyper order of solutions of some second order linear differential equations*, Acta Math. Sin. (Engl. Ser.), **18**, 1(2002), 79–88.

[6] Z.X. Chen, *The rate of growth of meromorphic solutions of higher-order linear differential equations*, Acta Math. Sinica (Chin. Ser.), **42**, 3(1999), 551–558. (Chinese)

[7] Z. X. Chen, C. C. Yang, *Some further results on the zeros and growths of entire solutions of second order linear differential equations*, Kodai Math. J., **22**, 2(1999), 273–285.

[8] G. G. Gundersen, *Estimates for the logarithmic derivative of a meromorphic function, plus similar estimates.* J. London Math. Soc., **(2) 37**, 1(1988), 88–104.

[9] G. G. Gundersen, *Finite order solutions of second order linear differential equations*, Trans. Amer. Math. Soc., **305**, 1(1988), 415–429.

[10] W. K. Hayman, *Meromorphic functions.* Oxford Mathematical Monographs Clarendon Press, Oxford 1964.

[11] S. Hellerstein, J. Miles, J. Rossi, *On the growth of solutions of certain linear differential equations*, Ann. Acad. Sci. Fenn. Ser. A I Math. **17**, 2(1992), 343–365.

[12] I. Laine, *Nevanlinna theory and complex differential equations.* de Gruyter Studies in Mathematics, 15. Walter de Gruyter & Co., Berlin, 1993.

[13] J. Wang, I. Laine, *Growth of solutions of nonhomogeneous linear differential equations*, Abstr. Appl. Anal. 2009, Art. ID 363927, 1-11.

[14] L. Wang, H. Liu, *Growth of meromorphic solutions of higher order linear differential equations*, Electron. J. Differential Equations, **2014**, 125(2014), 1-11.

[15] C. C. Yang, H. X. Yi, *Uniqueness theory of meromorphic functions*, Mathematics and its Applications, 557. Kluwer Academic Publishers Group, Dordrecht, 2003.

[16] C. Y. Zhang, J. Tu, *Growth of solutions to linear differential equations with entire coefficients of slow growth*, Electron. J. Differential Equations, 43(2010).

ROMAI J., v.14, no.1(2018), 129–141

WELL-POSEDNESS FOR A NONLINEAR REACTION-DIFFUSION EQUATION ENDOWED WITH NONHOMOGENEOUS CAUCHY-NEUMANN BOUNDARY CONDITIONS AND DEGENERATE MOBILITY

Alina Gavriluţ, Costică Moroşanu

"Al. I. Cuza" University of Iaşi, Iaşi, Romania

costica.morosanu@uaic.ro

Abstract The work is devoted to the study of a nonlinear parabolic equation with principal part in divergence form, endowed with non-homogeneous Cauchy-Neumann boundary conditions and non-constant thermal conductivity. The existence, uniqueness and regularity of solutions is established. Here we extend the results already proven by one of the authors for a nonlinearity of cubic type, making the present mathematical model to be more capable for description the complexity of certain wide classes of real physical phenomena (phase change, for instance).

Keywords: Qualitative properties of solutions, boundary value problems for nonlinear parabolic PDE, reaction-diffusion equations, Leray-Schauder principle, thermodynamics, heat transfer, phase-field models.

2010 MSC: 35Bxx, 35K55, 35K60, 35Qxx, 74A15, 80AXX.

1. INTRODUCTION

On a bounded domain $\Omega \subset I\!\!R^n$, $n \in \{1, 2, 3\}$, with a C^2 boundary $\partial\Omega \overset{not}{=} \Gamma$, and for a finite time $T > 0$, we consider the following second boundary value problem

$$\begin{cases} p_1 \dfrac{\partial}{\partial t} u - \dfrac{d}{dx_i}(k(u(t,x))\nabla u) = p_2(u - u^3) + f(t,x) & \text{in } Q = (0,T] \times \Omega \\ k(u(t,x)) \dfrac{\partial}{\partial \mathbf{n}} u + p_3 u = w(t,x) & \text{on } \Sigma = (0,T] \times \partial\Omega \\ u(0,x) = u_0(x) & \text{on } \Omega, \end{cases} \tag{1.1}$$

where:

- $u(t,x)$ is the unknown function and

$$\nabla u = \left(\frac{\partial}{\partial x_1}u, \frac{\partial}{\partial x_2}u, \cdots, \frac{\partial}{\partial x_n}u\right) \overset{not}{=} \left(u_{x_1}, u_{x_2}, \cdots, u_{x_n}\right) \overset{not}{=} u_x;$$

- p_1, p_2, p_3 are positive values;
- $k(u(t,x))$ - is the *degenerate mobility* [attached to the solution $u(t,x)$ of (1.1)], assumed to satisfy (see [5] and [27] for a detailed discussion of it):

$$0 < k_{min} \leq k(u(t,x)) \leq k_{max}, \qquad \forall (t,x) \in Q; \tag{1.2}$$

- $f(t,x) \in L^p(Q)$ is a given function, where p satisfies

$$p \geq 2; \tag{1.3}$$

- $w(t,x) \in W_p^{1-\frac{1}{2p}, 2-\frac{1}{p}}(\Sigma)$ is a given function depending on two variables, which also can be interpreted as *boundary control*,
- $u_0(x) \in W_\infty^{2-\frac{2}{p}}(\Omega)$ verifying $k(u_0(x))\dfrac{\partial}{\partial \mathbf{n}}u_0(x) + p_3 u_0(x) = w(0,x)$.
- $\mathbf{n}$=$\mathbf{n}$(x) is a vector of the outward (from Ω) unit normal to the surface Σ; $\frac{\partial}{\partial \mathbf{n}}$ denotes differentiation along $\mathbf{n}$.

Concerning equation $(1.1)_1$, we recall for reader's convenience that this is a quasi-linear one with principal part in divergence form of type (2.3) in [14, p. 11], with

$$a_i(t,x,u,u_x) = k(u(t,x))u_{x_i},\ i = 1,...,n$$

and

$$a(t,x,u,u_x) = -p_2(u - u^3) - f(t,x),$$

while the boundary conditions $(1.1)_2$ are of second type (7.2) in [14, p. 475], with

$$a_i(t,x,u)u_{x_i}cos(\mathbf{n},x_i) = k(u(t,x))\tfrac{\partial}{\partial \mathbf{n}}u(t,x),\ i = 1,...,n$$

and

$$\psi(t,x,u)|_\Sigma = p_3 u(t,x) - w(t,x).$$

If we write equation $(1.1)_1$ in the equivalent form (see [14, p.3])

$$p_1\frac{\partial}{\partial t}u - \frac{\partial}{\partial u_{x_j}}(k(u)u_{x_i})u_{x_ix_j} = A(t,x,u,u_{x_i}) + p_2(u-u^3) + f(t,x) \text{ in } Q, \tag{1.4}$$

with $u_{x_ix_j} = \frac{\partial^2}{\partial x_i \partial x_j}u$ and $A(t,x,u,u_{x_i}) = \frac{\partial}{\partial u}\left(k(u)u_{x_i}\right)u_{x_i} + \frac{\partial}{\partial x_i}(k(u)u_{x_i})$, then it is easy to recognize (1.4) as being a quasi-linear one of type (2.4) in [14, p. 11], with

$$a_{ij}(t,x,u,u_x) = \tfrac{\partial}{\partial u_{x_j}}a_i(t,x,u,u_x) = \tfrac{\partial}{\partial u_{x_j}}k(u(t,x))u_{x_i},\ i = 1,...,n$$

and

$$a(t,x,u,u_x) = -A(t,x,u,u_x) - p_2(u - u^3) - f(t,x).$$

In addition, unless otherwise stated, we assume that equations $(1.1)_1$ and (1.4) are *uniformly parabolic*, which means fulfilment of the conditions

$$\nu(|u|)\xi^2 \leq \frac{\partial a_i(t,x,u,p)}{\partial p_j}\xi_i\xi_j \leq \mu(|u|)\xi^2 \tag{1.5}$$

$$\nu(|u|)\xi^2 \leq a_{ij}(t,x,u,p)\xi_i\xi_j \leq \mu(|u|)\xi^2 \tag{1.6}$$

for arbitrary u and p and $\xi = (\xi_1, \cdots, \xi_n)$ an arbitrary real vector, where $\nu(r)$ and $\mu(r)$ are positive (nonincreasing and nondecreasing, respectively) continuous functions of $r \geq 0$.

Definition 1.1. $u(t,x)$ *is a classical solution of the second boundary value problem (1.1) if it is continuous in* $\bar{Q}$*, has continuous derivatives* u_t*,* u_x*,* u_{xx} *in* Q*, satisfies the equation (1.1)*$_1$ *at all points* $(t,x) \in Q$ *and satisfies conditions (1.1)*$_2$ *and (1.1)*$_3$ *for* $(t,x) \in \Sigma$ *and* $t = 0$*, respectively.*

In the present work we will investigate the solvability of the second boundary value problems of the form (1.1) in the class $W_p^{1,2}(Q)$. One proves the existence, the regularity and the uniqueness of solutions (Theorem 2.1 below) for the nonlinear parabolic problem (1.1) in the new mathematical formulation in which the principal part is in divergence form and considering the cubic nonlinearity $p_2(u-u^3)$ which verifies for $n \in \{1,2,3\}$ the assumption H_0 in [23], that is:

$H_0:\ (u-u^3)|u|^{3p-4}u \leq 1 + |u|^{3p-1} - |u|^{3p}.$

In the following we will denote by C several positive constants, with the remark that the extra dependencies will be set out on occurrence. In addition, every product is understood in the L^2-space, except when otherwise specified. In particular, the norm and the scalar product in $L^2(\Omega)$ are denoted by $\|\cdot\|$ and $< .,. >$, respectively.

2. WELL-POSEDNESS OF SOLUTIONS TO THE NONLINEAR EQUATION (1.1)

The main result of this Section establishes the dependence of the solution $u(t,x)$ in the nonlinear parabolic equation (1.1) on the terms $f(t,x)$ and $w(t,x)$.

Basic tools in our approach are the Leray-Schauder degree theory [11], the L^p-theory of linear and quasi-linear parabolic equations [14], as well as the Lions and Peetre embedding Theorem [16, p. 24] to ensure the existence of a continuous embedding $W_p^{1,2}(Q) \subset L^\mu(Q)$, where the number μ is defined as follows

$$\mu = \begin{cases} \text{any positive number} \geq 3p & \text{if } \frac{1}{p} - \frac{2}{n+2} \leq 0, \\ \left(\frac{1}{p} - \frac{2}{n+2}\right)^{-1} & \text{if} \frac{1}{p} - \frac{2}{n+2} > 0, \end{cases} \tag{2.1}$$

and, for a given positive integer k and $1 \leq p \leq \infty$, $W_p^{k,2k}(Q)$ denote the Sobolev space on Q:

$$W_p^{k,2k}(Q) = \left\{ y \in L^p(Q) :\ \frac{\partial^r}{\partial t^r}\frac{\partial^q}{\partial x^q}\ y \in L^p(Q),\ \text{for } 2r+q \leq k \right\},$$

i.e., the spaces of functions whose t-derivatives and x-derivatives up to the order k and $2k$, respectively, belong to $L^p(Q)$ (see [14, p. 5]).

Also, we shall use the set $C^{1,2}(\bar{Q})$ $(C^{1,2}(Q))$ of all continuous functions in $\bar{Q}$ (in Q) having continuous derivatives u_t, u_x, u_{xx} in $\bar{Q}$ (in Q), as well as the Sobolev spaces $W_p^l(\Omega)$, $W_p^{l,l/2}(\Sigma)$ with non integral l for the initial and boundary conditions, respectively (see [14, p. 8, p. 70 and p. 81]).

Our main result in studying the existence, estimate, uniqueness and regularity of solution in problem (1.1) is the following.

Theorem 2.1 *There exists a solution* $u \in W_p^{1,2}(Q)$ *to problem* (1.1) *which satisfies*

$$\|u\|_{W_p^{1,2}(Q)} \leq C \ \Bigg[1 + \|u_0\|_{W_\infty^{2-\frac{2}{p}}(\Omega)} + \|u_0\|_{L^{3p-2}(\Omega)}^{\frac{3p-2}{p}} + \|f\|_{L^p(Q)} \tag{2.2}$$

$$+\|w\|_{W_p^{1-\frac{1}{2p},2-\frac{1}{p}}(\Sigma)} + \|w\|_{L^{3p-2}(\Sigma_t)}^{\frac{3p-2}{p}}\Bigg],$$

where the constant C *depends on* $|\Omega|$, T, n, p, p_1, p_2 *and* p_3 *but is independent of* u, f *and* w.

If $u^1, u^2 \in W_p^{1,2}(Q)$ *are two solutions to* (1.1), *corresponding to* $\{f^1, w^1, u_0^1\}$ *and* $\{f^2, w^2, u_0^2\}$, *respectively, such that*

$$\|u^1\|_{W_p^{1,2}(Q)} \leq M, \qquad \|u^2\|_{W_p^{1,2}(Q)} \leq M \qquad for\ some \quad M \in (0,\infty), \tag{2.3}$$

and

$$k(u^1(t,x)) = k(u^2(t,x)) \qquad for \quad (t,x) \in \Sigma, \tag{2.3'}$$

then the following estimate holds

$$\underset{(t,x)\in Q}{max} |u^1 - u^2| \leq C_1 e^{CT} max \left\{ \underset{(t,x)\in Q}{max} |f^1 - f^2|, \underset{(t,x)\in\Sigma}{max} |w^1 - w^2|, \underset{(t,x)\in\Omega}{max} |u_0^1 - u_0^2| \right\}, \tag{2.4}$$

where the constants C_1 *and* C *are independent of* $\{u^1, f^1, w^1, u_0^1\}$ *and* $\{u^2, f^2, w^2, u_0^2\}$. *In particular, the uniqueness of solution to problem* (1.1) *holds.*

Proof. In order to use the Leray-Schauder degree theory we will choose as suitable Banach space

$$B = W_p^{0,1}(Q) \cap L^{3p}(Q),$$

endowed with the norm

$$\|u\|_B = \|u\|_{L^p(Q)} + \|u_x\|_{L^p(Q)},$$

and let us define the nonlinear operator $T : B \times [0,1] \to B$ as

$$u = u(v,\lambda) = T(v,\lambda) \quad \forall v \in B, \ \ \forall\, \lambda \in [0,1], \tag{2.5}$$

where u is the unique solution to the following linear second boundary value problem

$$\begin{cases} p_1 \dfrac{\partial}{\partial t} u - \left[\lambda \dfrac{\partial}{\partial v_{x_j}} (k(v) v_{x_i}) + (1-\lambda)\delta_i^j \right] u_{x_i x_j} \\ \qquad = \lambda \left[A(t,x,v,v_{x_i}) + p_2(v - v^3) + f(t,x) \right] & \text{in } \ Q \\ k(v) \dfrac{\partial}{\partial \mathbf{n}} u + p_3 u = \lambda w(t,x) & \text{on } \ \Sigma \\ u(0,x) = \lambda u_0(x) & \text{on } \ \Omega. \end{cases} \tag{2.6}$$

As regards the nonlinear operator T defined by (2.5), we have to check the properties **i-ii** listed below, i.e.:

i. T **is well-defined** (the problem (2.6) has a unique solution). From the right hand of $(2.6)_1$, it follows that, $\forall v \in W_p^{0,1}(Q) \cap L^{3p}(Q)$, then $v^3 \in L^p(Q)$ and thus $A(t,x,v,v_{x_i}) + p_2(v - v^3) + f(t,x) \in L^p(Q)$. Using now L^p-theory of linear parabolic equations (see [14, p. 341-342]), the solution u to problem (2.6) exists and is unique with

$$u = u(v,\lambda) \in W_p^{1,2}(Q) \quad \forall\, v \in W_p^{0,1}(Q) \cap L^{3p}(Q), \ \forall\, \lambda \in [0,1]. \tag{2.7}$$

Thanks to the continuous inclusions (see [16, p. 24])

$$W_p^{1,2}(Q) \subset W_p^{0,1}(Q) \cap L^{3p}(Q), \tag{2.8}$$

we derive that $T(v,\lambda) = u \in W_p^{0,1}(Q) \cap L^{3p}(Q)$ for all $v \in W_p^{0,1}(Q) \cap L^{3p}(Q)$ and $\forall\, l \in [0,1]$ which means that T is well defined.

ii. T **is continuous and compact.** Let $v_n \to v$ in $W_p^{0,1}(Q) \cap L^{3p}(Q)$ and $\lambda_n \to \lambda$ in $[0,1]$. Denote $u^{n,\lambda_n} = T(v^n,\lambda_n)$, $u^{n,\lambda} = T(v^n,\lambda)$ and $u^{1,\lambda} = T(v,\lambda)$. Relations (2.5) and (2.6) give

$$\begin{cases} p_1 \dfrac{\partial}{\partial t} U^{n,\lambda_n,\lambda} - \left[\lambda \dfrac{\partial}{\partial v^n_{x_j}} (k(v^n) v^n_{x_i}) + (1-\lambda)\delta_i^j \right] U^{n,\lambda_n,\lambda}_{x_i x_j} \\ \qquad = (\lambda_n - \lambda) \left\{ \left[\frac{\partial}{\partial v^n_{x_j}} (k(v^n) v^n_{x_i}) - \delta_i^j \right] u^{n,\lambda_n}_{x_i x_j} \right. \\ \qquad \left. + A(t,x,v^n,v^n_{x_i}) + p_2(v^n - (v^n)^3) + f(t,x) \right\} & \text{in } \ Q \\ k(v^n) \dfrac{\partial}{\partial \mathbf{n}} U^{n,\lambda_n,\lambda} + p_3 U^{n,\lambda_n,\lambda} = (\lambda_n - \lambda) w(t,x) & \text{on } \ \Sigma \\ u(0,x) = (\lambda_n - \lambda) u_0(x) & \text{on } \ \Omega, \end{cases} \tag{2.9}$$

where $U^{n,\lambda_n,\lambda} = u^{n,\lambda_n} - u^{n,\lambda}$.

The right-hand side of $(2.9)_1$ belongs to $L^p(Q)$ and thus we may apply the L^p-theory (see [14, p. 341-342]) to problem (2.9), which gives the estimate

$$\|U^{n,\lambda_n,\lambda}\|_{W_p^{1,2}(Q)} \leq C|\lambda_n-\lambda| \left[\|u_0\|_{W_\infty^{2-\frac{2}{p}}(\Omega)} + \left\| \left(\frac{\partial}{\partial v_{x_j}^n}(k(v^n)v_{x_i}^n) - \delta_i^j \right) u_{x_ix_j}^{n,\lambda_n} \right\|_{L^p(Q)} \right.$$

$$\left. +\|(v^n-(v^n)^3)\|_{L^p(Q)} + \left\|A(t,x,v^n,v_{x_i}^n)\right\|_{L^p(Q)} + \|f\|_{L^p(Q)} + \|w\|_{W_p^{2-\frac{1}{p},1-\frac{1}{2p}}(\Sigma)} \right].$$

Having v^n bounded in $W_p^{0,1}(Q) \cap L^{3p}(Q)$, we can derive that $(v^n)^3$ is bounded in $L^p(Q)$ (see, e.g., [11] or [14, p. 42]). Moreover, by virtue of the working hypothesis, we can easily deduce that the remaining terms on the right-hand side from the above inequality are also bounded. Thus, making use of the convergence $\lambda_n \to \lambda$, from the above inequality we get

$$\|u^{n,\lambda_n} - u^{n,\lambda}\|_{W_p^{1,2}(Q)} \to 0 \quad as \;\; n \to \infty. \tag{2.10}$$

From (2.5) and (2.6) we also obtain

$$\begin{cases} p_1 \dfrac{\partial}{\partial t}U^{n,1,\lambda} - \left[\lambda \dfrac{\partial}{\partial v_{x_j}^n}(k(v^n)v_{x_i}^n) + (1-\lambda)\delta_i^j \right] U_{x_ix_j}^{n,1,\lambda} & \\ \quad = \lambda \left\{ \left[\frac{\partial}{\partial v_{x_j}^n}(k(v^n)v_{x_i}^n) - \frac{\partial}{\partial v_{x_j}}(k(v)v_{x_i}) \right] u_{x_ix_j}^{\lambda} \right. & \\ \quad + \left[A(t,x,v^n,v_{x_i}^n) - A(t,x,v,v_{x_i}) \right] & \\ \quad \left. +p_2 \left[(v^n - v) - ((v^n)^3 - v^3) \right] \right\} & \text{in } Q \\ k(v^n)\dfrac{\partial}{\partial \nu}U^{n,1,\lambda} + p_3 U^{n,1,\lambda} = 0 & \text{on } \Sigma \\ u(0,x) = 0 & \text{on } \Omega, \end{cases} \tag{2.11}$$

where $U^{n,1,\lambda} = u^{n,\lambda} - u^{1,\lambda}$.

The L^p-theory applied to (2.11) (see [14, p. 341-342]), give us the estimate

$$\|u^{n,\lambda} - u^{1,\lambda}\|_{W_p^{1,2}(Q)} \leq C\lambda \left\{ \left\| \left[\frac{\partial}{\partial v_{x_j}^n}(k(v^n)v_{x_i}^n) - \frac{\partial}{\partial v_{x_j}}(k(v)v_{x_i}) \right] u_{x_ix_j}^{\lambda} \right\|_{L^p(Q)} \right.$$

$$\left. +\|A(t,x,v^n,v_{x_i}^n) - A(t,x,v,v_{x_i})\|_{L^p(Q)} + \|(v^n-v) - ((v^n)^3 - v^3)\|_{L^p(Q)} \right\},$$

for a positive constant C. Then, the convergence $v_n \to v$ in $W_p^{0,1}(Q) \cap L^{3p}(Q)$ and the continuity of the Nemytskij operator (see, e.g., [11]), as well as the continuity of $\frac{\partial}{\partial v^n_{x_j}}(k(v^n)v^n_{x_i})$ and $A(t,x,v^n,v^n_{x_i})$, allow us to conclude that

$$\|u^{n,\lambda} - u^{1,\lambda}\|_{W^{1,2}_p(Q)} \to 0 \quad as \;\; n \to \infty. \tag{2.12}$$

Making use of the continuous embedding (2.8) and relations (2.10), (2.12), we derive the continuity of the nonlinear operator T defined in (2.5). Furthermore, T is compact. Indeed, since $\mu > 3p$ (see (2.1)), the inclusion $W^{1,2}_p(Q) \hookrightarrow W^{0,1}_p(Q) \cap L^{3p}(Q)$ is compact (see [16, p. 21]). Moreover, writing T as the composition

$$B \times [0,1] \to W^{1,2}_p(Q) \hookrightarrow W^{0,1}_p(Q) \cap L^{3p}(Q) = B,$$

the compactness of T immediately follows. ▮

2.1. THE REGULARITY OF THE SOLUTION

We will establish now the existence of a number $\delta > 0$ such that (see (2.5))

$$(u,\lambda) \in B \times [0,1] \quad with \;\; u = T(u,\lambda) \quad \Longrightarrow \quad \|u\|_B < \delta. \tag{2.13}$$

Let $u \in W^{0,1}_p(Q) \cap L^{3p}(Q)$ solving the problem (see (2.6))

$$\begin{cases} p_1 \dfrac{\partial}{\partial t}u - \lambda \dfrac{d}{dx_i}(k(u)\nabla u) + (1-\lambda)\Delta u = \lambda\left[p_2(u-u^3) + f(t,x)\right] & \text{in } \; Q \\ k(u)\dfrac{\partial}{\partial \nu}u + p_3 u = \lambda w(t,x) & \text{on } \; \Sigma \\ u(0,x) = \lambda u_0(x) & \text{on } \; \Omega. \end{cases} \tag{2.14}$$

Multiplying the first equation in (2.14) by $|u|^{3p-4}u$, integrating over $Q_t := (0,t) \times \Omega$, $t \in (0,T]$ and using Green's Theorem, we get

$$\begin{aligned} &\frac{p_1}{3p-2}\int\limits_{\Omega} |u(t,x)|^{3p-2}dx + 3(p-1)(1-\lambda)\int\limits_{Q_t} |\nabla u|^2 |u|^{3p-4}\, d\tau dx + p_3 \int\limits_{\Sigma_t} |u|^{3p-2}\, d\tau d\gamma \\ &\leq l\frac{p_1}{3p-2}\int\limits_{\Omega} |u_0(x)|^{3p-2}\, dx + \lambda p_2 \int\limits_{Q_t} (u-u^3)|u|^{3p-4}u\, d\tau dx \\ &+\lambda \int\limits_{Q_t} f|u|^{3p-4}u\, d\tau dx + \lambda \frac{k_{min}+1}{k_{min}} \int\limits_{\Sigma_t} w|u|^{3p-4}u\, d\tau d\gamma \quad \text{for all } \; t \in (0,T]. \end{aligned} \tag{2.15}$$

The Hölder's and Cauchy's inequalities, applied to the last term in (2.15), give us

$$\lambda\frac{k_{min}+1}{k_{min}}\int\limits_{\Sigma_t} w|u|^{3p-4}u\ d\tau d\gamma \tag{2.16}$$

$$\leq p_3\frac{3p-3}{3p-2}\int\limits_{\Sigma_t}|u|^{3p-2}\ d\tau d\gamma+\lambda\frac{k_{min}+1}{k_{min}}\frac{1}{p_3}\frac{1}{3p-2}\int\limits_{\Sigma_t}|w|^{3p-2}\ d\tau d\gamma.$$

By H_0, relations (1.3), (2.16) and Young's inequality, from (2.15) we obtain

$$\frac{p_1}{3p-2}\int\limits_{\Omega}|u(t,x)|^{3p-2}dx+3(p-1)(1-\lambda)\int\limits_{Q_t}|\nabla u|^2|u|^{3p-4}\ d\tau dx \tag{2.17}$$

$$+\lambda p_2\int\limits_{Q_t}|u|^{3p}\ d\tau dx+\frac{p_3}{3p-2}\int\limits_{\Sigma_t}|u|^{3p-2}\ d\tau d\gamma$$

$$\leq\lambda\frac{p_1}{3p-2}\int\limits_{\Omega}|u_0(x)|^{3p-2}\ dx+\lambda\left(|\Omega|T+\frac{1}{3p}\varepsilon^{-3p}|\Omega|T+\frac{1}{p}\varepsilon^{-p}\|f\|^p_{L^p(Q)}\right)$$

$$+\lambda\left\{\frac{3p-1}{3p}\varepsilon^{\frac{3p}{3p-1}}+\frac{p-1}{p}\varepsilon^{\frac{p}{p-1}}\right\}\int\limits_{Q_t}|u|^{3p}\ dsdx+\lambda\frac{k_{min}+1}{k_{min}}\frac{1}{p_3}\frac{1}{3p-2}\int\limits_{\Sigma_t}|w|^{3p-2}\ d\tau d\gamma.$$

Taking ε small enough, inequality (2.17) yields

$$\lambda\||u|^3\|^p_{L^p(Q)}\leq C_1\left(1+\|u_0\|^{3p-2}_{L^{3p-2}(\Omega)}+\|f\|^p_{L^p(Q)}+\|w\|^{3p-2}_{L^{3p-2}(\Sigma_t)}\right), \tag{2.18}$$

for a constant $C_1=C(|\Omega|,T,n,p,k_{min},p_1,p_2,p_3)>0$.

Applying L^p-theory to problem (2.14) (see [14, p. 341-342]), we get

$$\|u\|_{W^{1,2}_p(Q)}\leq C_2\left(\|u_0\|_{W^{2-\frac{2}{p}}_{\infty}(\Omega)}+p_2\|(u-u^3)\|_{L^p(Q)}+\|f\|_{L^p(Q)}+\|w\|_{W^{1-\frac{1}{2p},2-\frac{1}{p}}_p(\Sigma)}\right), \tag{2.19}$$

for a constant $C_2=C(|\Omega|,T,n,p,p_1,p_2,p_3)>0$.

Taking into account Lemma 1.1 in [23] and relation (2.18), we deduce that

$$\|u-u^3\|_{L^p(Q)}\leq C_1\left(1+\|u_0\|^{\frac{3p-2}{p}}_{L^{3p-2}(\Omega)}+\|f\|_{L^p(Q)}+\|w\|^{\frac{3p-2}{p}}_{L^{3p-2}(\Sigma)}\right)$$

and then (2.19) becomes

$$\|u\|_{W^{1,2}_p(Q)}\leq C_2\left(1+\|u_0\|_{W^{2-\frac{2}{p}}_{\infty}(\Omega)}+\|u_0\|^{\frac{3p-2}{p}}_{L^{3p-2}(\Omega)}\right. \tag{2.20}$$

$$+\|f\|_{L^p(Q)} + \|w\|_{W_p^{1-\frac{1}{2p},2-\frac{1}{p}}(\Sigma)} + \|w\|^{\frac{3p-2}{p}}_{L^{3p-2}(\Sigma_t)}\Big) .$$

The continuous embedding in (2.8) ensures that

$$\|u\|_B \leq C\|u\|_{W_p^{1,2}(Q)}$$

which, owing to (2.20), ensures that a constant $\delta > 0$ can be found such that the property expressed in (2.13) is true.

Denoting

$$B_\delta := \Big\{u \in B : \ \|u\|_B < \delta\Big\},$$

relation (2.13) implies that

$$T(u, \lambda) \neq u \quad \forall u \in \partial B_\delta, \ \ \forall \lambda \in [0, 1],$$

provided that $\delta > 0$ is sufficiently large. Furthermore, following the same reasoning as in paper [6], we conclude that problem (1.1) has a solution $u \in W_p^{1,2}(Q)$ (see also [23, p. 195]). Estimate (2.2) follows directly from relation (2.20) and this completes the proof of the first part.

2.2. THE UNIQUENESS OF THE SOLUTION

Now, let us prove the second part of Theorem 2.1. Precisely, we will establish the estimate (2.4) and, as a consequence, the uniqueness of the solution to problem (1.1) or (1.4)-$(1.1)_{2,3}$.

By hypothesis, $u^1, u^2 \in W_p^{1,2}(Q)$ solve problem (1.1) corresponding to f^1, w^1, u_0^1 and f^2, w^2, u_0^2, respectively. Thus $u^1 - u^2 \in W_p^{1,2}(Q)$.

For convenience, we denote in what follow:

$u^\lambda(t, x) = \lambda u^1(t, x) + (1-\lambda)u^2(t, x), \qquad u_x^\lambda(t, x) = \lambda u_x^1(t, x) + (1-\lambda)u_x^2(t, x)$

and

$a_{ij}(t, x, u^1, u_x^1) = \frac{\partial}{\partial u^1_{x_j}} k(u^1)u^1_{x_i}, \qquad a_{ij}(t, x, u^2, u_x^2) = \frac{\partial}{\partial u^2_{x_j}} k(u^1)u^2_{x_i}.$

Also, following (5.3) in [14, p. 445], we write the increments of the a_{ij} and A in the form

$$a_{ij}(t, x, u^1, u_x^1) - a_{ij}(t, x, u^2, u_x^2) = \int_0^1 \frac{d}{d\lambda} a_{i,j}\left(t, x, u^\lambda, u_x^\lambda\right) d\lambda,$$

$$A(t, x, u^1, u_x^1) - A(t, x, u^2, u_x^2) = \int_0^1 \frac{d}{d\lambda} A\left(t, x, u^\lambda, u_x^\lambda\right) d\lambda,$$

and then

$$a_{ij}(t,x,u^1,u^1_x)u^1_{x_ix_j} - a_{ij}(t,x,u^2,u^2_x)u^2_{x_ix_j} = a_{ij}(t,x,u^1,u^1_x)U_{x_ix_j} \tag{2.22}$$

$$+u^2_{x_ix_j}\left[U_{x_i}\int\limits_0^1 \frac{\partial}{\partial u^\lambda_{x_j}}a_{i,j}\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda + U\int\limits_0^1 \frac{\partial}{\partial u^\lambda}a_{i,j}\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda\right],$$

$$A(t,x,u^1,u^1_x) - A(t,x,u^2,u^2_x) \tag{2.23}$$

$$= U_{x_i}\int\limits_0^1 \frac{\partial}{\partial u^\lambda_{x_j}}A\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda + U\int\limits_0^1 \frac{\partial}{\partial u^\lambda}A\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda,$$

where $U(t,x) = u^1(t,x) - u^2(t,x)$.

We subtract the equations (1.4)-$(1.1)_{2,3}$ for $u^2(t,x)$ from the equations (1.4)-$(1.1)_{2,3}$ for $u^1(t,x)$ and, owing to (2.3'), we obtain the linear equation

$$\begin{cases} p_1\dfrac{\partial}{\partial t}U - \hat{a}_{ij}(t,x)U_{x_ix_j} + \hat{a}_i(t,x)U_{x_i} + \hat{a}(t,x)U = f^1 - f^2 & \text{in } Q \\ k(u^1)\dfrac{\partial}{\partial \mathbf{n}}U + p_3U = w^1 - w^2 & \text{on } \Sigma \\ U(0,x) = u^1_0(x) - u^2_0(x) & \text{on } \Omega, \end{cases} \tag{2.24}$$

where

$$\begin{aligned} \hat{a}_{ij}(t,x) &= a_{ij}(t,x,u^1,u^1_x), \\ \hat{a}_i(t,x) &= -u^2_{x_ix_j}\int\limits_0^1 \frac{\partial}{\partial u^\lambda_{x_j}}a_{i,j}\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda + \int\limits_0^1 \frac{\partial}{\partial u^\lambda_{x_j}}A\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda, \\ \hat{a}(t,x) &= -u^2_{x_ix_j}\int\limits_0^1 \frac{\partial}{\partial u^\lambda}a_{i,j}\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda + \int\limits_0^1 \frac{\partial}{\partial u^\lambda}A\left(t,x,u^\lambda,u^\lambda_x\right)d\lambda \\ &\quad -p_2\left[1 - \left((u^1)^2 + u^1u^2 + (u^2)^2\right)\right]. \end{aligned}$$

By virtue of the relations (1.2), (1.6) and (2.3), the conditions of Theorem 2.3 in [14, p. 16] on linear equations are fulfilled. In view of this, it follows from (2.24) that estimate of type (2.4) is valid for U, which finishes the proof of Theorem 2.1.

As a consequence, the uniqueness of solution to problem (1.1) is valid.

Corollary 2.1. *For the same initial conditions, the problem* (1.1) *possesses a unique classical solution.*

Proof. Let $f^1 = f^2 = f$ and $w^1 = w^2 = w$ in the Theorem 2.1. Then (2.4) shows that the conclusion of the corollary is true (see Definition 1.1 and [14, Theorem 2.4, p. 17]). ∎

3. CONCLUSIONS

The problem addressed in this paper is a nonlinear reaction-diffusion equation with principal part in divergence form, endowed with non-homogeneous Cauchy-Neumann and degenerate mobility. Thus, by the presence of operator *div* in the diffusion term, the present paper extends the type of nonlinearities already studied (see [6], [8], [9], [17], [18], [20], [22], [23]).

Provided that the initial and boundary data meet appropriate regularity and compatibility conditions, we prove the existence, uniqueness and regularity of solutions. Precisely, the Leray-Schauder principle is applied to prove the existence result for the nonlinear problem in question, while the L^p theory of linear and quasi-linear parabolic equations is involved in order to derive regularity properties for the solutions. Moreover, the *a priori* estimates are made in $L^p(Q)$ which leads to a better estimates for unknown functions $u(t,x)$. This approach could be applied in future to study other kind of the first and second boundary value problems. For another technique regarding this topic, we can suggest for the readers the monographs [3] and [4].

The mathematical model (1.1) is linked, and not only, to the Allen-Cahn equation (see [1], [2]) and the Cahn-Hilliard equation (see and [7, Figure 1, p. 421]) which models, among other, the time evolution of the order parameter in the non-isothermal case and the phase-separation phenomenon, respectively. Recently, the Allen-Cahn equation has been widely applied to many complex moving interface problems, like: the mixture of two incompressible fluids, the nucleation of solids, vesicle membranes, etc. Also, the nonlinear parabolic equation $(1.1)_1$ occurs in the Caginalp's phase-field transition system (see [7]) describing the transition between the solid and liquid phases in the solidification process of a material occupying a region Ω. Regarding the latter very complex physical process, we wish to emphasize that our assumption in (1.2) is sustained by industrial applications (see [12], for example). In [26] the reader can find more details relative to a more extensive class of problems on the type those treated in this paper (reaction-diffusion equation), as well as different types of the nonlinear term.

Numerical analysis as well as various simulations regarding the physical phenomena described by the nonlinear parabolic problem (1.1) (in particular, the *separating region*), represent a matter for further investigation (see [10], [12], [13], [15], [19]-[21], [24], [25] and [28], for example). In addition, the qualitative results obtained here can be involved in the study of distributed and/or boundary nonlinear optimal control problems governed by the nonlinear problem (1.1). Amongst other things, we wish to exploit all this in our future works.

References

[1] S. M. Allen, J. W. Cahn, *Ground state structures in ordered binary alloys with second neighbor interactions*, Acta Metallurgica, **20**, 3(1972), 423-433.

[2] S. M. Allen, J. W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metallurgica, **27**(1979), 1084-1095.

[3] V. Barbu, *Partial differential equations and boundary value problems*, vol. 441 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, 1998.

[4] V. Barbu, *Nonlinear Differential Equations of Monotone Types in Banach Spaces*, Springer Monographs in Mathematics, DOI 10.1007/978-1-4419-5542-5_1 , 2010.

[5] J.W. Barrett, J.F. Blowey, H. Garcke, *Finite element aproximation of the Cahn-Hilliard equation with degenerate mobility*, SIAM J. Numer. Anal., **37**(1999), 286-318.

[6] T. Benincasa, C. Moroşanu, *Fractional steps scheme to approximate the phase-field transition system with non-homogeneous Cauchy-Neumann boundary conditions*, Numer. Funct. Anal. & Optimiz., **30**, 3-4(2009), 199-213.

[7] G. Caginalp, X. Chen, *Convergence of the phase field model to its sharp interface limits*, Euro. J. of Applied Mathematics, **9**(1998), 417-445.

[8] L. Calatroni, P. Colli, *Global solution to the Allen-Cahn equation with singular potentials and dynamic boundary conditions*, Nonlinear Analysis. TMA, **79**(2013), 12-27, \arXiv1206.6738

[9] O. Cârjă, A. Miranville, C. Moroşanu, *On the existence, uniqueness and regularity of solutions to the phase-field system with a general regular potential and a general class of nonlinear and non-homogeneous boundary conditions,* Nonlinear Analysis. TMA, **113**(2015), 190-208, http://dx.doi.org/10.1016/j.na.2014.10.003

[10] X. Feng, A. Prohl, *Numerical analysis of the Allen-Cahn equation and approximation for mean curvature flows*, Numer. Math., **94**, 1(2003), 33-65.

[11] I. Fonseca, W. Gangbo, Degree Theory in Analysis and Applications, Clarendon, Oxford, 1995.

[12] Gh. Iorga, C. Moroşanu, I. Tofan *Numerical simulation of the thickness accretions in the secondary cooling zone of a continuous casting machine*, Metalurgia International, **XIV**, 1(2009), 72-75.

[13] N. Kenmochi, M. Niezgódka, *Evolution systems of nonlinear variational inequalities arising from phase change problems*, Nonlinear Anal. TMA, **22**(1944), 1163–1180.

[14] O.A. Ladyzhenskaya, B.A. Solonnikov, N.N. Uraltzava, Linear and quasi-linear equations of parabolic type, Prov. Amer. Math. Soc., 1968.

[15] H. G. Lee, J.-Y. Lee, *A semi-analytical Fourier spectral method for the Allen-Cahn equation*, Comput. Math. Appl., **68**, 3(2014), 174-184.

[16] J.L. Lions, *Control of distributed singular systems*, Gauthier-Villars, Paris, 1985.

[17] A. Miranville, *Existence of solutions for a one-dimensional Allen-Cahn equation*, J. Appl. Anal. & Comput., **3**, 3(2013), 265-277.

[18] A. Miranville, C. Moroşanu, *On the existence, uniqueness and regularity of solutions to the phase-field transition system with non-homogeneous Cauchy-Neumann and nonlinear dynamic boundary conditions*, Appl. Math. Modell., **40**, 1(2016), 192-207, doi: 10.1016/j.apm.2015.04.039

[19] C. Moroşanu, *Approximation of the phase-field transition system via fractional steps method*, Numer. Funct. Anal. & Optimiz., **18**(5& 6)(1997), 623-648.

[20] C. Moroşanu, Analysis and optimal control of phase-field transition system: Fractional steps methods, Bentham Science Publishers, 2012, http://dx.doi.org/10.2174/97816080535061120101.

[21] C. Moroşanu, *Cubic spline method and fractional steps scheme to approximate the phase-field system with non-homogeneous Cauchy-Neumann boundary conditions*, ROMAI J., **8**, 1(2012), 73-91 .

[22] C. Moroşanu C., A. Croitoru, *Analysis of an iterative scheme of fractional steps type associated to the phase-field equation endowed with a general nonlinearity and Cauchy-Neumann boundary conditions*, J. Math. Anal. and Appl., **425**(2015), 1225-1239. http://dx.doi.org/10.1016/j.jmaa.2015.01.033

[23] C. Moroşanu, D. Motreanu, *The phase field system with a general nonlinearity*, International Journal of Differential Equations and Applications, **1**, 2(2000), 187-204.

[24] C. Moroşanu, A.-M. Moşneagu, *On the numerical approximation of the phase-field system with non-homogeneous Cauchy-Neumann boundary conditions, Case 1D*, ROMAI J., **9**, 1(2013), 91-110.

[25] A. A. Ovono, *Numerical approximation of the phase-field transition system with non-homogeneous Cauchy-Neumann boundary conditions in both unknown functions via fractional steps methods*, J. Appl. Anal. & Comput., **3**(2013), 4, 377-397 .

[26] C. V. Pao, *Nonlinear parabolic and elliptic equations*, ISBN 0-306-44343-0, Plenum Press, New York, 1992.

[27] O. Penrose, P. C. Fife, *Thermodynamically consistent models of phase-field type for kinetics of phase transitions*, Phys. D., **43**(1990), 44-62.

[28] J. Zhang, Q. Du, *Numerical studies of discrete approximations to the Allen-Cahn equation in the sharp interface limit*, SIAM J. Sci. Comput., **31**, 4(2009), 3042-3063.

ROMAI J., v.14, no.1(2018), 143–147

ON SEMICOMPACT OPERATORS

Omer Gok

Yildiz Technical University, Istanbul, Turkey

gok@yildiz.edu.tr

Abstract Let X be a Banach space and E be a Banach lattice.An operator $T : X \to E$ is called a semi-compact operator if every $\epsilon > 0$ there is an $0 \leq x$ in E such that $T(ball(X)) \subseteq [-x, x] + \epsilon(ball(E))$. In this study, we investigate semicompactness and b-semicompactness of an operator and its adjoint.

Keywords: Banach lattice, semicompact operators, order continuous.

2010 MSC: 46B25, 46B42, 47B60, 47B65.

1. INTRODUCTION

Let X be a Banach space and E be a Banach space. We denote E' by the set of all order bounded linear functionals on$E$, and E'' by the set of all second order dual of E. By X' we denote the set of all continuous linear functionals on X. Since E is a Banach lattice, order dual and continuous dual are coincide, $[1, 5]$.Let A be a subset of E. A is called a b-order bounded if A is an order bounded in the second order dual E'' of E. A Banach lattice is said to have b-property if every b-order bounded set is an order bounded set in E, [2,3].A Banach lattice E is called a KB-space if every positive increasing norm bounded sequence in E converges. A Banach lattice E is a KB space if and only if it has an order continuous norm and with property (b),[2].There are a lot of KB spaces in Banach lattices, $[1, 5]$.A Banach lattice E is said to have an order continuous norm if $x_n \downarrow 0$ in E implies $\| x_n \| \to 0$ as $n \to \infty$. For example, c_0 has an order continuous norm. Let E be a Banach lattice. E' is a KB space if and only if E has an order continuous norm.A continuous linear operator $T : X \to E$ is called a b-semicompact if for each $\epsilon > 0$ there exists some $0 \leq u \in E''$ such that $T(ball(X) \subseteq [-u, u] + \epsilon ball(E''))$, [4]. A continuous linear operator $T : X \to E$ is called a semicompact if for each $\epsilon > 0$ there exists some $0 \leq u \in E$ such that $T(ball(X) \subseteq [-u, u] + \epsilon ball(E)$,[7],[6] A continuous linear operator $T : E \to X$ is called b-order weakly compact if image of a b-order bounded set under T is relatively weakly compact set. Every weakly compact operator is b-weakly compact and every b-weakly compact operator is order weakly compact.In some special cases, converses of this known results are true.

In this paper, we study the link between semicompact operators, b-semicompact operators, order weakly compact operators and their adjoints.

In addition , we investigate the relation between b-semicompact operator and ring ideals generated by positive operators.

2. ON THE LINK BETWEEN B-SEMICOMPACT OPERATORS AND OTHER TYPE OPERATORS

Definition 2.1. *([1,5])Let X be a Banach space and E be a Banach lattice. A continuous linear operator $T : X \to E$ is called an L-weakly compact if $T(ball(X))$ is an L-weakly compact set. A subset A of E is called L-weakly compact if $\| x_n \| \to 0$ as $n \to \infty$ for every disjoint sequence (x_n) in the solid hull of A.*

Definition 2.2. *([1,5])A continuous linear operator $T : E \to X$ is called M-weakly compact if $lim_{n\to\infty} \| Tx_n \| = 0$ for every disjoint sequence (x_n) in the closed unit ball of E.*

Adjoint of an M-weakly compact operator is an L-weakly compact and adjoint of an L-weakly compact operaor is an M-weakly compact. Every L-weakly compact and M-weakly compact operators are weakly compact.

Theorem 2.1. *([1,4]) Let X be a Banach space and E be a Banach lattice. Assume that $T : X \to E$ is a linear operator. Then, the following assertions are true:*
(i) If T is a compact operator , then T is a b-semicompact operator.
(ii) If T is an L-weakly compact operator, then T is b-semicompact.
(iii) If T is a semicompact operator, then T is b-semicompact.

Proof. The proof is done directly by using the definitions. ∎

By $L(X, E)$ we denote the vector space of all continuous linear operators from a Banach space X into a Banach lattice E.

Theorem 2.2. *([1,4]) The collection of all b-semicompact operators from a Banach space X into a Banach lattice E form a closed subspace of $L(X, E)$.*

The collection of all b-semicompact operators form closed two- sided ideals in the space of all regular operators $L^r(E, E) = L^r(E)$.

Theorem 2.3. *([1]) Let $S, T : E \to F$ be positive linear operators between Banach lattices. Then, the following assertion are true:*
(i) If T is an M weakly compact operator, then T is b-semicompact.
(ii) If $0 \leq S \leq T$ and T is b-semicompact operator, then S is b-semicompact.

By $L_w(X, E)$ we denote the set of all L-weakly compact operators, by $L_s(X, E)$ we denote the set of all semicompact operators and by $L_{bs}(X, E)$ we denote the set of all b-semicompact operators. The following result is known from [4]: E is a KB space if and only if $L_w(X,E) = L_s(X,E) = L_{bs}(X,E)$.

Let X and Y be Banach spaces and $T : X \to Y$ be a continuous linear operator.The adjoint operator T' of T is defined from Y' into X' by $T'(f)(x) = f(Tx)$ for every $f \in Y'$ and for every $x \in X$.

Theorem 2.4. *Suppose that X is a Banach space and a Banach lattice E has an order continuous norm. Let $T : E \to X$ be a continuous linear operator such that adjoint T' of T is a b-semicompact operator . Then, T is an order weakly compact operator.*

Proof. Since the topological dual E' of E is KB space, $T' : X' \to E'$ is an L-weakly compact operator. So, T is an M-weakly compact operator. Every M-weakly compact operator is weakly compact and every weakly compact operator is an order weakly compact operator. ∎

Theorem 2.5. *Suppose that X is a Banach space and a Banach lattice E is a KB space. If $T : X \to E$ is a b-semicompact operator, then adjoint operator $T' : E' \to X'$ is an order weakly compact operator.*

Proof. Since E is a KB space, T is an L-weakly compact operator.Adjoint T' of T is an M-weakly compact operator. This implies that T' is a weakly compact operator. Since every weakly compact operator is an order weakly compact, we have proven the claim. ∎

Theorem 2.6. *([1])Let $T : X \to E$ be a b-semicompact operator from a Banach space X into a Banach lattice E. Then, there exists some y in the positive cone of E'' such that the ideal A_y generated by y satisfies $T(X) \subseteq \bar{A}_y$.*

Proof. We know that the ideal generated by y is

$$A_y = \{x :\mid x \mid \leq n \mid y \mid for \quad some \quad n\}.$$

For every n let us take $0 < u_n \in E''$ such that $\parallel (\mid Tx \mid -u_n)^+ \parallel < n^{-1}$ holds for all $x \in ball(X)$. Let $y = \sum_{n=1}^{\infty} 2^{-n} u_n / \parallel u_n \parallel$. Let A_y be the ideal generated by y.The sequence $[u_n)$ is in A_y. Let $x \in ball(X)$. It is easily seen that $\mid Tx \mid \in \bar{A}_y$.It is known that the closure of an ideal is again an ideal. So, $Tx \in \bar{A}_y$. Hence, the result follows from here. ∎

Definition 2.3. *([1,5])A Banach lattice E is called an abstract M-space(AM-space) if $\parallel x + y \parallel = max\{\parallel x \parallel, \parallel y \parallel\}$ for every $x, y \in E$ with $x \wedge y = 0$.*

A Banach lattice E is said to be an abstract L-space (AL space) if dual E' of E is an abstract M-space.

Every continuous linear operator from a Banach space into an AM-space with unit is b-semicompact. Every regular operator from an AM -space with unit into a Banach lattice is b-semicompact.Hence, we can give the following result.

Theorem 2.7. *If $T : E \to X$ is a continuous linear operator between AL-space E and Banach space X, then adjoint operator $T' : X' \to E'$ is a b-semicompact.*

Note that adjoint of a b-semicompact operator is not necessary to be a b-semicompact and we can not say that an operator is b-semicompact if adjoint of an operator is a b-semicompact operator.

Theorem 2.8. *Suppose that a Banach lattice E has an order continuous norm and let $T : E \to E$ be a positive linear operator.If $T' : E' \to E'$ is b-semicompact operator, then $T : E \to E$ is a b-semicompact operator.*

Proof. Since E has an order continuous norm, the topological dual E' is a KB space.By hypothesis,$T' : E' \to E'$ is a b-semicompact operator.From here, T' is an L weakly compact operator and so, T is M-weakly compact operator.Since a positive M-weakly compact operator is a semicompact operator, it is a b-semicompact operator. ∎

Theorem 2.9. *Let E be a KB-space and $T : E \to E$ be a positive linear operator. If $T : E \to E$ is a b-semicompact operator, then $T' : E' \to E'$ is a b-semicompact operator.*

Proof. Since T is an L-weakly compact operator, it follows that its adjoint $T' : E' \to E'$ is an M-weakly compact operator.Every positive M-weakly compact operator is a b-semicompact operator. Hence, the claim is true. ∎

Let $T : X \to Y$ be a continuous linear operator between Banach spaces X and Y.Assume that $Ring(T)$ is the norm closure in $L(X, Y)$ of the vector subspace consisting of all operators of the form $\sum_{i=1}^{n} R_i T S_i$ with $S_i \in L(X)$ and $R_i \in L(Y)$ for $i = 1, ..., n$. We say that the closed vector subspace $Ring(T)$ of $L(X, Y)$ is the ring ideal generated by T.

Theorem 2.10. *([1]) Let $S, T : E \to E$ be positive linear operators on KB space such that $0 \leq S \leq T$ holds.If S is a b-semicompact operator, then $S^3 \in Ring(T)$.*

Theorem 2.11. *([1])Let E be a KB space and $S, T; E \to E$ be positive linear operators such that $0 \leq S \leq T$. If T is an M-weakly compact operator, $S^3 \in Ring(T)$.*

Proof. Since T is a M-weakly compact operator, T is a b-semicompact operator. By the definirtion of b-semicompact operator, S is b-semicompact operator. If you pass to adjoint, then $0 \leq S' \leq T'$ holds. So, T' is an L-weakly compact operator. By the domination property, S' is a b-semicompact operator. Therefore, the result is true. ∎

References

[1] C.D. Aliprantis,O.Burkinshaw, *Positive Operators*, Academic Press, New York,1985.

[2] S.Alpay, B.Altin, C.Tonyali, *On property (b) of vector lattices*, Positivity 7(2003), 135-139.

[3] S.Alpay, B.Altin, *A note on b-weakly compact operators*, Positivity 11(2007), 575-582.

[4] N.Machrafi, K.El-Fahri,M.Moussa, *A note on b-semicompact sets and operators*, Rend.Circ.Mat.Palermo(2016), 47-53.

[5] P.Meyer-Nieberg, *Banach Lattices*, Springer, Berlin, 1991.

[6] A.R.Schep, *Semicompact operators*, Indag.Math.N.S.1(1)(1990), 115-125.

[7] A.C.Zaanen, *Riesz Spaces II*, North Holland, Amsterdam,1983.

ROMAI J., v.14, no.1(2018), 149–158

NUMERICAL STUDY OF THE EROSION PROCESS

Stelian Ion, Dorin Marinescu, Ştefan Gicu-Cruceanu

"Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy, Bucharest, Romania

ro_diff@yahoo.com, dorin.marinescu@ima.ro, stefan.cruceanu@ima.ro

Abstract The erosion is a complex process determined by different physical and biological factors. It is difficult to find a mathematical model that captures all the details of the sediment transport through water flow in the presence of vegetation. Consequently, we build a model that captures the essence of the phenomenon by coupling Saint-Venant type equations for water dynamics with a Hairsine-Rose type model for soil erosion, both taking into account the presence of the plants on the soil surface. For numerical purposes, we discretize the PDE system using a finite volume method for the space variables. To advance in time, we use a two-step fractionary method. Finally, several numerical results are presented.

Keywords: shallow water equation, soil erosion process, numerical approximation.
2010 MSC: 35K55, 65M08.

1. INTRODUCTION

Roughly speaking, the soil erosion is understood as the moving process, due to different agents like wind, water or gravitational force, of a certain quantity of soil from a soil surface point to another point. The detachment of the soil particles and their transportation are very complicated and hardly quantifiable processes. In addition, the presence of the vegetation increases the difficulty of the mathematical modeling problem: the plant stems strongly interact with the water motion and the roots modify the physico-chemical properties of the soil.

Modeling at a catchment scale involves large variations in the physico-mechanical properties of the soil, in its topography, as well as in its plant cover structure. Given the complexity of the erosion processes, there are plenty of models, each of them performing well for a narrowed class of factors that affects the soil erosion. A very brief classification of the erosion models divides them into empirical models (the most known in this class is RUSLE [6]) and physically based model. For a deeper classification and comprehensive review, the reader can see for example [2].

Among physical based models, the shallow water equation for water dynamics and Hairsine-Rose model for erosion process [1, 5, 9] are most suited to be used in order to analyze the plant influence on the water dynamics and soil erosion. Tu summarize this paper, we introduce such a mathematical model in section 2 and present an analytical solution in section 3. This solution can be used as a validation tool for the numerical method briefly presented in section 4. In the end, we present several scenarios concerning the environmental variables and give some conclusions.

2. MATHEMATICAL MODEL OF EROSION IN THE PRESENCE OF VEGETATION

To produce a useful model, one is forced to retain the driving factors of the process and find the best suited mathematical objects that model them and their interactions. Since our concern is to qualitatively analyze the plant effect on the erosion of the soil, we build our model based on the following assumptions:

– the main driving force of the erosion is the water flow;

– the soil surface is covered by plants which can be considered as rigid sticks that are taller than the water level;

– the sediment consists of two phases: suspended particles in water and deposited layer;

– there is a mass exchange between the suspended and deposited phases;

– the sediment in the deposited layer can only move on vertical direction by saltation process.

In this context, the model must take into account the water dynamics and the water interactions with plant stems and soil surface. Consequently, we propose a model that couples a variant of the shallow water equation model and a modified Hairsine-Rose model for soil erosion.

Both models differs from their classical formulations by the presence of a new function that takes into account the density of the vegetation. In this coupled model, the soil surface is modeled by the altitude function $z(x)$ and the vegetation is quantified by the porosity function $\theta(x)$. The reader should note that $\theta(x) = 1$ for bare soil and $\theta(x)$ plays the role of soil porosity in the porous media theory. Note also that these altitude and porosity functions are environmental variables which must be know from measurements.

The hydrodynamic variables that account for water moving on the soil surface are the water depth $h(t, x)$ and the components $v^a(t, x)$, $a = 1, 2$ of the water velocity $\boldsymbol{v}$.

The sediment is partitioned into N size classes and it consists of a suspended phase and a deposited phase. Let us denote the mass density of the suspended sediment of size class α by $\rho_\alpha(t,x)$ and the mass density of the deposited sediment of size class α by $m_\alpha(t,x)$.

The state variables of the model are $\{h(t,x),\, v^a(t,x),\, \rho_\alpha(t,x),\, m_\alpha(t,x)\}$ and their evolution is governed by the equations

$$\begin{gathered}\partial_t(\theta h) + \partial_a(\theta h v^a) = 0,\\ \partial_t(\theta h v^a) + \partial_b(\theta h v^a v^b) + \theta h g \delta^{ab}\partial_b(z+h) = \tau_v^a + \tau_s^a, \quad a = 1,2,\end{gathered} \tag{1}$$

$$\partial_t(\theta h \rho_\alpha) + \partial_a(\theta \rho_\alpha h v^a) = \theta(e_\alpha + e_\alpha^r - d_\alpha), \quad \alpha = \overline{1,N}, \tag{2}$$

$$\partial_t m_\alpha = \theta(d_\alpha - e_\alpha^r), \quad \alpha = \overline{1,N}. \tag{3}$$

where g denotes the gravitational acceleration. The terms τ_v^a and τ_s^a quantify the water-plant and water-soil interactions, respectively. The erosion and sedimentation processes are modeled through the terms e_α - entrainment rate, e_α^r - re-entrainment rate and d_α - deposition rate of the sediment from the size class α, respectively.

One assumes that the flow resistance exercised by plants and soil obeys laws (4) and (5), respectively

$$\tau_v^a = -\alpha_v h\,(1-\theta)\,|\boldsymbol{v}|v^a, \tag{4}$$

$$\tau_s^a = -\theta\alpha_s|\boldsymbol{v}|v^a, \tag{5}$$

where α_v and α_s are material parameters. The coefficient α_v depends on the geometry of the plants from the vegetation cover, while α_s depends on the soil roughness.

In the Hairsine-Rose model [1, 5, 9], the entrainment and deposition rates are given by

$$\begin{aligned} d_\alpha &= \nu_{s,\alpha}\cdot\rho_\alpha,\\ e_\alpha &= p_\alpha(1-H)\frac{F\,(\Omega-\Omega_{cr})_+}{J},\\ e_\alpha^r &= H\frac{m_\alpha}{m_t}\frac{\gamma_s}{\gamma_s-1}\frac{F\,(\Omega-\Omega_{cr})_+}{gh},\end{aligned} \tag{6}$$

where p_α is the proportion of the sediment in the original soil, $\nu_{s,\alpha}$ is the settling velocity of the sediment in the size class α, and γ_s is specific weight of sediment.

The parameters F - effective fraction of power stream, J - energy of soil particle detachment and Ω_{cr} - critical power stream are specific to a given type of soil.

The erosion processes are controlled by the water flow through the stream power Ω. In the present paper, we use the law

$$\Omega = \rho_{\mathrm{w}} |\tau_s| |\boldsymbol{v}|. \tag{7}$$

The function

$$H = \min \left\{ \frac{m_t}{m_t^\star}, 1 \right\} \tag{8}$$

plays the role of a protecting factor of the original soil to the erosion process. The terms

$$m_t = \sum_{a=1}^{N} m_a$$

and $m_t^\star$ from (8) are the total mass of sediment deposited on the soil and the mass required to protect the original soil from erosion, respectively.

3. ANALYTIC SOLUTIONS

Despite the complexity of the model equations (1-3), there are some configurations of the soil surface and vegetation density distribution that allow us to obtain the analytic solutions. Let us consider the case of the plain soil surface with constant vegetation density. For such case, the problem reduces to a 1-D model equation. Let $\partial_x z = -s_0$ be the constant gradient of the soil surface and $\theta(x) = \theta_0$ be the porosity of the cover plant. If h_0 and v_0 satisfy

$$v_0^2 = \frac{\theta_0 g h_0 s_0}{\alpha_v h_0 (1 - \theta_0) + \theta_0 \alpha_s}, \tag{9}$$

then

$$h(t, x) = h_0, \quad v(t, x) = v_0$$

is a solution of the shallow water equation (1).

In the case of uniform flow $h(t, x) = h_0$, $v(t, x) = v_0$, one can find analytic solutions of sediment equations for bare soil, see [7, 9]. Similarly, one can find analytic solution when vegetation is present, but uniform flow must be assumed.

We introduce the following notations:

$$\Gamma := \frac{\gamma_s}{\gamma_s - 1} \frac{F(\Omega(v_0) - \Omega_{crt})_+}{g h_0}, \quad \Lambda := \frac{F(\Omega(v_0) - \Omega_{crt})}{J}, \quad q := h_0 v_0. \tag{10}$$

Net erosion

For a numerical validation purpose, we introduce the analytic solution of the net-erosion process. But we must remark that for the uniform flow, the solutions of the sediment are similar for both cases: with or without vegetation. What differs is the equilibrium relations among the hydrodynamic variables h_0, v_0, θ_0 and s_0.

Water flow generates a net erosion of the soil if the total mass m_t of the deposited sediment is smaller than the m_t^*, i.e. $H := m_t/m_t^* < 1$. In order to have a steady state of m_α, the following must hold

$$\nu_{s,\alpha} \cdot \rho_\alpha = \Gamma \frac{m_\alpha}{m_t^*}. \tag{11}$$

The suspended sediment solves the equations

$$q \frac{\mathrm{d}\rho_\alpha}{\mathrm{d}x} = p_\alpha \Lambda \left(1 - \frac{1}{\Gamma} \sum_{\beta=1}^{N} \nu_{s,\beta} \cdot \rho_\beta \right), \tag{12}$$

Using (11) and (12), one obtains an equation for the ratio m_t/m_t^*

$$q \frac{\mathrm{d}}{\mathrm{d}x} \frac{m_t}{m_t^*} = \frac{\Lambda}{\Gamma} \sum_\alpha \nu_{s,\alpha} \cdot p_\alpha \left(1 - \frac{m_t}{m_t^*} \right) \tag{13}$$

which has the solution

$$\frac{m_t}{m_t^*}(x) = 1 + \left(\frac{m_t}{m_t^*} - 1 \right)_{x=0} \exp\left(-\frac{\Lambda}{q\Gamma} \sum_\alpha \nu_{s,\alpha} \cdot p_\alpha x \right). \tag{14}$$

The condition $H < 1$ is satisfied for all $x > 0$ if

$$\frac{m_t}{m_t^*}(0) = \frac{1}{\Gamma} \sum_\alpha \nu_{s,\alpha} \cdot \rho_\alpha(0) < 1. \tag{15}$$

Using the solution (14), one obtains

$$\rho_\alpha(x) = \rho_\alpha(0) + \frac{p_\alpha \Gamma}{\sum_\beta \nu_{s,\beta} \cdot p_\beta} \left(\frac{1}{\Gamma} \sum_\beta \nu_{s,\beta} \cdot \rho_\beta(0) - 1 \right) \left[\exp\left(-\frac{\Lambda}{q\Gamma} \sum_\beta \nu_{s,\beta} \cdot p_\beta x \right) - 1 \right]. \tag{16}$$

4. NUMERICAL SCHEME

The numerical scheme we use in this paper is obtained using a finite volume method to discretize the space variable and then a fractional time step method to integrate an ODE system.

Using a finite volume method to approximate (1-3), one gets an ODE system of the form

$$\partial_t U + \mathcal{F}(U) = \mathcal{R}(U), \tag{17}$$

where U is the vector of the unknowns, $\mathcal{F}$ is the "flux" term generated by the derivative of the space variable, and $\mathcal{R}$ is the "source" term.

The fractional time step is a method for approximating the solution of the ODEs by splitting the initial model into two sub-models, separately integrating each of these sub-models and then combining the two obtained solutions, [8, 10].

Let us consider

$$\partial_t U + \mathcal{F}(U) = \mathcal{S}_1(U) + \mathcal{S}_2(U), \tag{18}$$

and let $\mathcal{E}^1(t)$ be an approximating evolution operator of

$$\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}t} = \mathcal{S}_1(\mathcal{U})$$

and $\mathcal{E}^2(t)$ an approximating evolution operator of

$$\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}t} + \mathcal{F} = \mathcal{S}_2(\mathcal{U}).$$

Then, a second order approximation of the problem (18) is given by

$$\mathcal{U}(t + \triangle t) = \mathcal{E}^1(\triangle t/2)\mathcal{E}^2(\triangle t)\mathcal{E}^1(\triangle t/2)\mathcal{U}(t). \tag{19}$$

One can find the explicit form terms used for (17-19) in the paper [4]. Here, we note that $\mathcal{F}(U)$ consists of the discrete approximation of conservative flux in (1), $\mathcal{S}_1(U)$ counts for the entrainment rate and $\mathcal{S}_2(U)$ takes into account the remaining terms in (1-3).

5. NUMERICAL SIMULATION. 1-D CASE

In this section we present some numerical tests in order to illustrate the ability of our model and our numerical scheme to capture the effects of the variation of the soil surface geometry and the vegetation density on the erosion process. In all these numerical experiments, we will work with the $1-D$ domain $\Omega = (0, L)$.

The analysis is performed starting from a reference case given by a bare soil with constant slope. We also assume that there is only one size class of sediment, $N = 1$.

In all the cases we will work here, we consider that there is an inflow of water and sediment at the upper side ($x = 0$) of the slope

$$h|_{x=0} = h_0, \quad v|_{x=0} = v_0, \quad \rho_1|_{x=0} = \rho^0$$

and free discharge at the other boundary $x = L$.

The values of the hydrodynamic variables and of the parameters for the soil and sediment are given in Table 1.

Table 1 The data used in the reference case.

$h_0[m]$	$v_0[ms^{-1}]$	F	$J[Jkg^{-1}]$	$\Omega_{\mathrm{cr}}[Wm^{-2}]$	m_t^*	s_0	p_1	$\nu_1[ms^{-1}]$
$5e^{-2}$	0.2	0.01	0.2	0.007	100	0.0625	1	0.003
$q = 0.01\, m^2 s^{-1}$, $\rho^0 = 13.51\, kg\, m^{-3}$								

Comparison of the numerical and analytic solution for net erosion

For the reference case previously considered, the analytic solutions of $m_1(x)$ and $\rho_1(x)$ are known (given by (14) and (16), respectively). Figure 1 shows that the numerical solutions provided by our scheme follows closely the path of the exact solutions.

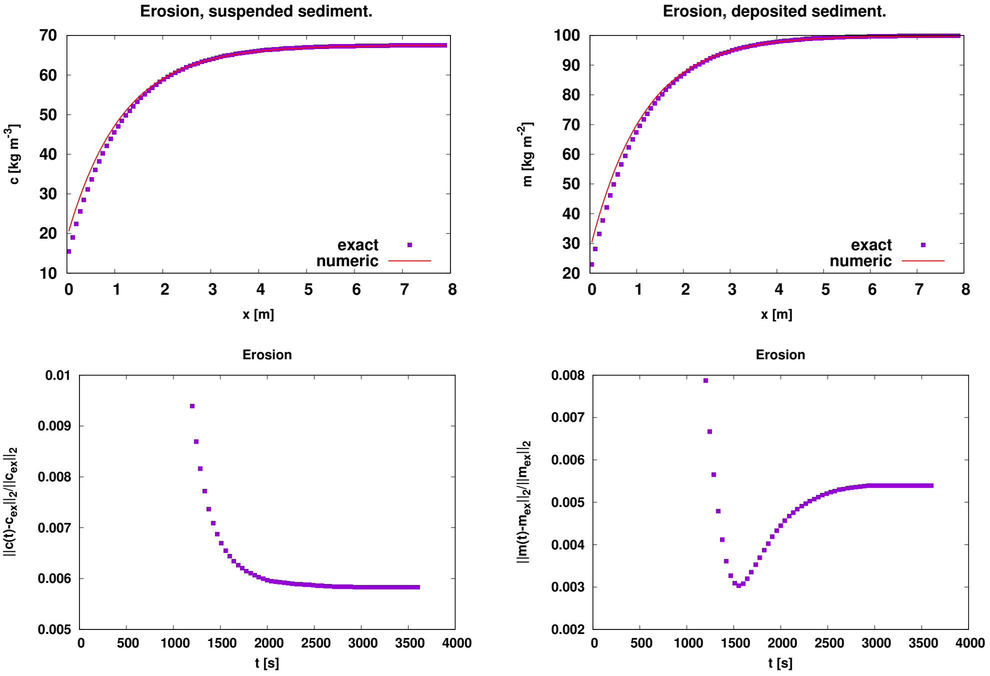


Fig. 1. Comparison of the exact and numerical solutions for the net erosion. The input parameters are given in Table 1. The distributions of the suspended and deposited sediment are illustrated in the left and right pictures of the first row, respectively. The asymptotic behavior of the relative error of the solutions is presented in the second row.

The influence of the soil surface topography on the erosion process
We now consider the case of a bare soil with piecewise constant slope. The soil surface $z(x)$ on the domain Ω is formed by the two segments of different slopes connected at $x = L/2$, segments resulting from the surface line of the previous case by moving the point $(L/2, z(L/2))$ to $(L/2, z(L/2)(1-\delta z))$. One can easily observe (see Figure 2) that a change in the soil surface topography will immediately bring an expected change in the erosion process: the higher the surface gradient is, the higher the erosion process is also.

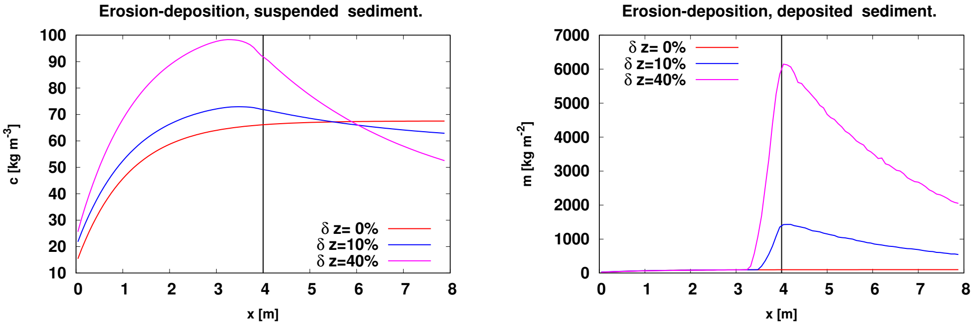


Fig. 2. The influence of soil gradient variation on the erosion process. The input parameters, except the slope s_0, are given in Table 1.

The influence of the vegetation density on the soil erosion
For this last case, let us consider a plant barrier on a bare soil modelled by

$$\theta(x) = \begin{cases} \theta_0, & x \in (x_0, x_1) \\ 1, & x \in (0, x_0) \cup (x_1, L) \end{cases} .$$

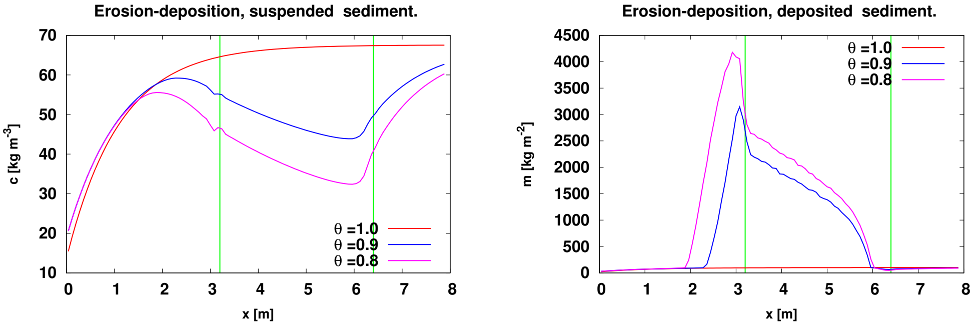


Fig. 3. The influence of vegetation density variation on the erosion process. The input parameters are given in Table 1.

The distributions of the hydrodynamic variables are shown in Figure 4. Note that plants act like a barrier against the water flow, the water level is raising in front of the filter and its speed decreases.

As expected, a decreasing level of velocity induces a reduction of the erosion rate and in the same time an increase of the deposition rate. The magnitude of these modification rates depends on the plant density in the filter, see Figure 3.

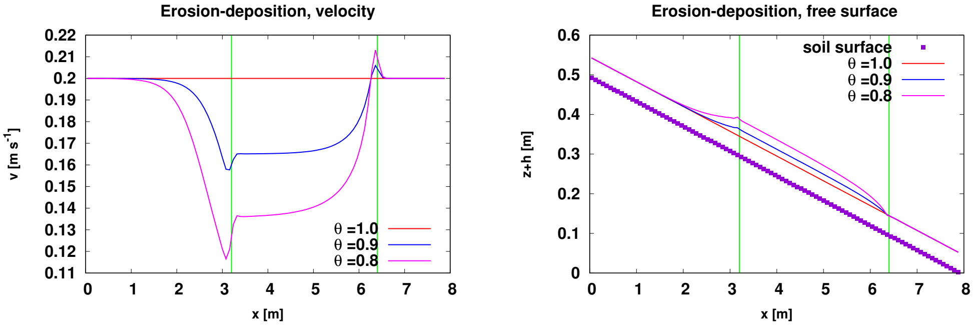


Fig. 4. The influence of vegetation density variation on hydrodynamic variables, velocity (left) and free surface (right). The input parameters are given in Table 1.

6. CONCLUSION

In this paper, we propose a mathematical model and a numerical scheme to integrate it for water motion and soil erosion on downhill. The model takes into account the presence of the plant cover. It was shown that the numerical solution is asymptotically convergent (in time) to the net erosion solution, see Figure 1. Our model is able to capture the essential features of the physical process induced by the variation of the plant cover density, see Figure 4, or the variation of the soil surface gradient, see Figure 3.

Acknowledgements. 1. This work was supported by a grant of Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, project PN-III-P1-1.2-PCCDI-2017-0721/34PCCCDI/2018 within PNCDI III.

2. The authors wish to thank to the anonymous referee for carefully reading the manuscript. His (her) pertinent observations helped us to improve the paper.

References

[1] P.B. Hairsine, C.W. Rose, *Modeling water erosion due to overland flow using physical principles: 1. Sheet flow*, Water Resour. Res., **28**(1)(1992), 237–243.

[2] Mohammad Hajigholizadeh, Assefa M. Melesse, Hector R. Fuentes, *Erosion and Sediment Transport Modelling in Shallow Waters: A Review on Approaches, Models and Applications*, Int. J. Environ. Res. Public Health 2018, **15**(3), 518, 1–24.

[3] S. Ion, D. Marinescu, S.G. Cruceanu, *Overland flow in the presence of vegetation*, Technical report, www.ima.ro/PNII_programme/ASPABIR/pub/report_ismma_aspabir_2013.pdf

[4] S. Ion, D. Marinescu, S.G. Cruceanu, *Mathematical modeling of the erosion processes in hydrographic basins*, Advanced Topics in Electrical Engineering (ATEE), 2015, 552–555.

[5] Jongho Kim, Valeriy Y. Ivanov, Nikolaos D. Katopodes, *Modeling erosion and sedimentation coupled with hydrological and overland flow processes at the watershed scale*, Water Resour. Res., **49**(2013), 5134–5154.

[6] K.G. Renard, G.R. Foster, G.A. Weesies, J.P. Porter, RUSLE*: Revised Universal soil Loss Equation*, Journal of Soil and Water Conservation, **46**(1)(1991), 30–33.

[7] G.C. Sander, P.B. Hairsine, L. Beuselinck, G. Govers, *Steady state sediment transport through an area of net deposition: Multisize class solutions*, Water Resour. Res., **38**(6)(2002), 1087, 23.1–23.8.

[8] R.L. LeVeque, *Time-Split Methods for Partial Differential Equations*, Phd. Thesis, Stanford University, 1982.

[9] G.C. Sander, J.-Y. Parlange, D.A. Barry, M.B. Parlange, W.L. Hogarth, *Limitation of the transport capacity approach in sediment transport modeling*, Water Resour. Res., **43**(2007), W02403, 1–9.

[10] G. Strang, *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal., **5**(1968), 506–517.

ROMAI J., v.14, no.1(2018), 159–167

ON CHARACTERIZATION OF BESSEL SYSTEMS

Al-jourany Khalid Hadi Hameed

Ph.D. Student, Department of Functions and Approximations Theory,

Faculty of Mathematics and Mechanics, Saratov State University, Saratov, Russia.

khalidaljourany@ymail.com

Abstract In some the previous work, we have studied an affine system of Walsh type generated by a periodic function in connection with a multishift in Hilbert space. In this paper, we give a new method for characterization of Bessel system. This method is based on the consideration of the question:under which necessary and sufficiently conditions on the function φ an affine system of functions of the Walsh type $\{\varphi_n\}_{n\geq 0}$ to be Bessel system in the space $L^2(0,1)$?Finally, some examples are given to explain our representation method.

Keywords: affine system of Walsh type, Rademacher system, Walsh-Paley system, Cuntz algebra Θ_n, Bessel system.

2010 MSC: 97U99.

1. INTRODUCTION

The new notion of affine system of Walsh type was introduced, studied and proved results about orthogonalizing and completion with preservation of structure of affine system by Terekhin P.A.[1]. Our results work in [2] about affine system of Walsh type can be classified in to three sections results:first, on the basis of the functional analytic structure of a multishift in a Hilbert space, which is a generalized analogue of the operator(simple, one-side)shift and closely related to the representations of the Cuntz C^*-algebra, the definition of an affine system of functions of the Walsh type was given, second, various criteria and signs of the completeness of affine systems of functions were given, finally, the minimality of the affine system is established as well as an explicit form of the biorthogonally conjugate system of functions was indicated and its completeness was established. Mironov V.A., Sarsenbi A.M., and Terekhin P.A.,[10] studied an affine Bessel sequences in connection with the spectral theory and the multishift structure in Hilbert space. They constructed a non-Besselian affine system $\{u_n(x)\}_{n=0}^{\infty}$ generated by continuous periodic function $u(x)$. Their results were based on Nikishin's example concerning convergence in measure, also they showed that affine systems $\{u_n(x)\}_{n=0}^{\infty}$ generated by any Lipchitz function $u(x)$ are Besselian.

Definition 1.1. *Let H be a Hilbert space, and*

$$W_0, W_1 : H \to H$$

isometric operators operating in space H. Let's say that the two isometrics W_0 and W_1 define the structure of multishifts, if there is a vector $e \in H$ such that:

$$W_{\alpha_1} \ldots W_{\alpha_{k-1}} e, \alpha_v \in \{0,1\}, 0 \leq v \leq k-1, k \geq 0,$$

forms an orthonormal basis of the space H.

Remark 1.1. *The concepts of multishift was introduced and studied in many works [3–5].*

Suppose that, the function $\varphi(s)$, $s \in \Re$, (where $\Re$ is a real number space), satisfied the condition:

$$\varphi(s) \in L^2[0,1], \int_0^1 \varphi(s) ds = 0, \varphi(s+1) = \varphi(s),$$

and let $L_0^2 = L_0^2(0,1)$ be a space such functions (where, L_0^2 is the space of square - integral and having a zero integral), as well as, we denote a linear operators in this space as:

$$W_0 \varphi(s) = \varphi(2s), W_1 \varphi(s) = r(s) \varphi(2s), \tag{1}$$

where $r(s)$ is the periodic function:Haar-Rademacher-Walsh.
For any $n \in N$, using the binary representation, $n = \sum_{v=0}^{k-1} \alpha_v 2^v + 2^k$ we set:

$$\varphi_n(t) = \varphi_\alpha(t) = \varphi_{kj}(t) = W^n \varphi(t) = W^\alpha \varphi(t) = W_{\alpha_1} \ldots W_{\alpha_k} \varphi(t),$$

where,
$k = 0, 1, ...; j = 0, 1, ..., 2^{k-1}, \alpha = (\alpha_1, \ldots, \alpha_k) \in \Omega, \Omega = \bigcup_{k=0}^{\infty} \{0,1\}^k$ Besides, we set $\varphi_0(t) \equiv 1$,

$$W_{\alpha_1} \ldots W_{\alpha_k},$$

denote the product of the operators:the operator W_{α_k} acts first, W_{α_1} acts last, and the empty product is set the equal to the identity operator I. For any function $\varphi \in L_0^2$, we have:

$$\varphi_\alpha(t) = W^\alpha \varphi(t) = W_{\alpha_0} \ldots W_{\alpha_{k-1}} \varphi(t) =$$

$$\varphi(2^k t) r^{\alpha_{k-1}}(2^{k-1} t) \ldots r^{\alpha_0}(t) = \varphi(2^k t) \prod_{v=0}^{k-1} r_v^{\alpha_v}(t),$$

where, $r_k(t) = r(2^k t), k = 0, 1, ...$ is Rademacher system .

Definition 1.2. *The system $\{\varphi_n\}_{n\geq 0} = \{W^{\alpha}\varphi\}$ is the affine system of Walsh type of the function φ without the constant $\varphi_0(t) \equiv 1$.*

If the generating function select $\omega(t) = r(t)$, then the system $\{\omega_n\}_{n=0}^{\infty}$ will the classical system of Walsh-Paley system. Walsh functions (without constant $\omega_0(t) \equiv 1$):

$$\omega_n(t) = \omega_\alpha(t) = W^{\alpha}\omega(t) = W_{\alpha_0} \dots W_{\alpha_{k-1}}\omega(t) = r_k(t) \prod_{v=0}^{k-1} r_v^{\alpha_v}(t),$$

forms an orthonormal basis of the space $H = L_0^2(0,1)$, therefore according to the definition(1.1) operators:

$$W_0\varphi(t) = \varphi(2t), W_1\varphi(t) = r(t)\varphi(2t),$$

define the structure of multishift[2].

Definition 1.3. [6]. *The Walsh-Paley system, $\omega = (\omega_n, n \in N)$ is defined as:if $n = \sum_{k=0}^{\infty} n_k 2^k \in N \cup \{0\}$ has binary coefficient $(n_k, k \in N \cup \{0\})$, then*

$$\omega_n = \prod_{k=0}^{\infty} r_k^{n_k}, \tag{2}$$

where,

$$r(x) = \begin{cases} 1, & x \in (0, 1/2) \\ -1, & x \in (1/2, 1) \end{cases}$$

$r\left(x + k\right) = r\left(x\right)$, $x \in (0,1)$, $k \in N$ and $r_k(x) = r(2^k x)$, $x \in \Re$, $k \in N \cup \{0\}$, where $r(x)$ is the Rademacher functions.

Definition 1.4. *A system (sequence) $\{\varphi_n\}_{n\in N}$ in Hilbert space H is called a Bessel system, if there exists a positive constant B for which*

$$\sum_{n=1}^{\infty} \left|(g, \varphi_n)\right|^2 \leq B \left\|g\right\|^2, \forall g \in H. \tag{3}$$

Definition 1.5. [7]. *Let $n \geq 0$, the Cuntz algebra Θ_n is the C^*-algebra generated by some isometries $(S_i)_{i\in Z_n}$ satisfying the Cuntz relations:*

$$S_i^* S_j = \delta_{ij} I, \sum_{i\in Z_n} S_i^* S_j = I, \tag{4}$$

where, $i, j \in Z_n$.

It should be noted that the extensions of operators $\{W_0, W_1\}$ to the space $L^2\,(0,1)$ of periodic function $\varphi(t)$ are defined by:

$$V_0\varphi(t) = \varphi(2t), V_1\varphi(t) = r(t)\varphi(2t). \tag{5}$$

From equation(1.4), we have representation of the Cuntz algebra Θ_2, which satisfy the Cuntz relations:

$$V_i^* V_j = \delta_{ij} I,$$
$$V_0 V_0^* + V_1 V_1^* = I.$$

Thus, the operators structure of the multishift $\{W_0, W_1\}$ is a restriction to the subspace $L^2(0,1)$ of the representation $\{V_0, V_1\}$ in the space $L^2(0,1)$ of the Banach C^*-algebra of Cuntz Θ_2.

2. THE MAIN RESULTS WITH EXAMPLES

Lemma 2.1. *The system* $\left\{\varphi_{k,j}\right\}_{j=0}^{2^{k-1}}$ *(k-fixed) is orthogonal block system.*

Proof The system $\left\{\varphi_{k,j}\right\}_{j=0}^{2^{k-1}} = \{W^\alpha\varphi\}_{\alpha\in\Omega}$. This is implies to that:

$$W^\alpha\varphi \in W^\alpha H.$$

Since,

$$W^\alpha H \perp W^\beta H, \alpha \neq \beta, |\alpha| = |\beta| = k.$$

Also,

$$W^\alpha\varphi \in W^\alpha H,$$
$$W^\beta\varphi \in W^\beta H.$$

Then, we have:

$$\left(W^\alpha\varphi, W^\beta\varphi\right) = 0.$$

From above, we have that: $\{W^\alpha\varphi\}_{|\alpha|=k}$ is orthogonal block system.

Lemma 2.2. *For all* $\alpha, \beta \in \Omega$, *we have:*

$$\left(\omega_\alpha, \varphi_\beta\right) = \begin{cases} (\omega_\alpha, \varphi)\,, & if \alpha = \beta\gamma \\ 0, & o.w. \end{cases}$$

Proof. Write the Fourier-Walsh series of the function φ as:

$$\varphi = \sum_{\gamma\in\Omega} (\varphi, \omega_\gamma)\, \omega_\gamma.$$

Also, we have:

$$\varphi_\beta = W^\beta \varphi = \sum_{\gamma \in \Omega} (\varphi, \omega_\gamma) W^\beta \omega_\gamma = \sum_{\gamma \in \Omega} (\varphi, \omega_\gamma) \omega_{\beta\gamma}.$$

On other hand

$$\varphi_\beta = \sum_{\gamma \in \Omega} \left(\varphi_\beta, \omega_\alpha\right) \omega_{\beta\alpha}, \alpha = \beta\gamma.$$

The coefficient of the Fourier-Walsh series are unique. Also, if $\alpha = \beta\gamma$ for some $\gamma \in \Omega$, then

$$\left(\varphi_\beta, \omega_\alpha\right) = (\varphi, \omega_\gamma).$$

It should be noted that, if α can not be expressed as $\beta\gamma, \forall\gamma \in \Omega$, then

$$\left(\varphi_\beta, \omega_\alpha\right) = 0.$$

∎

Theorem 2.1. *Let $\varphi \in L^2(0,1)$,$supp\varphi \subset [0,1]$, $\int_0^1 \varphi(t)dt = 0$. If the inequality:*

$$\sum_{k=0}^{\infty} \left(\sum_{j=0}^{2^{k-1}} |(\varphi, \omega_{kj})|^2 \right)^{1/2} = c < \infty.$$

Then the affine system of Walsh type $\{\varphi_n\}_{n \geq 0}$ is Bessel system with Bessel constant $B = max\{1, c\}^2$.

Proof. Write the Fourier-Walsh series of the function φ as:

$$\varphi = \sum_{\alpha \in \Omega} x_\alpha \omega_\alpha,$$

and write the polynomial of affine system $\{\varphi_n\}_{n \geq 1}$ finite sum as:

$$P = \sum_{\beta \in \Omega} c_\beta \varphi_\beta.$$

We consider for $k = 0, 1, \ldots$, the Walsh-Paley polynomials can be represented as:

$$P_k = \sum_{|\alpha|=k} x_\alpha \sum_{\beta \in \Omega} c_\beta \omega_{\beta\alpha}.$$

The system $\{\omega_{\beta\alpha} : |\alpha| = k(k - fixed), \beta \in \Omega\}$ is orthogonal system.

$$\omega_{\beta\alpha} = \omega_{\beta'\alpha'}, \left|\alpha'\right| = k, \beta'\alpha' \in \Omega, \alpha = \alpha', \beta = \beta', \beta' \in \Omega.$$

Now, if $\beta\alpha = \beta^{'}\alpha^{'}$, then:

$$|\alpha| + |\beta| = \left|\alpha^{'}\right| + \left|\beta^{'}\right|, |\alpha| = \left|\alpha^{'}\right| and \, |\beta| = \left|\beta^{'}\right|, \alpha = \alpha^{'} and \beta \; = \beta^{'}.$$

We can count:

$$\|P_k\| = \left(\sum_{|\alpha|=k, \beta\in\Omega} |x_\alpha c_\beta|^2 \right)^{1/2} = \left(\sum_{|\alpha|=k} |x_\alpha|^2 \right)^{1/2} \left(\sum_{\beta\in\Omega} |c_\beta|^2 \right)^{1/2}.$$

And

$$\sum_{k=0}^{\infty} \|P_k\| = \left(\sum_{\beta\in\Omega} |c_\beta|^2 \right)^{1/2} . \sum_{k=0}^{\infty} \left(\sum_{|\alpha|=k} |x_\alpha|^2 \right)^{1/2} < \infty.$$

We calculate:

$$(P, \omega_\gamma) = \sum_{\beta\in\Omega} c_\beta \left(\varphi_\beta, \omega_\gamma\right) = \sum_{\alpha,\beta:\gamma=\beta\alpha} c_\beta \left(\varphi, \omega_\alpha\right) = \sum_{\alpha,\beta:\gamma=\beta\alpha} x_\alpha c_\beta,$$

(By using Lemma(2.2)).

$$\left(\sum_{k=0}^{\infty} P_k, \omega_\gamma \right) = \sum_{k=0}^{\infty} (P_k, \omega_\gamma) = \sum_{k=0}^{\infty} \sum_{|\alpha|=k} x_\alpha \sum_{\beta\in\Omega} c_\beta \left(\omega_{\beta\alpha}, \omega_\gamma\right) = \sum_{\alpha,\beta:\gamma=\beta\alpha} x_\alpha c_\beta.$$

Since, $(\omega_{\beta\alpha}, \omega_\gamma) = \delta_{\beta\alpha,\gamma}$.
From the above, we have the following induction:$P = \sum_{k=0}^{\infty} P_k$!
Now:

$$\|P\| \le \sum_{k=0}^{\infty} \|P_k\| = \sum_{k=0}^{\infty} \left(\sum_{|\alpha|=k} |x_\alpha|^2 \right)^{1/2} \left(\sum_{\beta\in\Omega} |c_\beta|^2 \right)^{1/2},$$

we have:

$$\left\| \sum_{\beta\in\Omega} c_\beta \varphi_\beta \right\| \le \|\varphi\|^* \left(\sum_{\beta\in\Omega} |c_\beta|^2 \right)^{1/2},$$

where, $\|\varphi\|^* = \sum_{k=0}^{\infty} \left(\sum_{|\alpha|=k} |x_\alpha|^2 \right)^{1/2}$.
It is equivalent to Bessel inequality:

$$\left(\sum_{\beta\in\Omega} \left|\left(g, \varphi_\beta\right)\right|^2 \right)^{1/2} \le \|\varphi\|^* \|g\|,$$

$$\left(\sum_{k=0}^{\infty}|(g,\varphi_n)|^2\right)^{1/2} \leq \left(\left(\int_0^1 g(t)dt\right)^2 + \sum_{\beta\in\Omega}\left|(g,\varphi_\beta)\right|^2\right)^{1/2},$$

$$\leq max\left\{1,\|\varphi\|^*\right\}.\|g\|^2,$$

$$\sum_{n=0}^{\infty}|(g,\varphi_n)|^2 \leq B\|g\|^2, B = max\left\{1,\|\varphi\|^*\right\}^2.$$

Then, we have:if

$$\|\varphi\|^* = \sum_{k=0}^{\infty}\left(\sum_{|\alpha|=k}|x_\alpha|^2\right)^{1/2} < \infty.$$

Then the affine system of Walsh type $\{\varphi_n\}_{n\geq 0}$ is Bessel system. ▮

Remark 2.1. *Theorem (2.1) in this paper is an analog of some results obtained by the authors in [8–10].*

We are going to give some examples to apply theorem (2.1). These examples are based on consideration that:H^∞ is the Banach algebra of analytic functions on the open unit disk and $G(H^\infty)$ is the group of invertible elements of the algebra H^∞. Note that for ζ to be belong to $G(H^\infty)$, it is necessary and sufficient that the function $\zeta(z)$ be analytic on the disk $(|z| < 1)$ and that the following inequalities be valid:

$$0 < inf\left|\zeta(z)\right|, sup\left|\zeta(z)\right| < \infty.$$

Let $\varphi \in R(H)$, where $R(H)$ is the space of Rademacher. Let $R(H) = \overline{span}[r_k]$be linear closure of the span Rademacher system $\{r_k\}_{k=0}^{\infty}$. The space $R(H)$ invariant with respect to W_0 and the multishift operator $R(H)$ as:

$$r_k = W_0^k r, k = 0, 1, \ldots.$$

And

$$\varphi(t) = \sum_{k=0}^{\infty}a_k r_k, \sum_{k=0}^{\infty}|a_k|^2 < \infty. \tag{6}$$

We assign the analytic function:

$$\phi(z) = \sum_{k=0}^{\infty}a_k z^k. \tag{7}$$

In the unite disk $D = (|z| < 1)$ of Hardy space $H^2(D)$, with the coefficient a_k from equation(2.6). This mapping is an isometric isomorphism of $R(H)$ on to

Hardy space $H^2(D)$, and the restriction of W_0 to $R(H)$ is unitary equivalent by this mapping to the operator of multiplication by z, i.e. is a shift operator.

Theorem 2.2. [11].*Let $\{\omega_n\}_{n\geq 0}$ be the Walsh system, $\{r_k\}_{k\geq 0}$ be the Rademacher system and*

$$\varphi = \sum_{k=0}^{\infty} a_k r_k, \sum_{k=0}^{\infty} |a_k|^2 < \infty.$$

If the analytic function

$$\phi(z) = \sum_{k=0}^{\infty} a_k z^k, |z| < 1,$$

belong to $G(H^\infty)$, then the affine system of Walsh type $\{\varphi_n\}_{n\geq 0}$ is Riesz bases in $L^2(0,1)$.

Example 2.1. *The analytic function*

$$\phi(z) = \sum_{k=0}^{\infty} a_k z^k, |z| < 1,$$

has no zero in the closed unit circle, then affine system of Walsh type $\{\varphi_n\}_{n\geq 0}$forms Riesz bases and since any Riesz bases is Bessel system, then affine system of Walsh type $\{\varphi_n\}_{n\geq 0}$forms Bessel system too. Indeed of Wiener theorem an absolutely convergent series Tayler follows that $\phi \in G(H^\infty)$.

Example 2.2. *Let $\varphi(t) = 1 - 2t$, $0 < t < 1$ and satisfy the following condition $\int_0^1 \varphi(t)dt = 0$. $\varphi \in R(H)$, this meaning that, the function φ can be representation as Rademacher system:*

$$\varphi = \sum_{k=0}^{\infty} a_k r_k = \sum_{k=0}^{\infty} \frac{r_k}{2^{k+1}}.$$

Then the corresponding analytic function as:

$$\phi(z) = \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \frac{z^k}{2^{k+1}} = \frac{1}{2-z}.$$

It is observe that, $\phi \in G(H^\infty)$, then, we have the affine system of Walsh type $\{\varphi_n\}_{n\geq 0}$ forms Riesz bases and Bessel system.

Theorem 2.3. *Let $\{w_n\}_{n\geq 0}$ be the Walsh system, $\{r_n\}_{n=0}^{\infty}$ be the Rademacher system and $\varphi \in L_0^2, \varphi = \sum\limits_{k=0}^{\infty} a_k r_k,\ \sum\limits_{k=0}^{\infty} |a_k|^2 < \infty$. Then, affine system of Walsh* *type* $\{\varphi_n\}_{n\geq 1}$ *is Riesz bases iff*

$$0 < c_1 \leq |\phi(z)| \leq c_2 < \infty,$$

where,

$$\phi(z) = \sum_{k=0}^{\infty} a_k z^k, |z| < 1,$$

is analytic function.

References

[1] P. A. Terekhin, *Affine Systems of Walsh type.Orthogonalization and Completion*, Izv Saratov University(N.S).Ser.Math.,Mech.,Inform, **14**, 4(2014), 395-400.(in Russia)

[2] H. H. Khalid, V. A. Mironov, P. A. Terekhin, *Affine System of Walsh type. Completeness and Minimality*, Izv Saratov University(N.S).Ser.Math., Mech., Inform, **16**, 3(2016), 247-256. DOI:10.18500/18169791-2016-16-3-247-256.(in Russia)

[3] P. A. Terekhin, *On representation properties of a system of contractions and shift of functions on an interval*, Izv Tul'sk.Gos. Univ.,Ser.Math.,Mech.,Inform, **4**, 1(1998), 136-138.(in Russia)

[4] P. A. Terekhin, *On the multiplicative structure of the centeralizer of a multishift on a Hilbert space*, Mathem.,Mech.: Collection of Scientific Papers , Saratov,Saratov University, **2**, (Press2000), 119–122.(in Russia)

[5] P. A. Terekhin, *Multishift in Hilbert spaces*, Funct . Anal.Appl., **39**, 1 (2005), 57-67.DOI:10.1007/s10688-005-0017-5.

[6] F. W. R. Schipp, P. Simon, J. Pal, *Walsh series : an introduction to dyadic harmonic analysis*, Bristol ; N. Y. : Adam Hilger, 1990.

[7] C. Joachim, *Simple C^*-algebras generated by isometries*, Comm. Math. Phys., **57**, 2(1977), 173-185.

[8] P. A. Terekhin, *Riesz Bases Generated by Contractions and Translations of a Function on an Interval*, Mathematical Notes, **72**, 4(2002), 505-518.

[9] A. M. Sarsebi, P. A. Terekhin, *Riesz basicity for general systems of functions*, J. Function Spaces, **2014**, 1-3.article ID 860279.DOI:10.1155/2014/860279.

[10] V. A. Mironov, A. M. Sarsenbi, P. A. Terekhin, *Affine Bessel Sequences and Nikishin's Example*, **31**, 4(2017), 963-966.DOI 10.2298/FIL 1704963M.

[11] H. H. Khalid, P. A. Terekhin, *On construction of Riesz bases using Walsh type affine systems in the space $L^2(0,1)$*, Mathem.,Mech.: Collection of Scientific Papers , Saratov,Saratov University, **18** (Press2016), 3-4.(in Russia)

ROMAI J., v.14, no.1(2018), 169–185

SHANNON ENTROPY FOR IMPRECISE AND UNDER-DEFINED OR OVER-DEFINED INFORMATION

Vasile Pătraşcu

Research Center for Electronics and Information Technology,

"Valahia" University, Târgovişte, Romania

patrascu.v@gmail.com

Abstract Shannon entropy was defined for probability distributions and then its using was expanded to measure the uncertainty of knowledge for systems with complete information. In this article, it is proposed to extend the using of Shannon entropy to under-defined or over-defined information systems. To be able to use Shannon entropy, the information is normalized by an affine transformation. The construction of affine transformation is done in two stages: one for homothety and another for translation. Moreover, the case of information with a certain degree of imprecision was included in this approach. Besides, the article shows the using of Shannon entropy for some particular cases such as: neutrosophic information both in the trivalent and bivalent case, bifuzzy information, intuitionistic fuzzy information, imprecise fuzzy information, and fuzzy partitions.

Keywords: Shannon entropy, under-defined information, over-defined information, neutrosophic information, bifuzzy information, intuitionistic fuzzy information, imprecise fuzzy information, fuzzy partitions.

2010 MSC: 94A24, 94A17.

1. INTRODUCTION

The Shannon entropy [12] plays an important role in the information uncertainty computing. Thus, if the information vector is defined by formula:

$$p = (p_1, p_2, \dots, p_n) \in [0,1]^n \tag{1}$$

and it verifies the condition of partition of unity, namely,

$$\sum_{j=1}^{n} p_j = 1 \tag{2}$$

then, we compute the Shannon entropy using the well-known formula:

$$E_S(p) = -\frac{1}{\ln(n)} \sum_{i=1}^{n} p_i \ln(p_i) \tag{3}$$

The formula (3) can be used only and only the information vector verifies the condition of the partition of unity (2). But, what happens when the information is under-defined, when there exists the following inequality:

$$\sum_{j=1}^{n} p_j < 1. \tag{4}$$

Also, what happens when the information is over-defined, when there exists the following inequality:

$$\sum_{j=1}^{n} p_j > 1. \tag{5}$$

Usually, the degree of uncertainty for a vector information can have values in the interval $[0, 1]$. Consequently, it is evidently that there exist different vectors from the n-dimensional unit hypercube that have the same value for the degree of uncertainty. For any value from the interval $[0, 1]$ it can associated a class of vectors that have for the degree of uncertainty a specified value. Hence, it results the following idea: for each information vector p that verifies the conditions (4) or (5), we must find an equivalent information vector $\hat{p}$ that verifies the condition (2) and then we obtain the entropy $E_S(p)$ calculating entropy $E_S(\hat{p})$ using formula (3). The obtaining of the equivalent vector $\hat{p}$ will be done using a normalization transformation and in the end it results a vector that verifies the condition of partition of unity (2).

Usually, the equivalent vector $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ is obtained under the condition of the information proportionality, and it is determined a real and positive number λ, so that:

$$\hat{p} = \lambda \cdot p. \tag{6}$$

From the condition (2) applied to vector $\hat{p}$ it results the number λ :

$$\lambda = \left(\sum_{j=1}^{n} p_j \right)^{-1}. \tag{7}$$

Using the scaling factor, it is obtained the normalized vector $\hat{p}$:

$$\hat{p}_i = \frac{p_i}{\sum_{j=1}^{n} p_j}. \tag{8}$$

The formula (8) has a deficiency because it becomes instable when the sum of the components approaches to zero. In addition, we cannot use the normalization transformation defined by (8), if the information vector $p = (p_1, p_2, \ldots, p_n)$ has a degree of imprecision defined by the parameter $s \in [0, 1]$. In this context, we are faced to compute the Shannon entropy for the extended vector of information denoted by P and defined by:

$$P = (p_1, p_2, \ldots, p_n, s). \tag{9}$$

It is observable that the formula (8) does not take into account the degree of imprecision s and this is an additional disadvantage. In order to solve the problem of information normalization, we will construct an affine transformation having two steps: a translation transformation [8] and a homothetic one [7]. Next, the article has the following structure: section 2 shows the construction of homothetic transformation; section 3 shows the construction of translation transformation; section 4 shows the aggregation of homothetic and translation in an affine transformation; section 5 shows particular cases of using of the proposed entropy computing method; section 6 shows some conclusions while the last section is that of references.

2. HOMOTHETIC TRANSFORMATION FOR OVER-DEFINED INFORMATION

In this section, we will analyze the case of over-defined information and without having the degree of imprecision. In other words, the vector of information $p = (p_1, p_2, \ldots, p_n)$ verifies the inequality (5) and the degree of imprecision is zero, namely:

$$\sum_{j=1}^{n} p_j > 1 \tag{10}$$

and

$$s = 0. \tag{11}$$

We can write the Jensen inequality [5], [6]:

$$-\sum_{i=1}^{n} p_i \ln(p_i) \leq -\sum_{i=1}^{n} p_i \ln\left(\frac{1}{n}\sum_{j=1}^{n} p_j\right), \tag{12}$$

and the following equivalent forms:

$$-\sum_{i=1}^{n} \frac{p_i}{\sum_{j=1}^{n} p_j} \ln\left(\frac{p_i}{\sum_{j=1}^{n} p_j}\right) \le \ln(n), \tag{13}$$

$$-\frac{1}{\ln(n)} \sum_{i=1}^{n} \frac{p_i}{\sum_{j=1}^{n} p_j} \ln\left(\frac{p_i}{\sum_{j=1}^{n} p_j}\right) \le 1. \tag{14}$$

We obtained the Shannon entropy for over-defined information:

$$E_S(p) = -\frac{1}{\ln(n)} \sum_{i=1}^{n} \frac{p_i}{\sum_{j=1}^{n} p_j} \ln\left(\frac{p_i}{\sum_{j=1}^{n} p_j}\right). \tag{15}$$

We will denote:

$$\hat{p}_i = \frac{p_i}{\sum_{j=1}^{n} p_j}, \tag{16}$$

and (15) becomes:

$$E_S(p) = -\frac{1}{\ln(n)} \sum_{i=1}^{n} \hat{p}_i \ln(\hat{p}_i). \tag{17}$$

The formula (17) represents the Shannon entropy that is utilized for the normalized information obtained using the homothetic transformation (16). The information vector $\hat{p}$ describes normalized information and belongs to the polytope defined by (2). But, the formula (16) becomes quite instable when the sum $(\sum_{j=1}^{n} p_j)$ is approaching zero. Because of that, we will directly use this normalization only when the sum $(\sum_{j=1}^{n} p_j)$ is greater than one, namely when the information is over-defined. When the sum $(\sum_{j=1}^{n} p_j)$ is less than one, namely the information is under-defined, we firstly do a translation and secondly the homothety defined by (16). The translation is presented in the next section.

3. TRANSLATION TRANSFORMATION FOR UNDER-DEFINED INFORMATION

In the previous section, we have presented the normalization of over-defined information. This is reduced to a simple homothety [7]. As we have said earlier, the normalization of the under-defined information is done in two steps: a translation and then a homothety. We will construct the translation, starting from the assumption that two information vectors that have the same distances from the points with maximum certainty are equivalent and must have the same entropy or uncertainty. The points with maximum certainty are the vertices of the polytope described by (2). In other words, we will associate to each information vector p describing an under-defined information, a vector that describes an over-defined information and keeps the distances from the vertices of the polytope defined by (2). This condition ensures that we obtain an equivalent vector from the point of view of preserving the degree of uncertainty. In the next, we will consider two unit hypercubes: one in the n-dimensional space given by vector p defined by (1) and one in $(n+1)$-dimensional space given by vector P defined by (9). The vertices of the polytope (2) are the points where the Shannon entropy is zero and are described by vectors where a component is one and all the other $(n-1)$ components are zero.

We consider an n-dimensional vector where the j^{th} component is 1, namely:

$$u = (0, \ldots, 0, 1, 0, \ldots, 0), \tag{18}$$

and its extension in the $(n+1)$-dimensional space with zero on the last position for imprecision parameter s:

$$U = (u, 0) = (0, \ldots, 0, 1, 0, \ldots, 0, 0). \tag{19}$$

The vector obtained after the translation of vector p will be defined by formula:

$$\tilde{p} = (p_1 + \vartheta, p_2 + \vartheta, \ldots, p_n + \vartheta). \tag{20}$$

The translation parameter ϑ will be obtained solving the equation that preserves the distance:

$$d(\tilde{p}, u) = d(P, U). \tag{21}$$

Using the Euclidean distance, the equation (21) becomes:

$$(p_j + \vartheta - 1)^2 + \sum_{\substack{i=1\\ i\neq j}}^{n} (p_i + \vartheta)^2 = (p_j - 1)^2 + \sum_{\substack{i=1\\ i\neq j}}^{n} p_i^2 + s^2, \tag{22}$$

$$n\vartheta^2 + 2\vartheta(\sum_{j=1}^{n} p_j - 1) - s^2 = 0. \tag{23}$$

We define the index of definedness by formula:

$$\delta = \sum_{j=1}^{n} p_j - 1, \tag{24}$$

and we obtain the following equation from (23):

$$n\vartheta^2 + 2\delta\vartheta - s^2 = 0. \tag{25}$$

Of course, there are two solutions:

$$\vartheta_{1,2} = \frac{-\delta \pm \sqrt{\delta^2 + ns^2}}{n}. \tag{26}$$

Since we are interested in over-defined information, we will only consider the variant with plus and it results for the translation parameter the following value:

$$\vartheta = \frac{-\delta + \sqrt{\delta^2 + ns^2}}{n}. \tag{27}$$

It results the translated vector components:

$$\tilde{p}_i = p_i + \frac{\sqrt{\delta^2 + ns^2} - \delta}{n}. \tag{28}$$

The translated vector $\tilde{p}$ represents over-defined information because it verifies the condition (5), namely:

$$\sum_{i=1}^{n} \tilde{p}_i = \sum_{i=1}^{n} p_i + n\vartheta, \tag{29}$$

$$\sum_{i=1}^{n} \tilde{p}_i = \sum_{i=1}^{n} p_i - \delta, + \sqrt{\delta^2 + ns^2} \tag{30}$$

$$\sum_{i=1}^{n} \tilde{p}_i = 1 + \sqrt{\delta^2 + ns^2} \geq 1. \tag{31}$$

In the second step, because the vector information $\tilde{p}$ is over-defined, we can apply the homothetic transformation (16) and at the end, it results the normalized vector $\hat{p}$.

$$\hat{p}_i = \frac{\tilde{p}_i}{\sum_{j=1}^{n} \tilde{p}_j}. \tag{32}$$

At the end, we will show that the parameter of translation can be obtained using a second way. We consider the n-dimensional vector a that is the center of the polytope (2)

$$a = \left(\frac{1}{n}, \dots, \frac{1}{n}\right), \tag{33}$$

and its extension in the $(n+1)$-dimensional space with zero on the last position for imprecision parameter s:

$$A = (a, 0) = \left(\frac{1}{n}, \ldots, \frac{1}{n}, 0\right). \tag{34}$$

The translation parameter ϑ will be obtained solving the equation that preserves the following distances:

$$d(\tilde{p}, a) = d(P, A). \tag{35}$$

Using the Euclidean distance, the equation (35) becomes:

$$\sum_{i=1}^{n} \left(p_i - \frac{1}{n} + \vartheta\right)^2 = \sum_{i=1}^{n} \left(p_i - \frac{1}{n}\right)^2 + s^2. \tag{36}$$

Finally the equation (36) is the same with equation (25), namely:

$$n\vartheta^2 + 2\delta\vartheta - s^2 = 0. \tag{37}$$

4. THE AFFINE TRANSFORMATION FOR INFORMATION NORMALIZATION

After the presentation of the translation and homothetic transformations in the previous sections, we conclude that the normalized information vector is obtained applying an affine transformation [3], [4]:

$$\hat{p}_i = \alpha p_i + \beta, \tag{38}$$

where the two parameters (α, β) are defined by:

$$\alpha = \frac{1}{1 + \sqrt{\delta^2 + ns^2}}, \tag{39}$$

$$\beta = \frac{\frac{\sqrt{\delta^2 + ns^2} - \delta}{n}}{1 + \sqrt{\delta^2 + ns^2}}, \tag{40}$$

and after all it is obtained the following formula:

$$\hat{p}_i = \frac{p_i + \frac{\sqrt{\delta^2 + ns^2} - \delta}{n}}{1 + \sqrt{\delta^2 + ns^2}}. \tag{41}$$

In the next, we will consider the following two parameters:

the degree of under-definedness:

$$u = max(-\delta, 0), \tag{42}$$

the degree of over-definedness:

$$o = max(\delta, 0), \tag{43}$$

Combining formulas (42), (43) and (41) it results consequently:

$$\hat{p}_i = \frac{p_i + \dfrac{\sqrt{\delta^2 + ns^2} - o + u}{n}}{1 + \sqrt{\delta^2 + ns^2}}, \tag{44}$$

$$\hat{p}_i = \frac{p_i + \dfrac{2u}{n} + \dfrac{\sqrt{\delta^2 + ns^2} - o - u}{n}}{1 + \sqrt{\delta^2 + ns^2}}, \tag{45}$$

$$\hat{p}_i = \frac{p_i + \dfrac{2u}{n} + \dfrac{\sqrt{\delta^2 + ns^2} - |\delta|}{n}}{1 + \sqrt{\delta^2 + ns^2}}. \tag{46}$$

We define the cumulated imprecision:

$$h = \sqrt{\delta^2 + ns^2} - |\delta|. \tag{47}$$

As a final point, it results the formula for transformed vector components:

$$\hat{p}_i = \frac{p_i + \dfrac{2u + h}{n}}{1 + |\delta| + h}. \tag{48}$$

After this, we compute the Shannon entropy for under-defined or over-defined information and supplementary having a degree of imprecision:

$$E_S(p) = -\frac{1}{\ln(n)} \sum_{i=1}^{n} \left(\frac{p_i + \dfrac{2u + h}{n}}{1 + |\delta| + h} \right) \ln \left(\frac{p_i + \dfrac{2u + h}{n}}{1 + |\delta| + h} \right). \tag{49}$$

If the imprecision is zero, namely $s = 0$, it results $h = 0$ and one obtains the particular form:

$$\hat{p}_i = \frac{p_i + \dfrac{2u}{n}}{1 + |\delta|}. \tag{50}$$

$$E_S(p) = -\frac{1}{\ln(n)} \sum_{i=1}^{n} \left(\frac{p_i + \dfrac{2u}{n}}{1 + |\delta|} \right) \ln \left(\frac{p_i + \dfrac{2u}{n}}{1 + |\delta|} \right). \tag{51}$$

In addition we can compute the Onicescu informational energy [9], the Tsallis entropy [10], [15] or Renyi entropy [11]:

Onicescu informational energy:

$$E_O(p) = \sum_{i=1}^{n} \left(\frac{p_i + \dfrac{2u+h}{n}}{1 + |\delta| + h} \right)^2 . \tag{52}$$

Tsallis entropy:

$$E_T(p) = \frac{1 - \sum_{i=1}^{n} \left(\dfrac{p_i + \dfrac{2u+h}{n}}{1 + |\delta| + h} \right)^{\alpha}}{\alpha - 1}. \tag{53}$$

Renyi entropy:

$$E_R(p) = \frac{1 - \ln \left(\sum_{i=1}^{n} \left(\dfrac{p_i + \dfrac{2u+h}{n}}{1 + |\delta| + h} \right)^{\alpha} \right)}{1 - \alpha}, \tag{54}$$

where α is a positive real number with $\alpha \neq 1$. When $\alpha \to 1$ the Tsallis and Renyi entropies recover the Shannon entropy.

Observation. Usually, at practical level, we have $\delta \approx 0$ and we can take into account the following approximation for cumulated imprecision:

$$\sqrt{\delta^2 + ns^2} - |\delta| \approx s\sqrt{n}. \tag{55}$$

It is obtained:

$$\hat{p}_i \approx \frac{p_i + \dfrac{2u + s\sqrt{n}}{n}}{1 + |\delta| + s\sqrt{n}}. \tag{56}$$

On the other hand, (56) is the exact formula for imprecise and complete information.

5. SOME PARTICULAR CASES FOR SHANNON ENTROPY

In the following we will present some particular cases for using of the Shannon entropy: neutrosophic information, bifuzzy information, intuitionistic fuzzy information, imprecise fuzzy information, and fuzzy partitions.

5.1. THREE-VALUED SHANNON ENTROPY FOR NEUTROSOPHIC INFORMATION.

The neutrosophic information proposed by Smarandache [13], [14] is defined by three parameters: degree of truth $T \in [0,1]$, degree of falsity $F \in [0,1]$ and degree of neutrality $I \in [0,1]$. The vector $p = (T, I, F)$ represents the primary information. We define the neutrosophic definedness and under-definedness by following two formulas:

$$D = T + F + I - 1, \tag{57}$$

$$U = \max(-D, 0). \tag{58}$$

If $D < 0$ then the neutrosophic information is under-defined and if $D > 0$ then the neutrosophic information is over-defined. In this case for three-valued Shannon entropy, we consider three points where the certainty is maximum, namely $p_T = (1,0,0)$, $p_I = (0,1,0)$ and $p_F = (0,0,1)$. Using (50) it results the three-valued normalized information $\hat{p} = (\hat{T}, \hat{I}, \hat{F})$:

$$\hat{T} = \frac{T + \frac{2U}{3}}{1 + |D|}, \tag{59}$$

$$\hat{I} = \frac{I + \frac{2U}{3}}{1 + |D|}, \tag{60}$$

$$\hat{F} = \frac{F + \frac{2U}{3}}{1 + |D|}. \tag{61}$$

The neutrosophic information $(\hat{T}, \hat{I}, \hat{F})$ verifies the condition of the partition of unity:

$$\hat{T} + \hat{I} + \hat{F} = 1. \tag{62}$$

The Shannon entropy is calculated using formula (51) and it results:

$$E_S(p) = -\frac{\left(\dfrac{T+\dfrac{2U}{3}}{1+|D|}\right)\ln\left(\dfrac{T+\dfrac{2U}{3}}{1+|D|}\right)}{\ln(3)} - \frac{\left(\dfrac{I+\dfrac{2U}{3}}{1+|D|}\right)\ln\left(\dfrac{I+\dfrac{2U}{3}}{1+|D|}\right)}{\ln(3)} - \frac{\left(\dfrac{F+\dfrac{2U}{3}}{1+|D|}\right)\ln\left(\dfrac{F+\dfrac{2U}{3}}{1+|D|}\right)}{\ln(3)}. \tag{63}$$

5.2. BI-VALUED SHANNON ENTROPY FOR NEUTROSOPHIC INFORMATION

The neutrosophic information is described by parameters: degree of truth $\mu \in [0,1]$, degree of falsity $\nu \in [0,1]$ and degree of imprecision $\omega \in [0,1]$.

Whe define the following parameters:

the bifuzzy definedness:

$$\delta = \mu + \nu - 1, \tag{64}$$

the bifuzzy incompleteness:

$$\pi = \max(-\delta, 0), \tag{65}$$

the cumulated imprecision:

$$h = \sqrt{\delta^2 + 2\omega^2} - |\delta|. \tag{66}$$

In this case for bi-valued Shannon entropy, we consider two points where the certainty is maximum, namely $p_T = (1,0,0)$ and $p_F = (0,1,0)$. It results its equivalent fuzzy degree of truth $\hat{\mu}$ and its fuzzy degree of falsity $\hat{\nu}$:

$$\hat{\mu} = \frac{\mu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}, \tag{67}$$

$$\hat{\nu} = \frac{\nu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}. \tag{68}$$

The fuzzy information $\hat{p} = (\hat{\mu}, \hat{\nu})$ represents the bi-valued normalized form of the primary information $p = (\mu, \nu, \omega)$ and there exists the equality:

$$\hat{\mu} + \hat{\nu} = 1. \tag{69}$$

Using the fuzzy information $\hat{p} = (\hat{\mu}, \hat{\nu})$ that was associated to the neutrosophic information $p = (\mu, \nu, \omega)$ we will compute the bi-valued Shannon entropy by the following formula:

$$E_S(p) = -\frac{\left(\dfrac{\mu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}\right) \ln \left(\dfrac{\mu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}\right)}{\ln(2)} - \frac{\left(\dfrac{\nu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}\right) \ln \left(\dfrac{\nu + \pi + \dfrac{h}{2}}{1 + |\delta| + h}\right)}{\ln(2)}. \tag{70}$$

5.3. SHANNON ENTROPY FOR BIFUZZY INFORMATION

The bifuzzy information [1], [2] is described by two parameters: degree of truth $\mu \in [0,1]$ and degree of falsity $\nu \in [0,1]$. We define the bifuzzy definedness $\delta \in [-1,1]$ and bifuzzy incompleteness $\pi \in [0,1]$ by:

$$\delta = \mu + \nu - 1, \tag{71}$$

$$\pi = \max(-\delta, 0). \tag{72}$$

In this case we consider two points where the certainty is maximum, namely $p_T = (1,0)$ and $p_F = (0,1)$. We compute the fuzzy degree of truth $\hat{\mu}$ and fuzzy degree of falsity $\hat{\nu}$ using (50) and it results:

$$\hat{\mu} = \frac{\mu + \pi}{1 + |\delta|}, \tag{73}$$

$$\hat{\nu} = \frac{\nu + \pi}{1 + |\delta|}. \tag{74}$$

There exists the equality:

$$\hat{\mu} + \hat{\nu} = 1. \tag{75}$$

Using the associated fuzzy information $\hat{p} = (\hat{\mu}, \hat{\nu})$ to the bifuzzy information $p = (\mu, \nu)$ we will compute the Shannon entropy by the following formula derived from (51):

$$E_S(p) = -\frac{\left(\frac{\mu+\pi}{1+|\delta|}\right)\ln\left(\frac{\mu+\pi}{1+|\delta|}\right) + \left(\frac{\nu+\pi}{1+|\delta|}\right)\ln\left(\frac{\nu+\pi}{1+|\delta|}\right)}{\ln(2)}. \tag{76}$$

5.4. SHANNON ENTROPY FOR INTUITIONISTIC FUZZY INFORMATION

The intuitionistic fuzzy information [1], [2] is described by two parameters: degree of truth $\mu \in [0,1]$ and degree of falsity $\nu \in [0,1]$ verifying the following inequality $1 \geq \mu + \nu$.

We define the degree of incompleteness π by:

$$\pi = 1 - \mu - \nu. \tag{77}$$

The information is under-defined or incomplete and we will associate the following fuzzy information with degree of truth $\hat{\mu}$ and degree of falsity $\hat{\nu}$:

$$\hat{\mu} = \frac{\mu+\pi}{1+\pi}, \tag{78}$$

$$\hat{\nu} = \frac{\nu+\pi}{1+\pi}, \tag{79}$$

with:

$$\hat{\mu} + \hat{\nu} = 1. \tag{80}$$

Using the associated fuzzy information $\hat{p} = (\hat{\mu}, \hat{\nu})$ to the intuitionistic fuzzy information $p = (\mu, \nu)$, we will compute the Shannon entropy by the following formula derived from (51):

$$E_S(p) = -\frac{\left(\frac{\mu+\pi}{1+\pi}\right)\ln\left(\frac{\mu+\pi}{1+\pi}\right) + \left(\frac{\nu+\pi}{1+\pi}\right)\ln\left(\frac{\nu+\pi}{1+\pi}\right)}{\ln(2)}. \tag{81}$$

Equivalent with:

$$E_S(p) = -\frac{\left(\frac{\bar{\mu}}{\bar{\mu}+\bar{\nu}}\right)\ln\left(\frac{\bar{\mu}}{\bar{\mu}+\bar{\nu}}\right) + \left(\frac{\bar{\nu}}{\bar{\mu}+\bar{\nu}}\right)\ln\left(\frac{\bar{\nu}}{\bar{\mu}+\bar{\nu}}\right)}{\ln(2)}. \tag{82}$$

where the negation is calculated using formula:

$$\bar{x} = 1 - x. \tag{83}$$

5.5. SHANNON ENTROPY FOR IMPRECISE FUZZY INFORMATION

The fuzzy information [16] is described by the degree of truth $\mu \in [0,1]$ while the imprecise fuzzy information is described by the pair $p = (\mu, \sigma)$, where $\mu \in [0,1]$ is the degree of truth and $\sigma \in \left[0, \frac{1}{2}\right]$ is the degree of imprecision. We must mention that $\nu = 1 - \mu$ represents the degree of falsity. The imprecise fuzzy information can be seen as particular neutrosophic case where (T, I, F) are defined by:

$$T = \mu,$$

$$I = 2\sigma,$$

$$F = 1 - \mu.$$

In this framework, it results the following particular values for definedness δ, cumulated imprecision h, fuzzy degree of truth $\hat{\mu}$ and fuzzy degree of falsity $\hat{\nu}$:

$$\delta = \mu + \nu - 1 = 0, \tag{84}$$

$$h = 2\sigma\sqrt{2}, \tag{85}$$

$$\hat{\mu} = \frac{\mu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}, \tag{86}$$

$$\hat{\nu} = \frac{\nu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}, \tag{87}$$

with:

$$\hat{\mu} + \hat{\nu} = 1. \tag{88}$$

Using (51), we obtain Shannon entropy for imprecise fuzzy information:

$$E_S(p) = -\frac{\left(\frac{\mu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}\right) \ln\left(\frac{\mu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}\right) + \left(\frac{\nu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}\right) \ln\left(\frac{\nu + \sigma\sqrt{2}}{1 + 2\sigma\sqrt{2}}\right)}{\ln(2)}. \tag{89}$$

5.6. BI-VALUED SHANNON ENTROPY FOR FUZZY PARTITION

We consider the fuzzy partition $(w_1, w_2, \ldots, w_n)$ and there exists the equality,

$$w_1 + w_2 + \ldots + w_n = 1. \tag{90}$$

We order the membership functions and get the following decreasing set of values:

$$o_1 \geq o_2 \geq \ldots \geq o_n, \tag{91}$$

where

$$o_1 = max(w_1, w_2, \ldots, w_n), \tag{92}$$

and

$$o_n = min(w_1, w_2, \ldots, w_n). \tag{93}$$

Firstly, we construct an intuitionistic fuzzy representation where $\mu = o_1$, $\nu = o_2$ and $\pi = 1 - o_1 - o_2$. Secondly, we construct the fuzzy representation where $\hat{\mu}$ and $\hat{\nu}$ are defined by (78) and (79). It results:

$$\hat{\mu} = \frac{o_1 + \pi}{1 + \pi}, \tag{94}$$

$$\hat{\nu} = \frac{o_2 + \pi}{1 + \pi}. \tag{95}$$

Using formula (51) for associated fuzzy information $(\hat{\mu}, \hat{\nu})$, one obtains the bi-valued Shannon entropy for the fuzzy partition w:

$$E_S(w) = -\frac{\left(\frac{o_1 + \pi}{1 + \pi}\right) \ln\left(\frac{o_1 + \pi}{1 + \pi}\right) + \left(\frac{o_2 + \pi}{1 + \pi}\right) \ln\left(\frac{o_2 + \pi}{1 + \pi}\right)}{\ln(2)}, \tag{96}$$

with its equivalent form derived from (82):

$$E_S(w) = -\frac{\left(\frac{\bar{o_1}}{\bar{o_1} + \bar{o_2}}\right) \ln\left(\frac{\bar{o_1}}{\bar{o_1} + \bar{o_2}}\right) + \left(\frac{\bar{o_2}}{\bar{o_1} + \bar{o_2}}\right) \ln\left(\frac{\bar{o_2}}{\bar{o_1} + \bar{o_2}}\right)}{\ln(2)}. \tag{97}$$

There are other non-logarithmic formulas for bi-valued fuzzy partition entropy computing such as the following three:

$$E_K(w) = 1 - \frac{|o_1 - o_2|}{1 + \pi}, \tag{98}$$

$$E_E(w) = \sqrt{\frac{1 - 2o_1 + \sum_{i=1}^{n} o_i^2}{1 - 2o_2 + \sum_{i=1}^{n} o_i^2}}, \tag{99}$$

$$E_P(w) = \frac{1 - o_1}{1 - o_2} = \frac{\bar{o_1}}{\bar{o_2}}. \tag{100}$$

6. CONCLUSIONS

The article presents a method of using Shannon entropy for under-defined or over-defined information with a certain degree of imprecision. For this purpose, a two-step normalization procedure is proposed: a translation and a homothetic one. After the presentation, the procedure is used for calculating Shannon's entropy in the case of particular representations of information such as neutrosophic information, bifuzzy information, intuitionistic fuzzy information, imprecise fuzzy information and fuzzy partitions. In the case of neutrosophic information, two variants are possible: the first is the trivalent variant in which the certainty has three prototypes: true, neutral and false; the second is the bivalent variant in which the certainty has two prototypes: true and false. The article mentions that the presented method of normalization can be used for other formulas such as Onicescu information energy, Tsallis entropy or Renyi entropy.

References

[1] K. T. Atanassov, *Intuitionistic fuzzy sets*, Fuzzy Sets Syst. 20, 87-96, 1986.

[2] K. T. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications*, Studies in Fuzziness and Soft Computing, vol. 35, Physica-Verlag, Heidelberg ,1999.

[3] M. Berger, *Geometry I*, Berlin, Springer, ISBN 3-540-11658-3, 1987.

[4] M. Hazewinkel, *Affine transformation*, Enciclopedia of Mathematics, Springer, ISBN 978-1-75608-010-4, 2001.

[5] M. Hazewinkel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4, 2001.

[6] J. L. W. V. Jensen, *Sur les fonctions convexes et les inegalites entre les valeurs moyennes*, Acta Mathematica, **30**, 1(1906), 175-193, doi:10.1007 / BF02418571,.

[7] B. Meserve, *Homothetic transformations*, Fundamental Concept of Geometry, Addison-Wesley, pp. 166-169, 1955.

[8] W. Osgood, W. Graustein, P*lane and solid analytic geometry*, The Macmillan Company, P330, 1921.

[9] O. Onicescu, *Energie informationnelle*, Comptes Rendus Hebdomadaires des Sciences de l'Academie des Sciences, Serie A 263(1966), 841-842.

[10] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys., 52(1988), 479-487.

[11] A. Renyi, *On Measures of Entropy and Information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: Contributions to the Theory of Statistics, 547-561, University of California Press, Berkeley, Calif., 1961.

[12] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. 27(1948), 379-423.

[13] F. Smarandache, *A Unifying Field in Logics: Neutrosophic Logic*, Multiple valued logic, An international journal, 8, no. 3, 385-438, 2002.

[14] F. Smarandache, *Neutrosophic Set - A Generalization of the Intuitionistic Fuzzy Set*, International Journal of Pure and Applied Mathematics,24, no. 3, 287-297, 2005.

[15] J. Harvda, F. Charvat, *Quantification method of classification processes-concept of structural β-entropy*, Kybernetika (Prague) 3, 30-35, 1967.

[16] L. A. Zadeh, *Fuzzy sets*, Information Control 8, 338-353, 1965.

ROMAI J., v.14, no.1(2018), 187–200

ON THE JACOBSTHAL SEQUENCES AND THEIR APPLICATIONS IN RINGS WITH IDENTITY

Yasemin Taşyurdu, Devran Çifçi

Department of Mathematics, Faculty of Sciences and Art,
The University of Erzincan Binali Yıldırım, Erzincan, Turkey.

ytasyurdu@erzincan.edu.tr, devrancifci@hotmail.com

Abstract In this study, we define the Jacobsthal sequences over arbitrary rings with identity. Also, the terms of these sequences are derivated by the matrix. The generating function and Binet formula given n-th general term of these sequences are found by using recurrence relation of the new defined Jacobsthal sequences in rings. Finally, we obtain the some properties of terms of Jacobsthal sequences in rings.

Keywords: Jacobsthal sequences, ring, matrix, Binet formula.
2010 MSC: 11B37, 11B83, 11R04, 40C05.

1. INTRODUCTION

The Jacobsthal sequence $\{j_n\}$ defined by

$$j_{n+2} = j_{n+1} + 2j_n \ , \qquad n \geq 0$$

with initial $j_0 = 0$ and $j_1 = 1$. Horadam in [2], exhibited a plethora of identities for the second order Jacobsthal sequences and then went on to explore their relationships and those of a variety of associated and representative sequences. Most of the studies on recurrence sequences have been considered over groups and but there are very little studies on rings ([6], [7], [3]). The Jacobsthal sequences in rings have never been studied. DeCarli [3] gave a generalized Fibonacci sequence over an arbitrary ring in 1970. Let R be a ring with identity 1. The sequences $\{S_n\}$ of elements of R, recursively defined by

$$S_{n+2} = A_1 S_{n+1} + A_0 S_n \qquad \text{for} \qquad n \geq 0, 1, 2, ... \tag{1}$$

where S_0, S_1, A_0 and A_1 are abritrary elements of R [3]. Special cases of Fibonacci sequence over an arbitrary ring have been considered by Buschman [8], Horadam [1] and Vorobyov [4] where this ring was taken to be the set of integers. Wyler [5] also worked with such a sequence over a particular commutative ring with identity. Then, Tasyurdu and Gultekin obtained the period of generalized Fibonacci sequence in finite rings with identity and fields of order p^2 by using equality recursively defined by $F_{n+2} = A_1F_{n+1} + A_0F_n$ for $n \geq 0$, where $F_0 = 0$ (zero of the ring), $F_1 = 1$ (identity of the ring) and A_0, A_1 are generator elements of finite rings with identity and fields of order p^2 ([9], [10]). Also, Tasyurdu and Dilmen obtained the period of generalized Fibonacci sequence was defined over an arbitrary ring and the terms of this sequence are derivated by determinant of Tridiagonal matrix [11].

2. MAIN RESULTS

We present the following the definition which is a special case of equation (1), denoted by $\{J_n\}$.

Definition 2.1. *The Jacobsthal sequences $\{J_n\}$ in rings with identity are defined by recurrence relation*

$$J_{n+2} = BJ_{n+1} + 2AJ_n\ , \qquad n \geq 0 \tag{2}$$

where $J_0 = 0$ (zero of a ring), $J_1 = 1$ (identity of a ring) and A, B are arbitrary elements of the ring, J_n is n-th term of $\{J_n\}$.

From Definition 2.1, note that it can be considered $A_1 = B$ and $A_0 = 2A$ where A_0, A_1, A and B are abritrary elements of the ring. By using Definition 2.1, we can write a few terms of sequences $\{J_n\}$ as follows

$$\begin{aligned}
J_0 &= 0, \\
J_1 &= 1, \\
J_2 &= B(1) + 2A\,(0) = B, \\
J_3 &= B^2 + 2A\,(1) = B^2 + 2A, \\
J_4 &= B^3 + 2BA + 2AB = B^3 + 4AB, \\
J_5 &= B^4 + 6AB^2 + 4A^2, \\
&\ \ \vdots
\end{aligned}$$

The following theorem expresses matrix representation of the terms of the Jacobsthal sequences $\{J_n\}$ in rings.

Theorem 2.1. *The Jacobsthal sequences $\{J_n\}$ in rings are generated by a matrix $M = \begin{pmatrix} B & 1 \\ 2A & 0 \end{pmatrix}$, then*

$$M^n = \begin{pmatrix} J_{n+1} & J_n \\ 2AJ_n & 2AJ_{n-1} \end{pmatrix} \tag{3}$$

where $n \in \mathbb{Z}^+$.

Proof. We will use the induction method on n. If $n = 1$, then

$$M = \begin{pmatrix} J_2 & J_1 \\ 2AJ_1 & 2AJ_0 \end{pmatrix} = \begin{pmatrix} B & 1 \\ 2A & 0 \end{pmatrix}.$$

So the proof is complete for $n = 1$. Let the equation (3) be hold for $n = k$; then we will show that the equation holds for $n = k + 1$. For $J_0 = 0$, $J_1 = 1$ and $J_2 = B$, by our assumption

$$\begin{aligned} M^k M &= \begin{pmatrix} J_{k+1} & J_k \\ 2AJ_k & 2AJ_{k-1} \end{pmatrix} \begin{pmatrix} J_2 & J_1 \\ 2AJ_1 & 2AJ_0 \end{pmatrix} \\ &= \begin{pmatrix} J_{k+1}J_2 + J_k 2AJ_1 & J_{k+1}J_1 + J_k 2AJ_0 \\ 2AJ_kJ_2 + 4A^2J_{k-1}J_1 & 2AJ_kJ_1 + 4A^2J_{k-1}J_0 \end{pmatrix} \\ &= \begin{pmatrix} BJ_{k+1} + 2AJ_k\,(1) & J_{k+1}\,(1) + J_k 2A\,(0) \\ 2A(BJ_k + 2AJ_{k-1}\,(1)) & 2AJ_k\,(1) + 4A^2J_{k-1}\,(0) \end{pmatrix} \\ &= \begin{pmatrix} J_{k+2} & J_{k+1} \\ 2AJ_{k+1} & 2AJ_k \end{pmatrix} \\ &= M^{k+1} \end{aligned}$$

which is as desired. ∎

Example 2.1. *From Theorem 2.1, we can obtain*

$$\begin{aligned} M^1 &= \begin{pmatrix} B & 1 \\ 2A & 0 \end{pmatrix} = \begin{pmatrix} J_2 & J_1 \\ 2AJ_1 & 2AJ_0 \end{pmatrix} \\ M^2 &= \begin{pmatrix} B^2 + 2A & B \\ 2AB & 2A \end{pmatrix} = \begin{pmatrix} J_3 & J_2 \\ 2AJ_2 & 2AJ_1 \end{pmatrix} \\ M^3 &= \begin{pmatrix} B^3 + 4AB & B^2 + 2A \\ 2AB^2 + 4A^2 & 2AB \end{pmatrix} = \begin{pmatrix} J_4 & J_3 \\ 2AJ_3 & 2AJ_2 \end{pmatrix} \\ &\vdots \end{aligned}$$

From Theorem 2.1, we obtain that pairs of successive term of the sequence $\{J_n\} = \{..., J_n, J_{n+1}, J_{n+2}, ...\}$ can be considered pairs as 2-vectors, e.g. $(J_{n+1}, J_n)^T$. That is,

$$(B, 2A) \begin{pmatrix} J_{n+1} \\ J_n \end{pmatrix} = BJ_{n+1} + 2AJ_n$$

where A, B are arbitrary elements of the ring.

2.1. THE GENERATING FUNCTION FOR JACOBSTHAL SEQUENCES $\{J_n\}$

Now, we find the generating function for the Jacobsthal sequences $\{J_n\}$ in rings. We know that the power series of the ordinary generating function for $\langle J_0, J_1, J_2, J_3, J_4, \ldots\rangle$ are follows

$$G(X) = J_0 + J_1x + J_2x^2 + J_3x^3 + \ldots = \sum_{n=0}^{\infty} J_nx^n. \tag{4}$$

Theorem 2.2. *The generating function $G(x)$ of the Jacobsthal sequences $\{J_n\}$ in rings is as shown*

$$G(X) = \frac{x}{1 - Bx - 2Ax^2} \tag{5}$$

where A, B are arbitrary elements of the ring.

Proof. From equation (4), it can be obtained that

$$\begin{aligned} G(X) &= \sum_{n=0}^{\infty} J_nx^n \\ &= 0 + x + \sum_{n=2}^{\infty} J_nx^n \\ &= x + \sum_{n=2}^{\infty}(BJ_{n-1} + 2AJ_{n-2})x^n \\ &= x + xB\sum_{n=0}^{\infty} J_nx^n + x^2 2A\sum_{n=0}^{\infty} J_nx^n \\ &= x + xBG(x) + x^2 2AG(x) \end{aligned}$$

where $J_0 = 0$, $J_1 = 1$. Then it can be shown that $G(X) = \dfrac{x}{1 - Bx - 2Ax^2}$. ∎

The terms of the Jacobsthal sequences $\{J_n\}$ in rings can be obtained by using the Definition 2.1. The Binet formula known as the general formula can be used instead of this definition. The Binet formula allows us to easily find any of terms of the Jacobsthal sequences $\{J_n\}$ in rings without having to know all the terms before it. That is, Binet formula give us to find the n-th Jacobsthal number in rings without creating the Jacobsthal sequences $\{J_n\}$ in rings. Now, we produce the Binet formula for the Jacobsthal sequences $\{J_n\}$ in rings.

2.2. THE BINET FORMULA OF JACOBSTHAL SEQUENCES $\{J_n\}$

Let α and β be roots of the characteristic equation of the recurence relation equation (2) and let the eigenvector X corresponding to the eigenvalues λ is $X = (x_1, x_2)$,

$$|\lambda I - M| = \begin{vmatrix} \lambda - B & -1 \\ -2A & \lambda \end{vmatrix} = (\lambda - B)\lambda - 2A = \lambda^2 - B\lambda - 2A.$$

Then using the quadratic formula and form $\lambda^2 - B\lambda - 2A = 0$, we obtain

$$\alpha = \frac{B + \sqrt{B^2 + 8A}}{2} \qquad \text{and} \qquad \beta = \frac{B - \sqrt{B^2 + 8A}}{2}$$

where $\lambda = \frac{B \pm \sqrt{B^2+8A}}{2}$. Note that this roots satisfy the following relations:

$$\begin{aligned} \alpha^2 &= \alpha B + 2A & \beta^2 &= \beta B + 2A \\ \alpha - \beta &= \sqrt{B^2 + 8A} & \alpha + \beta &= B \\ \frac{1}{\alpha} &= \frac{2}{B + \sqrt{B^2 + 8A}} & \frac{1}{\beta} &= \frac{2}{B - \sqrt{B^2 + 8A}} \\ \alpha\beta &= -2A \end{aligned}$$

Let us find the eigenvectors. Firstly, for eigenvalue α

$$(\alpha I - M)X = 0 \Rightarrow \begin{pmatrix} \alpha - B & -1 \\ -2A & \alpha \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

so, the using $\alpha^2 = \alpha B + 2A$

$$\begin{aligned} \alpha I - M &= \begin{pmatrix} \alpha - B & -1 \\ -2A & \alpha \end{pmatrix} r_2 \to r_2 + \alpha r_1 \begin{pmatrix} \alpha - B & -1 \\ \alpha^2 - (\alpha B + 2A) & 0 \end{pmatrix} \\ &\to \begin{pmatrix} \alpha - B & -1 \\ \alpha^2 - \alpha^2 & 0 \end{pmatrix} \to \begin{pmatrix} \alpha - B & -1 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The eigenvector of the eigenvalue α is $(1, \alpha - B)^T$. As similary, the eigenvector of the eigenvalue β is $(1, \beta - B)^T$. Then we can find the diagonalization of M. So, we put these eigenvectors, into change of basis matrix

$$N = \begin{pmatrix} 1 & 1 \\ \alpha - B & \beta - B \end{pmatrix} \text{ and } N^{-1} = \frac{1}{-\sqrt{B^2 + 8A}} \begin{pmatrix} \beta - B & -1 \\ B - \alpha & 1 \end{pmatrix}$$

$$M = N\widetilde{M}N^{-1}. \tag{6}$$

From equation (6), we get

$$\widetilde{M} = \begin{pmatrix} \left(\frac{\beta\alpha-\alpha^2}{-\sqrt{B^2+8A}}\right) & 0 \\ 0 & \left(\frac{\beta^2-\alpha\beta}{-\sqrt{B^2+8A}}\right) \end{pmatrix}$$

where $\widetilde{M}$ is diagonalization of M. Now we can easily compute powers of M from equation (6)

$$\begin{aligned} M^n &= \left[N\widetilde{M}N^{-1}\right]^n \\ &= N(\widetilde{M})^n N^{-1} \\ &= \frac{1}{-\sqrt{B^2+8A}} \begin{pmatrix} 1 & 1 \\ \alpha-B & \beta-B \end{pmatrix} \begin{pmatrix} \left(\frac{\beta\alpha-\alpha^2}{-\sqrt{B^2+8A}}\right)^n & 0 \\ 0 & \left(\frac{\beta^2-\alpha\beta}{-\sqrt{B^2+8A}}\right)^n \end{pmatrix} \begin{pmatrix} \beta-B & -1 \\ B-\alpha & 1 \end{pmatrix} \end{aligned}$$

We know that $\binom{J_{n+1}}{2AJ_n} = M^n\binom{1}{0}$. Then

$$\begin{aligned} \begin{pmatrix} J_{n+1} \\ 2AJ_n \end{pmatrix} &= M^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} J_{n+1} \\ 2AJ_n \end{pmatrix} &= \frac{1}{-\sqrt{B^2+8A}} \begin{pmatrix} 1 & 1 \\ \alpha-B & \beta-B \end{pmatrix} \begin{pmatrix} \left(\frac{\beta\alpha-\alpha^2}{-\sqrt{B^2+8A}}\right)^n & 0 \\ 0 & \left(\frac{\beta^2-\alpha\beta}{-\sqrt{B^2+8A}}\right)^n \end{pmatrix} \begin{pmatrix} \beta-B & -1 \\ B-\alpha & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} J_{n+1} \\ 2AJ_n \end{pmatrix} &= \frac{1}{-\sqrt{B^2+8A}} \begin{pmatrix} 1 & 1 \\ \alpha-B & \beta-B \end{pmatrix} \begin{pmatrix} \left(\frac{\beta\alpha-\alpha^2}{-\sqrt{B^2+8A}}\right)^n & 0 \\ 0 & \left(\frac{\beta^2-\alpha\beta}{-\sqrt{B^2+8A}}\right)^n \end{pmatrix} \begin{pmatrix} \beta-B \\ B-\alpha \end{pmatrix} \\ \begin{pmatrix} J_{n+1} \\ 2AJ_n \end{pmatrix} &= \frac{1}{-\sqrt{B^2+8A}} \begin{pmatrix} 1 & 1 \\ \alpha-B & \beta-B \end{pmatrix} \begin{pmatrix} \left(\frac{\beta\alpha-\alpha^2}{-\sqrt{B^2+8A}}\right)^n (\beta-B) \\ \left(\frac{\beta^2-\alpha\beta}{-\sqrt{B^2+8A}}\right)^n (B-\alpha) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} 2AJ_n &= \frac{1}{-\sqrt{B^2+8A}} \frac{(\beta-\alpha)^n}{\left(-\sqrt{B^2+8A}\right)^n} (\beta-B)(\alpha-B)(\alpha^n-\beta^n) \\ 2AJ_n &= \frac{1}{\left(-\sqrt{B^2+8A}\right)^{n+1}} (\beta-\alpha)^n (\beta-B)(\alpha-B)(\alpha^n-\beta^n) \\ 2AJ_n &= \frac{(\alpha^n-\beta^n)}{\left(-\sqrt{B^2+8A}\right)^{n+1}} \left(-\sqrt{B^2+8A}\right)^n \left(\frac{B-\sqrt{B^2+8A}}{2} - B\right) \left(\frac{B+\sqrt{B^2+8A}}{2} - B\right) \end{aligned}$$

$$
\begin{aligned}
2AJ_n &= \frac{1}{\left(-\sqrt{B^2+8A}\right)^{n+1}}\left(-\sqrt{B^2+8A}\right)^n(-2A)(\alpha^n-\beta^n)\\
2AJ_n &= \frac{1}{-\sqrt{B^2+8A}}(-2A)(\alpha^n-\beta^n)\\
J_n &= \frac{1}{\alpha-\beta}\alpha^n-\beta^n\\
J_n &= \frac{\alpha^n-\beta^n}{\alpha-\beta}.
\end{aligned}
$$

That is,

$$
\begin{aligned}
J_n &= \frac{1}{\sqrt{B^2+8A}}(\alpha^n-\beta^n)\\
j_n &= \frac{1}{\sqrt{B^2+8A}}\left[\left(\frac{B+\sqrt{B^2+8A}}{2}\right)^n-\left(\frac{B-\sqrt{B^2+8A}}{2}\right)^n\right]
\end{aligned}
$$

and so we are done.

Example 2.2. *From Definition 2.1, we know the Jacobsthal sequences* $\{J_n\}$ *in rings*

$$\left\{0,1,B,\ B^2+2A,\ B^3+4AB,B^4+6AB^2+4A^2,...\ \right\}$$

where $n\geq 0$ *and by using Binet formula of* $\{J_n\}$, *we obtain*

$$
\begin{aligned}
J_0 &= \frac{\alpha^0-\beta^0}{\alpha-\beta}=0\\
J_1 &= \frac{\alpha-\beta}{\alpha-\beta}=1\\
J_2 &= \frac{\alpha^2-\beta^2}{\alpha-\beta}=\alpha+\beta=B\\
J_3 &= \frac{\alpha^3-\beta^3}{\alpha-\beta}=\alpha^2+\alpha\beta+\beta^2=B(\alpha+\beta)+2A+(-2A)+2A=B^2+2A\\
&\vdots
\end{aligned}
$$

2.2.1 The Binet Formula of $\{J_n\}$ with the Inductive Method.

We will use the induction method on n to prove $J_n=\frac{\alpha^n-\beta^n}{\alpha-\beta}$. If $n=0$, then

$$\frac{1}{\sqrt{B^2+8A}}(\alpha^0-\beta^0)=0=J_0$$

and $n = 1$,

$$\frac{1}{\sqrt{B^2+8A}}(\alpha^1 - \beta^1) = 1 = J_1$$

Let us suppose that the Binet formula holds for arbitrary $n = k$. Then we will show that the Binet formula holds for $n = k + 1$. By our assumption, we obtain

$$\begin{aligned} J_{k+1} &= BJ_k + 2AJ_{k-1} \\ &= B\left(\frac{\alpha^k - \beta^k}{\alpha - \beta}\right) + 2A\left(\frac{\alpha^{k-1} - \beta^{k-1}}{\alpha - \beta}\right) \\ &= \frac{1}{\alpha - \beta}(\alpha^{k-1}\alpha^2 - \beta^{k-1}\beta^2) \\ &= \frac{(\alpha^{k+1} - \beta^{k+1})}{\alpha - \beta} \end{aligned}$$

and so the proof is completed.

2.2.2 The Binet Formula of $\{J_n\}$ with the Generating Function Method. We will find the Binet formula of the Jacobsthal sequences $\{J_n\}$ in rings by using generating function. We know that

$$\begin{aligned} 2Ax^2 + Bx - 1 &= 2A\left(x + \frac{\alpha}{2A}\right)\left(x + \frac{\beta}{2A}\right) \\ 4A^2x^2 + 2ABx - 2A &= 4A^2\left(x + \frac{\alpha}{2A}\right)\left(x + \frac{\beta}{2A}\right). \end{aligned}$$

Thus, from equation (5) we obtain

$$\begin{aligned} G(x) &= \frac{x}{1 - Bx - 2Ax^2} \\ G(x) &= \frac{-x}{2Ax^2 + Bx - 1} \\ G(x) &= \frac{-2Ax}{4A^2x^2 + 2ABx - 2A} \end{aligned}$$

First, we factor the denominator:

$$4A^2x^2 + 2ABx - 2A = 2A\left(x + \frac{\alpha}{2A}\right)2A\left(x + \frac{\beta}{2A}\right)$$

where $\alpha = \frac{B+\sqrt{B^2+8A}}{2}$ and $\beta = \frac{B-\sqrt{B^2+8A}}{2}$. Next, we find U and V which satisfy:

$$\frac{-2Ax}{4A^2x^2+2ABx-2A} = \frac{U}{2A\left(x+\frac{\alpha}{2A}\right)} + \frac{V}{2A\left(x+\frac{\beta}{2A}\right)}.$$

We do this by plugging in various values of x to generate linear equations in U and V. We can then find U and V by solving a linear system.

This gives $U = \frac{-\alpha}{\alpha-\beta}$ and $V = \frac{\beta}{\alpha-\beta}$. By substituting into the equation above gives the partial fractions expansion of $2AG(x)$:

$$\begin{aligned}
2AG(x) &= \frac{\frac{-\alpha}{\alpha-\beta}}{\left(x+\frac{\alpha}{2A}\right)} + \frac{\frac{\beta}{\alpha-\beta}}{\left(x+\frac{\beta}{2A}\right)} \\
&= \frac{-\alpha}{\alpha-\beta}\left(\frac{1}{x+\frac{\alpha}{2A}}\right) + \frac{\beta}{\alpha-\beta}\left(\frac{1}{x+\frac{\beta}{2A}}\right) \\
&= \frac{-\alpha}{\alpha-\beta}\left(\frac{1}{x-\frac{1}{\beta}}\right) + \frac{\beta}{\alpha-\beta}\left(\frac{1}{x-\frac{1}{\alpha}}\right) \\
&= \frac{-\alpha}{\alpha-\beta}\frac{1}{\frac{1}{\beta}}\frac{1}{\beta x-1} + \frac{\beta}{\alpha-\beta}\frac{1}{\frac{1}{\alpha}}\frac{1}{\alpha x-1} \\
&= \frac{-\alpha\beta}{\alpha-\beta}\left(\sum_{n=0}^{\infty}\alpha^n x^n - \sum_{n=0}^{\infty}\beta^n x^n\right) \\
G(x) &= \sum_{n=0}^{\infty}\frac{1}{\alpha-\beta}(\alpha^n-\beta^n)x^n
\end{aligned}$$

where $\alpha\beta = -2A$.

By equating coefficients, we again conlude that

$$J_n = \frac{1}{\alpha-\beta}\alpha^n - \beta^n = \frac{1}{\sqrt{B^2+8A}}\left[\left(\frac{B+\sqrt{B^2+8A}}{2}\right)^n - \left(\frac{B-\sqrt{B^2+8A}}{2}\right)^n\right]$$

and so proof is completed.

2.3. SOME PROPERTIES OF TERMS OF JACOBSTHAL SEQUENCES $\{J_n\}$

Theorem 2.3. *If $J_{n+2} = BJ_{n+1} + 2AJ_n$ then $J_{n+2} = J_{n+1}B + J_n 2A$.*

Proof. We can complete the proof by induction method on n. For $n = 0$, if $J_2 = BJ_1 + 2AJ_0$ then

$$J_2 = BJ_1 + 2AJ_0 = B\,(1) + 2A\,(0) = (1)\,B + (0)\,2A = J_1B + J_0 2A.$$

Let is assume that is right for $n = k$. That is, if $J_{k+2} = BJ_{k+1} + 2AJ_k$ then $J_{k+2} = J_{k+1}B + J_k 2A$. Now, we will obtain that is right for $n = k + 1$. For $n = k + 1$, by using our assumption and the fact that a ring satisfies the associative law for multiplication, if $J_{k+3} = BJ_{k+2} + 2AJ_{k+1}$ then

$$\begin{aligned}
J_{k+3} &= BJ_{k+2} + 2AJ_{k+1} \\
&= B(BJ_{k+1} + 2AJ_k) + 2A(BJ_k + 2AJ_{k-1}) \\
&= B(J_{k+1}B + J_k 2A) + 2A(J_kB + J_{k-1}2A) \\
&= B\,(J_{k+1}B) + 2A(J_kB) + B(J_k 2A) + 2A\,(J_{k-1}2A) \\
&= (BJ_{k+1})B + (2AJ_k)B + (BJ_k)2A + (2AJ_{k-1})\,2A \\
&= (BJ_{k+1} + 2AJ_k)B + (BJ_k + 2AJ_{k-1})2A \\
&= J_{k+2}B + J_{k+1}2A.
\end{aligned}$$

Thus, the proof is completed. ∎

Theorem 2.4. *For* $n \geq 1$,
a) $J_{n+1}J_{n-1} - J_n^2 = J_{n-1}2AJ_{n-1} - J_n 2AJ_{n-2}$
b) $J_{n-1}J_{n+1} - J_n^2 = J_{n-1}2AJ_{n-1} - J_{n-2}2AJ_n$

Proof. a) From Theorem 2.3, if $J_{n+2} = BJ_{n+1} + 2AJ_n$ then $J_{n+2} = J_{n+1}B + J_n 2A$. Then

$$\begin{aligned}
J_{n+1}J_{n-1} - J_n^2 &= (J_nB + J_{n-1}2A)J_{n-1} - J_n(BJ_{n-1} + 2AJ_{n-2}) \\
&= J_nBJ_{n-1} + J_{n-1}2AJ_{n-1} - J_nBJ_{n-1} - J_n 2AJ_{n-2} \\
&= J_{n-1}2AJ_{n-1} - J_n 2AJ_{n-2}.
\end{aligned}$$

b) From Theorem 2.3, if $J_{n+2} = BJ_{n+1} + 2AJ_n$ then $J_{n+2} = J_{n+1}B + J_n 2A$. Then

$$\begin{aligned}
J_{n-1}J_{n+1} - J_n^2 &= J_{n-1}(BJ_n + 2AJ_{n-1}) - (J_{n-1}B + J_{n-2}2A)J_n \\
&= J_{n-1}BJ_n + J_{n-1}2AJ_{n-1} - J_{n-1}BJ_n - J_{n-2}2AJ_n \\
&= J_{n-1}2AJ_{n-1} - J_{n-2}2AJ_n.
\end{aligned}$$

So the proof is completed. ∎

From Definition 2.1 the Jacobsthal sequences $\{J_n\}$ in rings are

$$\left\{0, 1, B,\ B^2 + 2A,\ B^3 + 4AB, B^4 + 6AB^2 + 4A^2, ...\ \right\}$$

for $n \geq 0$. For the application of the Theorem 2.4 we can write the following example by using these sequences.

Example 2.3. *a) For $n = 3$, if $J_{n+1}J_{n-1} - J_n^2 = J_{n-1}2AJ_{n-1} - J_n2AJ_{n-2}$ then*

$$\begin{aligned} J_4J_2 - J_3^2 &= J_22AJ_2 - J_32AJ_1 \\ \left(B^3 + 4AB\right)B - (B^2 + 2A)^2 &= (B)2A(B) - (B^2 + 2A)(2A)(1) \\ -4A^2 &= -4A^2. \end{aligned}$$

b) For $n = 4$, if $J_{n-1}J_{n+1} - J_n^2 = J_{n-1}2AJ_{n-1} - J_{n-2}2AJ_n$ then

$$\begin{aligned} J_3J_5 - J_4^2 &= J_32AJ_3 - J_22AJ_4 \\ \left(B^2 + 2A\right)\left(B^4 + 6AB^2 + 4A^2\right) - \left(B^3 + 4AB\right)^2 &= \left(B^2 + 2A\right)(2A)\left(B^2 + 2A\right) \\ &- (B)\,2A\left(B^3 + 4AB\right) \\ 8A^3 &= 8A^3. \end{aligned}$$

So, Theorem 2.4 is provided.

There is a relation between the $\{S_n\}$ sequences from equation (1) and the $\{J_n\}$ sequences from equation (2). We can give the following theorem and corollary for this relation.

Theorem 2.5. *For $n \geq 1$, $r \geq 0$,*

$$S_{n+r} = J_r2AS_{n-1} + J_{r+1}S_n.$$

Proof. We can complete the proof by induction over r and by using Definition 2.1. For $r = 0$

$$\begin{aligned} S_{n+0} &= S_n \\ &= (0)\,2AS_{n-1} + (1)\,S_n \\ &= J_02AS_{n-1} + J_1S_n. \end{aligned}$$

That is, $S_n = J_02AS_{n-1} + J_1S_n$. For $r = 1$

$$\begin{aligned} S_{n+1} &= BS_n + 2AS_{n-1} \\ &= (1)\,2AS_{n-1} + B\,(1)\,S_n + 2A\,(0)\,S_n \\ &= J_12AS_{n-1} + (BJ_1 + 2AJ_0)S_n \\ &= J_12AS_{n-1} + J_2S_n \end{aligned}$$

where $J_0 = 0$, $J_1 = 1$. That is, $S_{n+1} = J_1 2AS_{n-1} + J_2 S_n$. Let is assume that it is true, for for $r = k$, that

$$S_{n+k} = J_k 2AS_{n-1} + J_{k+1} S_n. \tag{7}$$

From our assumption, we know

$$S_{n+(k-1)} = J_{k-1} 2AS_{n-1} + J_k S_n. \tag{8}$$

Now, we will shown

$$S_{n+k+1} = J_{k+1} 2AS_{n-1} + J_{k+2} S_n$$

for $r = k + 1$. From equations (7) and (8), we can write

$$\begin{aligned} BS_{n+k} + 2AS_{n+(k-1)} &= (BJ_k + 2AJ_{k-1})2AS_{n-1} + (BJ_{k+1} + 2AJ_k)S_n \\ &= J_{k+1} 2AS_{n-1} + J_{k+2} S_n \end{aligned}$$

where $A_1 = B$ and $A_0 = 2A$. Thus,

$$S_{n+k+1} = J_{k+1} 2AS_{n-1} + J_{k+2} S_n$$

so, the proof is completed. ∎

Corollary 2.1. *For $n \geq 1$,*

$$S_n = J_n S_1 + J_{n-1} 2AS_0.$$

Proof. Interchange r and n, replace n by $n - 1$ and set $r = 1$ in Theorem 2.5, we obtain

$$S_{n+r} = J_r 2AS_{n-1} + J_{r+1} S_n$$

then

$$\begin{aligned} S_{r+n} &= S_{1+(n-1)} \\ &= J_{n-1} 2AS_{1-1} + J_{(n-1)+1} S_1 \\ &= J_{n-1} 2AS_0 + J_n S_1 \end{aligned}$$

That is, $S_n = J_n S_1 + J_{n-1} 2AS_0$ and the proof is completed. ∎

For the Jacobsthal sequences $\{J_n\}$ in rings, Theorem 2.5 becomes

$$J_{n+r} = J_r 2AJ_{n-1} + J_{r+1} J_n \qquad n \geq 1 \tag{9}$$

If we replace n by $n + 1$ and r by n in equation (9), then we have

$$J_{n+1}^2 + J_n 2AJ_n = J_{2n+1}.$$

Theorem 2.6. *For $n \geq 1$, $r \geq 1$*

$$J_n J_{n+r} - J_{n+r} J_n = J_n J_r 2A J_{n-1} - J_{n-1} 2A J_r J_n.$$

Proof. If we replace n by $r+1$ and r by $n-1$ in equation (9), we have

$$\begin{aligned} J_{n+r} &= J_{(r+1)+(n-1)} \\ &= J_{n-1} 2A J_{r+1-1} + J_{n-1+1} J_{r+1} \\ &= J_{n-1} 2A J_r + J_n J_{r+1}. \end{aligned}$$

That is,

$$J_{n+r} = J_{n-1} 2A J_r + J_n J_{r+1}. \tag{10}$$

From equations (9) and (10) and the fact that a ring satisfies the associative law for multiplication, we have

$$\begin{aligned} J_n(J_{r+1}J_n) &= (J_n J_{r+1}) J_n \\ J_n(J_{r+1}J_n + J_r 2A J_{n-1} - J_r 2A J_{n-1}) &= (J_n J_{r+1} + J_{n-1} 2A J_r - J_{n-1} 2A J_r) J_n \\ J_n \left(\underbrace{J_r 2A J_{n-1} + J_{r+1} J_n}_{J_{n+r}} - J_r 2A J_{n-1} \right) &= \left(\underbrace{J_n J_{r+1} + J_{n-1} 2A J_r}_{J_{n+r}} - J_{n-1} 2A J_r \right) J_n \\ J_n(J_{n+r} - J_r 2A J_{n-1}) &= (J_{n+r} - J_{n-1} 2A J_r) J_n \\ J_n J_{n+r} - J_n J_r 2A J_{n-1} &= J_{n+r} J_n - J_{n-1} 2A J_r J_n \\ J_n J_{n+r} - J_{n+r} J_n &= J_n J_r 2A J_{n-1} - J_{n-1} 2A J_r J_n \end{aligned}$$

so, the proof is completed. ▮

Theorem 2.7. *For $n \geq 3$,*

$$J_n J_{n+1} - J_{n-1} J_{n+2} = J_{n-2} 2A J_{n+1} - J_{n-1} 2A J_n.$$

Proof. From Definition 2.1, we have

$$\begin{aligned} J_n J_{n+1} - J_{n-1} J_{n+2} &= (J_{n-1} B + J_{n-2} 2A) J_{n+1} - J_{n-1}(B J_{n+1} + 2A J_n) \\ &= J_{n-1} B J_{n+1} + J_{n-2} 2A J_{n+1} - J_{n-1} B J_{n+1} - J_{n-1} 2A J_n \\ &= J_{n-2} 2A J_{n+1} - J_{n-1} 2A J_n. \end{aligned}$$

That is,

$$J_n J_{n+1} - J_{n-1} J_{n+2} = J_{n-2} 2A J_{n+1} - J_{n-1} 2A J_n.$$

Thus, the proof is completed. ▮

3. CONCLUSIONS

Most of the studies on recurrence sequences in the literature were on groups. However, only Fibonacci sequences were studied on the rings and the Jacobsthal sequences in rings have never been studied. In this study, we defined the Jacobsthal sequences $\{J_n\}$ in arbitrary rings with identity and gave some properties of these new sequences. Also, it was obtained the matrix derivated the terms of Jacobsthal sequences $\{J_n\}$ in rings and generating function for $\{J_n\}$. It was introduced the Binet formula known as the general formula and given us to find the n-th Jacobsthal number in rings without creating the Jacobsthal sequence $\{J_n\}$ in rings.

This study fills the gap in the literature by providing recurrence sequences on rings using definitions of the Jacobsthal sequences $\{J_n\}$ in rings given for the first time.

Acknowledgements. The authors declare that there is no conflict of interest regarding the publication of this paper.
The authors express their sincere thanks to the referee for his/her careful reading and suggestions that helped to improve this paper.

References

[1] A. F. Horadam, *A Generalized Fibonacci Sequence*, Amer. Math. Monthly, **68**, 5(1961), 445-459.

[2] A.F. Horadam, *Jacobsthal Representation Numbers*, The Fibonacci Quarterly, **34**, 1(1996), 40–54.

[3] D. J. DeCARLI, *A Generalized Fibonacci Sequence Over An Arbitrary Ring*, Fibonacci Quart., **8**, 2(1970), 182-184,198.

[4] N. N. Vorobyov, *The Fibonacci Numbers*, translated from the Russian by Normal D. Whaland, Jr., and Olga A. Tittlebaum, D. C. Heath and Co., Boston, 1963.

[5] O. Wyler, *On Second-order Recurrences*, Amer. Math. Monthly, **72**, 5 (1965), 500-506.

[6] O. Deveci, E. Karaduman, *Recurrence sequences in groups*, Lap Lambert Academic Publishing, 2013.

[7] O. Deveci, E. Karaduman, G. Saglam, *The Jacobsthal sequences in finite groups*, Bulletin of the Iranian Mathematical Society, **42**, 1(2016), 79-89.

[8] R. G. Buschman, *Fibonacci Numbers, Chebyshev Polynomials, Generalizations and Difference Equations*, Fibonacci Quart. **1**, 4(1963), 1-7.

[9] Y. Tasyurdu, I. Gultekin, *On period of Fibonacci sequences in finite rings with identity of order p^2*, Journal of Mathematics and System Science, 3(2013), 349-352.

[10] Y. Tasyurdu, I. Gultekin, *The period of Fibonacci sequences over the finite field of order p^2*, New Trends in Mathematical Sciences, **4**, 1(2016), 248-255.

[11] Y. Tasyurdu, Z. Dilmen, O*n Period of Generalized Fibonacci Sequence Over Finite Ring and Tridiagonal Matrix*, Celal Bayar University Journal of Science, **13**, 1(2017), 165-169.