BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 3–22
ISSN 1024–7696

The intersection and the union of the asynchronous systems

Serban E. Vlad

Abstract. The asynchronous systems f are the models of the asynchronous circuits from digital electrical engineering. They are multi-valued functions that associate to each input $u : \mathbf{R} \to \{0,1\}^m$ a set of states $x \in f(u)$, where $x : \mathbf{R} \to \{0,1\}^n$. The intersection of the systems allows adding supplementary conditions in modeling and the union of the systems allows considering the validity of one of two systems in modeling, for example when testing the asynchronous circuits and the circuit is supposed to be 'good' or 'bad'. The purpose of the paper is that of analyzing the intersection and the union against the initial/final states, initial/final time, initial/final state functions, subsystems, dual systems, inverse systems, Cartesian product of systems, parallel connection and serial connection of systems.

Mathematics subject classification: 93C62, 94C10.
Keywords and phrases: Asynchronous system, asynchronous circuit, signal, testing, modeling.

1 Preliminary definitions

Definition 1. *The set* $\mathbf{B} = \{0,1\}$ *endowed with the laws: the complement '$\overline{}$', the union '$\cup$', the intersection '$\cdot$', the modulo 2 sum '$\oplus$' etc is called the binary Boole algebra.*

Definition 2. *We denote by* $\mathbf{R}$ *the set of the real numbers. The initial value* $x(-\infty+0) \in \mathbf{B}$ *and the final value* $x(\infty-0) \in \mathbf{B}$ *of the function* $x : \mathbf{R} \to \mathbf{B}$ *are defined by*

$$\exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = x(-\infty+0),$$

$$\exists t_f \in \mathbf{R}, \forall t > t_f, x(t) = x(\infty-0).$$

The definition and the notations are similar for the $\mathbf{R} \to \mathbf{B}^n$ *functions,* $n \geq 1$.

Definition 3. *The characteristic function* $\chi_A : \mathbf{R} \to \mathbf{B}$ *of the set* $A \subset \mathbf{R}$ *is defined by*

$$\forall t \in \mathbf{R}, \chi_A(t) = \begin{cases} 1, if\ t \in A, \\ 0, if\ t \notin A. \end{cases}$$

Definition 4. *The set* $S^{(n)}$ *of the* $n-$*signals consists by definition in the functions* $x : \mathbf{R} \to \mathbf{B}^n$ *of the form*

$$x(t) = x(-\infty+0) \cdot \chi_{(-\infty,t_0)}(t) \oplus x(t_0) \cdot \chi_{[t_0,t_1)}(t) \oplus x(t_1) \cdot \chi_{[t_1,t_2)}(t) \oplus \ldots$$

where $x(-\infty+0)\in\mathbf{B}^n$, $t_0<t_1<t_2<...$ *is some strictly increasing unbounded sequence of real numbers and the laws '*$\cdot$*', '*$\oplus$*' are induced by those from* $\mathbf{B}$.

Notation 1. For an arbitrary set H, we use the notation

$$P^*(H)=\{H'|H'\subset H,H'\neq\emptyset\}.$$

Definition 5. *The functions* $f:U\rightarrow P^*(S^{(n)}),U\in P^*(S^{(m)})$ *are called (asynchronous) systems. Any* $u\in U$ *is called (admissible) input of* f *and the functions* $x\in f(u)$ *are the (possible) states of* f.

Remark 1. In the paper $t\in\mathbf{R}$ represents time. The $n-$signals model the tensions in digital electrical engineering and the asynchronous systems are the models of the asynchronous circuits. They represent multi-valued associations between a cause u and a set $f(u)$ of effects because of the uncertainties that occur in modeling.

Definition 5 represents the definition of the systems given under the explicit form. In previous works (such as [1]) we used equations and inequalities for defining systems under the implicit form.

Definition 6. *We have the systems* $f:U\rightarrow P^*(S^{(n)})$, $g:V\rightarrow P^*(S^{(n)})$ *with* $U,V\in P^*(S^{(m)})$. *If* $\exists u\in U\cap V,f(u)\cap g(u)\neq\emptyset$, *the system* $f\cap g:W\rightarrow P^*(S^{(n)})$ *defined by*

$$W=\{u|u\in U\cap V,f(u)\cap g(u)\neq\emptyset\}, \tag{1.1}$$

$$\forall u\in W,(f\cap g)(u)=f(u)\cap g(u)$$

is called the intersection of f *and* g.

Remark 2. The intersection of the systems represents the gain of information (of precision) in the modeling of a circuit that results by considering the validity of two (compatible!) models at the same time.

We have the special case when $V=S^{(m)}$ and the system g is constant (such systems are called autonomous): $\forall u\in S^{(m)}$, $g(u)=X$ where $X\in P^*(S^{(n)})$. Then $f\cap X:W\rightarrow P^*(S^{(n)})$ is the system given by

$$W=\{u|u\in U,f(u)\cap X\neq\emptyset\},$$

$$\forall u\in W,(f\cap X)(u)=f(u)\cap X.$$

We interpret $f\cap X$ in the next manner. When f models a circuit, $f\cap X$ represents a gain of information resulting by the statement of a request that does not depend on u.

Example 1. We give some possibilities of choosing in the intersection $f\cap g$ the constant system $g=X$:

i) the initial value of the states is null;

ii) the coordinates $x_1,...,x_n$ of the states are monotonous relative to the order $0<1$ (this allows defining the so called hazard-freedom of the systems);

iii) at each time instant, at least one coordinate of the state should be 1:

$$X = \{x|x \in S^{(n)}, \forall t \in \mathbf{R}, x_1(t) \cup ... \cup x_n(t) = 1\};$$

iv) the state can switch[1] with at most one coordinate at a time (a special case when the so called technical condition of good running of the systems is satisfied):

$$X = \{x|x \in S^{(n)}, \forall t \in \mathbf{R}, x(t-0) \neq x(t) \Longrightarrow \exists! i \in \{1, ..., n\}, x_i(t-0) \neq x_i(t)\};$$

v) X represents a 'stuck at 1 fault':

$$\exists i \in \{1, ..., n\}, X = \{x|x \in S^{(n)}, \forall t \in \mathbf{R}, x_i(t) = 1\},$$

this last choice of X is interesting in designing systems for testability, respectively in designing redundant systems;

vi) X consists in all $x \in S^{(n)}$ satisfying the next 'absolute inertia' property: $\delta_r > 0, \delta_f > 0$ are given so that $\forall i \in \{1, ..., n\}, \forall t \in \mathbf{R}$,

$$\overline{x_i(t-0)} \cdot x_i(t) \leq \bigcap_{\xi \in [t, t+\delta_r]} x_i(\xi);$$

$$x_i(t-0) \cdot \overline{x_i(t)} \leq \bigcap_{\xi \in [t, t+\delta_f]} \overline{x_i(\xi)}.$$

The interpretation of these inequalities is the following: if x_i switches from 0 to 1, then it remains 1 for more than δ_r time units and if x_i switches from 1 to 0 then it remains 0 for more than δ_f time units.

Example 2. We show a possibility of choosing in the intersection $f \cap g$, g non-constant. The Boolean function $F : \mathbf{B}^m \to \mathbf{B}^n$ is given and f is the arbitrary model of a circuit that computes F. $V = S^{(m)}$ and the parameters $\delta_r > 0, \delta_f > 0$ exist so that

$$\forall u \in S^{(m)}, g(u) = \{x|x \in S^{(n)}, \forall i \in \{1, ..., n\}, \forall t \in \mathbf{R},$$

$$\overline{x_i(t-0)} \cdot x_i(t) \leq \bigcap_{\xi \in [t-\delta_r, t)} F_i(u(\xi)),$$

$$x_i(t-0) \cdot \overline{x_i(t)} \leq \bigcap_{\xi \in [t-\delta_f, t)} \overline{F_i(u(\xi))}\}$$

meaning that $g(u)$ contains all x with the property that, on all the coordinates i and at all the time instants t:

– x_i switches from 0 to 1 only if $F_i(u(\cdot))$ was 1 for at least δ_r time units;

– x_i switches from 1 to 0 only if $F_i(u(\cdot))$ was 0 for at least δ_f time units.

---

[1] The left limit $x(t-0)$ of $x(t)$ that occurs in some examples is defined like this:

$$\forall t \in \mathbf{R}, \exists \varepsilon > 0, \forall \xi \in (t-\varepsilon, t), x(\xi) = x(t-0);$$

x switches if $x(t-0) \neq x(t)$, i.e. if it has a (left) discontinuity.

Definition 7. *The union of the systems* $f : U \to P^*(S^{(n)})$ *and* $g : V \to P^*(S^{(n)}), U, V \in P^*(S^{(m)})$ *is the system* $f \cup g : U \cup V \to P^*(S^{(n)})$ *that is defined by*

$$\forall u \in U \cup V, (f \cup g)(u) = \begin{cases} f(u), if\ u \in U \setminus V, \\ g(u), if\ u \in V \setminus U, \\ f(u) \cup g(u), if\ u \in U \cap V. \end{cases}$$

If $U \cap V = \emptyset$, *then* $f \cup g$ *is called the disjoint union of* f *and* g.

Remark 3. The union of the systems is the dual concept to that of intersection representing the loss of information (of precision) in modeling that results in general by considering the validity of one of two models of the same circuit. The disjoint union means no loss of information however.

Another possibility is that in Definition 7 f, g model two different circuits, see Example 3.

We have the special case when in the union $f \cup g$ the system g is constant under the form $V = S^{(m)}$, $g : S^{(m)} \to P^*(S^{(n)})$, $\forall u \in S^{(m)}$, $g(u) = X$, with $X \subset S^{(n)}$. Then $f \cup X : S^{(m)} \to P^*(S^{(n)})$ is defined by:

$$\forall u \in S^{(m)}, (f \cup X)(u) = \begin{cases} X, if\ u \in S^{(m)} \setminus U, \\ f(u) \cup X, if\ u \in U. \end{cases}$$

The interpretation of $f \cup X$ is the next one: when f is the model of an asynchronous circuit, X represents perturbations that are independent on u.

Example 3. In the union $f \cup g$ we presume that $U \cap V \neq \emptyset$ and f, g model two different circuits, the first considered 'good, without errors' and the second 'bad, with a certain error'. The testing problem consists in finding an input $u \in U \cap V$ so that $f(u) \cap g(u) = \emptyset$; after its application to $f \cup g$ and the measurement of a state $x \in (f \cup g)(u)$, we can say if $x \in f(u)$ and the tested circuit is 'good' or perhaps $x \in g(u)$ and the tested circuit is 'bad'.

2 Initial states and final states

Remark 4. In the next properties of the system f:

$$\forall u \in U, \forall x \in f(u), \exists \mu \in \mathbf{B}^n, \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu, \tag{2.1}$$

$$\forall u \in U, \exists \mu \in \mathbf{B}^n, \forall x \in f(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu, \tag{2.2}$$

$$\exists \mu \in \mathbf{B}^n, \forall u \in U, \forall x \in f(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu, \tag{2.3}$$

$$\forall u \in U, \forall x \in f(u), \exists \mu \in \mathbf{B}^n, \exists t_f \in \mathbf{R}, \forall t \geq t_f, x(t) = \mu, \tag{2.4}$$

$$\forall u \in U, \exists \mu \in \mathbf{B}^n, \forall x \in f(u), \exists t_f \in \mathbf{R}, \forall t \geq t_f, x(t) = \mu, \tag{2.5}$$

$$\exists \mu \in \mathbf{B}^n, \forall u \in U, \forall x \in f(u), \exists t_f \in \mathbf{R}, \forall t \geq t_f, x(t) = \mu \tag{2.6}$$

we have replaced $t > t_f$ from Definition 2 with $t \geq t_f$ and on the other hand (2.1) is always true due to the way that the $n-$signals were defined. We remark the truth of the implications

$$(2.3) \Longrightarrow (2.2) \Longrightarrow (2.1),$$

$$(2.6) \Longrightarrow (2.5) \Longrightarrow (2.4).$$

Definition 8. *Because f satisfies (2.1), we use to say that it has initial states. The vectors μ are called (the) initial states (of f), or (the) initial values of the states (of f).*

Definition 9. *We presume that f satisfies (2.2). We say in this situation that it has race-free initial states and the initial states μ are called race-free themselves.*

Definition 10. *When f satisfies (2.3), we use to say that it has a (constant) initial state μ. We say in this case that f is initialized and that μ is its (constant) initial state.*

Definition 11. *If f satisfies (2.4), it is called absolutely stable and we also say that it has final states. The vectors μ have in this case the name of final states (of f), or of final values of the states (of f).*

Definition 12. *If f fulfills the property (2.5), it is called absolutely race-free stable and we also say that it has race-free final states. The final states μ are called in this case race-free.*

Definition 13. *We presume that the system f satisfies (2.6). Then it is called absolutely constantly stable or equivalently we say that it has a (constant) final state. The vector μ is called in this situation (constant) final state.*

Theorem 1. *Let $f : U \to P^*(S^{(n)})$ and $g : V \to P^*(S^{(n)})$ be some systems, $U, V \in P^*(S^{(m)})$. If $\exists u \in U \cap V, f(u) \cap g(u) \neq \emptyset$ and f has race-free initial states (constant initial state), then $f \cap g$ has race-free initial states (constant initial state).*

Proof. If one of the previous properties is true for the states in $f(u)$, then it is true for the states in the subset $f(u) \cap g(u) \subset f(u)$ also, $u \in U$. □

Theorem 2. *If f has final states (race-free final states, constant final state) and $f \cap g$ exists, then $f \cap g$ has final states (race-free final states, constant final state).*

Theorem 3. *a) If f, g have race-free initial states and $\forall u \in U \cap V, f(u) \cap g(u) \neq \emptyset$ then $f \cup g$ has race-free initial states.*

b) If f, g have constant initial states and $\bigcup_{u \in U} f(u) \cap \bigcup_{u \in V} g(u) \neq \emptyset$ then $f \cup g$ has constant initial states.

Proof. a) The hypothesis states the truth of the next properties

$$\forall u \in U, \exists \mu \in \mathbf{B}^n, \forall x \in f(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu,$$

$$\forall u \in V, \exists \mu \in \mathbf{B}^n, \forall x \in g(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu,$$
$$\forall u \in U \cap V, f(u) \cap g(u) \neq \emptyset. \tag{2.7}$$

If $(U \setminus V) \cup (V \setminus U) \neq \emptyset$, then $\forall u \in (U \setminus V) \cup (V \setminus U)$ the statement is true because it states separately for f and g that they have race-free initial states. And if $U \cap V \neq \emptyset$, then $\forall u \in U \cap V, \forall x \in f(u) \cup g(u)$, the initial value $\mu = x(-\infty + 0)$ depends on u only, not also on the fact that $x \in f(u)$ or $x \in g(u)$ due to (2.7). We have that

$$\forall u \in U \cup V, \exists \mu \in \mathbf{B}^n, \forall x \in (f \cup g)(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu$$

is true.

b) Because $\bigcup\limits_{u \in U} f(u) \cap \bigcup\limits_{u \in V} g(u) \neq \emptyset$, in the statements

$$\exists \mu \in \mathbf{B}^n, \forall u \in U, \forall x \in f(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu,$$

$$\exists \mu' \in \mathbf{B}^n, \forall u \in V, \forall x \in g(u), \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu'$$

the two constants μ and μ', whose existence is unique, coincide. □

Theorem 4. *a) If f, g have final states, then $f \cup g$ has final states.*

b) If f, g have race-free final states and $\forall u \in U \cap V, f(u) \cap g(u) \neq \emptyset$ then $f \cup g$ has race-free final states.

c) If f, g have constant final states and $\bigcup\limits_{u \in U} f(u) \cap \bigcup\limits_{u \in V} g(u) \neq \emptyset$ then $f \cup g$ has constant final states.

3 Initial time and final time

Notation 2. The set of the $n-$signals with final values is denoted by $S_c^{(n)}$. It consists in the functions $x : \mathbf{R} \to \mathbf{B}^n$ of the form

$$x(t) = x(-\infty + 0) \cdot \chi_{(-\infty, t_0)}(t) \oplus x(t_0) \cdot \chi_{[t_0, t_1)}(t) \oplus x(t_1) \cdot \chi_{[t_1, t_2)}(t) \oplus \dots$$
$$\dots \oplus x(t_k) \cdot \chi_{[t_k, t_{k+1})}(t) \oplus x(\infty - 0) \cdot \chi_{[t_{k+1}, \infty)}(t)$$

where $x(-\infty + 0), x(\infty - 0) \in \mathbf{B}^n$ and $t_0 < t_1 < ... < t_k < t_{k+1}$ is a finite family of real numbers, $k \geq 0$.

Remark 5. We state the next properties on the asynchronous system $f : U \to P^*(S^{(n)}), U \in P^*(S^{(m)})$:

$$\forall u \in U, \forall x \in f(u), \exists \mu \in \mathbf{B}^n, \exists t_0 \in \mathbf{R}, \forall t < t_0, x(t) = \mu, \tag{3.1}$$

$$\forall u \in U, \exists t_0 \in \mathbf{R}, \forall x \in f(u), \exists \mu \in \mathbf{B}^n, \forall t < t_0, x(t) = \mu, \tag{3.2}$$

$$\exists t_0 \in \mathbf{R}, \forall u \in U, \forall x \in f(u), \exists \mu \in \mathbf{B}^n, \forall t < t_0, x(t) = \mu, \tag{3.3}$$

$$\forall u \in U, \forall x \in f(u) \cap S_c^{(n)}, \exists \mu \in \mathbf{B}^n, \exists t_f \in \mathbf{R}, \forall t \geq t_f, x(t) = \mu, \tag{3.4}$$

$$\forall u\in U,\exists t_f\in\mathbf{R},\forall x\in f(u)\cap S_c^{(n)},\exists\mu\in\mathbf{B}^n,\forall t\geq t_f,x(t)=\mu, \tag{3.5}$$

$$\exists t_f\in\mathbf{R},\forall u\in U,\forall x\in f(u)\cap S_c^{(n)},\exists\mu\in\mathbf{B}^n,\forall t\geq t_f,x(t)=\mu. \tag{3.6}$$

The properties (3.1) and (3.4) are fulfilled by all the systems and the next implications hold:

$$(3.3)\Longrightarrow(3.2)\Longrightarrow(3.1),$$

$$(3.6)\Longrightarrow(3.5)\Longrightarrow(3.4).$$

Definition 14. *The fact that f satisfies (3.1) is expressed sometimes by saying that it has unbounded initial time and any t_0 satisfying this property is called unbounded initial time (instant).*

Definition 15. *Let f be a system that fulfills the property (3.2). We say that it has bounded initial time and any t_0 making this property true is called bounded initial time (instant).*

Definition 16. *When f satisfies (3.3), we use to say that it has fixed initial time and any t_0 fulfilling (3.3) is called fixed initial time (instant).*

Definition 17. *The fact that f satisfies (3.4) is expressed by saying that it has unbounded final time and any t_f satisfying this property is called unbounded final time (instant).*

Definition 18. *If f fulfills the property (3.5), we say that it has bounded final time. Any number t_f satisfying (3.5) is called bounded final time (instant).*

Definition 19. *We presume that the system f satisfies the property (3.6). Then we say that it has fixed final time and any number t_f satisfying (3.6) is called fixed final time (instant).*

Theorem 5. *If f has bounded initial time (fixed initial time) and $f\cap g$ exists, then $f\cap g$ has bounded initial time (fixed initial time).*

Proof. Like previously, if one of the above properties is true for the states in $f(u)$, then it is true for the states in $f(u)\cap g(u)\subset f(u),u\in U$. □

Theorem 6. *If f has bounded final time (fixed final time) and $f\cap g$ exists, then $f\cap g$ has bounded final time (fixed final time).*

Theorem 7. *If f,g have bounded initial time (fixed initial time), then $f\cup g$ has bounded initial time (fixed initial time).*

Proof. We presume that f,g have bounded initial time. If $(U\setminus V)\cup(V\setminus U)\neq\emptyset$, then $\forall u\in(U\setminus V)\cup(V\setminus U)$, $(f\cup g)(u)$ has the desired property, that refers to exactly one of f,g. We presume that $U\cap V\neq\emptyset$ and let $u\in U\cap V$ be arbitrary. $t_0',t_0"\in\mathbf{R}$ exist, depending on u, so that

$$\forall x\in f(u),\exists\mu\in\mathbf{B}^n,\forall t<t_0',x(t)=\mu,$$

$$\forall x \in g(u), \exists \mu \in \mathbf{B}^n, \forall t < t_0^", x(t) = \mu,$$

$t_0 = \min\{t_0', t_0^"\}$ satisfies

$$\forall x \in f(u) \cup g(u), \exists \mu \in \mathbf{B}^n, \forall t < t_0, x(t) = \mu.$$

□

Theorem 8. *If f, g have bounded final time (fixed final time), then $f \cup g$ has bounded final time (fixed final time).*

4 Initial state function and set of initial states. Final state function and set of final states

Definition 20. *Let $f : U \to P^*(S^{(n)}), U \in P^*(S^{(m)})$ be a system. The initial state function $\phi_0 : U \to P^*(\mathbf{B}^n)$ and the set of the initial states $\Theta_0 \in P^*(\mathbf{B}^n)$ of f are defined by*

$$\forall u \in U, \phi_0(u) = \{x(-\infty + 0) | x \in f(u)\},$$

$$\Theta_0 = \bigcup_{u \in U} \phi_0(u).$$

Definition 21. *If f has final states, i.e. if (2.4) is satisfied, the final state function $\phi_f : U \to P^*(\mathbf{B}^n)$ and the set of the final states $\Theta_f \in P^*(\mathbf{B}^n)$ of f are*

$$\forall u \in U, \phi_f(u) = \{x(\infty - 0) | x \in f(u)\},$$

$$\Theta_f = \bigcup_{u \in U} \phi_f(u).$$

Theorem 9. *For the systems f, g we have $(\phi \cap \gamma)_0 : W \to P^*(\mathbf{B}^n)$,*

$$\forall u \in W, (\phi \cap \gamma)_0(u) = \phi_0(u) \cap \gamma_0(u),$$

$$(\Theta \cap \Gamma)_0 = \bigcup_{u \in W} (\phi \cap \gamma)_0(u).$$

We have presumed that the domain W of $f \cap g$ is non-empty and we have denoted by $\phi_0, \gamma_0, (\phi \cap \gamma)_0$ the initial state functions of $f, g, f \cap g$ and respectively by $(\Theta \cap \Gamma)_0$ the set of initial states of $f \cap g$.

Proof. We can write that $\forall u \in W$,

$$(\phi \cap \gamma)_0(u) = \{x(-\infty + 0) | x \in (f \cap g)(u)\} = \{x(-\infty + 0) | x \in f(u) \cap g(u)\} =$$

$$= \{x(-\infty + 0) | x \in f(u)\} \cap \{x(-\infty + 0) | x \in g(u)\} = \phi_0(u) \cap \gamma_0(u).$$

□

Theorem 10. *If f, g have final states, then we have $(\phi \cap \gamma)_f : W \to P^*(\mathbf{B}^n)$,*

$$\forall u \in W, (\phi \cap \gamma)_f(u) = \phi_f(u) \cap \gamma_f(u),$$

$$(\Theta \cap \Gamma)_f = \bigcup_{u \in W} (\phi \cap \gamma)_f(u).$$

We have presumed that $W \neq \emptyset$ and the notations are obvious and similar with those from the previous theorem.

Theorem 11. *For the systems f, g we have $(\phi \cup \gamma)_0 : U \cup V \to P^*(\mathbf{B}^n)$,*

$$\forall u \in U \cup V, (\phi \cup \gamma)_0(u) = \begin{cases} \phi_0(u),\ u \in U \setminus V, \\ \gamma_0(u),\ u \in V \setminus U, \\ \phi_0(u) \cup \gamma_0(u),\ u \in U \cap V, \end{cases}$$

$$(\Theta \cup \Gamma)_0 = \bigcup_{u \in U \cup V} (\phi \cup \gamma)_0(u).$$

We have denoted by $(\phi \cup \gamma)_0$ the initial state function of $f \cup g$ and respectively by $(\Theta \cup \Gamma)_0$ the set of initial states of $f \cup g$.

Proof. Three possibilities exist for an arbitrary $u \in U \cup V : u \in U \setminus V, u \in V \setminus U$ and $u \in U \cap V$. If for example $u \in U \setminus V$, then:

$$(\phi \cup \gamma)_0(u) = \{x(-\infty + 0) | x \in (f \cup g)(u)\} = \{x(-\infty + 0) | x \in f(u)\} = \phi_0(u).$$

□

Theorem 12. *We presume that f, g have final states. We have $(\phi \cup \gamma)_f : U \cup V \to P^*(\mathbf{B}^n)$,*

$$\forall u \in U \cup V, (\phi \cup \gamma)_f(u) = \begin{cases} \phi_f(u),\ u \in U \setminus V, \\ \gamma_f(u),\ u \in V \setminus U, \\ \phi_f(u) \cup \gamma_f(u),\ u \in U \cap V, \end{cases}$$

$$(\Theta \cup \Gamma)_f = \bigcup_{u \in U \cup V} (\phi \cup \gamma)_f(u)$$

where the notations are obvious and similar with those from the previous theorem.

5 Subsystem

Definition 22. *Let $f : U \to P^*(S^{(n)})$ and $g : V \to P^*(S^{(n)}), U, V \in P^*(S^{(m)})$ be two systems. f is called a subsystem of g if*

$$U \subset V \text{ and } \forall u \in U, f(u) \subset g(u).$$

Remark 6. The subsystem of a system represents a more precise model of the same circuit, obtained perhaps after restricting the inputs set.

A special case in the inclusion $f \subset g$ is the one when f is uni-valued (it is called deterministic in this situation). This is considered to be non-realistic in modeling.

Example 4. Let f be a system and we take some arbitrary $\mu \in \Theta_0$. The subsystem $f_\mu : U_\mu \to P^*(S^{(n)})$ defined by

$$U_\mu = \{u | u \in U, \mu \in \phi_0(u)\},$$

$$\forall u \in U_\mu, f_\mu(u) = \{x | x \in f(u), x(-\infty + 0) = \mu\}$$

is called the restriction of f at μ. The next property is satisfied: for $\Theta_0 = \{\mu^1, ..., \mu^k\}$, we have $f = f_{\mu^1} \cup ... \cup f_{\mu^k}$ (the union is not disjoint).

Theorem 13. *Let $f : U \to P^*(S^{(n)})$, $f_1 : U_1 \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $g_1 : V_1 \to P^*(S^{(n)})$ be some systems with $U, U_1, V, V_1 \in P^*(S^{(m)})$. If $f \subset f_1, g \subset g_1$ and if $f \cap g$ exists, then $f_1 \cap g_1$ exists and the inclusion $f \cap g \subset f_1 \cap g_1$ is true.*

Proof. We denote by W the set from (1.1) and with W_1 the set

$$W_1 = \{u | u \in U_1 \cap V_1, f_1(u) \cap g_1(u) \neq \emptyset\}$$

From the fact that $U \subset U_1$, $\forall u \in U, f(u) \subset f_1(u)$, $V \subset V_1$, $\forall v \in V, g(v) \subset g_1(v)$ and $W \neq \emptyset$ we infer $W \subset W_1$, $W_1 \neq \emptyset$ and furthermore we have $\forall u \in W, (f \cap g)(u) = f(u) \cap g(u) \subset f_1(u) \cap g_1(u) = (f_1 \cap g_1)(u)$. □

Theorem 14. *We consider the systems $f : U \to P^*(S^{(n)})$, $f_1 : U_1 \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $g_1 : V_1 \to P^*(S^{(n)})$ with $U, U_1, V, V_1 \in P^*(S^{(m)})$. If $f \subset f_1$, $g \subset g_1$ then $f \cup g \subset f_1 \cup g_1$.*

Proof. From $U \subset U_1, V \subset V_1$ we infer that $U \cup V \subset U_1 \cup V_1$. It is shown that $\forall u \in U \cup V, (f \cup g)(u) \subset (f_1 \cup g_1)(u)$ is true in all the three situations $u \in U \setminus V, u \in V \setminus U$ and $u \in U \cap V$. For example if $u \in U \setminus V$, then two possibilities exist:

– $u \in U_1 \setminus V_1$, thus

$$(f \cup g)(u) = f(u) \subset f_1(u) = (f_1 \cup g_1)(u),$$

– $u \in U_1 \cap V_1$, when

$$(f \cup g)(u) = f(u) \subset f_1(u) \subset f_1(u) \cup g_1(u) = (f_1 \cup g_1)(u)$$

is true. We observe that $u \in V_1 \setminus U_1$ is impossible, since $u \notin U_1$ implies $u \notin U$, contradiction. □

6 Dual system

Notation 3. For $u \in S^{(m)}$, we denote by $\overline{u} \in S^{(m)}$ the complement of u satisfying

$$\forall t \in \mathbf{R}, \overline{u}(t) = (\overline{u_1(t)}, ..., \overline{u_m(t)})$$

Definition 23. *The dual system of f is the system $f^* : U^* \to P^*(S^{(n)})$ defined in the next way*

$$U^* = \{\overline{u} | u \in U\},$$

$$\forall u \in U^*, f^*(u) = \{\overline{x} | x \in f(\overline{u})\}.$$

Remark 7. For any $u \in U^*$, $\overline{u} \in U$ and Definition 23 is correct.

If f models a circuit, then f^* models the circuit that is obtained from the previous one after the replacement of the OR gates with AND gates and viceversa and respectively of the input and state tensions with their complements (the complement of the 'HIGH' tension is by definition the 'LOW' tension and viceversa).

Theorem 15. *If $f \cap g$ exists, then $(f \cap g)^*$, $f^* \cap g^*$ exist and*

$$(f \cap g)^* = f^* \cap g^*$$

Proof. We denote by W the domain (1.1) of $f \cap g$. The domain of $(f \cap g)^*$ is W^* and the domain W_1 of $f^* \cap g^*$ is:

$$W_1 = \{u|u \in U^* \cap V^*, f^*(u) \cap g^*(u) \neq \emptyset\} =$$

$$= \{u|\overline{u} \in U \cap V, \{\overline{x}|x \in f(\overline{u})\} \cap \{\overline{x}|x \in g(\overline{u})\} \neq \emptyset\} =$$

$$= \{\overline{u}|u \in U \cap V, \{\overline{x}|x \in f(u)\} \cap \{\overline{x}|x \in g(u)\} \neq \emptyset\} =$$

$$= \{\overline{u}|u \in U \cap V, \{x|x \in f(u)\} \cap \{x|x \in g(u)\} \neq \emptyset\} = W^*.$$

Moreover, for any $u \in W^*$ we infer

$$(f \cap g)^*(u) = \{\overline{x}|x \in (f \cap g)(\overline{u})\} = \{\overline{x}|x \in f(\overline{u}) \cap g(\overline{u})\} =$$

$$= \{\overline{x}|x \in f(\overline{u})\} \cap \{\overline{x}|x \in g(\overline{u})\} = f^*(u) \cap g^*(u) = (f^* \cap g^*)(u)$$

□

Theorem 16. *We have*

$$(f \cup g)^* = f^* \cup g^*.$$

Proof. We remark that the equal domains of the two systems are $(U \cup V)^* = U^* \cup V^*$. Let $u \in U^* \cup V^*$ be an arbitrary input. If $u \in U^* \setminus V^*$, then $f^*(u) = (f^* \cup g^*)(u)$ and the fact that $\overline{u} \in U \setminus V$ implies $(f \cup g)(\overline{u}) = f(\overline{u})$, thus

$$(f \cup g)^*(u) = \{\overline{x}|x \in (f \cup g)(\overline{u})\} = \{\overline{x}|x \in f(\overline{u})\} = f^*(u) = (f^* \cup g^*)(u).$$

If $u \in V^* \setminus U^*$, the situation is similar. We presume in this moment that $u \in U^* \cap V^*$, implying $f^*(u) \cup g^*(u) = (f^* \cup g^*)(u)$, $\overline{u} \in U \cap V, (f \cup g)(\overline{u}) = f(\overline{u}) \cup g(\overline{u})$ and we have:

$$(f \cup g)^*(u) = \{\overline{x}|x \in (f \cup g)(\overline{u})\} = \{\overline{x}|x \in f(\overline{u}) \cup g(\overline{u})\} =$$

$$= \{\overline{x}|x \in f(\overline{u})\} \cup \{\overline{x}|x \in g(\overline{u})\} = f^*(u) \cup g^*(u) = (f^* \cup g^*)(u).$$

In all the three cases the statement of the theorem was proved to be true. □

7 Inverse system

Definition 24. *The inverse system of f is defined by $f^{-1}: X \to P^*(S^{m)})$,*

$$X = \bigcup_{u \in U} f(u),$$

$$\forall x \in X, f^{-1}(x) = \{u | u \in U, x \in f(u)\}.$$

Remark 8. The inputs and the states of f become states and inputs of f^{-1}, meaning that f^{-1} inverts the causes and the effects in modeling: its aim is to answer the question "given an effect x, which are the causes u producing it?"

Theorem 17. *Let $f : U \to P^*(S^{(n)}), g : V \to P^*(S^{(n)}), U, V \in P^*(S^{(m)})$ be some systems. If $\exists u \in U \cap V, f(u) \cap g(u) \neq \emptyset$, then the systems $(f \cap g)^{-1}$, $f^{-1} \cap g^{-1}$ exist and they have the same domain:*

$$Y = \bigcup_{u \in W} (f(u) \cap g(u)).$$

Furthermore, we have

$$(f \cap g)^{-1} = f^{-1} \cap g^{-1}.$$

Proof. Y is obviously the domain of $(f \cap g)^{-1}$. We can write

$$Y = \bigcup_{u \in U \cap V} (f(u) \cap g(u)) = \{x | \exists u \in U \cap V, x \in f(u) \cap g(u)\} =$$

$$= \{x | x \in \bigcup_{v \in U} f(v) \cap \bigcup_{v \in V} g(v), \exists u, u \in U, x \in f(u) \text{ and } u \in V, x \in g(u)\} =$$

$$= \{x | x \in \bigcup_{v \in U} f(v) \cap \bigcup_{v \in V} g(v), \exists u, u \in f^{-1}(x) \text{ and } u \in g^{-1}(x)\} =$$

$$= \{x | x \in \bigcup_{v \in U} f(v) \cap \bigcup_{v \in V} g(v), f^{-1}(x) \cap g^{-1}(x) \neq \emptyset\}$$

thus Y is the domain of $f^{-1} \cap g^{-1}$ too. We have $\forall x \in Y$,

$$(f \cap g)^{-1}(x) = \{u | u \in U \cap V, x \in (f \cap g)(u)\} = \{u | u \in U \cap V, x \in f(u) \cap g(u)\} =$$

$$= \{u | u \in U \cap V, x \in f(u)\} \cap \{u | u \in U \cap V, x \in g(u)\} =$$

$$= (\{u | u \in U \setminus V, x \in f(u)\} \cup \{u | u \in U \cap V, x \in f(u)\}) \cap$$

$$\cap (\{u | u \in V \setminus U, x \in g(u)\} \cup \{u | u \in U \cap V, x \in g(u)\}) =$$

$$= \{u | u \in U, x \in f(u)\} \cap \{u | u \in V, x \in g(u)\} = f^{-1}(x) \cap g^{-1}(x) = (f^{-1} \cap g^{-1})(x).$$

□

Theorem 18. *We consider the systems* $f : U \to P^*(S^{(n)}), g : V \to P^*(S^{(n)}), U, V \in P^*(S^{(m)})$. *The systems* $(f \cup g)^{-1}$, $f^{-1} \cup g^{-1}$ *have the domain equal with*

$$Y' = \bigcup_{u \in U} f(u) \cup \bigcup_{u \in V} g(u)$$

and the next equality is true

$$(f \cup g)^{-1} = f^{-1} \cup g^{-1}$$

Proof. The domain of $(f \cup g)^{-1}$ is

$$\bigcup_{u \in U \cup V} (f \cup g)(u) = \bigcup_{u \in U \setminus V} (f \cup g)(u) \cup \bigcup_{u \in U \cap V} (f \cup g)(u) \cup \bigcup_{u \in V \setminus U} (f \cup g)(u) =$$

$$= \bigcup_{u \in U \setminus V} f(u) \cup \bigcup_{u \in U \cap V} (f(u) \cup g(u)) \cup \bigcup_{u \in V \setminus U} g(u) =$$

$$= \bigcup_{u \in U \setminus V} f(u) \cup \bigcup_{u \in U \cap V} f(u) \cup \bigcup_{u \in U \cap V} g(u) \cup \bigcup_{u \in V \setminus U} g(u) = \bigcup_{u \in U} f(u) \cup \bigcup_{u \in V} g(u)$$

and it coincides with Y', that is obviously the domain of $f^{-1} \cup g^{-1}$. For any $x \in Y'$ we have:

$$(f \cup g)^{-1}(x) = \{u | u \in U \cup V, x \in (f \cup g)(u)\} = \{u | u \in U \setminus V, x \in f(u)\} \cup$$

$$\cup \{u | u \in V \setminus U, x \in g(u)\} \cup \{u | u \in U \cap V, x \in f(u)\} \cup \{u | u \in U \cap V, x \in g(u)\} =$$

$$= \{u | u \in U, x \in f(u)\} \cup \{u | u \in V, x \in g(u)\} =$$

$$= \begin{cases} f^{-1}(x), x \in \bigcup\limits_{u \in U} f(u) \setminus \bigcup\limits_{u \in V} g(u) \\ g^{-1}(x), x \in \bigcup\limits_{u \in V} g(u) \setminus \bigcup\limits_{u \in U} f(u) \\ f^{-1}(x) \cup g^{-1}(x), x \in \bigcup\limits_{u \in U} f(u) \cap \bigcup\limits_{u \in V} g(u) \end{cases} = (f^{-1} \cup g^{-1})(x)$$

□

8 Cartesian product

Definition 25. *Let* $u \in S^{(m)}, u' \in S^{(m')}$ *be two signals. We define the Cartesian product* $u \times u' \in S^{(m+m')}$ *of the functions* u *and* u' *by*

$$\forall t \in \mathbf{R}, (u \times u')(t) = (u_1(t), ..., u_m(t), u'_1(t), ..., u'_{m'}(t))$$

Definition 26. *For any sets* $U \in P^*(S^{(m)}), U' \in P^*(S^{(m')})$ *we define the Cartesian product* $U \times U' \in P^*(S^{(m+m')})$,

$$U \times U' = \{u \times u' | u \in U, u' \in U'\}$$

Definition 27. *The Cartesian product of the systems f and $f' : U' \to P^*(S^{(n')}), U' \in P^*(S^{(m')})$ is $f \times f' : U \times U' \to P^*(S^{(n+n')})$,*

$$\forall u \times u' \in U \times U', (f \times f')(u \times u') = f(u) \times f'(u')$$

Remark 9. The Cartesian product of the systems models two circuits that are not interconnected and run under different inputs.

Theorem 19. *Let $f : U \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $U, V \in P^*(S^{(m)})$ and $f' : U' \to P^*(S^{(n')})$, $U' \in P^*(S^{(m')})$ be three systems. If $\exists u \in U \cap V$, $f(u) \cap g(u) \neq \emptyset$ then the systems $(f \cap g) \times f'$, $(f \times f') \cap (g \times f')$ are defined and $W \times U'$ is their common domain, where we have used again the notation*

$$W = \{u | u \in U \cap V, f(u) \cap g(u) \neq \emptyset\}.$$

The next equality is true

$$(f \cap g) \times f' = (f \times f') \cap (g \times f').$$

Proof. We show that $W \times U'$, that is the domain of $(f \cap g) \times f'$, is also the domain of $(f \times f') \cap (g \times f')$:

$$W \times U' = \{u \times u' | u \in W, u' \in U'\} =$$

$$= \{u \times u' | u \in U \cap V, u' \in U', f(u) \cap g(u) \neq \emptyset \text{ and } f'(u') \neq \emptyset\} =$$

$$= \{u \times u' | u \times u' \in (U \cap V) \times U', (f(u) \times f'(u')) \cap (g(u) \times f'(u')) \neq \emptyset\} =$$

$$= \{u \times u' | u \times u' \in (U \times U') \cap (V \times U'), (f \times f')(u \times u') \cap (g \times f')(u \times u') \neq \emptyset\}.$$

Furthermore for any $u \times u' \in W \times U'$ we have

$$((f \cap g) \times f')(u \times u') = (f \cap g)(u) \times f'(u') = (f(u) \cap g(u)) \times f'(u') =$$

$$= (f(u) \times f'(u')) \cap (g(u) \times f'(u')) = (f \times f')(u \times u') \cap (g \times f')(u \times u') =$$

$$= ((f \times f') \cap (g \times f'))(u \times u').$$

□

Theorem 20. *Let $f : U \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $U, V \in P^*(S^{(m)})$ and $f' : U' \to P^*(S^{(n')})$, $U' \in P^*(S^{(m')})$ be some systems. The common domain of $(f \cup g) \times f'$, $(f \times f') \cup (g \times f')$ is $(U \cup V) \times U' = (U \times U') \cup (V \times U')$ and the next equality holds*

$$(f \cup g) \times f' = (f \times f') \cup (g \times f').$$

Proof. $\forall u \times u' \in (U \cup V) \times U'$ we have one of the next possibilities:

Case $u \times u' \in (U \setminus V) \times U' = (U \times U') \setminus (V \times U')$

$$((f \cup g) \times f')(u \times u') = (f \cup g)(u) \times f'(u') = f(u) \times f'(u') = (f \times f')(u \times u') =$$
$$= ((f \times f') \cup (g \times f'))(u \times u');$$

Case $u \times u' \in (V \setminus U) \times U'$ is similar;

Case $u \times u' \in (U \cap V) \times U' = (U \times U') \cap (V \times U')$

$$((f \cup g) \times f')(u \times u') = (f \cup g)(u) \times f'(u') = (f(u) \cup g(u)) \times f'(u') =$$
$$= (f(u) \times f'(u')) \cup (g(u) \times f'(u')) = (f \times f')(u \times u') \cup (g \times f')(u \times u') =$$
$$= ((f \times f') \cup (g \times f'))(u \times u').$$

□

9 Parallel connection

Definition 28. *The parallel connection of f with $f_1' : U_1' \to P^*(S^{(n')}), U_1' \in P^*(S^{(m)})$ is $(f, f_1') : U \cap U_1' \to P^*(S^{(n+n')})$,*

$$\forall u \in U \cap U_1', (f, f_1')(u) = (f \times f_1')(u \times u).$$

Remark 10. The parallel connection models two circuits that are not interconnected and run under the same input.

Theorem 21. *We consider the systems $f : U \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $f_1' : U_1' \to P^*(S^{(n')})$, with $U, V, U_1' \in P^*(S^{(m)})$. We presume that $\exists u \in U \cap V \cap U_1'$ so that $f(u) \cap g(u) \neq \emptyset$. Then the set*

$$W' = \{u | u \in U \cap V \cap U_1', f(u) \cap g(u) \neq \emptyset\}$$

is the domain of the systems $(f \cap g, f_1')$, $(f, f_1') \cap (g, f_1')$ and the next equality holds

$$(f \cap g, f_1') = (f, f_1') \cap (g, f_1').$$

Proof. We observe that W' is non-empty, it is the domain of $(f \cap g, f_1')$ and we show that it is also the domain of $(f, f_1') \cap (g, f_1')$. We denote by

$$W" = \{u | u \in (U \cap U_1') \cap (V \cap U_1'), (f, f_1')(u) \cap (g, f_1')(u) \neq \emptyset\}$$

the domain of $(f, f_1') \cap (g, f_1')$ for which we have

$$W" = \{u | u \in U \cap V \cap U_1', (f(u) \times f_1'(u)) \cap (g(u) \times f_1'(u)) \neq \emptyset\} =$$
$$= \{u | u \in U \cap V \cap U_1', (f(u) \cap g(u)) \times f_1'(u) \neq \emptyset\} =$$

$$= \{u | u \in U \cap V \cap U_1', f(u) \cap g(u) \neq \emptyset\}$$

thus $W" = W'$. For any $u \in W'$ we have:

$$(f \cap g, f_1')(u) = ((f \cap g) \times f_1')(u \times u) \overset{Theorem\ 19}{=} ((f \times f_1') \cap (g \times f_1'))(u \times u) =$$

$$= (f \times f_1')(u \times u) \cap (g \times f_1')(u \times u) = (f, f_1')(u) \cap (g, f_1')(u) = ((f, f_1') \cap (g, f_1'))(u).$$

□

Remark 11. A similar result with the one from Theorem 19 states the truth of the formula

$$f \times (f' \cap g') = (f \times f') \cap (f \times g')$$

and then from Theorem 19 we get the next property

$$(f \cap g) \times (f' \cap g') = (f \times f') \cap (f \times g') \cap (g \times f') \cap (g \times g').$$

Like in Theorem 21 we can prove that

$$(f, f_1' \cap g_1') = (f, f_1') \cap (f, g_1')$$

is true and then from Theorem 21 we obtain

$$(f \cap g, f_1' \cap g_1') = (f, f_1') \cap (f, g_1') \cap (g, f_1') \cap (g, g_1').$$

Theorem 22. *Let $f : U \to P^*(S^{(n)})$, $g : V \to P^*(S^{(n)})$, $f_1' : U_1' \to P^*(S^{(n')})$ be three systems with $U, V, U_1' \in P^*(S^{(m)})$. If $U \cap U_1' \neq \emptyset$, $V \cap U_1' \neq \emptyset$, then the common domain of the systems $(f \cup g, f_1')$, $(f, f_1') \cup (g, f_1')$ is $(U \cup V) \cap U_1' = (U \cap U_1') \cup (V \cap U_1')$ and we have*

$$(f \cup g, f_1') = (f, f_1') \cup (g, f_1').$$

Remark 12. We observe the truth of the formulas

$$f \times (f' \cup g') = (f \times f') \cup (f \times g'),$$

$$(f \cup g) \times (f' \cup g') = (f \times f') \cup (f \times g') \cup (g \times f') \cup (g \times g')$$

and respectively of the formulas

$$(f, f_1' \cup g_1') = (f, f_1') \cup (f, g_1'),$$

$$(f \cup g, f_1' \cup g_1') = (f, f_1') \cup (f, g_1') \cup (g, f_1') \cup (g, g_1').$$

10 Serial connection

Definition 29. *The serial connection of* $h : X \rightarrow P^*(S^{(p)}), X \in P^*(S^{(n)})$ *with* $f : U \rightarrow P^*(S^{(n)})$, $U \in P^*(S^{(m)})$ *is defined whenever* $\bigcup\limits_{u \in U} f(u) \subset X$ *by*[2] $h \circ f : U \rightarrow P^*(S^{(p)})$,

$$\forall u \in U, (h \circ f)(u) = \bigcup_{x \in f(u)} h(x).$$

Remark 13. The serial connection of the systems models two circuits connected in cascade and it coincides with the usual composition of the multi-valued functions.

Theorem 23. *We consider the systems* $f : U \rightarrow P^*(S^{(n)})$, $g : V \rightarrow P^*(S^{(n)})$, $U, V \in P^*(S^{(m)})$ *and* $h : X \rightarrow P^*(S^{(p)})$, $h_1 : X_1 \rightarrow P^*(S^{(p)})$, $X, X_1 \in P^*(S^{(n)})$.

a) If $\bigcup\limits_{u \in U} f(u) \subset X$, $\bigcup\limits_{u \in V} g(u) \subset X$ *and* $\exists u \in U \cap V, f(u) \cap g(u) \neq \emptyset$ *then the sets*

$$W = \{u | u \in U \cap V, f(u) \cap g(u) \neq \emptyset\},$$

$$W_1 = \{u | u \in U \cap V, \bigcup_{x \in f(u)} h(x) \cap \bigcup_{x \in g(u)} h(x) \neq \emptyset\}$$

are non-empty and represent the domains of the systems $h \circ (f \cap g)$, $(h \circ f) \cap (h \circ g)$. *We have*

$$h \circ (f \cap g) \subset (h \circ f) \cap (h \circ g);$$

b) We ask that $\bigcup\limits_{u \in U} f(u) \subset \{x | x \in X \cap X_1, h(x) \cap h_1(x) \neq \emptyset\}$. U *is the domain of the systems* $(h \cap h_1) \circ f$, $(h \circ f) \cap (h_1 \circ f)$ *and the next inclusion is true:*

$$(h \cap h_1) \circ f \subset (h \circ f) \cap (h_1 \circ f)$$

Proof. a) From the hypothesis $f \cap g$ is defined and has the domain W. As

$$\bigcup_{u \in W} (f \cap g)(u) \subset \bigcup_{u \in W} f(u) \subset \bigcup_{u \in U} f(u) \subset X$$

we have obtained that $h \circ (f \cap g)$ is defined and has the domain W.

From the same hypothesis $h \circ f$ and $h \circ g$ are defined and have the domains U, V. Because $\emptyset \neq W \subset W_1$, the system $(h \circ f) \cap (h \circ g)$ is defined and has the domain W_1.

[2] We show a more general definition of the serial connection that was used in previous works: the request $\bigcup\limits_{u \in U} f(u) \subset X$ is replaced by $\exists u \in U, f(u) \cap X \neq \emptyset$ and $h \circ f : Z \rightarrow P^*(S^{(p)})$ is defined by

$$Z = \{u | u \in U, f(u) \cap X \neq \emptyset\},$$

$$\forall u \in Z, (h \circ f)(u) = \bigcup_{x \in f(u) \cap X} h(x).$$

$\forall u\in W$ we get

$$\bigcup_{x\in f(u)\cap g(u)} h(x)\subset \bigcup_{x\in f(u)} h(x),\quad \bigcup_{x\in f(u)\cap g(u)} h(x)\subset \bigcup_{x\in g(u)} h(x)$$

from where

$$\bigcup_{x\in f(u)\cap g(u)} h(x)\subset \bigcup_{x\in f(u)} h(x)\cap \bigcup_{x\in g(u)} h(x)$$

and we conclude that $\forall u\in W$,

$$(h\circ (f\cap g))(u)=\bigcup_{x\in f(u)\cap g(u)} h(x)\subset \bigcup_{x\in f(u)} h(x)\cap \bigcup_{x\in g(u)} h(x)=$$

$$=(h\circ f)(u)\cap (h\circ g)(u)=((h\circ f)\cap (h\circ g))(u).$$

b) The hypothesis $\bigcup\limits_{u\in U} f(u)\subset \{x|x\in X\cap X_1, h(x)\cap h_1(x)\neq\emptyset\}$ states that the domain $\{x|x\in X\cap X_1, h(x)\cap h_1(x)\neq\emptyset\}$ of $h\cap h_1$ is non-empty and that $(h\cap h_1)\circ f$ is defined. From the hypothesis we infer that $\bigcup\limits_{u\in U} f(u)\subset X$, $\bigcup\limits_{u\in U} f(u)\subset X_1$ and $h\circ f, h_1\circ f$ are defined themselves. The domain of $(h\cap h_1)\circ f$ is U. Moreover from $\forall u\in U, \forall x\in f(u), h(x)\cap h_1(x)\neq\emptyset$ we conclude that the domain $\{u|u\in U, \bigcup\limits_{x\in f(u)} h(x)\cap \bigcup\limits_{x\in f(u)} h_1(x)\neq\emptyset\}$ of $(h\circ f)\cap (h_1\circ f)$ is equal with U too.

Let $u\in U$ be arbitrary and fixed. From

$$\bigcup_{x\in f(u)} (h\cap h_1)(x)\subset \bigcup_{x\in f(u)} h(x),\quad \bigcup_{x\in f(u)} (h\cap h_1)(x)\subset \bigcup_{x\in f(u)} h_1(x)$$

we get

$$\bigcup_{x\in f(u)} (h\cap h_1)(x)\subset \bigcup_{x\in f(u)} h(x)\cap \bigcup_{x\in f(u)} h_1(x)$$

and eventually we obtain

$$((h\cap h_1)\circ f)(u)=\bigcup_{x\in f(u)} (h\cap h_1)(x)\subset$$

$$\subset \bigcup_{x\in f(u)} h(x)\cap \bigcup_{x\in f(u)} h_1(x)=(h\circ f)(u)\cap (h_1\circ f)(u)=((h\circ f)\cap (h_1\circ f))(u).$$

□

Theorem 24. *We have the systems* $f:U\to P^*(S^{(n)})$, $g:V\to P^*(S^{(n)})$, $U,V\in P^*(S^{(m)})$ *and* $h:X\to P^*(S^{(p)})$, $h_1:X_1\to P^*(S^{(p)})$, $X,X_1\in P^*(S^{(n)})$.

a) We presume that $\bigcup\limits_{u\in U} f(u)\subset X$, $\bigcup\limits_{u\in V} g(u)\subset X$; *the set* $U\cup V$ *is the common domain of* $h\circ (f\cup g)$, $(h\circ f)\cup (h\circ g)$ *and the next equality is true*

$$h\circ (f\cup g)=(h\circ f)\cup (h\circ g).$$

b) If $\bigcup_{u\in U} f(u)\subset X$, $\bigcup_{u\in U} f(u)\subset X_1$ *then* $(h\cup h_1)\circ f$, $(h\circ f)\cup(h_1\circ f)$ *have the domain* U *and*

$$(h\cup h_1)\circ f=(h\circ f)\cup(h_1\circ f).$$

Proof. a) The systems $h\circ f$ and $h\circ g$ are defined from the hypothesis and because (see the proof of Theorem 18)

$$\bigcup_{u\in U\cup V}(f\cup g)(u)=\bigcup_{u\in U} f(u)\cup\bigcup_{u\in V} g(u)\subset X$$

we infer that $h\circ(f\cup g)$ is defined too. The common domain of $h\circ(f\cup g)$ and $(h\circ f)\cup(h\circ g)$ is $U\cup V$.

Let $u\in U\cup V$ be arbitrary. We can prove the statement of the theorem in the three cases: $u\in(U\setminus V)$, $u\in(V\setminus U)$, $u\in(U\cap V)$. For example in the last case we have:

$$(h\circ(f\cup g))(u)=\bigcup_{x\in(f\cup g)(u)} h(x)=\bigcup_{x\in f(u)\cup g(u)} h(x)=$$

$$=\bigcup_{x\in f(u)} h(x)\cup\bigcup_{x\in g(u)} h(x)=(h\circ f)(u)\cup(h\circ g)(u)=((h\circ f)\cup(h\circ g))(u).$$

b) The hypothesis implies $\bigcup_{u\in U} f(u)\subset X\cup X_1$ thus $(h\cup h_1)\circ f$ is defined and on the other hand $h\circ f$ and $h_1\circ f$ are defined too. The systems $(h\cup h_1)\circ f$, $(h\circ f)\cup(h_1\circ f)$ have the same domain $U=U\cup U$.

For any $u\in U$ fixed, we have

$$f(u)=f(u)\cap(X\cup X_1)=f(u)\cap((X\setminus X_1)\cup(X_1\setminus X)\cup(X\cap X_1))=$$
$$=(f(u)\cap(X\setminus X_1))\cup(f(u)\cap(X_1\setminus X))\cup(f(u)\cap(X\cap X_1))$$

thus

$$((h\cup h_1)\circ f)(u)=\bigcup_{x\in f(u)}(h\cup h_1)(x)=$$

$$=\bigcup_{x\in(f(u)\cap(X\setminus X_1))\cup(f(u)\cap(X_1\setminus X))\cup(f(u)\cap(X\cap X_1))}(h\cup h_1)(x)=$$

$$=\bigcup_{x\in f(u)\cap(X\setminus X_1)}(h\cup h_1)(x)\cup\bigcup_{x\in f(u)\cap(X_1\setminus X)}(h\cup h_1)(x)\cup\bigcup_{x\in f(u)\cap X\cap X_1}(h\cup h_1)(x)=$$

$$=\bigcup_{x\in f(u)\cap(X\setminus X_1)} h(x)\cup\bigcup_{x\in f(u)\cap(X_1\setminus X)} h_1(x)\cup\bigcup_{x\in f(u)\cap X\cap X_1} h(x)\cup\bigcup_{x\in f(u)\cap X\cap X_1} h_1(x)=$$

$$=\bigcup_{x\in(f(u)\cap(X\setminus X_1))\cup(f(u)\cap X\cap X_1)} h(x)\cup\bigcup_{x\in(f(u)\cap(X_1\setminus X))\cup(f(u)\cap X\cap X_1)} h_1(x)=$$

$$=\bigcup_{x\in f(u)\cap X} h(x)\cup\bigcup_{x\in f(u)\cap X_1} h_1(x)=\bigcup_{x\in f(u)} h(x)\cup\bigcup_{x\in f(u)} h_1(x)=$$
$$=(h\circ f)(u)\cup(h_1\circ f)(u)=((h\circ f)\cup(h_1\circ f))(u).$$

□

11 Final remarks

The intersection and the union of the systems are dual concepts and their properties, as expressed by the previous theorems, are similar.

On the other hand, let us remark the roots of our interests in the Romanian mathematical literature represented by the works in schemata with contacts and relays from the 50's and the 60's of Grigore Moisil. Modeling is different there, but the modelled switching phenomena are exactly the same like ours.

References

[1] VLAD S.E. *Real Time Models of the Asynchronous Circuits.* The Delay Theory in New Developments in Computer Science Research, Susan Shannon (Editor), Nova Science Publishers, Inc., 2005

Oradea, Romania
Received September 21, 2006
E-mail: *serban_e_vlad@yahoo.com*

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 23–34
ISSN 1024–7696

LiScNLE – a Matlab package for some nonlinear partial differential evolution equations

Titus Petrila, Damian Trif

Abstract. We will present a MATLAB package for nonlinear evolution equations, based on the *Lyapunov-Schmidt* (LS) method. The eigenfunctions basis of the linear part is used to represent the solution at every time level (or for every value of the parameters in the case of bifurcation analysis). These eigenfunctions are calculated in a preprocessing stage or are given by the user.

Mathematics subject classification: 34B15, 49M15, 65M70.
Keywords and phrases: Lyapunov-Schmidt method, evolution equations, MATLAB package.

1 Introduction

Much of the work of an engineer or scientist is that of formulating suitable mathematical models for a particular physical system. For a dynamical system in continuous time, the model is often some system of ordinary differential equations or partial differential equations. When formulating such models, one of the goals is to maximize qualitative correctness in representing the dynamics of the physical system. However, in many cases, correctly representing the dynamics is not the sole objective in formulating mathematical models. In particular, the model needs to be useful for its intended application. For example, if a model is required in some kind of real-time feedback control scheme, then a model that is computationally intensive may be unsuitable for this purpose. We may wish to sacrifice some of the correctness of the model in order to make the equations easier to solve or to allow faster computation of the trajectories. In other words, given a model of a dynamical system that is known to correctly represent the system dynamics, how do we formulate a model of reduced complexity which retains as much of the original predictive capability as possible?

Many of the mathematicians of the twentieth century devoted their efforts to studying boundary value problems for *linear* differential equations. However, in many cases problems arising in biology, mechanics, chemistry, may be seen as non-linear perturbations of linear ones. All these can be represented in the abstract form $Lu = Nu$ where $L : X \to Y$ (linear) and $N : Y \to Y$ (nonlinear) are suitable operators between Banach spaces X, Y where $X \subset Y$ compactly.

When L is invertible, $Lu = Nu$ can be rewritten as a fixed point equation $u = [L^{-1}N]u$. In case $L^{-1} : Y \to X$ and $N : Y \to Y$ are continuous and carry bounded sets into bounded sets, $L^{-1}N : Y \to Y$ is completely continuous. Thus, the Schauder's Fixed Point Theorem (which extends the well-known Brower's Fixed Point Theorem to completely continuous operators on infinite dimensional Banach spaces) could be used in the treatment of such problems.

Schauder's Fixed Point Theorem had little impact outside the scope of nonlinear perturbations of *invertible* operators. Often we must treat some problems where the equation is a nonlinear perturbation of a linear operator with *nontrivial* kernel (problems at resonance). A useful tool for studying such type of problems is the *Lyapunov-Schmidt reduction method*.

The Lyapunov-Schmidt (LS) method, elaborated in the years 1906-1908 and reformulated in a modern mathematical language by L. Cesari [1] after 1963 applies to some nonlinear equations of the type $Lu = Nu$, in the presence of boundary conditions, considered on the domain of the linear operator L.

As a simple example (following [2]), let us consider the problem

$$\begin{gathered} -u'' - \alpha u' - \lambda_1(\alpha)u + g(u) = 0, \quad t \in [0, \pi], \\ u(0) = u(\pi) = 0 \end{gathered} \tag{1}$$

where α is a given real number, $\lambda_1(\alpha) = 1 + \alpha^2/4$ is the first eigenvalue of the linear problem

$$\begin{gathered} -u''(t) - \alpha u'(t) = \lambda u(t), \quad t \in [0, \pi], \\ u(0) = u(\pi) = 0 \end{gathered}$$

and g is a continuous and $T-$periodic function with zero mean.

In order to apply the Lyapunov-Schmidt reduction method, we consider the linear differential operator

$$L : W_0^{2,1}(0, \pi) \to L^1(0, \pi), \quad Lu = -u'' - \alpha u' - \lambda_1(\alpha)u$$

and the Nemytskii operator

$$N : W_0^{2,1}(0, \pi) \to L^1(0, \pi), \quad Nu(t) = -g(u(t)), \quad \forall t \in [0, \pi]$$

so that (1) is equivalent to the operator equation $Lu = Nu$.

It is well known that L is a linear Fredholm operator of zero index, $\ker L = sp(\varphi)$, $im(L) = \psi^\perp$, where

$$\varphi(t) = \frac{e^{-\frac{\alpha}{2}t} \sin t}{\sqrt{\int_0^\pi \left(e^{-\frac{\alpha}{2}s} \sin s\right)^2}}, \quad \psi(t) = \frac{e^{\frac{\alpha}{2}t} \sin t}{\sqrt{\int_0^\pi \left(e^{\frac{\alpha}{2}s} \sin s\right)^2}}, \quad t \in [0, \pi].$$

The splitting $W_0^{2,1}(0, \pi) = sp(\varphi) \oplus \varphi^\perp$ leads us to rewrite any element $u \in W_0^{2,1}(0, \pi)$ as $u = \widetilde{u} + \overline{u}\varphi$, where $\overline{u} \in \mathbb{R}$ and $\widetilde{u} \in \varphi^\perp$ and to observe that $L : \varphi^\perp \to \psi^\perp$

is a topological isomorphism. Let us denote $K : \psi^{\perp} \to \varphi^{\perp}$ the inverse of this isomorphism and define the projection

$$Q : L^1(0,\pi) \to L^1(0,\pi), \quad h \mapsto \left(\int_0^{\pi} h(s)\psi(s)ds \right) \psi.$$

This way, equation $Lu = Nu$ becomes equivalent to the Lyapunov-Schmidt system

$$\widetilde{u} = K(I-Q)N(\widetilde{u} + \overline{u}\varphi), \tag{2}$$

$$\int_0^{\pi} g(\widetilde{u}(s) + \overline{u}\varphi(s))\psi(s)ds = 0. \tag{3}$$

From the *auxiliary equation* (2) we observe that, being N bounded and K compact, the Schauder Fixed Point Theorem implies the existence, for any $\overline{u} \in \mathbb{R}$, of the fixed point $\widetilde{u}(\overline{u}) \in \varphi^{\perp}$. Consequently, the *bifurcation equation* (3) becomes an equation for $\overline{u} \in \mathbb{R}$,

$$\int_0^{\pi} g(\widetilde{u}\left(\overline{u}\right)(s) + \overline{u}\varphi(s))\psi(s)ds = 0.$$

and the solvability of $Lu = Nu$ comes from the solvability of this one-dimensional equation.

This method could be easily extended to the case of a nonlinear evolution equation on a Hilbert space H (usually an L^2 space) of the form $\frac{du}{dt} = F(u) \equiv Lu + Nu$ where the domain of F is dense in H. We assume that $\{\varphi_i, i = 0, 1, \dots\}$ forms a complete orthogonal basis for H (for example the eigenfunctions of L).

Fix $m \in \mathbb{N}$ and let $P \equiv P_m : H \to X_m \equiv X$ be the orthogonal projection from H onto the finite dimensional subspace spanned by $\{\varphi_1, \dots, \varphi_m\}$. Let $Q \equiv Q_m = (I-P) : H \to Y \equiv Y_m$ be the complementary orthogonal projection.

Given $u \in H$, let $Pu = p$ and $Qu = q$. The equation can be rewritten as

$$\frac{dp}{dt} = PF(p,q), \tag{4}$$

$$\frac{dq}{dt} = QF(p,q). \tag{5}$$

The strategy is fairly simple: study the dynamics of the low dimensional Galerkin projection (4) (where $q = q(p)$ from (5)) to draw conclusions about the dynamics of the given equation.

Although it has been used for a long time only for the theoretical demonstration of the existence and branching of the solutions of such problems, the LS method (or the *alternative method*, following Cesari) is also very useful for the effective approximation of these solutions.

We will present *LiScNLE*, a MATLAB package for dynamical systems, based on the LS method. The eigenfunctions basis of the linear part L of the system is used to represent the solution at every time level, or for every value of the parameters in the case of bifurcation analysis. These eigenfunctions are calculated in a preprocessing stage [3] or are given by the user. Also, other functions could be used as basis. The package extends a preliminary steady version [4].

The advantage of the LS method consists of the important reduction of the dimension of the nonlinear system to be solved together with the possibility to oversee the approximating errors. This advantage can be remarked in some examples, which prove that the LS method behaves better than other known methods, such as `bvp4c` or `sbvp`.

The first two sections present the basic theory and the implementation of LiScNLE. The last two sections present examples and conclusions.

2 The LS method

We assume that the linear part L of the equation $Lu = Nu$ is a Sturm-Liouville operator

$$Ly \equiv \frac{1}{w(x)}\left[-(p(x)y')' + q(x)y\right], \quad x \in [a,b],$$

$$y(a)\cos\alpha + (py')(a)\sin\alpha = 0, \quad y(b)\cos\beta + (py')(b)\sin\beta = 0$$

where $1/p$, q, w are real-valued functions on $[a,b]$, $p(x) > 0$, $w(x) > 0$ on $[a,b]$, $p \in C^1[a,b]$, $q, w \in C[a,b]$. It is well known that the eigenvalues of L form an increasing sequence $\lambda_0 < \lambda_1 < \ldots$ converging to infinity and the corresponding eigenfunctions φ_n form an orthogonal (orthonormal) basis of the Hilbert space $L^2_w(a,b)$. We remark the asymptotic behaviour of the eigenvalues $\lambda_n \in O(n^2)$.

A theoretical but constructive variant of the LS method could be found in [5, 6]. We are looking for an approximate solution of the equation $Lu = Nu$ of the form $u = \sum_{i=1}^N c_i\varphi_i$ (eigenfunction expansion) which leads to the following equation for the unknowns c_i

$$\sum_{i=1}^N c_i L\varphi_i = N\left(\sum_{i=1}^N c_i\varphi_i\right).$$

We obtain the equation

$$\sum_{i=1}^m c_i\lambda_i\varphi_i + \sum_{i=m+1}^N c_i\lambda_i\varphi_i = N\left(\sum_{i=1}^N c_i\varphi_i\right) \tag{6}$$

where m is a positive integer, less than N. By applying the partial inverse H_m of L,

$$H_m u = \sum_{i=m+1}^N \frac{c_i}{\lambda_i}\varphi_i$$

to (6), we are led to

$$\sum_{i=m+1}^{N} c_i\varphi_i = H_m N\left(\sum_{i=1}^{N} c_i\varphi_i\right) = \sum_{i=m+1}^{N} C_i\varphi_i$$

so that we have

$$c_i = C_i(c_1, \ldots, c_N), i = m+1, \ldots, N.$$

For a sufficiently great m we may calculate $c_{m+1}, \ldots, c_N$ as functions of $c_1 \ldots c_m$, using Banach Fixed Point Theorem.

By applying the projection P_m to (6) we obtain the determining equation

$$\sum_{i=1}^{m} c_i\lambda_i\varphi_i = P_m N\left(\sum_{i=1}^{N} c_i\varphi_i\right)$$

which is a small finite dimensional system for $c_1, \ldots, c_m$.

In fact, in LS methods, the true unknowns are $c_1, \ldots, c_m$; the other coefficients $c_{m+1}, \ldots, c_N$ are calculated as coefficients of the associated fixed point.

The first version of our package applies only to the Sturm-Liouville case for the linear operator L, in the form

$$Lu = \frac{1}{w(x)}\left[\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + g(x)u\right],$$

$$au'(0) + bu(0) = 0, \quad cu'(1) + du(1) = 0.$$

There exists a Matlab package `MATSLISE` of V. Ledoux (2004) [7], based on the works of L. Ixaru which uses the so called CP methods to calculate the eigenfunctions of Sturm-Liouville or Schrodinger operators but this package works slowly. A more interesting package is `MWRtools` of R.A. Adomaitis (1998–2001) [8] which uses spectral methods to calculate the eigenfunctions of the Sturm-Liouville operator in order to solve some *linear* boundary value problems.

We remark that in the case of *Galerkin's method,* the approximating solutions are being looked for under the form $u^* = \sum_{k=1}^{N} c_k\varphi_k$, where the coefficients $c_k, k = 1, \ldots, N$, are determined from the equations $(Lu^* - Nu^*, \varphi_k) = 0,\ k = 1, \ldots, N$, i.e.

$$(\lambda_k u^* - Nu^*, \varphi_k) = 0,\ k = 1, \ldots, N.$$

These equations are got from the determining equations for $m = N$. If $m = 0$ the system of the determining equations disappears. The associate function to a certain u^* verifies the equation $y = L^{-1}Ny$, so the algorithm is reduced, in this case, to the transformation of the equation $Lu = Nu$ into a fixed point problem. Obviously, this case arises only when the inverse L^{-1} exists and $L^{-1}N$ is a contraction.

3 Implementation

In this section we propose a Chebyshev-tau method to solve the Sturm-Liouville problem in order to get a good basis φ_i and we present the corresponding Matlab package.

Let us consider the problem

$$p_2(x)u'' + p_1(x)u' + p_0(x)u = g(x) \quad x \in (a, b)\,, \tag{7}$$

$$\begin{aligned} \alpha_{11}u(x_{11}) + \alpha_{12}u'(x_{12}) &= \beta_1, \\ \alpha_{21}u(x_{12}) + \alpha_{22}u'(x_{22}) &= \beta_2 \end{aligned} \tag{8}$$

and let us suppose for the moment $a = -1$, $b = 1$. A powerful methods to solve (7) is to express u as a Chebyshev series $u(x) = c_0\dfrac{T_0(x)}{2} + c_1T_1(x) + \dots$ where $T_i(x) = \cos(i\cos^{-1}(x))$ is the standard Chebyshev polynomial of order i. For the practical implementation, we define the vectors c and t by $c^T = (c_0, c_1, c_2, \dots)$, $t^T = \left(\dfrac{T_0}{2}, T_1, T_2, \dots\right)$ so that $u(x) = c^Tt = t^Tc$.

There exists a matrix X for which $x \cdot u(x) = (Xc)^Tt$, see [9, 10],

$$X_{0,1} = 1, X_{ii} = 0, X_{i,i-1} = X_{i,i+1} = \frac{1}{2}.$$

Then, in general, $x^mu(x) = (X^mc)^Tt$ and $f(x)u(x) = (f(X)c)^Tt$ for analytical functions f, i.e.

$$f(x) = \sum_{k=0}^{\infty} f_k\frac{x^k}{k!}.$$

Moreover, $\dfrac{u(x)}{x^m} = (X^{-m}c)^T\,t$ if the l.h.s. has no singularity at the origin. Of course, X is a tri-diagonal matrix, X^2 is a penta-diagonal matrix and so on but, generally, the matrix version `funm(X)`of the scalar function $f(x)$ or $X^{-m} = [inv(X)]^m$ are no longer sparse matrices.

Similarly, let D be the differentiation matrix giving $\dfrac{d^mu}{dx^m} = (D^mc)^Tt$. D is an upper triangular matrix with $D_{ii} = 0$, $D_{ij} = 0$ for $(j - i)$ even and $D_{ij} = 2j$ otherwise.

Applying these formulae to equation (7), we get

$$(p_2(X)D^2 + p_1(X)D + p_0(X))c = g$$

where G are the coefficients of the r.h.s. function $g(x)$.

$$g(x) = g_0\frac{T_0(x)}{2} + g_1T_1(x) + \dots\ .$$

The condition (8) can be formulated in a similar manner. We define

$$t_{ij} = \left(\frac{T_0(x_{ij})}{2}, T_1(x_{ij}), T_2(x_{ij}), \dots\right)^T$$

so that it can be written in the form $h_i^T c = \beta_i, i = 1, 2$, where

$$h_i^T = \sum_{j=1}^{2} \alpha_{ij} t_{ij}^T D^{j-1}, \quad i = 1, 2.$$

Now we define the matrices $A = \sum_{i=0}^{2} P_i(X) D^i$ and $H = (h_1, h_2)^T$. Then the vector c satisfies

$$\begin{pmatrix} H \\ A \end{pmatrix} c = \begin{pmatrix} \beta \\ q \end{pmatrix} \tag{9}$$

of the form $\mathcal{A}c = b$, where $\beta = (\beta_1, \beta_2)^T$.

Of course, in reality we cannot work with infinite matrices but only with finite portions ($N \times N$) of them. For the initial conditions, we restrict t_i to have N components and use the truncation D_N instead of D, so that the computed matrix will be $\begin{pmatrix} H^* \\ A^* \end{pmatrix}$. We then take the first N rows and columns of $\begin{pmatrix} H^* \\ A^* \end{pmatrix}$ as the matrix to use, together with the first N elements of $\begin{pmatrix} \beta \\ q \end{pmatrix}$.

If we have another interval $[a, b]$ instead of $[-1, 1]$ for x, we use the change of coordinates $x = \alpha\xi + \beta$ where $\alpha = \dfrac{b-a}{2}$ and $\beta = \dfrac{b+a}{2}$ so that $\xi \in [-1, 1]$. We must change X to $\alpha X + \beta I$ and D to D/α.

`LiScNLE` (`Li`apunov-`Sc`hmidt `N`on-`L`inear `E`volution) is a Matlab package for the study of some nonlinear differential evolution equations for the unknown function $u(x, t)$

$$\frac{\partial u}{\partial t} + Lu = Nu, \quad x \in (a, b), t > 0$$

where L is a Sturm-Liouville operator

$$Lu \equiv \frac{1}{w(x)} \left[-\frac{\partial}{\partial x} \left(p(x) \frac{\partial u}{\partial x} \right) + q(x) u \right]$$

and Nu is a nonlinear (differential) operator

$$Nu \equiv N(x, u, \frac{\partial u}{\partial x}).$$

We have also boundary value conditions

$$a_{11} u(a, t) + a_{12} \frac{\partial u}{\partial x}(a, t) = 0,$$

$$a_{12} u(b, t) + a_{22} \frac{\partial u}{\partial x}(b, t) = 0$$

and initial condition $u(x, 0) = u_0(x)$.

We perform a time semi-discretization by Crank-Nicolson method (for example)

$$\frac{u^{n+1} - u^n}{\Delta t} = \frac{1}{2} \left(-Lu^{n+1} + Nu^{n+1} - Lu^n + Nu^n \right)$$

i.e.

$$u^{n+1} + \frac{\Delta t}{2} L u^{n+1} = \frac{\Delta t}{2} N u^{n+1} + u^n - \frac{\Delta t}{2} L u^n + \frac{\Delta t}{2} N u^n$$

where

$$u^n = u(x, n\Delta t),\ \ u^0 = u_0(x)$$

and Δt is the time step. For each n, this problem is of the form

$$\mathcal{L} u^{n+1} = \mathcal{N} u^{n+1}$$

where

$$\mathcal{L} u = \left(I + \frac{\Delta t}{2} L \right) u,$$

$$\mathcal{N} u = \frac{\Delta t}{2} N u + F, \quad F = u^n - \frac{\Delta t}{2} L u^n + \frac{\Delta t}{2} N u^n$$

so that the numerical steady Lyapunov-Schmidt method *LiScNLS* [4] could be applied.

Remark 1. If we have a second order in time equation,

$$\frac{\partial^2 u}{\partial t^2} + L u = N u, \quad x \in (a, b),\ t > 0,$$

with the same boundary conditions and initial conditions

$$u(x, 0) = u_0(x), \quad u_t(x, 0) = v_0(x),$$

the Crank-Nicolson discretization looks like

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2} = \frac{1}{2} \left(-L u^{n+1} + N u^{n+1} - L u^{n-1} + N u^{n-1} \right)$$

i.e.

$$u^{n+1} + \frac{\Delta t^2}{2} L u^{n+1} = \frac{\Delta t^2}{2} N u^{n+1} + u^{n-1} - \frac{\Delta t^2}{2} L u^{n-1} + \frac{\Delta t^2}{2} N u^{n-1} + 2(u^n - u^{n-1}).$$

Remark 2. The backward-Euler method is

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2} = -L u^{n+1} + N u^{n+1}$$

i.e.

$$u^{n+1} + \Delta t^2 L u^{n+1} = \Delta t^2 N u^{n+1} + 2u^n - u^{n-1}.$$

If $\lambda_k, \varphi_k,\ \ k = 1, 2, \ldots,$ are the eigenvalues and the eigenfunctions of the Sturm-Liouville operator L, then $1 + \dfrac{\Delta t}{2} \lambda_k, \varphi_k,\ \ k = 1, 2, \ldots,$ are the eigenvalues and the eigenfunctions of the operator $\mathcal{L}$. Let us suppose that we know the first n eigenfunctions and eigenvalues of $\mathcal{L}$,

$$\mathcal{L}\Phi_k = \lambda_k \Phi_k, k = 1, \ldots, n$$

where

$$\int_a^b \Phi_k \Phi_j w dx = \delta_{kj}, \quad k, j = 1, \ldots, n.$$

Then, we search for the solution of the nonlinear steady problem

$$\mathcal{L}u = \mathcal{N}u \tag{10}$$

of the form (see [5],[6] for the hypotheses on $\mathcal{L}$ and $\mathcal{N}$)

$$u = \sum_{i=1}^{n} c_i \Phi_i = \Phi \cdot c$$

The nonlinear part is

$$\mathcal{N}(u) = \mathcal{N}\left(\sum_{i=1}^{n} c_i \Phi_i\right) = \sum_{i=1}^{n} C_i \Phi_i$$

where

$$C_i = \int_a^b \mathcal{N}(u) \cdot \Phi_i \cdot w dx, \quad i = 1, \ldots, n.$$

Let us choose an index m and project the equation $\mathcal{L}u = \mathcal{N}u$ on $span\,\{\Phi_{m+1}, \ldots, \Phi_n\}$, i.e.

$$c_i = \frac{1}{\lambda_i} C_i(c_1, \ldots, c_n), \quad i = m+1, \ldots, n. \tag{11}$$

For a sufficiently great m, for fixed $c_1, \ldots, c_m$, the above operator becomes a contraction so we can iterate until a fixed point

$$c^* = \left(c_1, \ldots, c_m, c^*_{m+1}, \ldots, c^*_n\right)$$

which is a solution of the equations (11). Of course, $c^*_i,\ i = m+1, \ldots, N$, depend on $c_i, i = 1, \ldots, m$.

Now we project the equation $\mathcal{L}u = \mathcal{N}u$ on $span\,\{\Phi_1, \ldots, \Phi_m\}$, i.e.

$$\lambda_i c_i = C_i(c_1, \ldots, c_m, c^*_{m+1}, \ldots, c^*_n) \tag{12}$$

which represents a nonlinear algebraic system for $c_1, \ldots, c_m$. Given $c_1, \ldots, c_m$, each evaluation of $C_i\left(c_1, \ldots, c_m, c^*_{m+1}, \ldots, c^*_n\right)$ means the fixed point iterations (11). We solve this system by a Newton method and finally we obtain the solution

$$c^* \equiv \left(c^*_1, \ldots, c^*_m, c^*_{m+1}, \ldots, c^*_n\right)$$

(i.e. $u = \Phi \cdot c^*$) of the problem (10).

This problem has a natural extension for a nonlinear part of the form $N(x, u(x), u'(x))$, that is

$$N(x, \sum_{i=1}^{n} c_i \Phi_i, \sum_{i=1}^{n} c_i \Phi_i')$$

The main function `of` *LiScNLE* is the function `evol` for the first order (in time) problems:

```
function [lam,phi,phip,x,C,kod]=...
evol(n,errtol,Lfun,m,Nfun,ICfun,dt,K,scene)
```

Here `n` is the dimension of the discretized problem, `errtol` is the tolerance used in the stopping criteria, `Lfun` describes the linear part of the equation (see *LiScEig Tutorial* [3]), `m` is the truncation parameter.

The nonlinear r.h.s. of the problem (10) is coded in `Nfun (see` *LiScNLS Tutorial* `[4]),ICfun` describes the initial condition $u_0(x)$, `dt` is the time step, `K` is the number of time steps to be performed and `scene` is used for the plot of the solution.

For the second order (in time) problems the corresponding file is `evol2`.

The output parameters of `evol` are:

`lam` – the eigenvalues of the linear part

`phi, phip` – the eigenfunctions and their derivatives

`x` – the grid

`C` – the coefficients of the numerical solution with respect to the eigenfunctions `phi`, column $n + 1$ for the $n-$ time level, $n = 0, 1, \ldots, K$.

`kod` – indicates the status of the solution.

We remark that, given the coefficients C of the solution with respect to the eigenfunctions `phi`, the values of the solutions at the Legendre grid points x are $phi * C$ and a plot of the solution could be obtained using the command `plot(x,phi*C);`

More details about the implementation could be found in the tutorial of LiScNLE [11].

4 Examples

The tutorial of *LiScNLE* [11] contains many difficult examples:

– the Burgers equation, which exhibits a near shock,

– a steady solution of a nonlinear reaction-difusion problem,

– the blowing-up behavior for a forced heat equation,

– the Allen-Cohn equation,

– periodical stable and unstable solutions for Kuramoto-Sivashinski equation,

– a moving step solution for Fisher equation,

– an example from electrodynamics (system, also in MATLAB demo),

– the sine-Gordon equation (second order in time).

Let us present here the Fisher equation example,

$$u_t - u_{xx} = u(1-u), x \in \mathbb{R}, \mathrm{t} > 0,$$
$$u(t,-\infty) = 1, u(t,\infty) = 0,$$
$$u(0,x) = \frac{1}{\left(1+e^{\frac{x}{\sqrt{6}}}\right)^2}$$

with the exact solution

$$u(t,x) = \frac{1}{\left(1+e^{-\frac{5t}{6}+\frac{x}{\sqrt{6}}}\right)^2}.$$

First, the spatial domain is truncated to $[-50, 50]$. Next, the boundary values are homogenized by using the function $u_0(x) = (50-x)/100$. The basic command is

```
[lam,phi,phip,x,C,kod]=evol(256,1.e-5,'LFisher',...
  0,'NFisher','ICFisher',0.01,1000,[-50 50 0 1]);
```

The absolute error is $2.5x10^{-3}$ in the region of the step and about 10^{-5} in general, due to truncation of the spatial domain.

5 Conclusions

The comparison between `LiScNLE` and `SBVP 1.0` of Auzinger [12] or `bvp4c` of Matlab (see Matlab help) shows an important reduction of the computing time for LiScNLE.. The Matlab profile reports show that about 75% of the computing time was spent on computation of the eigenfunctions and only about 6% on the effective calculations of the numerical solution. We have good reasons to use LS method.

1. We can build a database with known eigenfunctions.

2. In the problems with parameters, where we have (for example) bifurcations, or in evolution problems, we can use repeatedly the same eigenfunctions.

3. The eigenfunctions carry physical information, so that our LS solution has a better structure for studies.

4. LS method could be easily extended to 2D or 3D (evolution) problems, with non-invertible linear part.

5. In all the cases, we finally have to solve a very small nonlinear system, usually with $m = 0, 1, 2$, values which also carry information about bifurcation behaviour.

References

[1] Cesari L. *Functional Analysis and Galerkin's Method*. Mich. Math. J., 1964, **11**, N 3, p. 383–414.

[2] Cañada A., Ureña A.J. *Asymptotic Behaviour of the Solvability Set for Pendulum Type Equations with Linear Dampuing and Homogeneous Dirichlet Conditions*. Electron. J. Diff. Eqns., Conf. 06, 2001, p. 55–64, http://ejde.math.unt.edu.

[3] Trif D. *LiScEig Tutorial 2005*. MATLAB Central > File Exchange > Mathematics > Differential Equations > LiScEig 1.0, http://www.mathworks.nl/matlabcentral/fileexchange.

[4] Trif D. *LiScNLS Tutorial 2005*. MATLAB Central > File Exchange > Mathematics > Differential Equations > LiScNLS 1.0, http://www.mathworks.nl/matlabcentral/fileexchange.

[5] Petrila T., Trif D. *Basics of Fluid Mechanics and Introduction to Computational Fluid Dynamics*. Springer, 2005.

[6] Trif D. *The Lyapunov-Schmidt method for two-point boundary value problems*. Fixed Point Theory, 2005, **6**, N 1, p. 119–132.

[7] Ledoux V. *MATSLISE package*. http://users.ugent.be/~vledoux/MATSLISE/.

[8] Adomaitis R.A., Lin Yi-hung. *A Collocation/Quadrature Based Sturm-Liouville Problem Solver*. ISR Technical Research Report TR 99-01.

[9] MacLeod A.J. *An automatic matrix approach to the Chebyshev series solution of differential equations*. http://maths.paisley.ac.uk/allanm/amcltech.htm.

[10] Liefvendahl M. *A Chebyshev tau spectral method for the calculation of Eigenvalues and pseudospectra*. http://www.nada.kth.se/~mli/research.html.

[11] Trif D. *LiScNLE Tutorial 2006*. MATLAB Central > File Exchange > Mathematics > Differential Equations > LiScNLE 1.0, http://www.mathworks.nl/matlabcentral/fileexchange.

[12] Auzinger W., Kneisl G., Koch O., Weinmuller E.B. *SBVP 1.0 A Matlab solver for singular boundary value problems*. Technische Universitat Wien, ANUM preprint no. 02/02.

Titus Petrila
Department of Applied Mathematics
Babes-Bolyai University of Cluj-Napoca
Romania

E-mail: *tituspetrila@yahoo.com*

Received September 21, 2006

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 35–48
ISSN 1024–7696

A characterization of the solutions of the Darboux Problem for third order hyperbolic inclusions

Georgeta Teodoru

Abstract. In this paper we consider the Darboux Problem for a third order hyperbolic inclusion of the form $u_{xyz} \in F(x, y, z, u)$ and we prove a characterization of the solutions of the considered problem using the Aumann integral defined for multifunctions.

Mathematics subject classification: 35L30, 35R70, 47H10.
Keywords and phrases: Multifunction, hyperbolic inclusion, upper semi-continuity, initial values, absolutely continuous in Carathéodory's sense function, Aumann integral.

1 Introduction

In this paper we consider the Darboux Problem for a third order hyperbolic inclusion of the form

$$\frac{\partial^3 u(x, y, z)}{\partial x\, \partial y\, \partial z} \in F(x, y, z, u)\,,\ (x, y, z) \in D = [0, a] \times [0, b] \times [0, c]\,,\ u \in \Omega \subset \mathbb{R}^n \tag{1.1}$$

with initial values

$$\begin{cases} u(x, y, 0) = \varphi(x, y)\,, & (x, y) \in D_1 = [0, a] \times [0, b]\,, \\ u(0, y, z) = \psi(y, z)\,, & (y, z) \in D_2 = [0, b] \times [0, c]\,, \\ u(x, 0, z) = \chi(x, z)\,, & (x, z) \in D_3 = [0, a] \times [0, c] \end{cases} \tag{1.2}$$

where φ, ψ, χ are absolutely continuous in Carathéodory's sense functions $[2, \S 565 - 570]$, $\varphi \in C^*(D_1; \mathbb{R}^n)$, $\psi \in C^*(D_2; \mathbb{R}^n)$, $\chi \in C^*(D_3; \mathbb{R}^n)$ and they satisfy the conditions

$$\begin{cases} u(x, 0, 0) = \varphi(x, 0) = \chi(x, 0) = v^1(x)\,, & x \in [0, a]\,, \\ u(0, y, 0) = \varphi(0, y) = \psi(y, 0) = v^2(y)\,, & y \in [0, b]\,, \\ u(0, 0, z) = \psi(0, z) = \chi(0, z) = v^3(z)\,, & z \in [0, c]\,, \\ u(0, 0, 0) = v^1(0) = v^2(0) = v^3(0) = v^0, & \end{cases} \tag{1.3}$$

where $F : D \times \Omega \to 2^{\mathbb{R}^n}$ is a multifunction with compact, convex and non-empty values and $\Omega \subset \mathbb{R}^n$ is an open subset.

Under suitable assumptions, we proved in [16] an existence theorem for a local solution of the Darboux Problem (1.1) + (1.2) and that the set of its solutions is compact in Banach space $C(D_0;\mathbb{R}^n)$, $D_0=[0,x_0]\times[0,y_0]\times[0,z_0]\subseteq D$; moreover, as a function of the initial values this set defines an upper semi-continuous multifunction.

In [17] we proved a theorem of prolongation for the solutions of the considered problem and also an existence theorem for a saturated solution.

In this paper we prove a characterization of the solutions of Darboux Problem (1.1) + (1.2) using the Aumann integral defined for multifunctions.

This study has been suggested by [15] and it provides an extension of the results in that article.

2 Preliminaries

The definitions and Theorems 2.1–2.5 plus Propositions 2.1–2.4 in this section are taken from [1, 2, 5–14].

Definition 2.1. Let X and Y be two non-empty sets. A *multifunction* $\Phi : X \to 2^Y$ is a function from X into the family of all non-empty subsets of Y.

To each $x \in X$, a subset $\Phi(x)$ of Y is associated by the multifunction Φ. The set $\bigcup\limits_{x\in X} \Phi(x)$ is the *range* of Φ. $\Phi(X) = \{\bigcup \Phi(x) \mid x \in X\}$.

Definition 2.2. Let us consider $\Phi : X \to 2^Y$.

a) If $A \subset X$, the *image* of A by Φ is $\Phi(A) = \bigcup\limits_{x\in A} \Phi(x)$;

b) If $B \subset Y$, the *counterimage* of B by Φ is

$$\Phi^-(B) = \{x \in X \mid \Phi(x) \cap B \neq \emptyset\};$$

c) The *graph* of Φ, denoted graph Φ, is the set

$$\operatorname{graph}\Phi = \{(x,y) \in X \times Y \mid y \in \Phi(x)\}.$$

Definition 2.3. Let us now take $\Phi : X \to 2^Y$. An element $x \in X$ with the property $x \in \Phi(x)$ is called a *fixed point* of the multifunction Φ.

Definition 2.4. A univalued function $\varphi : X \to Y$ is said to be a *selection* of $\Phi : X \to 2^Y$ if $\varphi(x) \in \Phi(x)$ for all $x \in X$.

Definition 2.5. Let X and Y be two topological spaces. The multifunction $\Phi : X \to 2^Y$ is *upper semi-continuous* if, for any closed $B \subset Y$, $\Phi^-(B)$ is closed in X.

Definition 2.6. If $(X,\mathcal{F})$ is a measurable space and Y is a topological space, the multifunction $\Phi : X \to 2^Y$ is *measurable* if $\Phi^-(B) \in \mathcal{F}$ for every closed subset $B \subset Y$, $\mathcal{F}$ being the σ-algebra of the measurable sets of X, i.e. $\Phi^-(B)$ is measurable.

Theorem 2.1 [13]. *Let X and Y be two metric spaces, Y compact and $\Phi : X \to 2^Y$ a multifunction with the property that $\Phi(x)$ is a closed subset of Y for any $x \in X$. The following assertions are equivalent:*

i) the multifunction Φ *is upper semi-continuous;*

ii) the graph of Φ *is a closed subset of* $X \times Y$*;*

iii) any would be the seguences $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$*, from* $x_n \to x$*,* $y_n \in \Phi(x_n)$ *and* $y_n \to y$ *it follows that* $y \in \Phi(x)$*.*

Definition 2.7 [2, 7, 8]. The function $u : \triangle \to \mathbb{R}^n, \triangle \subset \mathbb{R}^2$, is *absolutely continuous in Carathéodory's sense* $[2, \S 565 - 570]$ if and only if it is continuous on $\triangle$, absolutely continuous in x (for any y), absolutely continuous in y (for any x), $u_x(x,y)$ is (possibly after a suitable definition on a two-dimensional set of zero measure) absolutely continuous in y (for any x) and u_{xy} is Lebesgue-integrable on $\triangle$.

Theorem 2.2 [2, 6, 14]. *The function* $u : \triangle \to \mathbb{R}^n$*,* $\triangle = [0,a] \times [0,b] \subset \mathbb{R}^2$*, is absolutely continuous in Carathéodory's sense on* $\triangle$ *if and only if there exist* $f \in L^1(\triangle; \mathbb{R}^n)$*,* $g \in L^1([0,a]; \mathbb{R}^n)$*,* $h \in L^1([0,b]; \mathbb{R}^n)$ *such that*

$$u(x,y) = \int_0^x \int_0^y f(s,t)\, ds\, dt + \int_0^x g(s)\, ds + \int_0^y h(t)\, dt + u(0,0).$$

We denote the class of absolutely continuous functions in Carathéodory's sense by $C^*(\triangle; \mathbb{R}^n)$ [7, 8]. In [6], this space is denoted by $AC(\triangle; \mathbb{R}^n)$.

Theorem 2.3 [6]. *The space* $C^*(\triangle; \mathbb{R}^n)$ *endowed with the norm*

$$\|u(\cdot,\cdot)\| = \int_0^a \int_0^b \|u_{xy}(s,t)\|\, ds\, dt + \int_0^a \|u_x(s,0)\|\, ds + \int_0^b \|u_y(0,t)\|\, dt + \|u(0,0)\|,$$

$\triangle = [0,a] \times [0,b] \subset \mathbb{R}^2$*, where* $\|\cdot\|$ *is the Euclidean norm, is a Banach space.*

Definition 2.8 [2, 9]. The function $u : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^3$, is absolutely *continuous in Carathéodory's sense* $[2, \S 565 - 570]$ if and only if $u(x,y,z)$ is continuous on D, absolutely continuous in each variable (for any pair of the other two variables) and similarly for $u_x(x,y,z)$, $u_y(x,y,z)$, $u_z(x,y,z)$, $u_{xy}(x,y,z)$, $u_{yz}(x,y,z)$, $u_{xz}(x,y,z)$, and u_{xyz} is Lebesgue-integrable on D.

Theorem 2.4 [6]. *The function* $u : D \to \mathbb{R}^n$*,* $D = [0,a] \times [0,b] \times [0,c] \subset \mathbb{R}^3$*, is absolutely continuous in Carathéodory's sense on* D *if and only if there exist* $f \in L^1(D; \mathbb{R}^n)$*,* $g_1 \in L^1(D_1; \mathbb{R}^n)$*,* $g_2 \in L^1(D_2; \mathbb{R}^n)$*,* $g_3 \in L^1(D_3; \mathbb{R}^n)$*,* $h_1 \in L^1([0,a]; \mathbb{R}^n)$*,* $h_2 \in L^1([0,b]; \mathbb{R}^n)$*,* $h_3 \in L^1([0,c]; \mathbb{R}^n)$*, such that*

$$\begin{aligned} u(x,y,z) = &\int_0^x \int_0^y \int_0^z f(r,s,t)\, dr\, ds\, dt + \int_0^x \int_0^y g_1(r,s)\, dr\, ds + \\ &+ \int_0^y \int_0^z g_2(s,t)\, ds\, dt + \int_0^x \int_0^z g_3(r,t)\, dr\, dt + \\ &+ \int_0^x h_1(r)\, dr + \int_0^y h_2(s)\, ds + \int_0^z h_3(t)\, dt + u(0,0,0). \end{aligned}$$

We denote the class of absolutely continuous functions in Carathéodory's sense on D by $C^*(D;\mathbb{R}^n)$ [9].

Theorem 2.5 [6]. *The space $C^*(D;\mathbb{R}^n)$ endowed with the norm*

$$\begin{aligned}\|u(\cdot,\cdot,\cdot)\| = &\int_0^a\int_0^b\int_0^c \|u_{xyz}(r,s,t)\|\,dr\,ds\,dt + \int_0^a\int_0^b \|u_{xy}(r,s,0)\|\,dr\,ds+\\ &+\int_0^b\int_0^c \|u_{yz}(0,s,t)\|\,ds\,dt + \int_0^a\int_0^c \|u_{xz}(r,0,t)\|\,dr\,dt+\\ &+\int_0^a \|u_x(r,0,0)\|\,dr + \int_0^b \|u_y(0,s,0)\|\,ds+\\ &+\int_0^c \|u_z(0,0,t)\|\,dt + \|u(0,0,0)\|,\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm, is a Banach space.

We denote by $d(x,y)$ the Euclidean distance from x to y, $x,y\in\mathbb{R}^n$, $\mathbb{R}^n$ is the Euclidean space. If $A\subset\mathbb{R}^n$, $d(x,A)=\inf\{d(x,y)\mid y\in A\}$.

$B[x,r]$ is the open ball of radius $r>0$ centered at $x\in\mathbb{R}^n$, $\operatorname{Conv} A$ is the convex covering of $A\subset\mathbb{R}^n$ and

$$|A| = \sup\{\|\zeta\| \mid \zeta\in A\}.$$

$\mathcal{C}(\mathbb{R}^n)$ is the set of compact and non-empty subsets of $\mathbb{R}^n$. Similarly with [1, 5, 15], we define the Aumann integral for multifunctions of three variables.

Definition 2.9. Let $D=[0,a]\times[0,b]\times[0,c]\subset\mathbb{R}^3$. For each $(x,y,z)\in D$, let $H(x,y,z)$ be a non-empty subset of $\mathbb{R}^n$. Let $\mathcal{H}$ be the set of functions $h: D\to\mathbb{R}^n$ integrable on D and $h(x,y,z)\in H(x,y,z)$ for each $(x,y,z)\in D$. Then, by the integral of the multifunction $H: D\to 2^{\mathbb{R}^n}$ we mean the set

$$\iiint\limits_D H(x,y,z)\,dx\,dy\,dz = \left\{\iiint\limits_D h(x,y,z)\,dx\,dy\,dz \mid h\in\mathcal{H}\right\}.$$

In what follows we list some properties of the integral defined above.

Proposition 2.1. *If $H: D\to 2^{\mathbb{R}^n}$ is an upper semi-continuous multifunction and there exists a positive real number C such that*

$$|H(x,y,z)| = \sup\{\|\zeta\| \mid \zeta\in H(x,y,z)\}\le C$$

for each $(x,y,z)\in D$, then

$$\iiint\limits_D H(x,y,z)\,dx\,dy\,dz = \iiint\limits_D \operatorname{conv} H(x,y,z)\,dx\,dy\,dz.$$

Proposition 2.2. *If $H_k: D\to 2^{\mathbb{R}^n}$, $k\in\mathbb{N}$, are upper semi-continuous multifunctions and there exists a positive real number C such that $|H_k(x,y,z)|\le C$ for each*

$(x, y, z) \in D$ and $k \in \mathbb{N}$, then

$$\iiint\limits_{D} \underline{\lim} H_k(x, y, z)\, dx\, dy\, dz \subset \underline{\lim} \iiint\limits_{D} H_k(x, y, z)\, dx\, dy\, dz.$$

Taking into account Definition 2 in [5], we have $(x, y, z) \in \underline{\lim} H_k(x, y, z)$ iff each neighbourhood of (x, y, z) intersects all the sets $H_k(x, y, z)$ with k large enough.

Proposition 2.3. *If A is a compact subset of $\mathbb{R}^n$, independent of (x, y, z), then*

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} A\, dx\, dy\, dz = (x_2 - x_1)(y_2 - y_1)(z_2 - z_1) \operatorname{conv} A,$$

where $(x_1, y_1, z_1), (x_2, y_2, z_2) \in D$.

Moreover, we need the following proposition.

Proposition 2.4. *If K is a convex set in a Banach space X, then the set $K_\varepsilon = \bigcup\limits_{x \in K} B[x, \varepsilon]$ is convex.*

3 Results

In [16] the notion of a *local solution* for the Darboux Problem (1.1) + (1.2) is defined and is proved an existence theorem for a local solution of this problem, together with some properties of the set of its solutions, namely that this set is a compact subset in Banach space $C(D_0; \mathbb{R}^n)$ and, as a function of initial values, it defines an upper semi-continuous multifunction on $D_0 = [0, x_0] \times [0, y_0] \times [0, z_0] \subseteq D$.

Let the following hypotheses be satisfied:

(H_1) $F : D \times \Omega \to 2^{\mathbb{R}^n}$ is a multifunction with compact convex non-empty values in $\mathbb{R}^n$, $D = [0, a] \times [0, b] \times [0, c] \subset \mathbb{R}^3$, and $\Omega \subset \mathbb{R}^n$ is an open subset.

(H_2) For any $(x, y, z) \in D$, the mapping $u \to F(x, y, z, u)$ is upper semi-continuous on Ω.

(H_3) For any $u \in \Omega$, the mapping $(x, y, z) \to F(x, y, z, u)$ is Lebesgue-measurable on D.

(H_4) There exists a function $k : D \to \mathbb{R}_+, k \in \mathcal{L}^1(D; \mathbb{R}^n)$ such that

$$\|\zeta\| \le k(x, y, z), (\forall)\, \zeta \in F(x, y, z, u), \quad (\forall)\,(x, y, z) \in D, \quad (\forall)\, u \in \Omega.$$

(H_5) The functions $\varphi \in C^*(D_1; \mathbb{R}^n)$, $\psi \in C^*(D_2; \mathbb{R}^n)$, $\chi \in C^*(D_3, \mathbb{R}^n)$ are absolutely continuous in Carathéodory's sense functions and satisfy condition (1.3).

Remark 1. The function $\alpha : D \to \mathbb{R}^n$ defined by

$$\begin{aligned}\alpha(x,y,z) = {} & \varphi(x,y) + \psi(y,z) + \chi(x,z) - \varphi(x,0) - \\ & - \varphi(0,y) - \psi(0,z) + \psi(0,0) = \\ = {} & \varphi(x,y) + \psi(y,z) + \chi(x,z) - v^1(x) - v^2(y) - v^3(z) + v^0, \end{aligned} \tag{3.1}$$

is an absolutely continuous in Carathéodory's sense function on D, $\alpha \in C^*(D;\mathbb{R}^n)$ $[2, \S 565 - 570]$.

Remark 2. Denote by $M \subset \Omega$ the convex compact set in which the function $\alpha : D \to \mathbb{R}^n$, defined by (3.1), takes its values for all $(x,y,z) \in D_0$.

Remark 3. Let $(x_0, y_0, z_0) \in]0,a] \times]0,b] \times]0,c]$ be a point such that

$$\int_0^{x_0} \int_0^{y_0} \int_0^{z_0} k(r,s,t)\, dr\, ds\, dt < d(M, C_\Omega),$$

where $d(M, C_\Omega)$ is the distance from M to $C_\Omega = \mathbb{R}^n - \Omega$, an inequality immediately resulting from the integrability of function k.

Definition 3.1 [16]. The *Darboux Problem* for the hyperbolic inclusion (1.1) means to determine a *solution* of this inclusion which satisfies the initial conditions (1.2).

Definition 3.2 [16]. A *local solution* of Darboux Problem $(1.1)+(1.2)$ is defined as a function $U : D_0 \to \Omega, U \in C^*(D_0;\mathbb{R}^n)$, absolutely continuous in Carathéodory's sense $[2, \S 565 - 570]$, which satisfies (1.1) for a.e $(x,y,z) \in D_0$, and also initial conditions (1.2) for all $(x,y) \in [0,x_0] \times [0,y_0]$, all $(y,z) \in [0,y_0] \times [0,z_0]$, all $(x,z) \in [0,x_0] \times [0,z_0]$.

In [16] we proved the following

Theorem 3.1 [16]. *Let the hypotheses* $(H_1) - (H_5)$ *be satisfied. Then:*

- (i) *there exists at least a local solution* U *of Darboux Problem* $(1.1)+(1.2)$*;*
- (ii) *the set* S_α *of the local solutions* U *is compact in Banach space* $C(D_0;\mathbb{R}^n)$*;*
- (iii) *the multifunction* $\alpha \to S_\alpha$ *is upper semi-continuous on* $C^*(D_0;\mathbb{R}^n)$*, taking values in* $C(D_0;\mathbb{R}^n)$*.*

The solution U is a fixed point of a suitable multifunction which satisfies the Kakutani-Ky Fan fixed point theorem and it is of the form

$$U(x,y,z) = \alpha(x,y,z) + \int_0^x \int_0^y \int_0^z \beta(r,s,t)\, dr\, ds\, dt, \quad (x,y,z) \in D_0, \tag{3.2}$$

where

$$\beta(x,y,z) \in \Gamma(x,y,z) \subset F(x,y,z,U(x,y,z)) \text{ for a.e. } (x,y,z) \in D_0, \tag{3.3}$$

β is a measurable selection of the multifunction $\Gamma : D_0 \to \mathcal{C}(\mathbb{R}^n)$ [3, 4, 16].

Definition 3.3 [17]. A local solution for the Darboux Problem (1.1) + (1.2) $U : D_0 \to \Omega$ is *prolongable* (or *non-saturated*) if there exists a solution $\widetilde{U} : \widetilde{D} \to \mathbb{R}^n$ for the Darboux Problem (1.1) + (1.2) such that

$$\begin{cases} D_0 \subseteq \widetilde{D}, \quad D_0 \neq \widetilde{D}, \\ \widetilde{U}(x,y,z) = U(x,y,z), \quad (x,y,z) \in D_0, \end{cases}$$

where $\widetilde{D} \subseteq D$ is a union of D_0 with a finite number of adjacent parallelepipeds.

In [17] we proved the following theorems:

Theorem 3.2 [17]. *Let the hypotheses $(H_1) - (H_5)$ be satisfied together with the hypotheses:*

(H_6) *The set Ω is bounded, that is there exists a constant $C \in \mathbb{R}_+$ such that $\|u\| \leq C,$, $(\forall)\, u \in \Omega$.*

(H_7) *The multifunction F maps bounded sets onto bounded sets, hence a constant $K \in \mathbb{R}_+$ exists such that*

$$\sup\{\|\zeta\| \mid \zeta \in F(x,y,z,u)\} \leq K,$$

for any $(x,y,z,u) \in D \times \Omega$.

Then the local solution U is prolongable.

Theorem 3.3 [17]. *We assume the hypotheses $(H_1) - (H_7)$ to be satisfied. If $U : D_0 \to \Omega$ is a local solution of Darboux Problem (1.1) + (1.2) that is non-saturated, hence prolongable, then there exists a saturated solution $U^* : D^* \to \Omega$ of the Darboux Problem (1.1) + (1.2) such that*

$$\begin{cases} D_0 \subseteq D^*, \quad D_0 \neq D^*, \quad D^* \subseteq D, \\ U^*(x,y,z) = U(x,y,z), \quad (x,y,z) \in D_0, \end{cases}$$

hence U^* is a prolongation of U onto D^* that has been built by joining D_0 with a union of parallelepipeds adjacent to D_0.

Theorem 3.4 [17]. *Let the hypotheses $(H_1) - (H_7)$ be satisfied. If the saturated solution U^* is bounded on D^*, then $D^* = D$.*

Theorem 3.5 [17]. *Let the hypotheses $(H_1) - (H_7)$ be satisfied together with the hypothesis:*

(H_8) *The multifunction $F : D \times \Omega \to 2^{\mathbb{R}^n}$ is sub-linear, hence two constants $k_1 > 0$ and $k_2 \in \mathbb{R}$ exist with the property*

$$\sup\{\|\zeta\| \mid \zeta \in F(x,y,z,u)\} \leq k_1 \|u\| + k_2, \quad \text{for a.e. } (x,y,z) \in D, \quad u \in \Omega. \tag{3.4}$$

Then the saturated solution $U^*: D \to \Omega$ *is bounded on* D.

The saturated solution U^* has the form, by Theorem 3.1 [16],

$$U^*(x,y,z) = \alpha(x,y,z) + \int_0^x \int_0^y \int_0^z \beta^*(r,s,t)\, dr\, ds\, dt, \quad (x,y,z) \in D, \tag{3.5}$$

where $\alpha(x,y,z)$ is given by (3.1) and β^* is a measurable selection of the multivalued mapping Γ^* [3, 4, 16], defined on D with compact non-empty values in $\mathbb{R}^n$, i.e. $\Gamma^*: D \to \mathcal{C}(\mathbb{R}^n)$, such that

$$\beta^*(x,y,z) \in \Gamma^*(x,y,z) \subseteq F(x,y,z,U^*(x,y,z)) \text{ for a.e. } (x,y,z) \in D. \tag{3.6}$$

Definition 3.4. A function $U: D \to \mathbb{R}^n$ is called a *solution* of the Darboux Problem (1.1)+(1.2) if it is absolutely continuous in Carathéodory's sense on $D, U \in C^*(D;\mathbb{R}^n)$ [2, §565 – 570] and satisfies (1.1) for a.e. $(x,y,z) \in D$, and also initial conditions (1.2) for all $(x,y) \in D_1$, all $(y,z) \in D_2$, all $(x,z) \in D_3$.

Similarly with [5, 15] in this paper we prove a theorem of characterization of the solutions for Darboux Problem (1.1) + (1.2).

Theorem 3.6. *Let the hypotheses* (H_1'), (H_3), (H_4), (H_5) *of Theorem 3.1 be satisfied:*

(H_1') $F: D \times \Omega \to 2^{\mathbb{R}^n}$ *is an upper semi-continuous multifunction with compact convex non-empty values in* $\mathbb{R}^n$, $D = [0,a] \times [0,b] \times [0,c] \subset \mathbb{R}^3$ *and* $\Omega \subset \mathbb{R}^n$ *is an open bounded set.*

The hypothesis (H_1') *includes the hypothesis* (H_6).

Then, the continuous function $U: D \to \mathbb{R}^n$ *is a solution of Darboux Problem* (1.1) + (1.2) *if and only if for each* $(x_1,y_1,z_1), (x_2,y_2,z_2) \in D$ *the membership relation*

$$\begin{aligned} &[U(x_2,y_2,z_2) - U(x_1,y_2,z_2) - U(x_2,y_1,z_2) + U(x_1,y_1,z_2)] - \\ &- [U(x_2,y_2,z_1) - U(x_1,y_2,z_1) - U(x_2,y_1,z_1) + U(x_1,y_1,z_1)] \in \\ &\in \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} F(x,y,z,U(x,y,z))\, dx\, dy\, dz, \end{aligned} \tag{3.7}$$

holds, and U *satisfies the conditions* (1.2).

The difference in (3.7) is an extension of hyperbolic difference for functions in two variables.

Proof. The *necessity* of (3.7) is a consequence of the following arguments. Let U be a solution of (1.1) + (1.2) on D. It exists from Theorem 3.4 and has the form (3.5). We denote $U^* = U$.

$$U(x,y,z) = \alpha(x,y,z) + \int_0^x \int_0^y \int_0^z \beta(r,s,t)\, dr\, ds\, dt, \quad (x,y,z) \in D, \tag{3.8}$$

$\beta^* = \beta$ is a measurable selection of multivalued mapping $\Gamma^* = \Gamma$ [3, 4, 16] defined on D with compact non-empty values in $\mathbb{R}^n, \Gamma : D \to \mathcal{C}(\mathbb{R}^n)$,

$$\beta(x,y,z) \in \Gamma(x,y,z) \subseteq F(x,y,z,U(x,y,z)) \quad \text{for a.e} \quad (x,y,z) \in D. \tag{3.9}$$

We denote $\delta = [x_1, x_2] \times [y_1, y_2] \times [z_1, z_2] \subseteq D$. By (3.8) it follows that

$$\frac{\partial^3 U(x,y,z)}{\partial x \partial y \partial z} = \beta(x,y,z) \in \Gamma(x,y,z) \subseteq F(x,y,z,U(x,y,z)) \quad \text{for a.e.} \quad (x,y,z) \in D \tag{3.10}$$

and U satisfies the conditions (1.2).

Choosing two points $(x_1, y_1, z_1), (x_2, y_2, z_2) \in D$ and integrating the equation (3.10) on δ we get

$$\begin{aligned}
&\int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \frac{\partial^3 U(x,y,z)}{\partial x \partial y \partial z} dx\, dy\, dz = \int_{x_1}^{x_2}\int_{y_1}^{y_2} \frac{\partial^2 U(x,y,z)}{\partial x \partial y}\Big|_{z=z_1}^{z=z_2} dx\, dy = \\
&= \int_{x_1}^{x_2}\int_{y_1}^{y_2} \left[\frac{\partial^2 U(x,y,z_2)}{\partial x \partial y} - \frac{\partial^2 U(x,y,z_1)}{\partial x \partial y}\right] dx\, dy = \\
&= \int_{x_1}^{x_2}\int_{y_1}^{y_2} \frac{\partial^2 U(x,y,z_2)}{\partial x \partial y} d\dot{x}\, dy - \int_{x_1}^{x_2}\int_{y_1}^{y_2} \frac{\partial^2 U(x,y,z_1)}{\partial x \partial y} dx\, dy = \\
&= \int_{x_1}^{x_2} \frac{\partial U(x,y,z_2)}{\partial x}\Big|_{y=y_1}^{y=y_2} dx - \int_{x_1}^{x_2} \frac{\partial U(x,y,z_1)}{\partial x}\Big|_{y=y_1}^{y=y_2} dx = \\
&= \int_{x_1}^{x_2} \left[\frac{\partial U(x,y_2,z_2)}{\partial x} - \frac{\partial U(x,y_1,z_2)}{\partial x}\right] dx - \\
&- \int_{x_1}^{x_2} \left[\frac{\partial U(x,y_2,z_1)}{\partial x} - \frac{\partial U(x,y_1,z_1)}{\partial x}\right] dx = \\
&= \left(U(x,y_2,z_2)\Big|_{x=x_1}^{x=x_2} - U(x,y_1,z_2)\Big|_{x=x_1}^{x=x_2}\right) - \\
&- \left(U(x,y,z_1)\Big|_{x=x_1}^{x=x_2} - U(x,y_1,z_1)\Big|_{x=x_1}^{x=x_2}\right) = \\
&= [(U(x_2,y_2,z_2) - U(x_1,y_2,z_2)) - (U(x_2,y_1,z_2) - U(x_1,y_1,z_2))] - \\
&- [(U(x_2,y_2,z_1) - U(x_1,y_2,z_1)) - (U(x_2,y_1,z_1) - U(x_1,y_1,z_1))] = \\
&= [U(x_2,y_2,z_2) - U(x_1,y_2,z_2) - U(x_2,y_1,z_2) + U(x_1,y_1,z_2)] - \\
&- [U(x_2,y_2,z_1) - U(x_1,y_2,z_1) - U(x_2,y_1,z_1) + U(x_1,y_1,z_1)] = \\
&= \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \beta(x,y,z)\, dx\, dy\, dz \in \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \Gamma(x,y,z)\, dx\, dy\, dz \subseteq \\
&\subseteq \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} F(x,y,z,U(x,y,z))\, dx\, dy\, dz.
\end{aligned} \tag{3.11}$$

According to (3.11), we have (3.7) satisfied it was stated.

In order to prove the *sufficiency* of (3.7), we firstly prove that the continuous function U, satisfying (3.7) and (1.2), has the derivative $\frac{\partial^3 U(x,y,z)}{\partial x \partial y \partial z}$ for a.e $(x,y,z) \in D$. For this, we prove that U is absolutely continuous in Carathéodory's sense on D. We associate to the continuous function U, the interval function $[2, \S 453, 565]$,

$$\begin{aligned}\Phi(\delta) = &[U(x_2,y_2,z_2) - U(x_1,y_2,z_2) - U(x_2,y_1,z_2) + U(x_1,y_1,z_2)] - \\ &- [U(x_2,y_2,z_1) - U(x_1,y_2,z_1) - U(x_2,y_1,z_1) + U(x_1,y_1,z_1)]. \qquad (3.12)\end{aligned}$$

We prove that $\Phi(\delta)$ is absolutely continuous, using the Theorem 1 in [2, §453]. From (3.7) and (3.12) we get

$$\Phi(\delta) \in \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} F(x,y,z,U(x,y,z))\, dx\, dy\, dz. \qquad (3.13)$$

In view of Definition 2.9 and (3.11), the relation (3.7) holds for $(x,y,z) \in \delta$. Then (3.7), (3.11), (3.13) yield

$$\begin{aligned}\Phi(\delta) = &\int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} \beta(x,y,z)\, dx\, dy\, dz \in \\ &\in \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} F(x,y,z,U(x,y,z_1))\, dx\, dy\, dz. \qquad (3.14)\end{aligned}$$

According to the hypothesis (H_4), we obtain

$$\begin{aligned}\|\Phi(\delta)\| \leq &\int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} \|\beta(x,y,z)\|\, dx\, dy\, dz \leq \\ &\leq \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} k(x,y,z)\, dx\, dy\, dz. \qquad (3.15)\end{aligned}$$

We set

$$\eta(\lambda) = \sup_{\mu(\delta) \leq \lambda} \|\Phi(\delta)\|, \text{ for any } \lambda \in \mathbb{R}_+. \qquad (3.16)$$

In view of the absolute continuity of the integral, for each $\varepsilon > 0$ there exists a $\delta_1(\varepsilon) > 0$ such that

$$\iiint_{\delta} k(x,y,z)\, dx\, dy\, dz = \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} k(x,y,z)\, dx\, dy\, dz < \varepsilon, \qquad (3.17)$$

whenever $\mu(\delta) < \delta_1(\varepsilon)$.

Let $\lambda < \delta_1(\varepsilon)$. According to (3.15), (3.16), (3.17) we obtain

$$\eta(\lambda) \leq \sup \iiint_{\delta} k(x,y,z)\, dx\, dy\, dz = \int_{x_1}^{x_2} \int_{y}^{y_2} \int_{z_1}^{z_2} k(x,y,z)\, dx\, dy\, dz < \varepsilon, \qquad (3.18)$$

whenever $\mu(\delta) \le \lambda < \delta_1(\varepsilon)$, or

$$\lim_{\lambda \to 0} \eta(\lambda) = 0. \tag{3.19}$$

According to Theorem in [2, §453] the interval function $\Phi(\delta)$ is absolutely continuous. Since the continuous function U satisfies the conditions (1.3) the hypothesis (H_5) holds too. In view of [2, §567]the function U is absolutely continuous in Carathéodory's sense. From Theorems 5, 6 [2, §569 − 570] the function U has the derivative $\dfrac{\partial^3 U(x,y,z)}{\partial x\, \partial y\, \partial z}$ for a.e. $(x,y,z) \in D$.

It remains to prove that the function U satisfies the inclusion (1.1).

Taking into account the hypothesis (H_1) and the continuity of the function U, it follows that the multifunction$\widetilde{F} : D \to 2^{\mathbb{R}^n}$, given by

$$\widetilde{F}(x,y,z) = F(x,y,z,U(x,y,z)), \quad (x,y,z) \in D, \tag{3.20}$$

is upper semi-continuous on D. Then by Theorem 9.3.1 [13] and [5], Definition 1, we deduce

$$\widetilde{F}([B(x,y,z),\delta_2]) \subset B\left[\widetilde{F}(x,y,z),\varepsilon\right], \quad (x,y,z) \in D, \tag{3.21}$$

where $B[(x,y,z),\delta_2]$ is the open ball centered at $(x,y,z) \in D$ of radius $\delta_2 = \delta_2(\varepsilon) > 0$ and

$$B\left[\widetilde{F}(x,y,z),\varepsilon\right] = \left\{\omega \in \mathbb{R}^n \mid d\left(\omega, \widetilde{F}(x,y,z)\right) < \varepsilon\right\}. \tag{3.22}$$

Fix $(x,y,z) \in D$. If $(x',y',z') \in B[(x,y,z),\delta_2]$, then

$$\widetilde{F}\left(x',y',z'\right) \subset B\left[\widetilde{F}(x,y,z),\varepsilon\right] \tag{3.23}$$

because by Definition 2.1, and by Definition 9.1.2 [13, p.67) and also [5, 2] we have

$$\widetilde{F}(B[(x,y,z),\delta_2]) = \left\{\bigcup \widetilde{F}\left(x',y',z'\right) \mid \left(x',y',z'\right) \in B[(x,y,z),\delta_2]\right\}. \tag{3.24}$$

The condition (3.7) may be rewritten as

$$\begin{aligned} &\left[U\left(x',y',z'\right) - U\left(x,y',z'\right) - U\left(x',y,z'\right) + U\left(x,y,z'\right)\right] - \\ &- \left[U\left(x',y',z\right) - U\left(x,y',z\right) - U\left(x',y,z\right) + U(x,y,z)\right] \in \\ &\in \int_x^{x'} \int_y^{y'} \int_z^{z'} F(r,s,t,U(r,s,t))\, dr\, ds\, dt, \end{aligned} \tag{3.25}$$

for the domain $[x,x'] \times [y,y'] \times [z,z'] \subseteq D$.

According to (3.20), we deduce from (3.25) that

$$\begin{aligned} &\left[U\left(x',y',z'\right) - U\left(x,y',z'\right) - U\left(x',y,z'\right) + U\left(x,y,z'\right)\right] - \\ &- \left[U\left(x',y',z\right) - U\left(x,y',z\right) - U\left(x',y,z\right) + U(x,y,z)\right] \in \end{aligned}$$

$$\in \int_x^{x'} \int_y^{y'} \int_z^{z'} \widetilde{F}(r,s,t,)\, dr\, ds\, dt. \tag{3.26}$$

By $(x', y', z') \in B[(x,y,z), \delta_2]$, we obtain $|x - x'| < \delta_2$, $|y - y'| < \delta_2$, $|z - z'| < \delta_2$. Moreover $|r - x| < \delta_2$, $|s - y| < \delta_2$, $|t - z| < \delta_2$ for $x \le r \le x'$, $y \le s \le y'$, $z \le t \le z'$.

By (3.23) we have

$$\widetilde{F}(r,s,t) \subset B\left[\widetilde{F}(r,s,t), \varepsilon\right]. \tag{3.27}$$

Then, by (3.27), the relation (3.26) yields

$$\begin{gathered}\left[U\left(x',y',z'\right) - U\left(x,y',z'\right) - U\left(x',y,z\right) + U\left(x,y,z'\right)\right] - \\ - \left[U\left(x',y',z\right) - U\left(x,y',z\right) - U\left(x',y,z\right) + U\left(x,y,z\right)\right] \in \\ \in \int_x^{x'} \int_y^{y'} \int_z^{z'} B\left[\widetilde{F}(r,s,t,), \varepsilon\right] dr\, ds\, dt.\end{gathered} \tag{3.28}$$

As the multifunction $\widetilde{F}$, given by (3.20), is upper semi-continuous on D, the set $B\left[\widetilde{F}(x,y,z), \varepsilon\right]$ is closed in $\mathbb{R}^n$.

In view of (3.22) it follows that $B\left[\widetilde{F}(x,y,z), \varepsilon\right]$ is also bounded in $\mathbb{R}^n$ and therefore it is a compact set. Then we can use Proposition 2.3, setting $A = B\left[\widetilde{F}(x,y,z), \varepsilon\right]$ and $[x,x'] \times [y,y'] \times [z,z']$ instead of $\delta = [x_1,x_2] \times [y_1,y_2] \times [z_1,z_2]$, we obtain

$$\begin{gathered}\int_x^{x'} \int_y^{y'} \int_z^{z'} B\left[\widetilde{F}(x,y,z), \varepsilon\right] dr\, ds\, dt = \\ = \left(x' - x\right)\left(y' - y\right)\left(z' - z\right) \operatorname{conv} B\left[\widetilde{F}(x,y,z), \varepsilon\right].\end{gathered} \tag{3.29}$$

According to (3.29), the relation (3.28) yields

$$\begin{gathered}\left[U\left(x',y',z'\right) - U\left(x,y',z'\right) - U\left(x',y,z'\right) + U\left(x,y,z'\right)\right] - \\ - \left[U\left(x',y',z\right) - U\left(x,y',z\right) - U\left(x',y,z\right) + U\left(x,y,z\right)\right] \in \\ \in \left(x' - x\right)\left(y' - y\right)\left(z' - z\right) \operatorname{conv} B\left[\widetilde{F}(x,y,z), \varepsilon\right].\end{gathered} \tag{3.30}$$

By Proposition (2.4), the set $B\left[\widetilde{F}(x,y,z), \varepsilon\right]$ is convex and therefore

$$\operatorname{conv} B\left[\widetilde{F}(x,y,z), \varepsilon\right] = B\left[\widetilde{F}(x,y,z), \varepsilon\right]. \tag{3.31}$$

Using (3.31), the relation (3.30) yields

$$\begin{gathered}\left[U\left(x',y',z'\right) - U\left(x,y',z'\right) - U\left(x',y,z'\right) + U\left(x,y,z'\right)\right] - \\ - \left[U\left(x',y',z\right) - U\left(x,y',z\right) - U\left(x',y,z\right) + U\left(x,y,z\right)\right] \in\end{gathered}$$

$$\in \left(x'-x\right)\left(y'-y\right)\left(z'-z\right) B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right]. \tag{3.32}$$

From (3.32) we get

$$\left[\frac{U\left(x',y',z'\right)-U\left(x,y',z'\right)}{x'-x}-\frac{U\left(x',y,z'\right)-U\left(x,y,z'\right)}{x'-x}\right]- \\ -\left[\frac{U\left(x',y',z\right)-U\left(x,y',z\right)}{x'-x}-\frac{U\left(x',y,z\right)-U\left(x,y,z\right)}{x'-x}\right]\in \\ \in \left(y'-y\right)\left(z'-z\right) B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right]. \tag{3.33}$$

Taking into account that $B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right]$ is closed, the relation (3.33) yields

$$\lim_{x'\to x}\left\{\left[\frac{U\left(x',y',z'\right)-U\left(x,y',z'\right)}{x'-x}-\frac{U\left(x',y,z'\right)-U\left(x,y,z'\right)}{x'-x}\right]-\right. \\ \left.-\left[\frac{U\left(x',y',z\right)-U\left(x,y',z\right)}{x'-x}-\frac{U\left(x',y,z\right)-U\left(x,y,z\right)}{x'-x}\right]\right\}= \\ =\left\{\left[\frac{\partial U}{\partial x}\left(x,y',z'\right)-\frac{\partial U}{\partial x}\left(x,y,z'\right)\right]-\left[\frac{\partial U}{\partial x}\left(x,y',z\right)-\frac{\partial U}{\partial x}\left(x,y,z\right)\right]\right\}\in \\ \in \left(y'-y\right)\left(z'-z\right) B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right] \tag{3.34}$$

and

$$\lim_{y'\to y}\left[\frac{\frac{\partial U}{\partial x}\left(x,y',z'\right)-\frac{\partial U}{\partial x}\left(x,y,z'\right)}{y'-y}-\frac{\frac{\partial U}{\partial x}\left(x,y',z\right)-\frac{\partial U}{\partial x}\left(x,y,z\right)}{y'-y}\right]\in \\ \in \left(z'-z\right) B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right] \tag{3.35}$$

or

$$\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z'\right)-\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z\right)\in \left(z'-z\right) B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right]. \tag{3.36}$$

It results

$$\frac{\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z'\right)-\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z\right)}{z'-z}\in B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right] \tag{3.37}$$

and

$$\lim_{z'\to z}\frac{\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z'\right)-\frac{\partial^2 U}{\partial x\partial y}\left(x,y,z\right)}{z'-z}=\frac{\partial^3 U}{\partial x\partial y\partial z}\left(x,y,z\right)\in B\left[\widetilde{F}\left(x,y,z\right),\varepsilon\right]. \tag{3.38}$$

Since $\widetilde{F}\left(x,y,z\right)$ is closed and F is an upper semi-continuous multifunction, the relation (3.38) yields, for $\varepsilon\to 0$,

$$\frac{\partial^3 U}{\partial x\partial y\partial z}\left(x,y,z\right)\in\widetilde{F}\left(x,y,z\right)=F\left(x,y,z,U\left(x,y,z\right)\right) \text{ for a.e. } \left(x,y,z\right)\in D. \tag{3.39}$$

Therefore, U satisfies the inclusion (1.1) as stated.

References

[1] Aumann R.J. *Integrals of Set – Valued Functions*. Journal of Mathematical Analysis and Applications, 1956, **12**, p. 1–12.

[2] Carathéodory C. *Vorlesungen über Reelle Funktionen*. Chelsea Publishing Company, New York, 1968, 3 Ed.

[3] Castaing Gh. *Sur les équations differentielles multivoques*. Comptes Rendus Acad. Sci. Paris, 1966, **263**, N 2, Série A, P. 63–66.

[4] Castaing Ch. *Quelques problèmes de mésurabilité liés à la théorie de la commande*. Comptes Rendus Acad. Sci. Paris, 1966, **262**, N 7, Série A, p. 409–411.

[5] Cellina A. *Multivalued differential equations and ordinary differential equations*. SIAM J. Appl. Math., 1970, **18**, N 2, p. 533–538.

[6] Cernea A. *Incluziuni diferenţiale hiperbolice şi control optimal*. Editura Academiei Române, Bucureşti, 2001.

[7] Deimling K. *A Carathéodory theory for systems of integral equations*. Ann. Mat. Pura Appl., 1970,**4**, N 86, p. 217–260.

[8] Deimling K. *Das Picard-Problem für* $u_{xy} = f(x, y, u, u_x, u_y)$ *unter Carathéodory-Voraussetzungen*. Math. Z., 1970, **114**, p. 303–312.

[9] Deimling K. *Das charakteristiche Anfangswertproblem für* $u_{x_1x_2x_3} = f$ *unter Carathéodory-Voraussetzungen*. Arch. Math. (Basel), 1971, **22**, p. 514–522.

[10] Marano S. *Generalized Solutions of Partial Differential Inclusions Depending on a Parameter*. Rend. Acad. Naz. Sc. XL., Mem. Mat., 1989, **13**, p. 281–295.

[11] Marano S. *Classical Solutions of Partial Differential Inclusions in Banach space*. Appl. Anal., 1991, **42**, p. 127–143.

[12] Marano S. *Controllability of Partial Differential Inclusions Depending on a Parameter and Distributed Parameter Control Process*. Le Matematiche, 1990, **XLV**, p. 283–300.

[13] Rus I.A. *Principii şi aplicaţii ale teoriei punctului fix*. Editura Dacia, Cluj-Napoca, 1979.

[14] Sosulski W. *On neutral partial functional-differential inclusions of hyperbolic type*. Demonstratio Mathematica, 1990, **23**, p. 893–909.

[15] Teodoru G. *A characterization of the solutions of the Darboux Problem for the equation* $\frac{\partial^2 z}{\partial x \partial y} \in F(x, y, z)$. Analele Ştiinţifice ale Universităţii "Al. I. Cuza" Iaşi, 1987, **33**, s. I a, Matematică, f. 1, p. 33–38.

[16] Teodoru G. *The Darboux Problem for third order hyperbolic inclusions*. Libertas Matematica, 2003, **23**, p. 119–127.

[17] Teodoru G. *Prolongation of solutions of the Darboux Problem for third order hyperbolic inclusions*. Libertas Matematica, 2006, **26**, p. 83–96.

Department of Mathematics
Technical University "Gh. Asachi" Iaşi
11 Carol I Blvd., RO-700506
Iaşi 6, România
E-mail: *teodoru@math.tuiasi.ro*

Received October 4, 2006

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 49–54
ISSN 1024–7696

Maximization methods of turbo-machines performances

Mircea Dimitrie Cazacu

Abstract. Due to the important role of the turbo-machines efficiency, concerning the energy economy and environment pollution diminishing [1, 2], we shall present three original methods to maximize their performances, establishing the optimum blade profile and its setting angles at different blade radii.

Mathematics subject classification: ?.
Keywords and phrases: ?.

1 Maximum extracted power by a run-of-river hydraulic turbine or wind turbine

Projecting the two components of hydrodynamic resultant on the rotational peripheral direction, we shall obtain the mechanical power expression

$$P = UF_u = U(F_y \sin\beta - F_x \cos\beta) = \frac{\rho}{2}V^3 bl \left[c_y(i)\frac{\cos\beta}{\sin^2\beta} - c_x(i)\frac{\cos^2\beta}{\sin^2\beta}\right], \qquad (1)$$

and cancelling the partial derivative

$$\frac{\partial P}{\partial \beta} = -c_y(i)\frac{1+\cos^2\beta}{\sin^3\beta} + c_x(i)\frac{\cos\beta(2+\cos^2\beta)}{\sin^4\beta} = 0, \qquad (2)$$

introducing the notation $\sin^2\beta = x$, we must solve the algebraic equation

$$P(x) = \left[f^2(i)+1\right]x^3 - \left[4f^2(i)+7\right]x^2 + \left[4f^2(i)+15\right]x - 9 = 0, \qquad (3)$$

from which the sub-unit solution maximizes really the power, for any chosen profile-shape. Once more, introducing these values i and β in the power expression (1), the maximal power value will indicate the best profile to use [3]. Applying the relation $V = U\mathrm{tg}\beta$ at the outskirts, we obtain the optimal angular velocity, which being the same for all the blade, determines the rotation velocity at any other radius R_j and because the relative angle is thus known, the power maximization will be obtained only by the variation of the incident angle in case of considered profile.

1.1. The best incidence angle of blade profile for other radii

For other flow channel, placed at radius $R_j \neq R_p$, the peripheral radius, we obtain the maximization of the extracted power

$$P_j = \frac{V}{\mathrm{tg}\beta_j}\frac{\rho}{2}V^2 bl_j(R_j)\left[c_y(i)\frac{1}{\sin\beta_j} - c_x(i)\frac{\cos\beta_j}{\sin^2\beta_j}\right] =$$

$$= \frac{\rho}{2} V^3 bl[A(R_j)c_y(i) - B(R_j)c_x(i)]. \tag{4}$$

From the fluid by annulling its partial derivative with respect to the incidence angle i

$$\frac{\partial \lfloor A(R)c_y(i) - B(R)c_x(i) \rfloor}{\partial i} = 0 = \left[A(c_{y1} - 4i^3 c_{y4}) - B(c_{x1} + 2ic_{x2}\right], \tag{5}$$

in which we considered the usual expressions of variation with the incidence angle of the lift and drag coefficients, supposed of the form: $c_y(i) = c_{y0} + ic_{y1} - -i^4 c_{y4}$ and $c_x(i) = c_{x0} + ic_{x1} + i^2 c_{x2}$, finally obtaining the calculation formula of the best incidence angle for any radius

$$i^3 + i\frac{\omega_{\text{opt}} c_{x2}}{2V c_{y4}} R + \frac{\omega_{\text{opt}} c_{x1}}{4V c_{y4}} R - \frac{c_{y1}}{4c_{y4}} = 0, \tag{6}$$

with the interesting remark that the optimum incidence angle rises at the same time with the radius decreasing, to obtain a greater velocity around the profile [3].

The good performance of power coefficient $C_p = 0,42$ obtained for a three blade rotor [4] and $C_p = 0,56$ for a four blade rotor have put into the evidence the validity of this maximization method.

2 Maximization of the propulsion force for an aircraft or ship propeller

The problem of propulsion force increasing for the same consumed mechanical power at the shaft, is very important not only concerning the operation radius enlargement of an aircraft or ship, but also by the fossil fuel savings and environmental protection, being of a greater importance for the ecological boats, which use the solar energy by means of photovoltaic cells [5, 6].

2.1. The determination of the best peripheral relative angle

Taking into account the expressions of the lift and drag forces, exerted on the profiled blade, laid at the incidence angle i with respect to the relative angle β, corresponding to the relative velocity W from the velocity triangle, we can calculate the axial component of these forces, representing the propulsion force

$$F_a = F_y \cos\beta - F_x \sin\beta = \frac{\rho}{2} V^2 bl(R) \left[c_y(i)\frac{\cos\beta}{\sin^2\beta} - c_x(i)\frac{1}{\sin\beta}\right] \tag{7}$$

and also the expression of the shaft driving mechanical power

$$P_m = U(F_y \sin\beta + F_x \cos\beta) \quad \text{or} \quad p_m = \frac{2P_m}{\rho V^3 bl} = c_y(i)\frac{\cos\beta}{\sin^2\beta} + c_x(i)\frac{\cos^2\beta}{\sin^3\beta}. \tag{8}$$

By annulling the partial differential of the axial force (7) with respect to the relative angle β

$$\partial F_a/\partial\beta = 0 = -c_y(1 + \cos^2\beta) + c_x \sin\beta\cos\beta, \tag{9}$$

one obtains the condition to maximize the propulsion axial force (denoting by $x = \sin^2\beta$ and the profile fineness $f(i) = c_y(i)/c_x(i)$ as function of the incidence angle i) given by the following algebraic relation

$$(f^2+1)x^2 - (4f^2+1)x + 4f^2 = 0, \tag{10}$$

having two real solutions and putting into the evidence the relative best and respectively worst angle β as function of the fineness of the aerodynamic or hydrodynamic profiles, for the positive value under the root expression, necessary to assure the non-imaginary solutions

$$x = \frac{4f^2+1 \pm \sqrt{1-8f^2}}{2f^2+2}, \quad \text{for} \quad 1-8f^2 \geq 0 \to f(i) = \frac{c_y}{c_x} \leq 0.3536\ldots \tag{11}$$

which condition eliminates a lot of profiles too curved and prefers these that have the lift force near by zero for a certain incidence angle i [7].

2.2. The determination of the optimum profile setting angle for other radii

For the other radii, because the peripheral relative angle β_j is already determined by the relation $V = U_j \mathrm{tg}\beta_j$, the power maximization will be obtained only by the election of the optimum incidence angle in case of considered profile, as we shall see below. We have determined the blade profile angle $\beta_b = \beta_j - i$ annulling the expression of the axial force with respect to the incidence angle i of the profile [3], obtaining the relation

$$F_j = \frac{\rho}{2} V^2 bl(R_j) \left[(c_{y0} + ic_{y1} \frac{\cos\beta_j}{\sin^2\beta_j} - (c_{x0} + ic_{x1} + i^2 c_{x2}) \frac{1}{\sin\beta_j} \right] \tag{12}$$

the blade spread being $b = \delta R =$ constant and the blade depth l as function of radius R_j having no importance, we can annul the axial propulsion force with respect to the incidence angle to obtain the optimal incidence for each relative radius

$$\frac{\partial F_a}{\partial i} = 0 = \frac{c_{y1}}{\mathrm{tg}\beta_j} - c_{x1} - 2ic_{x2} \to i_{\text{opt}} = \frac{1}{2c_{x2}} (\frac{c_{y1}}{\mathrm{tg}\beta_p} \frac{R_j}{R_p} - c_{x1}), \tag{13}$$

considering the variation approximately linear of the lift coefficient of the profile (for example of the symmetric profile Gö 445 [3, 4]) as function of the incidence angle $C_y(i) \simeq C_{y0} + C_{y1} i = 0.002i$ and the parabolic approximately variation of the drag coefficient of the profile

$$C_x(i) \simeq C_{x0} + C_{x1} i - C_{x2} i^2 = 0.005 + 0.004,5i - 0.000,5i^2. \tag{14}$$

In this manner we can establish the airfoil profile, which realises the best propulsion axial force, as also the value of the relative mechanical driving power.

For the smaller relative radius $r = R_j/R_p < 1$, where we have already the relative angle β_j imposed, to maximize the axial force F_a one calculates the values of the optimal incidence angle i_{opt} given by the relation (13).

2.3. Maximization of the ratio between the axial force and consumed power

In this case [8], by annulling the partial differential with respect to the relative angle $0 \leq \beta \leq \pi/2$

$$\frac{\partial\left(f_a/p_m\right)}{\partial\beta} = \frac{f\mathrm{ctg}^2\beta - 2\mathrm{ctg}\beta - f}{\cos^2\beta(f^2 + 2f\mathrm{ctg}\beta + \mathrm{ctg}^2\beta)} = 0 \rightarrow f\mathrm{ctg}^2\beta - 2\mathrm{ctg}\beta - f = 0, \quad (15)$$

one obtains the maximization condition, that by introducing the notation $x = \mathrm{ctg}\beta$, leads us to the solving of the algebraic equation of 2^{nd} degree

$$f(i)x^2 - 2x - f(i) = 0, \quad (16)$$

having always two real solutions, one positive and other negative

$$x_{1,2} = \frac{1 \pm \sqrt{1+f^2}}{f(i)}, \quad (17)$$

as one can see for the case of Göttingen 450 profile [3], which are vindicated again as the best performing, and where we put also the value of the ratio f_a/p_m for the confirmation of the maximal value of the axial force, obtained at the approximate incidence angle $i \approx 1°$.

3 The obtaining of the fluid current maximal velocity

This optimization method presents a special importance in the problem of optimal profiling of the axial rotor blades for a mixer, ventilator or pump. To maximize the fluid current velocity V, we shall present two possibilities to solve this problem: using the velocity relation deduced by the axial force expression (7) or from that of the rotor driving power (8).

3.1. The fluid velocity maximizing using the axial force relation

We shall consider the mathematical problem of linked maximum, corresponding to the obtaining of maximal velocity of the axial fluid current using the axial force expression (7).

3.1.1. The optimal profiling of the blade and the optimal peripheral setting angle

Considering the relation (7), we can write

$$\frac{V^2\rho bl}{2F_a} = \frac{1}{c_y\frac{\cos\beta}{\sin^2\beta} - c_x\frac{1}{\sin\beta}} \rightarrow \frac{V}{\sqrt{2F_a/\rho bl}} = \left(c_y\frac{\cos\beta}{\sin^2\beta} - c_x\frac{1}{\sin\beta}\right)^{-1/2}, \quad (18)$$

from that by annulment of its partial derivation, we shall obtain the value of the relative angle β, introducing the profile fineness $f = c_y/c_x$ and denoting $\sin^2\beta = x$, the problem reduces to the solving of the same algebraic equation (10).

3.1.2. The optimal setting angle of the profile to the other blade radii

Because for the other blade radii the relative angle is already determined, we shall maximize the current velocity by annulment of its derivative with respect to profile incidence angle, obtaining the new expression of the fluid velocity

$$\frac{V}{\sqrt{2F_a/\rho bl}} = \left[(c_{y0} + ic_{y1} - i^2 c_{y2})\frac{\cos\beta_j}{\sin^2\beta_j} - (c_{x0} + ic_{x1} + i^2 c_{x2})\frac{1}{\sin\beta_j}\right]^{-1/2}, \quad (19)$$

which by annulment of its partial derivation with respect to the incidence angle i, gives us the necessary relation

$$i_{\text{opt}} = \frac{c_{y1}\text{ctg}\beta_j - c_{x1}}{c_{y2}\text{ctg}\beta_j + c_{x2}}. \quad (20)$$

3.2. The obtaining of the maximal velocity for the minimum consumed power

We solved this problem reporting the fluid velocity to the rotor driving mechanical power

$$\frac{V}{\sqrt[3]{2P_m/\rho bl}} = \left[c_y(i)\frac{\cos\beta}{\sin^2\beta} + c_x(i)\frac{\cos^2\beta}{sin^3\beta}\right]^{-1/3}, \quad (21)$$

in which we shall annul the partial derivative

$$\frac{\partial V}{\partial\beta} \approx \frac{c_y\sin\beta(2-\sin^2\beta) + c_x(3-\sin^2\beta)\cos\beta}{3[c_y\sin\beta\cos\beta + c_x(1-\sin^2\beta)]^{4/3}} = 0 \quad (22)$$

and because the denominator can never become infinite, the annulment of the numerator leads us to a same algebraic equation as (3).

References

[1] Cazacu M.D. *Tehnologii pentru o dezvoltare durabilă.* Acad. Oamenilor de Ştiinţă din România. Congresul "Dezvoltarea în pragul mileniului al III-lea", Secţia "Dezvoltarea durabilă", 27–29 septembrie 1998, Bucureşti. Editura Europa Nova, 1999, p. 533–539.

[2] Cazacu M.D. *A survey of the technics which might preserve the biosphere reservation Danube Delta.* The 8th Symposium "Technologies, installations and equipments for improvement of environment quality", 9–12 November 1999, Bucharest.

[3] Cazacu M.D., Nicolaie S. *Micro-hydroturbine for run-of-river power station.* A 2-Conferinţă a Hidroenergeticienilor din România, 24–25 mai 2002, Univ. Politehnica, Bucureşti, Vol. II, p. 443–448.

[4] Cazacu M.D. *Microagregat hidroelectric pentru asigurarea autonomiei energetice a balizelor luminoase sau a unor bărci fluviale.* Rev. Ştiinţă, Industrie, Tehnologie, Nr. 2, Bucureşti, 2005, p. 44–45.

[5] Tudor G., Mocanu Z. *Solar energy propelled craft intended for the ecological tourism and transport in "Danube Delta Biosphere" and other protected zones.* The 1st International Symposium "Renewable energies and Sustainable development", 23–25 September 2004, Tulcea.

[6] Aucouturier J.L., Henry H., Stempin E. *Electro solar boats for passengers transport: a reality now.* The 2nd Internat. Symposyum ”Renewable energies and Sustainable development”, 22–24 September 2005, Tulcea.

[7] Cazacu M.D. *The maximization of the propulsion force for an aircraft or ship propeller.* The 30th ”Caius Iacob” Conference on Fluid Mechanics and its Technical Applications, Bucharest, 25–26 November 2005, CD.

[8] Cazacu M.D. *Optimization of the axial propulsion force.* A 31-a Conferinţă Naţională ”Caius Iacob” de Mecanica fluidelor şi Aplicaţiile ei tehnice, Univ. Transilvania din Braşov, 19–21 oct. 2006, Bull. of Transilvania University, series B1 – Mathematics, Informatics, Physics, 2006, Vol. 13(48), p. 71–76.

University Politehnica of Bucharest
Romania

Received March 16, 2007

E-mail: *cazacumircea@yahoo.com*

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 55–63
ISSN 1024–7696

Modelling of explosive magnetorotational phenomena: from 2D to 3D

G.S. Bisnovatyi-Kogan, S.G. Moiseenko, B.P. Rybakin, G.V. Secrieru

Abstract. In the paper we describe the results of mathematical modelling of magnetorotational(MR) supernova explosion in 1D and 2D approach and formulate the problems and features for the numerical approach to simulations of the MR supernovae in 3D case.

Mathematics subject classification: 85A30.
Keywords and phrases: Partial differential equations, numerical methods, magnetohydrodynamics.

1 Introduction

Explosions of supernovae are very spectacular event in the Universe. Explanation of mechanism of core collapse supernova explosion is one of the most interesting and complicated problems of modern astrophysics. At the initial stage of core collapse supernova research the mechanisms of explosion had been connected with neutrino deposition, and bounce shock propagation. Spherically symmetrical numerical simulations have shown that the bounce shock appears at the distance 10-30km from the center, then it moves to the radius of about 100-200 km, and stalls, not giving an explosion. Farther investigation of this problem was an extension of the same mechanism to 2D and 3D cases. Numerical simulations of 2D and 3D models have an additional feature connected with a development of neutrino driven convection deep inside, and behind the shock. The extensive calculations have shown that this mechanism does not give supernova explosions either with a sufficient level of confidence. Recently improved models of the core collapse, where the neutrino transport was simulated by solving the Bolzmann equation, also do not explode [12].

The MR mechanism for core collapse supernova explosion was suggested by Bisnovatyi-Kogan in 1970 [9], see also [10]. The main idea of the MR mechanism is to transform part of the rotational energy of presupernova into the radial kinetic energy (explosion energy). During collapse the star rotates differentially. Differential rotation leads to the appearing and amplification of the toroidal component of the magnetic field. Growth of the magnetic field means amplification of the magnetic pressure with time. A compression wave appears near the region of the extremum of the magnetic field. This compression wave moves outwards along steeply decreasing density profile. In a short time it transforms to the fast MHD shock wave. When the shock reaches the surface of the collapsing star it ejects part of the matter and

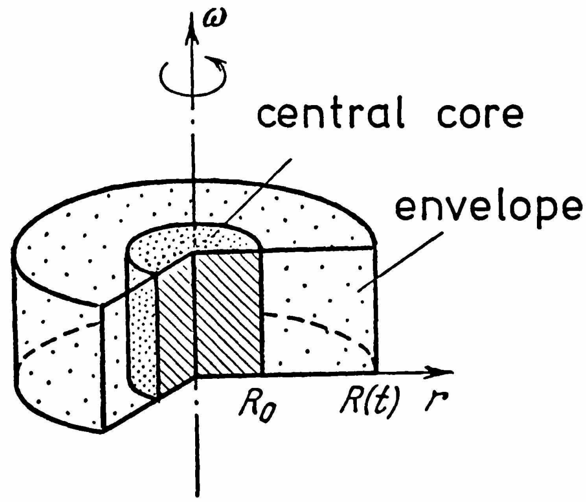

Figure 1. Model of MR presupernova in 1D case [11]

energy to the infinity. This ejection can be interpreted as explosion of the core collapse supernova.

The first 2D simulations of the collapse of the rotating magnetized star were presented in the paper [15], with unrealistically large values of the magnetic field. The differential rotation and amplification of the magnetic field resulted in the formation of the axial jet.

2 Results of 1D and 2D MR supernova simulations

The 1D simulations of MR supernova had been made in papers [4, 11]. In 1D case a star was represented as an infinite cylinder (Fig.1). For the simulations a set of ideal MHD equations with self gravitation in Lagrangian variables was used. Initial magnetic field had only r component. Differential rotation led to appearance and amplification of the toroidal φ component of the magnetic field. Numerical simulations of 1D MR supernova had shown that amplified due to the differential rotation toroidal field produced MHD shock wave which moved outwards. Part of the matter was ejected by the shock wave. The amount of the ejected energy $\approx 10^{51} erg$ is enough for the explanation of the supernova event. 1D simulations show that time of the evolution of MR supernova t_{expl} depends on the relation of the initial magnetic E_{mag} and gravitational E_{grav} energies $\alpha = \dfrac{E_{mag}}{E_{grav}}$ as $t_{expl} \sim \dfrac{1}{\sqrt{\alpha}}$. It means that for real values of the magnetic field ($\alpha \approx 10^{-6-8}$) t_{expl} becomes rather large. Parameter α characterizes a stiffness of the MHD equations describing MR supernova. The smallness of the parameter α is one of the main difficulties for the numerical simulation of MR supernova. From the physical point of view small α

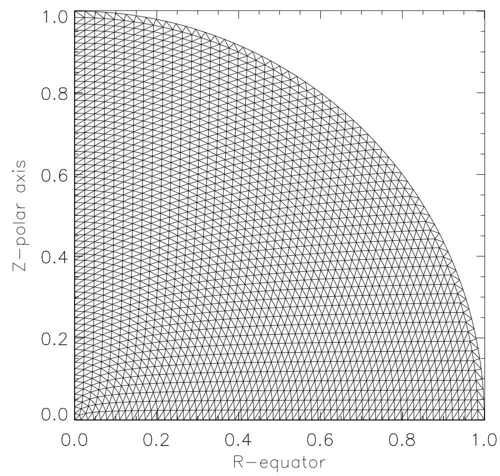

Figure 2. Triangular Lagrangian grid for 2D simulations of the magnetorotational supernova explosion

means existence two significantly different time scales. Very small acoustic time scale and huge time scale proportional to the time of the magnetic field amplification.

More realistic model of magnetorotational supernova was calculated in 2D approximation. The star was represented by a rotating self-gravitating gaseous body. The basic set of equations is a set of ideal MHD equations with self gravity in Lagrangian variables:

$$\begin{gathered} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v}, \qquad \frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\nabla\cdot\mathbf{v} = 0, \\ \rho\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\mathrm{grad}\left(P + \frac{\mathbf{H}\cdot\mathbf{H}}{8\pi}\right) + \frac{\nabla\cdot(\mathbf{H}\otimes\mathbf{H})}{4\pi} - \rho\nabla\Phi, \\ \rho\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathbf{H}}{\rho}\right) = \mathbf{H}\cdot\nabla\mathbf{v},\ \Delta\Phi = 4\pi G\rho, \\ \rho\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} + P\nabla\cdot\mathbf{v} + \rho F(\rho,T) = 0, \\ P = P(\rho,T),\ \varepsilon = \varepsilon(\rho,T), \end{gathered} \tag{1}$$

where $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla$ is the total time derivative, $\mathbf{x} = (r,\varphi,z)$, $\mathbf{v} = (v_r, v_\varphi, v_z)$ is the velocity vector, ρ is the density, P is the pressure, $\mathbf{H} = (H_r,\ H_\varphi,\ H_z)$ is the magnetic field vector, Φ is the gravitational potential, ε is the internal energy, G is gravitational constant, $\mathbf{H}\otimes\mathbf{H}$ is the tensor of rank 2, and $F(\rho,T)$ is the rate of neutrino losses.

Spatial Lagrangian coordinates are r, φ and z, i.e. $r = r(r_0,\varphi_0,z_0,t)$, $\varphi = \varphi(r_0,\varphi_0,z_0,t)$, and $z = z(r_0,\varphi_0,z_0,t)$, where r_0,φ_0,z_0 are the initial coordinates of material points of the matter.

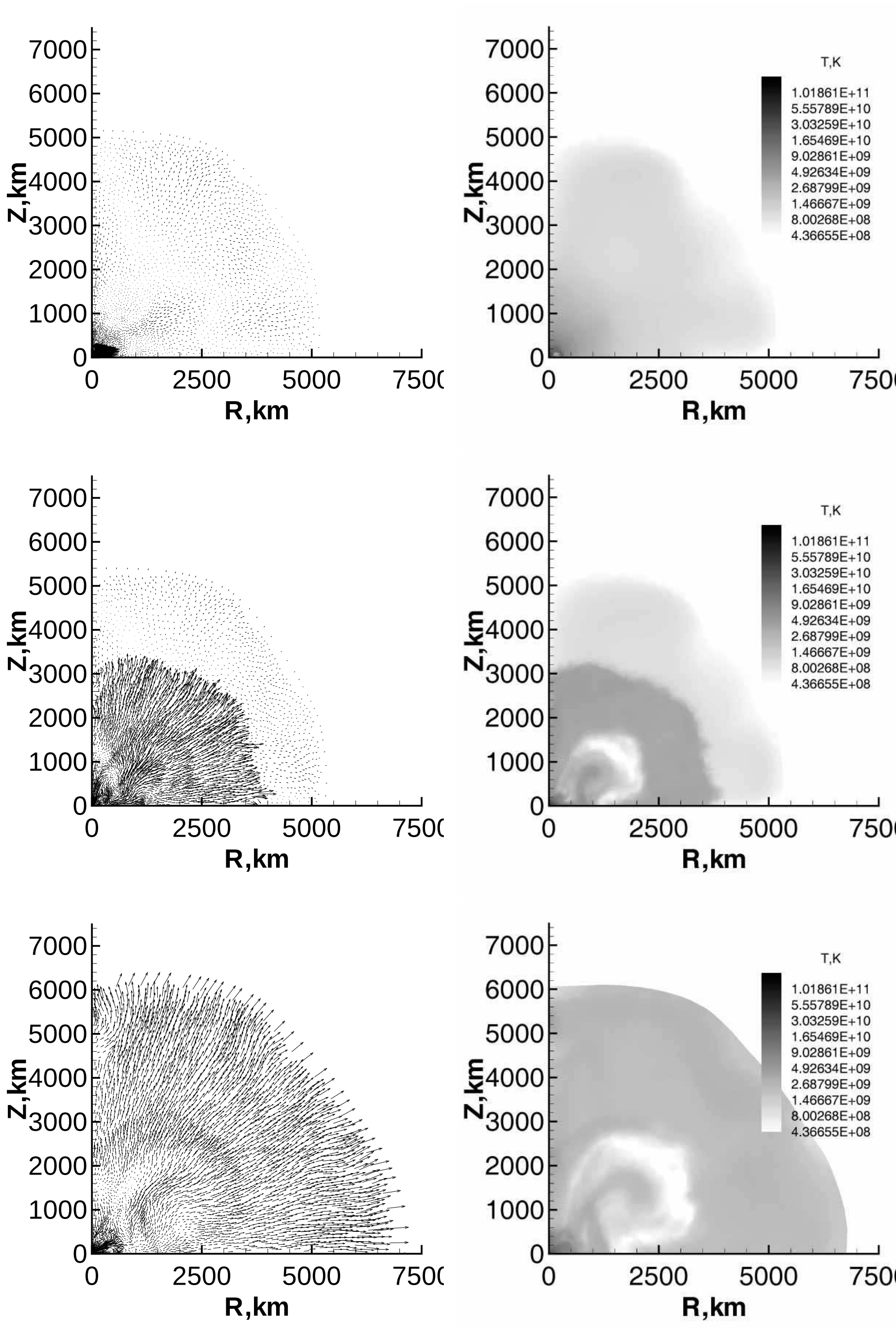

Figure 3. The distribution of the velocity field (left column) and the temperature (right column) for the time moments $t = 0.07,\ 0.2,\ 0.3s$ for the initial *quadrupole*-like magnetic field

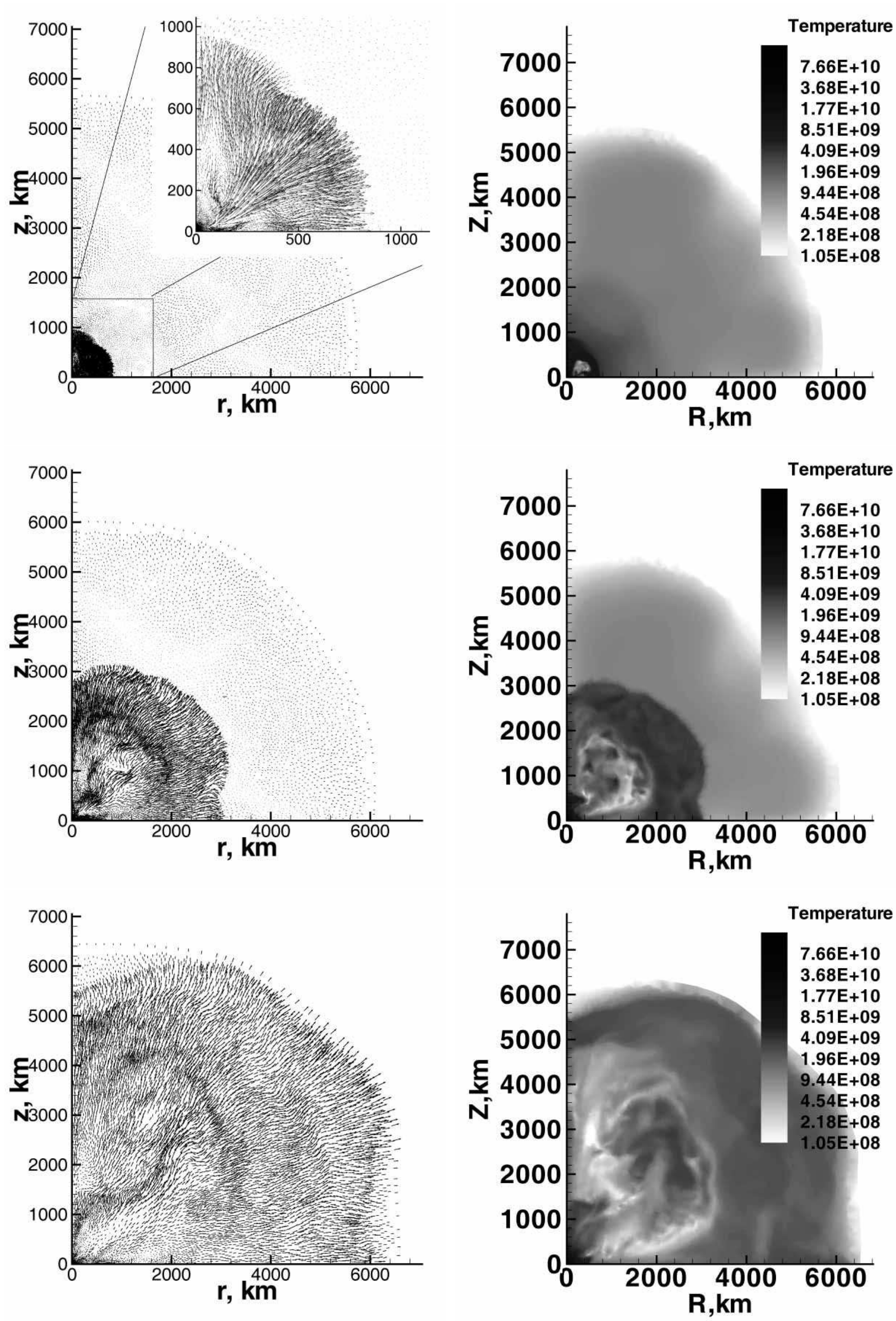

Figure 4. The distribution of the velocity field (left column) and the temperature (right column) for the time moments $t = 0.075,\ 0.1,\ 0.25s$ for the initial *dipole*-like magnetic field

Taking into account symmetry assumptions $(\dfrac{\partial}{\partial\varphi} = 0)$, the divergency of the tensor $\mathbf{H}\otimes\mathbf{H}$ can be presented in the following form:

$$\nabla\cdot(\mathbf{H}\otimes\mathbf{H}) = \begin{pmatrix} \dfrac{1}{r}\dfrac{\partial(rH_rH_r)}{\partial r} + \dfrac{\partial(H_zH_r)}{\partial z} - \dfrac{1}{r}H_\varphi H_\varphi \\ \dfrac{1}{r}\dfrac{\partial(rH_rH_\varphi)}{\partial r} + \dfrac{\partial(H_zH_\varphi)}{\partial z} + \dfrac{1}{r}H_\varphi H_r \\ \dfrac{1}{r}\dfrac{\partial(rH_rH_z)}{\partial r} + \dfrac{\partial(H_zH_z)}{\partial z} \end{pmatrix}.$$

Axial symmetry $(\dfrac{\partial}{\partial\varphi} = 0,\quad r \geq 0)$ and symmetry to the equatorial plane $(z \geq 0)$ are assumed. The problem is solved in the restricted domain [5]. At $t = 0$ the domain is restricted by the rotational axis $r \geq 0$, equatorial plane $z \geq 0$, and outer boundary of the star where the density of the matter is zero, while poloidal components of the magnetic field H_r and H_z can be non-zero.

We assume axial and equatorial symmetry ($r \geq 0,\ z \geq 0$). On the rotational axis ($r = 0$) the following boundary conditions are defined: $(\nabla\Phi)_r = 0$, $v_r = 0$. On the equatorial plane ($z = 0$) the boundary conditions are: $(\nabla\Phi)_z = 0$, $v_z = 0$. On the outer boundary (boundary with vacuum) the following condition is defined: $P_{\text{outer boundary}} = 0$.

The equation of state, expression for the internal energy and formula for neutrino losses are the same as in [3].

At the initial moment we start with rigidly rotating sphere of $1.2M_\odot$ mass without magnetic field [2]. As first stage we calculate a rotating core collapse and formation of the protoneutron star. The ratios between the initial rotational and gravitational energies and between the internal and gravitational energies of the star are the following:

$$\frac{E_{rot}}{E_{grav}} = 0.0057, \qquad \frac{E_{int}}{E_{grav}} = 0.727.$$

During the collapse the bounce shock appears and moves outwards. The shock leads to the ejection of $\approx 2.9\times10^{48}$erg of energy. The amount of the ejected energy is too small for the explanation of the supernova explosion.

For the simulations we used completely conservative operator-difference scheme on triangular Lagrangian grid (Fig.2) of variable structure [6].

Results of the 2D simulations of the magnetorotational supernova are qualitatively different from 1D results. In the 2D case the magnetorotational instability (MRI) appears, leading to the exponential growth of all components of the magnetic field. MRI significantly reduce the time for the magnetorotational explosion. In the paper [3] a toy model for the explanation of MRI development in the magnetorotational supernova was suggested.

Due to MRI the dependence of the explosion time on the strength of the initial magnetic field can be expressed by the approximate formula: $t_{expl} \approx |log(\alpha)|$, where $\alpha = \dfrac{E_{mag}}{E_{grav}}$ is a relation between initial magnetic and gravitational energies.

In the 1D case the development of MRI is not possible due to the restricted number of the freedom degrees.

The results of 2D simulations [3,17] show that the magnetorotational mechanism allows to produce $0.5-0.6\cdot10^{51}$ ergs energy of explosion. These values of SN explosion energy correspond to estimations made from core collapse SN observations.

The shape of the magnetorotational explosion qualitatively depends on the configuration of the initial magnetic field. For the initial quadrupole like configuration [3] the explosion develops mainly near equatorial plane (Fig.3). The dipole like initial magnetic field [17] leads to the formation and development of mildly collimated axial jet (Fig.4).

3 Simulation of MR supernovae in 3D case

3D models of the magnetorotational supernova are the more realistic, and have no constraints connected with the symmetry assumptions.

3D models allow us to simulate the magnetorotational supernova explosion in the case when rotational axis and axis of dipole magnetic field (if dipole is taken as initial magnetic field) do not coincide (inclined rotator).

The application of numerical method in Lagrangian variables, similar to the method used for the 2D case, leads in 3D case to serious difficulties.

In 2D case the matter of the star is slipping in φ direction. To produce the magnetorotational explosion the protoneutron star has to make thousands of revolutions. The rotation of the matter in the outer layers of the protoneutron star is highly differential. If 3D Lagrangian grid consisting of tetrahedrons would be applied for the simulations, then in the region of strong differential rotation the grid would require reconstruction almost at every time step. The reconstruction of the grid leads to the interpolation of the grid functions to a new grid structure. Frequent application of the grid reconstruction procedure and interpolation of grid functions for the same parts of the Lagrangian grid can lead to the significant perturbation of the solution of initial set of MHD equations with self gravitation.

One of the possible ways to simulate magnetorotational supernova in 3D case is to apply methods based on the unstructured grids of Dirichlet cells (see for example [18]). This type of methods can be effectively applied for the simulations of the different types of gas dynamical flows, but the procedure of the construction of the grid of Dirichlet cells is rather expensive, especially in 3D case.

Another method widely applied for the simulations of astrophysical problems is Smooth Particle Hydrodynamics (SPH) [13, 16] method. Codes based on the SPH approach can be easily developed, but to achieve a high accuracy in simulations SPH requires huge number of particles. The simulation of the problems of gravitational gas dynamics using SPH leads to the concentration of the particles near the gravitational center, on the periphery of the computational domain the number of particles is rather small and it leads to the significant loss of the accuracy of the results of simulations.

One of the most suitable approaches for the simulations of the explosive magnetorotational phenomena in 3D case is an application of the numerical methods in Eulerian variables based on the solution of the decomposition of discontinuity (Riemann solver) problem. This type of methods was successfully applied for the solution of the different astrophysical magnetorotational problems. Application of the Eulerian grid does not require grid reconstruction and interpolation of grid functions. The methods of this type are described in [14]. The methods based on the approximate MHD Riemann solver in Eulerian variables are the most suitable for the simulations of the explosive magnetorotational phenomena

For the simulations of astrophysical magnetorotational explosive phenomena it is important to calculate gravitational potential with sufficient accuracy. The procedure of the calculation of the gravitational potential is rather expensive (up to 40% of the computer time for the time step).

For our simulations we plan to apply Adaptive Mesh Refinement (AMR) approach. The adaptive refinement and rarefaction of the grid can increase the accuracy of the calculations significantly with the reasonable number of the grid points. We expect to apply AMR using two approaches. First one is a construction of the hierarchical tree which root is our initial 3D grid [1,7]. The second approach consists in construction of the rectangular (for the 2D case) [8] or parallelepiped (for the 3D case) patches consisting of specially chosen association of the cells of one level.

4 Acknowledgments

GSBK and SGM would like to thank Russian Foundation for Basic Research(RFBR) for the partial support of the work in the frame of the grants 05-02-17697, 06-02-91157, 06-02-90864. The work of BPR and GVS was supported by the joint project between Academy of Sciences of Moldova and RFBR in the frame of the project 06.02CRF.

References

[1] Aftersmoth M.J., Berger M.J., Melton J.E. J. AIAA, 1998, N 36(6), p. 952.

[2] Ardeljan N.V., Bisnovatyi-Kogan G.S., Kosmachevskii K.V., Moiseenko S.G. Astrophysics, 2004, **47**, 1.

[3] Ardeljan N.V., Bisnovatyi-Kogan G.S., Moiseenko S.G. MNRAS, 2005, **359**, p. 333.

[4] Ardelyan N.V., Bisnovatyi-Kogan G.S., Popov Yu.P. Astron. Zh., 1979, **56**, p. 1244.

[5] Ardelyan N.V., Bisnovatyi-Kogan G.S., Popov Yu.P., Chernigovsky S.V. Astron. Zh., 1987, **64**, p. 761.

[6] Ardeljan N.V., Kosmachevskii K.V. Comput. Math. Modelling, 1995, **6**, p. 209.

[7] Balsara D.S., Norton, G.D. Parallel Computing, 2001, **27**, p. 37.

[8] Berger M., Rigoutsos I. IEEE Trans. System Man Cybernetics, 1991, **21**, p. 61.

[9] Bisnovatyi-Kogan G.S. Astron. Zh., 1970, **47**, p. 813.

[10] Bisnovatyi-Kogan G.S. Stellar Physics, 2 volumes, Springer, 2001.

[11] Bisnovatyi-Kogan G.S., Popov Yu.P., Samochin, A.A. Astrophys. and Space Sci., 1976, **41**, p. 321.

[12] Buras R., Rampp M., Janka H.Th., Kifonidis K. Phys. Rev. Lett., 2003, **90**, p. 241101.

[13] Gingold R.A., Monaghan J.J. MNRAS, 1977, **181**, p. 375.

[14] Kulikovsky A.G., Pogorelov N.V., Semenov A.Yu. *Mathematical problems of the numerical solution of sets of hyperbolic equations*. Moskva, FIZMATLIT, 2001.

[15] Leblanc L.M., Wilson J.R. ApJ, 1970, **161**, p. 541.

[16] Lucy L.B. ApJ, 1977, **82**, p. 1013.

[17] Moiseenko S.G., Bisnovatyi-Kogan G.S., Ardeljan N.V. MNRAS, 2006, **370**, p. 501.

[18] Pushkina I.G., Tishkin V.F. Mathematical modeling, 2000, **12**, p. 4.

G.S. Bisnovatyi-Kogan, S.G. Moiseenko
Space Research Institute
Russian Academy of Sciences
Profsoyuznaya str. 84/32
Moscow, Russia
E-mail: *gkogan@iki.rssi.ru, moiseenko@iki.rssi.ru*

Received April 23, 2007

B.P. Rybakin, G.V. Secrieru
Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
str. Academiei 5, Chisinau MD-2028 Moldova
E-mail: *rybakin@math.md, secrieru@cc.acad.md*

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 64–72
ISSN 1024–7696

About Quasiconformal Maps in Finsler Spaces

Veronica-Teodora Borcea

Abstract. We consider a constant C which measures the deviation of the Finsler metric from a Riemannian metric and we prove that the problem of the existence of quasiconformal mappings between Finsler spaces can be reduced to the same problem between Riemann spaces.

Mathematics subject classification: 30C65.
Keywords and phrases: Characteristic function, quasiconformal map, Finsler metric, regular atlas.

1 Introduction

The quasiconformal mappings represent a generalization of the conformal transformations. It is known that there exist different equi- valent definitions for the conformal transformations, most of these using some conformal invariants (modulus of the ring or a family of arcs, angles, infinitesimal circles,...) or, as the solutions of a Cauchy-Riemann system.

The conformal transformations were used for the modelling, sometimes with approximation, of some phenomena. For example, in the hydrodynamic, where were considered "ideal fluids" (incompressible and not viscous) and their flow was without whirlpools.

The definitions of quasiconformal mappings appeared, naturally, from the corresponding definitions of the conformal transformations, for example, by substituting quasi-invariance for the invariance.

K. Suominen extends the study of the quasiconformality to the finite dimensional Riemannian manifolds [1], and P. Caraman to the Riemann-Wiener manifolds [2].

The study of quasiconformality was extended by us to the infinit dimensional Riemannian manifolds and to the Finsler spaces [3, 4].

In 1982 M. Nakai and H. Tanaka proved the existence of quasiconformal mappings between finite dimensional Riemannian manifolds [5].

In this paper we associate to a Finsler space a constant C, which measures the deviation of the Finslerian metric from a riemannian metric. By using this constant we establish an inequality between the Finslerian and Riemannian characteristic functions and we prove that the problem of the existence of quasiconformal mappings between finite dimensional Finsler spaces can be reduced to the same problem between finite dimensional Riemann manifolds. The main result is

Theorem A. *A homeomorphism f is Finslerian quasiconformal iff f is Riemannian quasiconformal.*

2 Regular atlases

Let us consider M a n-dimensional, connected, paracompact, orientable, C^∞-differentiable manifold and $L : TM \to \mathbb{R}$ a Finsler metric on M ($TM = \bigcup_{x\in M} T_xM$ denotes the tangent bundle of M and T_xM the tangent space at $x \in M$).

The restriction of L to T_xM, $L(x,.) : T_xM \to \mathbb{R}$, is a norm, generally non-Hilbertian, denoted by $\|\cdot\|$ and

$$L^2(x,X) = a_{ij}(x,X)\,X^iX^j,$$

for every $X = X^i\frac{\partial}{\partial x^i} \in T_xM$, where

$$a_{ij}(x,X) = \frac{1}{2}\frac{\partial^2 L^2(x,X)}{\partial X^i \partial X^j}$$

are homogeneous functions of degree zero with respect to X. We have

$$\|X\| = L(x,X) = \sqrt{a_{ij}(x,X)\,X^iX^j}.$$

The manifold M is a metric space with the geodesic metric

$$d(x,y) = \inf\{\ell(\gamma)\,/\,\gamma \in \Gamma\},$$

where Γ is the set of all differentiable arcs joining x with y and $\ell(\gamma)=\int_0^1 \left\|\frac{d\gamma}{dt}\right\| dt.$

The geodesics of M are the autoparalleles of nonlinear Cartan connection ∇ and their equation is $\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0$.

If $\gamma_X(t)$ is the geodesic with the initial condition (x,X), then $\gamma_X(t) = \gamma_{\alpha X}(\alpha^{-1}t)$ for every $\alpha \in \mathbb{R}\setminus\{0\}$ and the map $\exp_x : V_x \to M$, $\exp_x X = \gamma_X(1)$ satisfies $\|Y\| = d(x, \exp_x Y)$ for every $Y \in V_x$ (V_x is the maximum domain where $\exp_x$ is a diffeomorphism).

Lemma 1. *If M is a Finsler space, then for every $\varepsilon \in (0,\infty)$ there exists $r_x = r(x,\varepsilon) \in (0,\infty)$ such that $\exp_x$ is a $(1+\varepsilon)$-isometry on the ball $B(0_x, r_x) \subset T_xM$, that is*

$$(1+\varepsilon)^{-1}\|Y - Z\| \le d(\exp_x Y, \exp_x Z) \le (1+\varepsilon)\|Y - Z\|, \tag{1}$$

for every $Y, Z \in B(0_x, r_x)$.

Proof. For every $\varepsilon \in (0,\infty)$ let us consider $r_x \in (0,\infty)$ such that the inequality $\|T_x \exp_x\| \le 1+\varepsilon$ is satisfied, for every $X \in B(0_x, r_x)$. Let us consider $\gamma : [0,1] \to T_xM$, $\gamma(t) = tY + (1-t)Z$ and $\gamma_1(t) = (\exp_x \circ \gamma)(t)$. We have

$$
\begin{aligned}
d\left(\exp_x Y, \exp_x Z\right) &\leq \ell\left(\gamma_1\right) = \int_0^1 \left\| \frac{d\gamma_1}{dt} \right\| dt \leq \int_0^1 \left\| \frac{d\gamma}{dt} \right\| \left\| T_{\gamma(t)} \exp_x \right\| dt \leq \\
&\leq (1+\varepsilon)\,\ell\left(\gamma\right) = (1+\varepsilon)\left\| Y - Z \right\|.
\end{aligned}
$$

Analogously, results the left-hand side of (1).

Remark 1. If $Z = 0_x$ and $Y = \exp_x^{-1} y$, we obtain

$$
d\left(x, y\right) = d\left(\exp_x 0_x, \exp_x Y\right) = \left\| Y \right\|
$$

and hence $\exp_x \left(B\left(0_x, r_x\right)\right)$ is the geodesic ball $B\left(x, r_x\right)$, consequently $\exp_x^{-1}$ is a $(1+\varepsilon)$-isometry, too.

Let us consider the homeomorphism $\varphi_x : B\left(0, \alpha_x\right) \to B\left(x, r_x\right)$, $\alpha_x \in (0, \infty)$, such that $\varphi_x(0) = x$ ($B\left(0, \alpha_x\right) = B_x$ is the ball with center 0 and radius α_x in $\mathbb{R}^n$).

The pair $h_x = \left(B_x, \varphi_x\right)$ is called *φ-chart* at x and the set $A = \{h_x / \ x \in M\}$ is called *φ-atlas* on M.

Obviously, for every $\varepsilon \in (0, \infty)$, $h_x^\varepsilon = \left(B\left(0_x, r_x\right), \exp_x\right)$ is an exp-chart and $A^\varepsilon = \{h_x^\varepsilon / \ x \in M\}$ is an exp-atlas, called the *atlas of geodesic balls*.

To any φ-atlas we can associate the function $k_A : M \to [1, \infty]$

$$
k_A(x) = \limsup_{\alpha \to 0} \frac{\sup\left\{d\left(x, y\right) / \ y \in \varphi_x\left(S\left(0, \alpha\right)\right)\right\}}{\inf\left\{d\left(x, y\right) / \ y \in \varphi_x\left(S\left(0, \alpha\right)\right)\right\}},\ \alpha \in \left(0, \alpha_x\right),
$$

called the *parameter* of A; $k_A(x)$ is called the *parameter* of the φ-chart h_x (we shall sometimes omit the subscript A if the choice of the atlas is clear from context).

If $k(x) < \infty$ we say that the φ-chart h_x is *k-regular*. If all the φ-charts of A are k-regular, we say that the atlas A is *k-regular*.

If φ_x is a conformal homeomorphism we say that h_x is a *conformal chart*. The atlas A is said to be *conformal* if its charts are conformal. In this case we obtain $k(x) = 1$ and so, any conformal atlas has the parameter $k = 1$. Particularly, the atlas of geodesic balls A^ε has the parameter $k = 1$.

Let $f : D \to \tilde{D}$ be a homeomorphism, where $D,\ \tilde{D}$ are domains in M.

If A is a φ-atlas on D, we can consider the $\tilde{\varphi}$-atlas $\tilde{A} = \left\{\tilde{h}_{\tilde{x}} \ / \ \tilde{x} \in \tilde{M}\right\}$, on $\tilde{D}$, with $\tilde{h}_{\tilde{x}} = \left(B_{\tilde{x}}, \tilde{\varphi}_{\tilde{x}}\right)$, $B_{\tilde{x}} = B\left(0, \alpha_{\tilde{x}}\right)$, $\tilde{\varphi}_{\tilde{x}} = f \circ \varphi_x$, $\tilde{x} = f(x)$ and $\alpha_{\tilde{x}}$ chosen such that $\tilde{\varphi}_{\tilde{x}}\left(B_{\tilde{x}}\right) \subset B\left(\tilde{x}, r_{\tilde{x}}\right)$, ($B\left(\tilde{x}, r_{\tilde{x}}\right)$ is the geodesic ball in $\tilde{D}$ where $\exp_{\tilde{x}}^{-1}$ is $(1+\varepsilon)$-isometry).

$\tilde{A}$ and $\tilde{h}_{\tilde{x}}$ are called, respectively, *$\tilde{\varphi}$-atlas* and *$\tilde{\varphi}$-chart induced by f*.

The parameter of $\tilde{A}$ will be

$$
\tilde{k}_{\tilde{A}}(\tilde{x}) = \limsup_{\alpha \to 0} \frac{\sup\left\{\tilde{d}\left(\tilde{x}, \tilde{y}\right) / \ \tilde{y} \in \tilde{\varphi}_{\tilde{x}}\left(S\left(0, \alpha\right)\right)\right\}}{\inf\left\{\tilde{d}\left(\tilde{x}, \tilde{y}\right) / \ \tilde{y} \in \tilde{\varphi}_{\tilde{x}}\left(S\left(0, \alpha\right)\right)\right\}},\ \alpha \in \left(0, \alpha_{\tilde{x}}\right).
$$

Generally, if A is k-regular, it does not result that $\tilde{A}$ is $\tilde{k}$-regular.

The homeomorphism f is called $k\tilde{k}-$*regular* if there exists a $\varphi-$atlas A, $k-$regular, on D such that the $\tilde{\varphi}-$atlas, $\tilde{A}$, induced by f is $\tilde{k}-$regular on $\tilde{D}$.

The function

$$q_f : D \to [1, \infty], \; q_f(x) = \inf \left\{ k(x) \cdot \tilde{k}(\tilde{x}) \right\},$$

where infimum is taken over all $k-$regular $\varphi-$atlases on D, is called the *Finslerian characteristic function* of f.

It follows that f is $k\tilde{k}-$regular if $q_f(x) < \infty$, for every $x \in D$.

Let us consider a $f-$isomorphism of vector bundles $T : TD \to T\tilde{D}$. The restriction, T_x, of T to $T_x D$, $T_x : T_x D \to T_{\tilde{x}} \tilde{D}$, $\tilde{x} = f(x)$, is an isomorphism of linear spaces, hence the image by T_x of $B(0_x, \alpha_x)$ is an ellipsoid $\tilde{E}_0(T_x) \subset B(0_{\tilde{x}}, r_{\tilde{x}}) \subset T_{\tilde{x}} \tilde{D}$, where $r_{\tilde{x}} = \alpha_x \|T_x\|$. We can consider α_x such that $\exp_{\tilde{x}}$ is $(1+\varepsilon)$-isometry on $B(0_{\tilde{x}}, r_{\tilde{x}})$.

It follows that $\tilde{h}_{\tilde{x}} = (B(0_x, \alpha_x), T_x)$ is a T-chart at $0_{\tilde{x}} \in T_{\tilde{x}} \tilde{D}$ and so, we can consider a $\tilde{T}$-chart on $\tilde{D}$, induced by $\exp_{\tilde{x}}$, $\tilde{H}_{\tilde{x}} = \left(B(0_x, \alpha_x), \tilde{T}_{\tilde{x}} \right)$, $\tilde{T}_{\tilde{x}} = \exp_{\tilde{x}} \circ T_x$, and, in such a way, we obtain a $\tilde{T}$-atlas $\tilde{A} = \left\{ \tilde{H}_{\tilde{x}} \;/\; \tilde{x} \in \tilde{D} \right\}$, called *atlas of geodesic ellipsoids*.

The geodesic ellipsoid $E_0(T_x) = \exp_{\tilde{x}} \left(\tilde{E}_0(T_x) \right)$ has the same extreme semiaxes as $\tilde{E}_0(T_x)$ ($\exp_{\tilde{x}}$ behaves as an isometry for the distances measured from $0_{\tilde{x}}$).

Let us consider $\tilde{E}_\alpha(T_x) = T_x(S(0_x, \alpha))$, $\alpha \in (0, \alpha_x)$, and

$$P_\alpha = \left\{ \tilde{d}\left(0_{\tilde{x}}, \tilde{Y}\right) \;/\; \tilde{Y} \in \tilde{E}_\alpha(T_x) \right\} = \left\{ \left\| \tilde{Y} \right\| \;/\; \|Y\| = \alpha \right\}.$$

The extreme semiaxes of $\tilde{E}_\alpha(T_x)$ are given by

$$\tilde{a}_0(\alpha, \tilde{x}) = \inf P_\alpha = \alpha \left\| T_x^{-1} \right\|^{-1},$$

$$\tilde{a}_1(\alpha, \tilde{x}) = \sup P_\alpha = \alpha \|T_x\|.$$

The function

$$p_T : \tilde{M} \to \mathbb{R}, \; p_T(\tilde{x}) = \frac{\tilde{a}_1(\alpha, \tilde{x})}{\tilde{a}_0(\alpha, \tilde{x})} = \|T_x\| \left\| T_x^{-1} \right\|$$

is called *the principal characteristic parameter* of the atlas of geodesic ellipsoids.

Arguing as above for f^{-1} and T^{-1}, we obtain

$$p_{T^{-1}}(x) = \|T_x\| \left\| T_x^{-1} \right\| = p_T(\tilde{x}).$$

The parameter of the atlas of geodesic ellipsoids is

$$\tilde{k}(\tilde{x}) = \limsup_{\alpha \to 0} \frac{\tilde{a}_1(\alpha, \tilde{x})}{\tilde{a}_0(\alpha, \tilde{x})} = p_T(\tilde{x}).$$

Lemma 2. *If $f : D \to \tilde{D}$ is a differentiable homeomorphism at $x \in D$ with $T_x f$ bijective, then:*

a) f is $k\tilde{k}$-regular on D iff for every $x \in D$, $F_x = \exp_{\tilde{x}}^{-1} \circ f \circ \exp_x$ is $k\tilde{k}$-regular at $0_x \in T_x D$;

b) F_x is $k\tilde{k}$-regular at 0_x iff $T_x f$ is $k\tilde{k}$-regular at 0_x;

c) $q_f(x) = q_{T_x f}(0_x) = q_{F_x}(0_x)$.

Proof. a) Let A be a φ-atlas k-regular on D, such that the induced $\tilde{\varphi}$-atlas $\tilde{A}$ is $\tilde{k}$-regular on $\tilde{D}$. We consider $h_x = (B_x, \varphi_x) \in A$ and $\tilde{h}_{\tilde{x}} = (B_{\tilde{x}}, \tilde{\varphi}_{\tilde{x}}) \in \tilde{A}$, where $\tilde{\varphi}_{\tilde{x}} = f \circ \varphi_x$. It follows that $H_x = (B_x, \phi_x)$, $\phi_x = \exp_x^{-1} \circ \varphi_x$ is k-regular and $\tilde{H}_{\tilde{x}} = \left(B_{\tilde{x}}, \tilde{\phi}_{\tilde{x}}\right)$, $\tilde{\phi}_{\tilde{x}} = \exp_{\tilde{x}}^{-1} \circ \tilde{\varphi}_{\tilde{x}}$ is $\tilde{k}$-regular.

Because

$$\tilde{\phi}_{\tilde{x}} = \exp_{\tilde{x}}^{-1} \circ \tilde{\varphi}_{\tilde{x}} = \exp_{\tilde{x}}^{-1} \circ f \circ \exp_x \circ \exp_x^{-1} \circ \varphi_x = F_x \circ \phi_x,$$

it follows that $\tilde{H}_{\tilde{x}}$ is the $\tilde{\phi}_{\tilde{x}}$-chart induced by F_x and so, F_x is $k\tilde{k}$-regular.

For $f = \exp_{\tilde{x}} \circ F_x \circ \exp_x^{-1}$, arguing as above, we obtain the sufficiency. In addition, we obtain

$$q_f(x) = q_{F_x}(0_x). \tag{2}$$

b) Let us consider the k-regular chart $H_x = (B_x, \phi_x)$ with the parameter

$$k(x) = \limsup_{\alpha \to 0} \frac{\sup P_\alpha}{\inf P_\alpha} < \infty, \ P_\alpha = \{\|X\| \ / \ X \in \phi_x(S(0,\alpha))\} \tag{3}$$

and the chart, $\bar{H}_{\tilde{x}} = \left(B_x, \bar{\phi}_{\tilde{x}}\right)$, $\bar{\phi}_{\tilde{x}} = T_x f \circ \phi_x$, induced by $T_x f$, for which

$$\bar{k}(\tilde{x}) = \limsup_{\alpha \to 0} \frac{\sup P'_\alpha}{\inf P'_\alpha}, \ P'_\alpha = \{\|(T_x f)(X)\| \ / \ X \in \phi_x(S(0,\alpha))\}. \tag{4}$$

For the $\tilde{\phi}$-chart, $\tilde{H}_{\tilde{x}} = \left(B_x, \tilde{\phi}_{\tilde{x}}\right)$, $\tilde{\phi}_{\tilde{x}} = F_x \circ \phi_x$, induced by F_x, we have

$$\tilde{k}(\tilde{x}) = \limsup_{\alpha \to 0} \frac{\sup P''_\alpha}{\inf P''_\alpha}, \ P''_\alpha = \{\|F_x(X)\| \ / \ X \in \phi_x(S(0,\alpha))\}. \tag{5}$$

Since $T_x f = DF_x(0_x)$, it follows that

$$F_x(X) = (T_x f)(X) + \varepsilon_x(X)\|X\|, \ \varepsilon_x : T_x M \to T_x M, \ \lim_{X \to 0_x} \|\varepsilon_x(X)\| = 0.$$

We have

$$\limsup_{\alpha \to 0} \left(\sup \bar{P}_\alpha\right) = 0, \ \bar{P}_\alpha = \{\|\varepsilon_x(X)\| \ / \ X \in \phi_x(S(0,\alpha))\}. \tag{6}$$

Since $T_x f$ is an isomorphism of topological vector spaces, we get

$$\|(T_x f)(X)\| \geq \frac{\|X\|}{\left\|(T_x f)^{-1}\right\|}.$$

It follows that

$$\sup P'_\alpha \geq \frac{\sup P_\alpha}{\left\|(T_x f)^{-1}\right\|};\ \ \inf P'_\alpha \geq \frac{\inf P_\alpha}{\left\|(T_x f)^{-1}\right\|}. \tag{7}$$

We have

$$\|T_x f(X)\| - \|\varepsilon_x\| \|X\| \leq \|F_x(X)\| \leq \|T_x f(X)\| + \|\varepsilon_x\| \|X\|$$

and so

$$\begin{aligned} \sup P'_\alpha - \sup P_\alpha \sup \bar{P}_\alpha &\leq \sup P''_\alpha \leq \sup P'_\alpha + \sup P_\alpha \sup \bar{P}_\alpha, \\ \inf P'_\alpha - \sup P_\alpha \sup \bar{P}_\alpha &\leq \inf P''_\alpha \leq \inf P'_\alpha + \sup P_\alpha \sup \bar{P}_\alpha. \end{aligned}$$

We obtain

$$\begin{cases} \bar{k}(\tilde{x}) \limsup\limits_{\alpha \to 0} \dfrac{1 - \left\|(T_x f)^{-1}\right\| \sup \bar{P}_\alpha}{1 + \dfrac{\sup P_\alpha}{\inf P_\alpha} \dfrac{\inf P_\alpha}{\inf P'_\alpha} \sup \bar{P}_\alpha} \leq \tilde{k}(\tilde{x}), \\ \tilde{k}(\tilde{x}) \leq \bar{k}(\tilde{x}) \limsup\limits_{\alpha \to 0} \dfrac{1 + \left\|(T_x f)^{-1}\right\| \sup \bar{P}_\alpha}{1 - \dfrac{\sup P_\alpha}{\inf P_\alpha} \dfrac{\inf P_\alpha}{\inf P'_\alpha} \sup \bar{P}_\alpha}. \end{cases} \tag{8}$$

From (3), (6), (7) and (8) it follows that $\tilde{k}(\tilde{x}) = \bar{k}(\tilde{x})$, which proves the assertion b) and we have

$$q_{F_x}(0_x) = q_{T_x f}(0_x). \tag{9}$$

c) It results from (2) and (9).

Lemma 3. *If $T : V \to \tilde{V}$ is an isomorphism of n-dimensional normed vector spaces, then T is $k\tilde{k}$-regular with $k(X) = \|T\| \left\|T^{-1}\right\|$, $\tilde{k}\left(\tilde{X}\right) = 1$ and $q_T(X) = p_{T^{-1}}(X)$, for every $X \in V$.*

Proof. Let us consider the $\tilde{k}$-regular $\tilde{\phi}$-chart, $\tilde{H}_{\tilde{X}} = \left(B\left(0_{\tilde{X}}, 1\right), \tilde{\phi}_{\tilde{X}}\right)$, $\tilde{\phi}_{\tilde{X}} \tilde{Y} = \tilde{X} + \tilde{Y}$, $\tilde{X} = TX$. It follows that $\tilde{k}\left(\tilde{X}\right) = 1$. The map T^{-1} induces a ϕ-chart $H_X = \left(B\left(0_{\tilde{X}}, 1\right), \phi_X\right)$, $\phi_X = T^{-1} \circ \tilde{\phi}_{\tilde{X}}$, with $k(X) = \|T\| \left\|T^{-1}\right\| < \infty$ and so H_X is k-regular. Thus, we obtain that T is $k\tilde{k}$-regular with $k(X) = \|T\| \left\|T^{-1}\right\|$, $\tilde{k}\left(\tilde{X}\right) = 1$. We have $p_{T^{-1}}(X) = k(X) \tilde{k}\left(\tilde{X}\right) = \|T\| \left\|T^{-1}\right\|$ and so

$$q_T(X) \le p_{T^{-1}}(X). \tag{10}$$

Let us consider a k-regular φ-chart, $h_X = (B(0,1), \varphi_X)$ and the $\tilde{\varphi}$-chart induced by T, $\tilde{h}_{\tilde{X}} = \left(B(0,1), \tilde{\varphi}_{\tilde{X}}\right)$, $\tilde{\varphi}_{\tilde{X}} = T \circ \varphi_X$, with the parameter $\tilde{k}\left(\tilde{X}\right)$. We have two cases:

1) $k(X) \ge p_{T^{-1}}(X)$, which implies that $k(X)\tilde{k}\left(\tilde{X}\right) \ge p_{T^{-1}}(X)$. It follows that $q_T(X) \ge p_{T^{-1}}(X)$ and from (10) we obtain $q_T(X) = p_{T^{-1}}(X)$.

2) $k(X) < p_{T^{-1}}(X)$. We denote by $\sigma_\alpha = \varphi_X(S(0,\alpha))$ and

$$r_0 = \inf\{\|Y - X\| \ / \ Y \in \sigma_\alpha\}, r_1 = \sup\{\|Y - X\| \ / \ Y \in \sigma_\alpha\};$$

it follows that $\sigma_\alpha \subset \bar{B}(X, r_1) - B(X, r_0)$. Taking $t_0 = r_0\|T\|$ and $t_1 = r_1\left\|T^{-1}\right\|^{-1}$, it follows that the ellipsoid $E_{t_0}\left(T^{-1}\right) = T^{-1}\left(S\left(\tilde{X}, t_0\right)\right)$ has the minimum semiaxis $a_0(t_0, X) = r_0 = t_0\|T\|^{-1}$ and $E_{t_1}\left(T^{-1}\right) = T^{-1}\left(S\left(\tilde{X}, t_1\right)\right)$ has the maximum semiaxis $a_1(t_1, X) = r_1 = t_1\left\|T^{-1}\right\|$. We obtain

$$k(X) = r_1 r_0^{-1} = t_1 t_0^{-1}\|T\|\left\|T^{-1}\right\| = t_1 t_0^{-1} p_{T^{-1}}(X)$$

and since $k(X) < p_{T^{-1}}(X)$ it follows that $t_1 < t_0$. We have

$$\begin{aligned} t_0 &\le \sup\left\{\left\|TY - \tilde{X}\right\| \ / \ Y \in \sigma_\alpha\right\}, \\ t_1 &\ge \inf\left\{\left\|TY - \tilde{X}\right\| \ / \ Y \in \sigma_\alpha\right\}, \end{aligned}$$

hence $\tilde{k}\left(\tilde{X}\right) \ge t_0 t_1^{-1} = (k(X))^{-1} p_{T^{-1}}(X)$ and then $k(X)\tilde{k}\left(\tilde{X}\right) \ge p_{T^{-1}}(X)$, which implies $q_T(X) \ge p_{T^{-1}}(X)$. By using (10) we obtain $q_T(X) = p_{T^{-1}}(X)$.

Remark 2. From Lemmas 2 and 3, it follows that if $f : D \to \tilde{D}$ is a differentiable homeomorphism at x with $J_f(x) \ne 0$, then f is $k\tilde{k}$-regular at x and $q_f(x) = \|T_x f\|\left\|(T_x f)^{-1}\right\|$.

3 The proof of main result

We consider $\mathcal{X}(M)$ the Lie algebra of the tangent fields on M and $\mathcal{X}_0(M) = \{V \ / \ V \in \mathcal{X}(M), \ \|V(x)\| = 1, \ \forall x \in M\}$.

The matrix $a_V = [a_{ij}(x, V)]$, for a fixed $V \in \mathcal{X}_0(M)$, is a Riemannian metric on M and the map

$$\|\cdot\|_V : T_x M \to \mathbb{R}, \ \|X\|_V = \sqrt{a_{ij}(x, V) X^i X^j}$$

is an Euclidean norm in $T_x M$.

Because the norms $\|\cdot\|$ and $\|\cdot\|_V$ are equivalent, there exists the map $C_V : M \to [1, \infty)$ such that

$$C_V^{-1}(x)\|X(x)\|_V \leq \|X(x)\| \leq C_V(x)\|X(x)\|_V, \ \forall X \in \mathcal{X}(M).$$

For every $V \in \mathcal{X}_0(M)$ we consider

$$P(x,V) = \{C_V(x) \ / \ C_V^{-1}(x)\|X(x)\|_V \leq \|X(x)\| \leq C_V(x)\|X(x)\|_V, \ \forall X \in \mathcal{X}(M)\}$$

and the map

$$C : M \to [1,\infty), \ C(x) = \inf\{Px,V)/V \in \mathcal{X}_0(M)\}.$$

It follows that for every $\varepsilon > 0$, there exists $V_\varepsilon \in \mathcal{X}_0(M)$ such that $C(x) \leq C_{V_\varepsilon}(x) < C(x) + \varepsilon$ and so, we can find $V_0 \in \mathcal{X}_0(M)$ which satisfies

$$C^{-1}(x)\|X(x)\|_{V_0} \leq \|X(x)\| \leq C(x)\|X(x)\|_{V_0}, \ \forall X \in \mathcal{X}(M).$$

If we consider $C = \sup\{C(x) \ / \ x \in M\} \in [1, \infty]$, we have

$$C^{-1}\|X(x)\|_{V_0} \leq \|X(x)\| \leq C\|X(x)\|_{V_0}, \ \forall X \in \mathcal{X}(M), \ \forall x \in M$$

hence, if $C = 1$ then L is a Riemannian metric and if $C > 1$, it is a Finsler metric, that is C measures the deviation of the Finsler metric from a riemannian metric.

In the following we suppose that $C \in (1, \infty)$ and we denote by $\|X(x)\|_0$ the norm $\|X(x)\|_{V_0}$.

If $f : D \to \tilde{D}$ is a non-degenerate differentiable homeomorphism at $x \in M$, then between the Riemannian characteristic function $q_f^0(x) = \|T_x f\|_0 \left\|(T_x f)^{-1}\right\|_0$ and the Finslerian characteristic function $q_f(x)$ we have the relation

$$C^{-4} q_f^0(x) \leq q_f(x) \leq C^4 q_f^0(x). \tag{11}$$

Lemma 4. *If $f : D \to \tilde{D}$ is a homeomorphism with q_f bounded in D, then f is almost everywhere (a.e.) differentiable (with respect to the Lebesgue measure) and $J_f(x) \neq 0$ a.e. in D.*

Proof. Let us consider the atlas of geodesic balls A^ε on D and $F_x = \exp_{\tilde{x}}^{-1} \circ f \circ \exp_x : B(0_x, r_x) \to T_{\tilde{x}}\tilde{D}$, $\tilde{x} = f(x)$. It results that $q_{F_x}(Y) \leq (1+\varepsilon)^4 q_f(y)$ for every $Y \in B(0_x, r_x)$, $y = \exp_x Y$. We obtain that q_{F_x} is bounded on $B(0_x, r_x)$, hence it is differentiable a.e. with $J_{F_x} \neq 0$ a.e.(see [6]). It follows that f is differentiable a.e. on $B(x, r_x)$ with $J_f \neq 0$ a.e. Since M is paracompact the assertion of theorem follows.

Definition. A homeomorphism $f : D \to \tilde{D}$ is called $K-$*Finslerian quasiconformal* in D, $(K - FQC), 1 \leq K < \infty$, if q_f is bounded in D and $q_f(x) \leq K$ a.e. in D.

If the Finsler metric on M is a Riemannian metric, we say that f is $K-$Riemannian quasiconformal in D, $(K-RQC)$.

From (11) we obtain:

If f is $K-FQC$ in D, then $C^{-4}q_f^0(x) \leq q_f(x) \leq K$ a.e. in D and hence $q_f^0(x) \leq C^4 K$. We obtain that f is $K_0 - RQC$ in D, with $K_0 = C^4 K$.

Analogously, we obtain that if f is $K-RQC$ in D, then f is $K_0 - FQC$ in D, with $K_0 = C^4 K$, hence the *Theorem A* is proved.

Remark 3. From *Theorem A* it follows that the existence of the quasiconformal mappings in Finsler spaces can be reduced to the existence of the quasiconformal mappings in Riemann spaces.

References

[1] Suominen K. *Quasiconformal Mappings in Manifolds.* Ann. Acad.Sci. Fenn., 1966, **393**, p. 5–39.

[2] Caraman P. *Module and p-module in an abstract Wiener space.* Rev. Roum. Math. Pures Appl., 1982, **27**, p. 551–599.

[3] Borcea V.T., Neagu A. *p-modulus and p-capacity in a Finsler space.* Math. Report, 2000, **52**, p. 431–439.

[4] Borcea V.T., Neagu A. *A class of homeomorphisms between the riemannian manifolds.* Rev. Roum. Math. Pures Appl., 1991, **36**, p. 323–332.

[5] Nakai M., Tanaka N. *Existence of quasiconformal mappings between riemannian manifolds.* Kodai Math. J., 1982, **5**, p. 122–131.

[6] Caraman P. *n-dimensional quasiconformal mappings.* Ed. Acad. Române, Bucureşti and Abacus Press, Tunbridge Wells (Kent) England, 1974.

Department of Mathematics
Technical University "Gh. Asachi" Iaşi
Romania

E-mail: *alneagu@gmail.com*

Received September 11, 2006

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 73–90
ISSN 1024–7696

On mixed LCA groups with commutative rings of continuous endomorphisms

Valeriu Popa

Abstract. Let $\mathcal{L}$ be the class of locally compact abelian (LCA) groups. For $X \in \mathcal{L}$, let $E(X)$ denote the ring of continuous endomorphisms of X. In this paper, we determine for certain subclasses $\mathcal{S}$ of $\mathcal{L}$ the groups $X \in \mathcal{S}$ such that $E(X)$ is commutative. The main results concern the case of mixed LCA groups.

Mathematics subject classification: Primary: 22B05; Secondary: 16W80.
Keywords and phrases: LCA groups, ring of continuous endomorphisms, commutativity.

1 Introduction

This paper is in continuation to the papers [14, 15] and [16] relating to LCA groups with commutative rings of continuous endomorphisms. We shall be mainly concerned with the case of mixed groups. The motivation for our work comes from a result of T. Szele and J. Szendrei. In [17], they have given among others a complete description of discrete mixed abelian groups without nonzero elements of infinite p-height for all relevant primes p, which have commutative endomorphism rings.

The main objective of the paper is to extend this result to the more general framework of all LCA groups. We also derive information about bounded order-by-discrete LCA groups with commutative rings of continuous endomorphisms.

2 Notation

In what follows we use the notation and terminology of [14, 15] and [16]. In addition, if $p \in \mathbb{P}$, $n \in \mathbb{N}_0$, and V is a closed subgroup of a group $X \in \mathcal{L}$, we let

$$p^{-n}V = \{x \in X | p^n x \in V\}.$$

For a subset S of $\mathbb{P}$, let

$$w_S(X) = \bigcap_{p \in S} \bigcap_{n \in \mathbb{N}} \overline{p^n X}.$$

Further, let $(X_i)_{i \in I}$ be a collection of topological groups. For $i \in I$, let U_i be an open subgroup of X_i. We denote by $\prod_{i \in I}^{loc}(X_i; U_i)$ the local product of $(X_i)_{i \in I}$

with respect to $(U_i)_{i\in I}$. Recall that, by definition, $\prod_{i\in I}^{loc}(X_i; U_i)$ is the cartesian product of the family $(X_i)_{i\in I}$, topologized by declaring all neighborhoods of zero in the topological group $\prod_{i\in I} U_i$ to be a fundamental system of neighborhoods of zero in $\prod_{i\in I}^{loc}(X_i; U_i)$ [3, Ch. III, §2, Exercice 26]. Clearly, the local direct product $\prod_{i\in I}(X_i; U_i)$ is open in $\prod_{i\in I}^{loc}(X_i; U_i)$. It is also clear that if each U_i is compact, then $\prod_{i\in I}^{loc}(X_i; U_i)$ is locally compact.

3 Groups with no elements of infinite topological S-height

In [17], T. Szele and J. Szendrei gave among other results a complete description of discrete, mixed, abelian groups with no elements of actually infinite height, which have commutative endomorphism rings. Their theorem reads:

Theorem 3.1 ([17], Theorem 2)**.** *Let X be a discrete mixed group in $\mathcal{L}$ with no elements of infinite $S(X)$-height, i. e. such that*

$$\bigcap_{p\in S(X)} \bigcap_{n\in\mathbb{N}} p^n X = \{0\}.$$

Then $E(X)$ is commutative if and only if X is isomorphic to an $S(X)$-pure subgroup of

$$\prod_{p\in S(X)}^{loc} \big(\mathbb{Z}(p^{n_p}); \{0\}\big)$$

containing

$$\bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p}),$$

where $n_p \in \mathbb{N}_0$ for all $p \in S(X)$.

Our aim here is to extend this theorem to more general groups in $\mathcal{L}$. But first we use it to obtain the solution to our problem in the case of compact groups in $\mathcal{L}$ having nontrivial connected component and dense torsion subgroup.

Corollary 3.2. *Let X be a compact group in $\mathcal{L}$ with $X \neq c(X) \neq \{0\}$ and $\overline{\sum_{p\in S(X)} t_p(X)} = X$. The endomorphism ring $E(X)$ is commutative if and only if X is topologically isomorphic to a quotient group of*

$$\Big(\prod_{p\in S(X)}^{loc} \big(\mathbb{Z}(p^{n_p}); \{0\}\big)\Big)^*$$

by a closed $S(X)$-pure subgroup contained in

$$c\Big(\Big(\prod_{p\in S(X)}^{loc} \big(\mathbb{Z}(p^{n_p}); \{0\}\big)\Big)^*\Big),$$

where $n_p \in \mathbb{N}_0$ for all $p \in S(X)$.

Proof. Since X is compact with $X \neq c(X) \neq \{0\}$ and $A(X^*; c(X)) = t(X^*)$ [8, (24.24)], it follows that X^* is discrete and mixed. Also, since $\overline{\sum_{p\in S(X)} t_p(X)} = X$, we conclude by [4, Proposition 3.3.3] and [8, (24.22)] that

$$\begin{aligned}\bigcap_{p\in S(X)} \bigcap_{n\in\mathbb{N}} p^n X^* &= A(X^*; \overline{\sum_{p\in S(X)} \sum_{n\in\mathbb{N}} X[p^n]}) \\ &= A(X^*; \overline{\sum_{p\in S(X)} t_p(X)}) = \{0\},\end{aligned}$$

so that X^* has no elements of infinite $S(X)$-height.

Let $G = \prod_{p\in S(X)}^{loc} (\mathbb{Z}(p^{n_p}); \{0\})$, and let Γ be a closed subgroup of G^* For $p \in S(X)$ and $k \in \mathbb{N}$, we have

$$\begin{aligned}A(G; p^k\Gamma) &= \{x \in G \mid p^k\gamma(x) = 0 \quad \text{for all} \quad \gamma \in \Gamma\} \\ &= \{x \in G \mid \gamma(p^k x) = 0 \quad \text{for all} \quad \gamma \in \Gamma\} \\ &= \{x \in G \mid p^k x \in A(G;\Gamma)\} = p^{-k}A(G;\Gamma).\end{aligned}$$

Since

$$p^k G^* \cap \Gamma = A(G^*; G[p^k]) \cap A(G^*; A(G;\Gamma)) = A(G^*; G[p^k] + A(G;\Gamma)),$$

it then follows that $p^k G^* \cap \Gamma = p^k\Gamma$ if and only if

$$A(G^*; G[p^k] + A(G;\Gamma)) = A(G^*; A(G; p^k\Gamma)) = A(G^*; p^{-k}A(G;\Gamma)),$$

or equivalently if $G[p^k] + A(G;\Gamma) = p^{-k}A(G;\Gamma)$, which in its turn is equivalent to $p^k G \cap A(G;\Gamma) = p^k A(G;\Gamma)$. Consequently, Γ is $S(X)$-pure in G^* if and only if $A(G;\Gamma)$ is $S(X)$-pure in G. Finally, observing that a closed subgroup of G^* is contained in $c(G^*)$ if and only if its annihilator in G contains $t(G) = \bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$, the assertion follows from Theorem 3.1 and duality. □

Definition 3.3. *Let S be a nonempty subset of $\mathbb{P}$. A group $X \in \mathcal{L}$ is said to have no elements of infinite topological S-height in case $w_S(X) = \{0\}$.*

We can prove the following generalization of Theorem 3.1.

Theorem 3.4. *Let X be a mixed group in $\mathcal{L}$ with no elements of infinite topological S-height, where $S = S_0(X)$. The following statements are equivalent:*

(i) *The subgroups $p^n X$ with $p \in S$ and $n \in \mathbb{N}$ are open in X, and $E(X)$ is commutative.*

(ii) *The cyclic, pure, p-subgroups of X, where $p \in S$, split topologically from X, and $E(X)$ is commutative.*

(iii) *S is infinite and X is topologically isomorphic to an S-pure subgroup of*

$$\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)$$

containing

$$\prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

where $l_p, n_p \in \mathbb{N}$ and $n_p \neq 0$ for all $p \in S$.

Proof. First observe that since X has no elements of infinite topological S-height, the subgroups $t_p(X)$ are reduced for all $p \in S$, so that X contains nonzero, cyclic, pure, p-subgroups for all $p \in S$ [5, Corollary 27.3].

Assume X satisfies (i), and let A be a cyclic, pure, p-subgroup of X, where $p \in S$. Then $A \cong \mathbb{Z}(p^n)$ for some $n \in \mathbb{N}_0$. Moreover, A splits algebraically from X [5, Proposition 27.1], and hence we can write $X = A \dotplus G$ for some subgroup G of X. It follows that $p^nX = p^nG \subset G$. As p^nX is open in X, we deduce that G is open in X too, so $X = A \oplus G$ by [1, Corollary 6.8]. This proves that (i) implies (ii).

Now assume (ii) holds. Letting $p \in S$, choose an arbitrary nonzero, cyclic, pure, p-subgroup $B(p)$ of X. Then $B(p) \cong \mathbb{Z}(p^{n_p})$ for some $n_p \in \mathbb{N}_0$. By hypothesis, there exists a closed subgroup $C(p)$ of X such that $X = B(p) \oplus C(p)$. We first show that $t_p(C(p)) = \{0\}$ and $\overline{pC(p)} = C(p)$. To do this, observe that since $E(X)$ is commutative, we must have by [14, Lemma 3.5]

$$H(B(p), C(p)) = \{0\} = H(C(p), B(p)).$$

Now, if $t_p(C(p))$ were nonzero, it would clearly follow that $H(B(p), C(p)) \neq \{0\}$, a contradiction. Thus $t_p(C(p)) = \{0\}$, and hence $t_p(X) = B(p)$. Suppose further that $\overline{pC(p)} \neq C(p)$ and pick an arbitrary element $a \in C(p) \setminus \overline{pC(p)}$. Then $\pi(a)$ is a nonzero element of $C(p)/\overline{pC(p)}$, where $\pi \in H\big(C(p), C(p)/\overline{pC(p)}\big)$ denotes the canonical projection. By [13, (3.8)], we can write $C(p)/\overline{pC(p)} = \langle\pi(a)\rangle \oplus \Gamma$ for some closed subgroup Γ of $C(p)/\overline{pC(p)}$. Let φ denote the canonical projection of $C(p)/\overline{pC(p)}$ onto $\langle\pi(a)\rangle$. Since $\langle\pi(a)\rangle$ is a nonzero cyclic p-group, $H(\langle\pi(a)\rangle, B(p)) \neq \{0\}$. Choosing an arbitrary nonzero $h \in H(\langle\pi(a)\rangle, B(p))$, it is clear that $h \circ \varphi \circ \pi$ is a nonzero element of $H(C(p), B(p))$, a contradiction. This shows that $\overline{pC(p)} = C(p)$, and hence for all $n \in \mathbb{N}$, $\overline{p^nC(p)} = C(p)$. As $\overline{p^{n_p}X} = \overline{p^{n_p}C(p)}$, it follows in particular that $\bigcap_{n\in\mathbb{N}} \overline{p^nX} = C(p)$. We next proceed to establish the topological isomorphism whose existence is asserted in (iii). For every $p \in S$, fix an arbitrary isomorphism f_p from $B(p)$ onto $\mathbb{Z}(p^{n_p})$, and let $g_p \in H(X, B(p))$ denote the canonical projection of X onto $B(p)$ with kernel $C(p)$. Also pick an arbitrary compact open subgroup U of X. Clearly, we have $f_p(g_p(U)) = \mathbb{Z}(p^{n_p})[p^{l_p}]$ for some $l_p \in \mathbb{N}$. Define

$$\alpha : X \to \prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)$$

by setting $\alpha(x) = (f_p g_p(x))_{p\in S}$ for all $x \in X$. Then α is a group homomorphism and $\alpha(U) \subset \prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$. Moreover, α is injective because

$$\ker(\alpha) = \bigcap_{p\in S} \ker(f_p g_p) = \bigcap_{p\in S} C(p) = \bigcap_{p\in S}\bigcap_{n\in\mathbb{N}} \overline{p^n X} = \{0\}.$$

Further, since every $f_p g_p$ is continuous, it follows that the homomorphism $x \to (f_p g_p(x))_{p\in S}$ from U to $\prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$ is continuous [2, Ch. I, §4, Proposition 1]. As U is open in X, it then follows that α is continuous as well [3, Ch. III, §2, Proposition 23]. In particular, $\alpha(U)$ is compact and hence closed in $\prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$. Taking into account that $\bigoplus_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$ is dense in $\prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$ [3, Ch. III, §2, Proposition 25] and contained in $\alpha(U)$, we conclude that $\alpha(U) = \prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$. This implies that α is open because U is compact in X [2, Ch. I, §9, Théorème 2, Corollaire 2] and $\prod_{p\in S} \mathbb{Z}(p^{n_p})[p^{l_p}]$ is open in $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$. Consequently, α establishes a topological isomorphism from X onto $\alpha(X)$. Also, since $\bigoplus_{p\in S} \mathbb{Z}(p^{n_p}) \subset \alpha(X)$ and

$$\overline{\bigoplus_{p\in S} \mathbb{Z}(p^{n_p})} = \prod_{p\in S}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$$

[3, Ch. III, §2, Exercice 26], we have

$$\prod_{p\in S}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big) \subset \alpha(X).$$

Finally, it is clear that for each $p \in S$ the multiplication by p is an open mapping on $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$, and hence on X. To show that $\alpha(X)$ is S-pure in $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$, pick any $q \in S$ and $n \in \mathbb{N}$, and let $x \in X$ be such that

$$\alpha(x) \in q^n \prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big).$$

Letting $\alpha(x) = q^n (y_p)_{p\in S}$ with $(y_p)_{p\in S} \in \prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$, we set

$$y'_p = \begin{cases} y_q, & \text{if } p = q, \\ 0, & \text{if } p \neq q, \end{cases} \qquad \text{and} \qquad y''_p = \begin{cases} 0, & \text{if } p = q, \\ y_p, & \text{if } p \neq q. \end{cases}$$

Clearly $\alpha(x) = q^n (y'_p)_{p\in S} + q^n (y''_p)_{p\in S}$. As $X = B(q) \oplus C(q)$, we can write $x = b_q + c_q$ for some $b_q \in B(q)$ and $c_q \in C(q)$. Since for $p \neq q$ we have $f_p g_p(b_q) = 0$ (because $H(\mathbb{Z}(q^{n_q}), \mathbb{Z}(p^{n_p})) = \{0\}$), and since $f_q g_q(c_q) = 0$ (because $c_q \in \ker(g_q)$), we conclude that $\alpha(b_q) = q^n (y'_p)_{p\in S}$ and $\alpha(c_q) = q^n (y''_p)_{p\in S}$. Remembering that $f_q : B(q) \to \mathbb{Z}(q^{n_q})$ is an isomorphism, choose $b'_q \in B(q)$ such that $f_q(b'_q) = y_q$. As $b_q - q^n b'_q \in \ker(\alpha)$, we have $b_q = q^n b'_q$. Also, since the multiplication by q is an open map and $C(q)$ is an open subgroup, we have $qC_q = \overline{qC_q} = C_q$, so that $q^n C_q = C_q$. Hence there exists $c'_q \in C_q$ such that $q^n c'_q = c_q$. It follows that

$$\alpha(x) = \alpha(b_q) + \alpha(c_q) = q^n(\alpha(b'_q) + \alpha(c'_q)),$$

so that

$$\alpha(X) \cap q^n \prod_{p\in S}^{loc} \big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big) \subset q^n \alpha(X).$$

As the converse inclusion clearly holds, we have

$$\alpha(X) \cap q^n \prod_{p\in S}^{loc} \big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big) = q^n \alpha(X),$$

so $\alpha(X)$ is S-pure in $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$. Consequently, (ii) implies (iii).

Next assume (iii) holds. We already mentioned that the multiplication by $p \in S$ is open on $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$, and hence on X. Let $X(d)$ denote the group X taken discrete. It then follows from our hypotheses that $X(d)$ is isomorphic to an S-pure subgroup of $\prod_{p\in S(X)}^{loc}(\mathbb{Z}(p^{n_p}); \{0\})$ containing $\bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$, so that $E(X(d))$ is commutative by Theorem 3.1. As $E(X) \subset E(X(d))$, this proves that (iii) implies (i). □

To state the dual analog of Theorem 3.4, a few definitions are in order. In the first one, we reconsider the notion of comixed LCA group, introduced in [16, Definition 6.5]. The reason for this modification is that we want an LCA group to be comixed if and only if its dual is mixed.

Definition 3.5. *A group $X \in \mathcal{L}$ is said to be comixed if either* (1) $\bigcap_{n\in\mathbb{N}_0} \overline{nX}$ *is a nontrivial subgroup of X, i. e. $\{0\} \neq \bigcap_{n\in\mathbb{N}_0} \overline{nX} \neq X$,* or (2) $\bigcap_{n\in\mathbb{N}_0} \overline{nX} = \{0\}$ *and X has no compact subgroups of the form $\overline{mX}$, where $m \in \mathbb{N}_0$.*

Definition 3.6. *Let $p \in \mathbb{P}$. A closed subgroup G of a group $X \in \mathcal{L}$ is said to be p-copure if, for each $n \in \mathbb{N}$, one has $p^{-n}G = \overline{G + X[p^n]}$. Given a nonempty subset S of $\mathbb{P}$, we say G is S-copure in case it is p-copure for all $p \in S$. G is called copure if it is $\mathbb{P}$-copure.*

As is easy to see, p-purity and p-copurity coincide for discrete and for compact groups.

Definition 3.7. *Let $p \in \mathbb{P}$. A subgroup G of an abelian group X is said to be p-submaximal if X/G is a cyclic p-group.*

Our next definition is inspired by one in [1, (4.34)].

Definition 3.8. *Let S be a nonempty subset of $\mathbb{P}$. A group $X \in \mathcal{L}$ is said to be S-power-proper if for each $p \in S$ and $n \in \mathbb{N}$ the multiplication by p^n is a proper map, i. e. for each open subset U of X, p^nU is open in p^nX, taken with its topology induced from X.*

We have

Corollary 3.9. *Let X be a comixed group in $\mathcal{L}$, and let $S = \{p \in \mathbb{P} \mid \overline{pX} \neq X\}$. If $\overline{\sum_{p\in S} t_p(X)} = X$, the following statements are equivalent:*

(i) *X is an S-power-proper group with commutative ring $E(X)$, and the subgroups $X[p^n]$ are compact for all $p \in S$ and $n \in \mathbb{N}$.*

(ii) *The closed, copure, p-submaximal subgroups of X, where $p \in S$, split topologically from X, and $E(X)$ is commutative.*

(iii) *S is infinite and X is topologically isomorphic to a quotient group of*

$$\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big)^*$$

by a closed S-copure subgroup, contained in

$$c\Big(\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big)^*\Big),$$

where $n_p, l_p \in \mathbb{N}$ and $n_p \neq 0$ for all $p \in S$.

Proof. Since X is a comixed group, X^* is mixed. It is also easy to see that $S = S_0(X^*)$, and $\overline{\sum_{p\in S} t_p(X)} = X$ if and only if X^* has no elements of infinite topological S-height.

Assume (i). Since X is S-power-proper, X^* is S-power-proper too [1, P.23(d)]. It follows that, for any $p \in S$ and $n \in \mathbb{N}$, the subgroup $p^n X^*$ is closed and hence open in X^* (because $X[p^n]$ is compact).

Pick any $p \in S$, and let G be a closed, copure, p-submaximal subgroup of X. Since $A(X^*, G) \cong (X/G)^*$, we see that $A(X^*, G)$ is a cyclic, p-subgroup of X^*. Moreover, since G is p-copure in X, we also have $p^{-n}G = \overline{G + X[p^n]}$. Passing to annihilators, we obtain

$$p^n A(X^*, G) = A(X^*, G) \cap p^n X^*,$$

so that $A(X^*, G)$ is p-pure and thus pure in X^* [5, p. 114, (g)]. It then follows from Theorem 3.4 that $A(X^*, G)$ splits topologically from X^*, and hence G splits topologically from X [1, Corollary 6.10]. Thus (i) implies (ii).

Now assume (ii), and pick any $p \in S$ and any cyclic, pure, p-subgroup Γ of X^*. It is easy to see that $A(X, \Gamma)$ is a closed, copure, p-submaximal subgroup of X. By hypothesis, $A(X, \Gamma)$ splits topologically from X, so that Γ splits topologically from X^*. Consequently, X^* satisfies condition (ii) and hence (iii) of Theorem 3.4. Observing that

$$k\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big) = \prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

and passing to duals, we deduce that (ii) implies (iii).

Assume (iii). It follows that X^* satisfies condition (iii) of Theorem 3.4, so that $p1_{X^*}$ is an open mapping on X^* for all $p \in S$. By using duality, it is then easy to see that (i) holds. □

We recall from [7] the following definition.

Definition 3.10. *Let X be a discrete, torsionfree group in $\mathcal{L}$. An independent subset M of X is said to be quasi-pure independent if $\langle M\rangle_*$ is the internal direct sum of subgroups $\langle x\rangle_*$ with $x \in M$, and $\langle x\rangle = \langle x\rangle_*$ whenever $\langle x\rangle_*$ is cyclic and $x \in M$.*

By Zorn's lemma, any quasi-pure independent subset of a discrete, torsionfree group $X \in \mathcal{L}$ is contained in a maximal quasi-pure independent subset of X [7, Proposition 123].

We now state and prove the main theorem of this section, which extends Theorem 3.4.

Theorem 3.11. *Let X be a group in $\mathcal{L}$ such that $t(X/c(X)) \neq \{0\}$, and let $S = S_0(X/c(X))$. Suppose, in addition, the following conditions hold:*

(i) *$w_S(X/c(X))$ is densely divisible and contains no compact elements;*

(ii) *The cyclic, pure, p-subgroups of X, where $p \in S$, and the compact, connected subgroups of X split topologically from X.*

Then $E(X)$ is commutative if and only if for each $p \in S$ there exist $n_p, l_p \in \mathbb{N}$ with $n_p \neq 0$ such that X is topologically isomorphic either to an S-pure subgroup of

$$\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$$

containing

$$\prod_{p\in S}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

or to a group of the form

$$D \times \prod_{p\in S}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

where D is topologically isomorphic to either $\mathbb{R}$, $\mathbb{Q}$, or an S-torsionfree quotient of $\mathbb{Q}^*$ by a closed subgroup.

Proof. Assume $E(X)$ is commutative. By [16, Theorem 4.6], there are two cases to consider:

(a) X is residual;

(b) $X \cong D \times Y$, where D is topologically isomorphic with either $\mathbb{R}$, $\mathbb{Q}$, or $\mathbb{Q}^*$, and Y is a topological torsion group with $t(Y) \neq \{0\}$.

Assume (a) holds. If $c(X) = \{0\}$, we deduce from (i) that $w_S(X)$ is densely divisible and contains no compact elements. As $d(X) \subset k(X)$, it follows that $w_S(X) = \{0\}$. Consequently, if X is mixed, we have by Theorem 3.4 that S is infinite and X is topologically isomorphic to an S-pure subgroup of $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$ containing $\prod_{p\in S}\big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big)$, where $l_p, n_p \in \mathbb{N}$ and $n_p \neq 0$ for all $p \in S$.

In case X is torsion, we deduce from [14, Corollary 5.7] that $X \cong \bigoplus_{p\in S}\mathbb{Z}(p^{n_p})$. It remains to observe that $\bigoplus_{p\in S}\mathbb{Z}(p^{n_p})$ is S-pure in $\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\{0\}\big)$.

Next suppose that $C = c(X)$ is nonzero. Since X is residual, C is compact [8, (24.24)], and hence, in view of (ii), we can write $X = C \oplus Z$ for some closed subgroup Z of X. In particular, $E(C)$ and $E(Z)$ are commutative [14, Lemma 3.2]. We also must have $C = \overline{t(C)}$. For if not, it would follow from [1, Proposition 6.12] that $C \cong (\mathbb{Q}^*)^\nu \times \overline{t(C)}$ for some cardinal number $\nu \geq 1$, contradicting the fact that X is residual. Thus $C = \overline{t(C)}$. Next we shall show that C is topologically isomorphic to a quotient of $\mathbb{Q}^*$ by a closed, necessarily nonzero subgroup. To see this, it is enough to show that C^* is topologically isomorphic to a subgroup of $\mathbb{Q}$. First observe that, being the character group of a compact and connected group, C^* is discrete and torsionfree [8, (23.17) and (24.25)]. Moreover, C^* is reduced because otherwise it would contain a direct summand isomorphic to $\mathbb{Q}$, and hence C would contain a topological direct summand topologically isomorphic to $\mathbb{Q}^*$, in contradiction with the fact that $C = \overline{t(C)}$. Now, if A is a closed, pure subgroup of C, then A is compact and connected [12, Corollary, p. 369]. Consequently, we can write $X = A \oplus B$ for some closed subgroup B of X. It is then clear that $C = A \oplus (B \cap C)$ [1, Proposition 6.5]. Since a subgroup L of the discrete group C^* is pure in C^* if and only if $A(C, L)$ is pure in C [1, Corollary 7.6], we deduce from [1, Corollary 6.10] that every pure subgroup of C^* splits from C^*. Now, let M be a maximal quasi-pure independent subset of C^*, and hence

$$\langle M\rangle_* \cong \bigoplus_{x\in M}\langle x\rangle_*.$$

Since $\langle M\rangle_*$ splits from C^*, we conclude by the maximality of M that $C^* = \langle M\rangle_*$, so C^* is completely decomposable. Further, since $E(C^*)$ is commutative, it follows from [10, Theorem 3] that the groups $\langle x\rangle_*$, where $x \in M$, have incomparable types. Assume by way of contradiction that $|M| > 1$, and pick any distinct elements $a, b \in M$. Then

$$G = \langle a\rangle_* \oplus \langle b\rangle_* \tag{3.1}$$

is pure in C^*, has rank two, and is completely decomposable. For $g \in G$, let $\tau(g)$ denote the type of g. We have $\tau(a+b) = \inf\big(\tau(a), \tau(b)\big)$ [6, §85, C)]. As $\tau(a)$ and $\tau(b)$ are incomparable, we also have $\tau(a+b) < \tau(a)$ and $\tau(a+b) < \tau(b)$. Further, since $\langle a+b\rangle_*$ splits from C^*, we clearly have

$$G = \langle a+b\rangle_* \oplus \Gamma \tag{3.2}$$

for some subgroup of rank one Γ of G [5, §16, Exercise 3(d)]. Since the number of summands of a given type in some decomposition of a discrete, completely decomposable group as a direct sum of groups of rank one is an invariant of that group [6, Proposition 86.1], (3.1) and (3.2) lead to a contradiction. Therefore we must have $|M| = 1$, so that C^* is isomorphic to a subgroup of $\mathbb{Q}$, and hence C is topologically isomorphic to a quotient of $\mathbb{Q}^*$ by a closed subgroup. On the other hand, since $X/c(X) \cong Z$, it is clear from (i) that $w_S(Z) = \{0\}$. Therefore, in case

Z is mixed, we deduce from Theorem 3.4 that Z is topologically isomorphic to an S-pure subgroup of $\prod^{loc}_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)$ containing $\prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)$. As $k(Z)=Z$ by [16, Lemma 4.4] and

$$k\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big)=\prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

it then follows that $Z\cong\prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)$. In the case when Z is torsion, we have $X\cong\bigoplus_{p\in S}\mathbb{Z}(p^{n_p})$ [14, Corollary 5.7]. Finally, if C were not S-torsionfree, we would clearly have $H(\mathbb{Z}(p^{n_p}),C)\neq\{0\}$ for some $p\in S$. Then, combining the canonical projection of Z onto Z_p with an arbitrary isomorphism from Z_p onto $\mathbb{Z}(p^{n_p})$ and with any nonzero $h\in H(\mathbb{Z}(p^{n_p}),C)$, we would obtain a nonzero element of $H(Z,C)$, in contradiction with [14, Lemma 3.5]. Thus C must be S-torsionfree.

Now assume (b) holds. If D is topologically isomorphic with either $\mathbb{R}$ or $\mathbb{Q}^*$, we must have $c(Y)=\{0\}$ since otherwise it would follow from [8, (25.20)] respectively [1, Corollary 4.10] that $H(D,Y)\neq\{0\}$, which is in contradiction with [14, Lemma 3.5]. As $k(Y)=Y$, we then see from (i) that $w_S(Y)=\{0\}$. In case $D\cong\mathbb{Q}$, we deduce by using as above [14, Lemma 3.5] that $d(Y)=\{0\}$. It follows that $w_S(X)\cong D$, and hence again $w_S(Y)=\{0\}$. Since $E(Y)$ is commutative [14, Lemma 3.2], we conclude as for Z in the case when $X\cong C\times Z$ that

$$Y\cong\prod_{p\in S}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

where $n_p,l_p\in\mathbb{N}$ and $n_p\neq 0$ for all $p\in S$.

The converse is clear. □

By use of duality, we obtain the following

Corollary 3.12. *Let X be a group in $\mathcal{L}$ such that $k(X)$ is not densely divisible, and let $S=\{p\in\mathbb{P}|\overline{p\cdot k(X)}\neq k(X)\}$. Suppose, in addition, the following conditions hold:*

(i) *$k(X)/\overline{\sum_{p\in S}t_p(X)}$ is torsionfree and connected;*

(ii) *The closed, copure, p-submaximal subgroups of X, where $p\in S$, and the open subgroups of X relative to which X has torsionfree quotients split topologically from X.*

Then $E(X)$ is commutative if and only if for each $p\in S$ there exist $n_p,l_p\in\mathbb{N}$ with $n_p\neq 0$ such that X is topologically isomorphic either to a quotient of

$$\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big)^*$$

by a closed, S-copure subgroup contained in

$$c\Big(\Big(\prod_{p\in S}^{loc}\big(\mathbb{Z}(p^{n_p});\mathbb{Z}(p^{n_p})[p^{l_p}]\big)\Big)^*\Big),$$

or to a group of the form

$$D \times \prod_{p \in S} \big(\mathbb{Z}(p^{n_p}); \mathbb{Z}(p^{n_p})[p^{l_p}]\big),$$

where D *is topologically isomorphic to either* $\mathbb{R}$, $\mathbb{Q}^*$, *or an* S*-divisible subgroup of* $\mathbb{Q}$.

Proof. As is easy to see, $k(X)$ is not densely divisible if and only if $t(X^*/c(X^*)) \neq \{0\}$ [1, Theorem 4.15]. It is also clear that $S_0(X^*/c(X^*)) = S$. Let $\Gamma = X^*/c(X^*)$. If $W = w_S(\Gamma)$, then

$$A(\Gamma^*; W) = \overline{\sum_{p \in S} t_p(\Gamma^*)},$$

so that

$$\begin{aligned} W^* \cong \Gamma^*/A(\Gamma^*; W) &\cong k(X)/\overline{\sum_{p \in S} t_p(k(X))} \\ &= k(X)/\overline{\sum_{p \in S} t_p(X)}. \end{aligned}$$

It follows that X satisfies condition (i) if and only if X^* satisfies condition (i) of Theorem 3.11. Similarly, X satisfies condition (ii) if and only if X^* satisfies condition (ii) of Theorem 3.11. The assertion follows from Theorem 3.11 and duality. □

4 Bounded order-by-discrete groups and their duals

In this section we will be dealing with bounded order-by-discrete groups and compact-by-bounded order groups, which were introduced in [14]. We begin with a characterization of bounded order-by-discrete groups.

Theorem 4.1. *A group* $X \in \mathcal{L}$ *is bounded order-by-discrete if and only if* $c(X) = \{0\}$ *and* $k(X) = t(X)$.

Proof. Assume $X \in \mathcal{L}$ is bounded order-by-discrete, and pick an arbitrary closed subgroup of bounded order B of X such that X/B is discrete. Since B is then open in X [8, (5.6)] and $t(X) \supset B$, it follows that $t(X)$ is open in X too. In particular, $t(X)$ is locally compact and $c(X) \subset t(X)$. As every torsion group in $\mathcal{L}$ is totally disconnected [1, Theorem 3.5], we must have $c(X) = \{0\}$, so that $k(X)$ is a topological torsion group. Letting $x \in k(X)$ be arbitrary, we then have $\lim_{n\to\infty}(n!x) = 0$, so $n!x \in t(X)$ for sufficiently large $n \in \mathbb{N}$, and hence $x \in t(X)$. It follows that $k(X) = t(X)$.

For the converse, observe that since $c(X) = \{0\}$, $k(X)$ and hence $t(X)$ is open in X [4, Proposition 3.3.6]. It follows that $t(X)$ is locally compact. Since $t(X) = \bigcup_{n\in\mathbb{N}_0} X[n]$, it then follows by Baire Category Theorem [8, (5.28)] that there is an

$n_0 \in \mathbb{N}_0$ such that $X[n_0]$ has nonempty interior, so that $X[n_0]$ is open in $t(X)$ and hence in X. Consequently, X is bounded order-by-discrete. □

Dualizing Theorem 4.1 gives the following characterization of compact-by-bounded order groups.

Corollary 4.2. *A group $X \in \mathcal{L}$ is compact-by-bounded order if and only if $c(X) = \bigcap_{n\in\mathbb{N}_0} \overline{nX}$ and $k(X) = X$.*

Proof. It is easy to see that X is compact-by-bounded order if and only if X^* is bounded order-by-discrete. The assertion follows then from Theorem 4.1 and duality. □

The following lemma considers a special case of bounded order-by-discrete groups having commutative rings of continuous endomorphisms.

Lemma 4.3. *Let $X \in \mathcal{L}$ be a bounded order-by-discrete, reduced group with primary components of bounded order. If $E(X)$ is commutative, then the following conditions hold:*

(i) *X is discrete;*

(ii) *$t(X) \cong \bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$, where the n_p's are positive integers;*

(iii) *$X/t(X)$ is $S(X)$-divisible;*

(iv) *$\bigcap_{p\in S(X)} p^{n_p}X$ is $S(X)$-divisible and torsionfree, and $X/\bigcap_{p\in S(X)} p^{n_p}X$ is isomorphic to an $S(X)$-pure subgroup of $\prod^{loc}_{p\in S(X)}\big(\mathbb{Z}(p^{n_p}); \{0\}\big)$ containing $\bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$.*

Proof. As we saw in Theorem 4.1, $c(X) = \{0\}$ and $k(X) = t(X)$, so $t(X)$ is a topological torsion group. Moreover, $t(X)$ is open in X. Fix any $p \in S(X)$ and any compact open subgroup U of X such that $U \subset t(X)$. By [1, Theorem 3.13], we have $t(X) \cong \prod_{q\in S(X)}(t_q(X); t_q(U))$, so that

$$t(X) = t_p(X) \oplus t_p(X)^{\#}, \tag{4.1}$$

where $t_p(X)^{\#} = \overline{\sum_{q\in S(X)\setminus\{p\}} t_q(X)}$. Since $t_p(X)$ is of bounded order, we also have

$$X = t_p(X) \dotplus C(p) \tag{4.2}$$

for some subgroup $C(p)$ of X [5, Corollary 27.4]. Our first task is to show that the last direct sum is topological. For $q \in S(X)$, let

$$q^{n_q} = \max\big\{o(x) \mid x \in t_q(X)\big\}.$$

It follows from decomposition (4.1) that $p^{n_p}t(X) \subset t_p(X)^{\#}$. In a similar way, writing $U = t_p(U) \oplus t_p(U)^{\#}$, where $t_p(U)^{\#} = \overline{\sum_{q\in S(X)\setminus\{p\}} t_q(U)}$, we obtain

$p^{n_p}U = p^{n_p}t(U) \subset t_p(U)^{\#}$. On the other hand, letting $q \in S(X) \setminus \{p\}$, we can choose $a(q), b(q) \in \mathbb{Z}$ such that $a(q)p^{n_p} + b(q)q^{n_q} = 1$. For $x \in t_q(X)$, we then have

$$x = a(q)p^{n_p}x + b(q)q^{n_q}x = p^{n_p}a(q)x \in p^{n_p}t(X),$$

so that $t_q(X) \subset p^{n_p}t(X)$. In a similar way, for $x \in t_q(U)$ we have $x \in p^{n_p}U$, and hence $t_q(U) \subset p^{n_p}U$. It follows that

$$t_p(X)^{\#} = \overline{\sum_{q \in S(X)\setminus\{p\}} t_q(X)} \subset \overline{p^{n_p}t(X)}$$

and

$$t_p(U)^{\#} = \overline{\sum_{q \in S(X)\setminus\{p\}} t_q(U)} \subset \overline{p^{n_p}U} = p^{n_p}U,$$

so that $t_p(X)^{\#} = \overline{p^{n_p}t(X)}$ and $t_p(U)^{\#} = p^{n_p}U$. As $t_p(U)^{\#} = U \cap t_p(X)^{\#}$, $p^{n_p}U$ is open in $t_p(X)^{\#}$, so $p^{n_p}t(X)$ is open in $t_p(X)^{\#}$ too, and hence $t_p(X)^{\#} = p^{n_p}t(X)$. Let $\varphi_p \in E(t(X))$ be the canonical projection of $t(X)$ onto $t_p(X)$ given by (4.1), and $\psi_p : X \to X$ be the canonical projection of X onto $t_p(X)$ given by (4.2). Since

$$t_p(X)^{\#} = p^{n_p}t(X) \subset p^{n_p}X \subset C(p),$$

it is clear that $\psi_p|_{t(X)} = \eta \circ \varphi_p$, where η is the canonical injection of $t(X)$ into X. Further, since $t(X)$ is open in X, it follows that ψ_p is continuous on X [3, Ch. III, §2, Proposition 23], and thus $X = t_p(X) \oplus C(p)$ by [3, Ch.III, §6, Proposition 2].

Now, taking account of [14, Lemma 3.2], we conclude that $E(t_p(X))$ is commutative, and so $t_p(X) \cong \mathbb{Z}(p^{n_p})$ by [14, Theorem 5.2]. Since in view of [14, Lemma 3.5] we must have $H(C(p), t_p(X)) = \{0\}$, it can be shown as in the proof of Theorem 3.4 that $\overline{pC(p)} = C(p)$.

Finally, since $p \in S(X)$ was arbitrarily chosen, we conclude that $t(X)$ is countable, and hence discrete [11, Ch. I, Theorem 2, Corollary]. But $t(X)$ is open in X, so X is discrete too. In particular, $\overline{qC(q)} = qC(q)$ for all $q \in S(X)$, and so, for all $q \in S(X)$, $X/t_q(X)$ is q-divisible as an isomorphic copy of $C(q)$. Since

$$X/t(X) \cong (X/t_q(X))/(t(X)/t_q(X))$$

for all $q \in S(X)$, it follows that $X/t(X)$ is $S(X)$-divisible. Thus X satisfies (i), (ii) and (iii).

To establish the first part of (iv), let $X_\infty = \bigcap_{p \in S(X)} p^{n_p}X$, and pick any $s \in S(X)$ and $x \in X_\infty$. Since $s^{n_s}X$ is s-divisible, there exists $y \in s^{n_s}X$ such that $x = sy$. Letting $r \in S(X) \setminus \{s\}$, choose $a(r), b(r) \in \mathbb{Z}$ such that $a(r)s + b(r)r^{n_r} = 1$. We have

$$y = a(r)sy + b(r)r^{n_r}y = a(r)x + b(r)r^{n_r}y \in X_\infty + r^{n_r}X \subset r^{n_r}X,$$

so that $y \in X_\infty$. As $x \in X_\infty$ and $s \in S(X)$ were arbitrary, it follows that X_∞ is $S(X)$-divisible. Moreover, since $X_\infty \cap t(X) = \{0\}$, X_∞ is also torsionfree. Now we proceed to establish the second part of (iv). For each $p \in S(X)$, let $g_p \in H(X, t_p(X))$ denote the canonical projection of X onto $t_p(X)$ with kernel $C(p)$, and f_p an isomorphism from $t_p(X)$ onto $\mathbb{Z}(p^{n_p})$. The mapping $\alpha : X \to \prod_{p\in S(X)}^{loc}\big(\mathbb{Z}(p^{n_p}); \{0\}\big)$, given by $\alpha(x) = (f_p g_p(x))_{p\in S(X)}$ for all $x \in X$, is then a group homomorphism with kernel X_∞, so that X/X_∞ is isomorphic with $\alpha(X)$. It is also clear that, for all $q \in S(X)$, α maps $t_q(X)$ onto the subgroup of $\prod_{p\in S(X)}^{loc}\big(\mathbb{Z}(p^{n_p}); \{0\}\big)$ consisting of all elements with zero p-components for $p \neq q$, whence we deduce that

$$\bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p}) \subset \alpha(X).$$

Finally, it can be seen, following the same way as in the proof of Theorem 3.4, that $\alpha(X)$ is $S(X)$-pure in $\prod_{p\in S(X)}^{loc}\big(\mathbb{Z}(p^{n_p}); \{0\}\big)$.

The proof is complete. □

Let us recall from [1] the following definition.

Definition 4.4. *Let $p \in \mathbb{P}$. A group $X \in \mathcal{L}$ is called p-thetic in case there exists $h \in H(\mathbb{Z}(p^\infty), X)$ such that $h(\mathbb{Z}(p^\infty))$ is dense in X.*

We are now ready to prove the main theorem of this section.

Theorem 4.5. *Let X be a bounded order-by-discrete group in $\mathcal{L}$. If $E(X)$ is commutative, then X is discrete and satisfies exactly one of the following conditions:*

(i) *X is isomorphic with either*

$$\bigoplus_{p\in S_1} \mathbb{Z}(p^\infty) \times \bigoplus_{p\in S_2} \mathbb{Z}(p^{n_p}) \qquad or \qquad \mathbb{Q} \times \bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p}),$$

where $S_1 \cup S_2 = S(X)$, $S_1 \cap S_2 = \varnothing$, and the n_p's are positive integers.

(ii) *$S(X)$ is finite and $X = t(X) \oplus W$, where W is a reduced, $S(X)$-divisible subgroup of X, and $t(X) \cong \bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$ for some positive integers n_p.*

(iii) *X is reduced, $S(X)$ is infinite, $X/t(X)$ is $S(X)$-divisible, and there exist positive integers n_p, one for each $p \in S(X)$, such that the following conditions hold:*

1) *$t(X) \cong \bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$;*

2) *$\bigcap_{p\in S(X)} p^{n_p} X$ is $S(X)$-divisible and torsionfree;*

3) *$X/\bigcap_{p\in S(X)} p^{n_p} X$ is isomorphic to an $S(X)$-pure subgroup of $\prod_{p\in S(X)}^{loc}\big(\mathbb{Z}(p^{n_p}); \{0\}\big)$ containing $\bigoplus_{p\in S(X)} \mathbb{Z}(p^{n_p})$.*

Proof. First assume X contains a subgroup D algebraically isomorphic to $\mathbb{Z}(p^\infty)$ for some $p \in S(X)$. Since $\overline{D}$ is then p-thetic, it follows from [1, Proposition 5.20 and Proposition 5.21] that either $\overline{D} \cong \mathbb{Z}(p^\infty)$ or else $\overline{D}$ is compact and connected. But X is totally disconnected by Theorem 4.1, so that the latter cannot occur, and hence $\overline{D} \cong \mathbb{Z}(p^\infty)$. Now, since $\mathbb{Z}(p^\infty)$ is splitting in the class of totally disconnected LCA groups [1, Proposition 6.21], we can write $X = \overline{D} \oplus A$ for some closed subgroup A of X. If A were not a torsion group, it would follow by Theorem 4.1 that $t(A)$ is open in A, so $A/t(A)$ is nonzero, discrete and torsionfree. Hence we would have $H(A/t(A), \overline{D}) \neq \{0\}$, whence $H(A, \overline{D}) \neq \{0\}$, contradicting by [14, Lemma 3.5] the commutativity of $E(X)$. Consequently, A must be torsion. In particular, X is torsion as a direct sum of two torsion groups. It then follows from [14, Corollary 5.7] that

$$X \cong \bigoplus_{p \in S_1} \mathbb{Z}(p^\infty) \times \bigoplus_{p \in S_2} \mathbb{Z}(p^{n_p}),$$

where $S_1 \cup S_2 = S(X)$, $S_1 \cap S_2 = \varnothing$, and the n_p's are positive integers.

Next assume $d(t(X)) = \{0\}$ but still $d(X) \neq \{0\}$, and pick a subgroup V of X algebraically isomorphic to $\mathbb{Q}$. Since $t(X)$ is open in X and $V \cap t(X) = \{0\}$, it follows that V is discrete and hence closed in X [8, (5.10)]. We can write $X = V \oplus B$ for some closed subgroup B of X, because $\mathbb{Q}$ is splitting in the class of totally disconnected LCA groups [1, Proposition 6.21]. As above, we make use of [14, Lemma 3.5] to deduce that $H(B, V) = \{0\} = H(V, B)$, which implies $B = t(B)$ and $d(B) = \{0\}$. Since $E(B)$ is clearly commutative, it follows from [14, Corollary 5.7] that $B \cong \bigoplus_{p \in S(X)} \mathbb{Z}(p^{n_p})$, where the n_p's are positive integers.

Next assume X is reduced. If $t(X)$ is of bounded order, it follows that $t(X)$ splits algebraically from X [5, Theorem 27.5], and since by Theorem 4.1 $t(X)$ is open in X, this splitting is topological [1, Corollary 6.8], i. e. $X = t(X) \oplus W$ for some discrete subgroup W of X. As $E(t(X))$ must be commutative, we conclude from [14, Corollary 5.7] that $t(X)$ is discrete and isomorphic to $\bigoplus_{p \in S(X)} \mathbb{Z}(p^{n_p})$, where the n_p's are positive integers. It follows, in particular, that X is discrete. Moreover, since $t(X)$ is of bounded order, $S(X)$ must be finite. Finally, by [16, Theorem 6.1], we must also have $pW = W$ for all $p \in S(X)$.

It remains to consider the case when $t(X)$ is not of bounded order. We shall show that then X has primary components of bounded order. Since X is bounded order-by-discrete, there is $n \in \mathbb{N}_0$ such that $X/X[n]$ is discrete. Pick any $p \in S(X)$, and write $n = p^{k_p} n'$, where $k_p \in \mathbb{N}$ and $p \nmid n'$. To see that $t_p(X)$ is of bounded order, it is enough to show that $t_p(X/X[n])$ is of bounded order. Suppose not. Then either $X/X[n]$ has a direct summand isomorphic to $\mathbb{Z}(p^\infty)$ [5, Theorem 21.2], or $t_p(X/X[n])$ is reduced and has direct summands of arbitrarily high orders [5, §27, Exercise 1]. Since $t_p(X/X[n])$ is pure in $X/X[n]$, we deduce from [5, Lemma 26.1 and Theorem 27.5] that in the second case $X/X[n]$ has as direct summands cyclic p-subgroups of arbitrarily high orders. Hence we can write

$$X/X[n] = T \oplus G, \tag{4.3}$$

where T is isomorphic to either $\mathbb{Z}(p^\infty)$, or $\mathbb{Z}(p^{l_p})$ with $l_p > k_p$. Here we must have $G \neq \{0\}$. This is clear in case $T \cong \mathbb{Z}(p^{l_p})$ because otherwise $t_p(X/X[n])$ would be of bounded order, contrary to our assumption. On the other hand, if we had $G = \{0\}$ and $T \cong \mathbb{Z}(p^\infty)$, then X would be torsion. As $E(X)$ is commutative, we would conclude from [14, Corollary 5.7] that X is also discrete and, for each $q \in S(X)$, $t_q(X)$ is isomorphic to either $\mathbb{Z}(q^\infty)$ or $\mathbb{Z}(q^{n_q})$ for some $n_q \in \mathbb{N}$. In particular, by [9, Corollary 8.11(ii)] we would have

$$X/X[n] \cong \bigoplus_{q \in S(X)} \Big(t_q(X)/t_q(X[n])\Big).$$

Since in the considered case $X/X[n] \cong \mathbb{Z}(p^\infty)$, this would imply $t_p(X) \cong \mathbb{Z}(p^\infty)$, contrary to the assumption that X is reduced. Thus $G \neq \{0\}$. Now, passing to duals in (4.3), we deduce that $\overline{nX^*} = T' \oplus G'$, where $T' \cong T^*$ and $G' \cong G^*$ [8, (23.18)]. As by [14, Lemma 3.1] $E(X^*)$ is commutative, we must have

$$H(G', T') = H(G', T')[n] \quad \text{and} \quad H(T', G' = H(T', G')[n],$$

since otherwise an application of [14, Lemma 3.5] with $\omega = n1_{X^*}$ and any $h \in H(G', T') \cup H(T', G')$ satisfying $nh \neq 0$ would produce a contradiction. Since for any $L, M \in \mathcal{L}$, $H(M^*, L^*) \cong H(L, M)$ [12, Corollary 2, p. 377], it follows that

$$H(T, G) = H(T, G)[n] \quad \text{and} \quad H(G, T) = H(G, T)[n].$$

Now we can show that either of the cases $T \cong \mathbb{Z}(p^\infty)$ or $T \cong \mathbb{Z}(p^{l_p})$ leads to a contradiction. Suppose first $T \cong \mathbb{Z}(p^\infty)$. We must have $G = t(G)$. For, if G contained an element, say a, of infinite order, then, choosing any $b \in T$ with $o(b) > p^{k_p}$, we could define $f \in H(\langle a \rangle, T)$ by the rule $f(a) = b$. Since T is divisible, there would exist $f_0 \in H(G, T)$ such that $f_0|_{\langle a \rangle} = f$, and hence $nf_0 \neq 0$, a contradiction. Thus $G = t(G)$, so $X/X[n] = t(X/X[n])$, and hence $X = t(X)$. Since by the assumption X is reduced, it follows from [14, Corollary 5.7] that $X \cong \bigoplus_{q \in S(X)} \mathbb{Z}(q^{n_q})$, where the n_q's are positive integers. But then $X/X[n]$ is reduced, contrary to the assumption that $T \cong \mathbb{Z}(p^\infty)$. Next suppose $T \cong \mathbb{Z}(p^{l_p})$. If there existed $c \in t_p(G)$ with $o(c) > p^{k_p}$, then we could find $c' \in t_p(G)$ such that $p^{k_p} < o(c') \leq p^{l_p}$. It would follow that there exists $g \in H(\mathbb{Z}(p^{l_p}), G)$ given by $g(1 + p^{l_p}\mathbb{Z}) = c'$ such that $ng \neq 0$. Since this would imply $H(T, G) \neq H(T, G)[n]$, we arrive at a contradiction. Hence we must have $p^{k_p} t_p(G) = \{0\}$, which implies $t_p(X/X[n])$ is of bounded order, a contradiction.

Consequently, our assumption that $t_p(X/X[n])$ is not of bounded order leads to a contradiction, so $t_p(X/X[n])$ must be of bounded order, whence we deduce that $t_p(X)$ is of bounded order too. As $p \in S(X)$ was arbitrary, it follows that X has primary components of bounded order. Moreover, since $t(X)$ is not of bounded order, $S(X)$ has to be infinite. Then, an application of Lemma 4.3 gives us (iii).

The proof is complete. □

We conclude this section by stating the dual analog of Theorem 4.5.

Corollary 4.6. *Let X be a compact-by-bounded order group in $\mathcal{L}$. If $E(X)$ is commutative, then X is compact and satisfies exactly one of the following conditions:*

(i) *X is topologically isomorphic with either*

$$\prod_{p\in S_1} \mathbb{Z}_p \times \prod_{p\in S_2} \mathbb{Z}(p^{n_p}) \qquad \text{or} \qquad \mathbb{Q}^* \times \prod_{p\in S(X)} \mathbb{Z}(p^{n_p}),$$

where $S_1 \cup S_2 = S(X)$, $S_1 \cap S_2 = \varnothing$, and the n_p's are positive integers.

(ii) *$S(X)$ is finite and $X = c(X) \oplus M$, where $c(X)$ is $S(X)$-torsionfree with $m(c(X)) = c(X)$, and $M \cong \prod_{p\in S(X)} \mathbb{Z}(p^{n_p})$ for some positive integers n_p.*

(iii) *$X = m(X)$, $S(X)$ is infinite, $c(X)$ is $S(X)$-torsionfree, and there exist positive integers n_p, one for each $p \in S(X)$, such that the following conditions hold:*

1) $X/c(X) \cong \prod_{p\in S(X)} \mathbb{Z}(p^{n_p})$;

2) $X/\overline{\sum_{p\in S(X)} X[p^{n_p}]}$ *is densely divisible and $S(X)$-torsionfree;*

3) $\overline{\sum_{p\in S(X)} X[p^{n_p}]}$ *is topologically isomorphic to a quotient group of $\left(\prod^{loc}_{p\in S}(\mathbb{Z}(p^{n_p}); \{0\})\right)^*$ by a closed, $S(X)$-pure subgroup contained in $c\left(\left(\prod^{loc}_{p\in S}(\mathbb{Z}(p^{n_p}); \{0\})\right)^*\right)$.*

Proof. Since a group $X \in \mathcal{L}$ is compact-by-bounded order if and only if X^* is bounded order-by-compact, the assertion follows from Theorem 4.5 and duality. □

References

[1] Armacost D.L. *The structure of locally compact abelian groups.* Pure and Applied Mathematics Series, Vol. 68, New York, Marcel Dekker, 1981.

[2] Bourbaki N. *Topologie generale, Chapter 1-2, Éléments de mathematique.* Moscow, Nauka, 1968.

[3] Bourbaki N. *Topologie generale, Chapter 3-8, Éléments de mathematique* Moscow, Nauka, 1969.

[4] Dikranjan D., Prodanov I., Stoyanov L. *Topological groups.* Pure and Applied Mathematics Series, Vol. 130, New York and Basel, Marcel Dekker, 1990.

[5] Fuchs L. *Infinite abelian groups, Vol. 1.* New York and London, Academic Press, 1970.

[6] Fuchs L. *Infinite abelian groups, Vol. 2.* New York and London, Academic Press, 1973.

[7] Griffith Ph. *Infinite abelian group theory.* Chicago and London, The University of Chicago Press, 1970.

[8] Hewitt E., Ross K. *Abstract Harmonic Analysis, Vol. 1.* Moscow, Nauka, 1975.

[9] Hungerford T. *Algebra.* New York, Springer-Verlag, 1974.

[10] Van Leeuwen L.C.A. *On the endomorphism rings of direct sums of groups.* Acta Scient. Math., 1967, **28**, N 1-2, p. 21–29.

[11] Morris S. *Pontryagin duality and the structure of locally compact abelian groups.* Cambridge, Cambridge University Press, 1977.

[12] Moskowitz M. *Homological algebra in locally compact abelian groups.* Trans. Amer. Math. Sos., 1967, **127**, p. 361–404.

[13] Popa V. *Units, idempotents and nilpotents of an endomorphism ring, II.* Bul. Acad. Şt. R. Moldova, Matematica, 1997, N 1(23), p. 93–105.

[14] Popa V. *On topological torsion LCA groups with commutative ring of continuous endomorphisms.* Bul. Acad. Şt. R. Moldova, Matematica, 2006, N 3(52), p. 87–100.

[15] Popa V. *On LCA groups in which some closed subgroups have commutative rings of continuous endomorphisms.* Bul. Acad. Şt. R. Moldova, Matematica, 2007, N 1(53), p. 83–94.

[16] Popa V. *On torsionfree LCA groups with commutative rings of continuous endomorphisms.* Bul. Acad. Şt. R. Moldova, Matematica, 2007, N 2(54), p. 81–100.

[17] Szele T., Szendrei J. *On abelian groups with commutative endomorphism ring.* Acta Math. Acad. Sci. Hung., 1951, **2**, p. 309–324.

Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
Academiei str. 5, MD-2028 Chişinău
Moldova

E-mail: *vpopa@math.md*

Received October 25, 2007

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 91–101
ISSN 1024–7696

Parametrical Approach for Bilinear Programming and its Application for solving Integer and Combinatorial Optimization Problems

Dmitrii Lozovanu

Abstract. A parametrical approach for bilinear programming is proposed and new algorithms on the basis of such approach for solving linear boolean and resource allocation problems are developed. Computational complexity of the proposed algorithms is discussed.

Mathematics subject classification: 65K05, 68W25.
Keywords and phrases: Integer programming, computational complexity.

1 Introduction and Problem Formulation

We consider the following bilinear programming problem (BPP) [1, 10, 11]:
to minimize

$$z = xCy + c'x + c''y \tag{1}$$

on subject

$$Ax \le a, \quad x \ge 0; \tag{2}$$

$$By \le b, \quad y \ge 0, \tag{3}$$

where C, A, B are matrices of size $n \times m, q \times n, l \times m$, respectively, and $c', x \in R^n$; $c'', y \in R^m$; $a \in R^q$, $b \in R^l$. In order to simplify the notations we will omit transposition sign for vectors.

This bilinear model generalizes a large class of integer and combinatorial optimization problems [6, 10]. An important particular case of BPP (1)–(3) represents the linear boolean programming problem:
to minimize

$$z = \sum_{i=1}^{n} c_i x_i \tag{4}$$

on subject

$$\begin{cases} \displaystyle\sum_{i=1}^{n} a_{ji} x_i \le a_{j0}, & j = \overline{1, q}; \\ x_i \in \{0, 1\}, & i = \overline{1, n}. \end{cases} \tag{5}$$

In [10] it is shown that this problem can be replaced by the following BPP:

to minimize

$$z = \sum_{i=1}^{n} c_i x_i + M \sum_{i=1}^{n} (x_i y_i + (1 - x_i)(1 - y_i)) \tag{6}$$

on subject

$$\begin{cases} \sum_{i=1}^{n} a_{ji} x_i \le a_{j0}, \quad j = \overline{1,q}; \\ 0 \le x_i \le 1, \quad i = \overline{1,n}, \end{cases} \tag{7}$$

$$0 \le y_i \le 1, \quad i = \overline{1,n}, \tag{8}$$

where $M > \sum_{i=1}^{n} |c_i|$. Another important case of BPP (1)–(3) represents the piecewise linear concave programming problem:

to minimize

$$z = \sum_{i=1}^{l} \min\{c^{ik} x + c_0^{ik}, \quad k = \overline{1, r_i}\} \tag{9}$$

on subject determined by (2), where $x \in R^n$, $c^{ik} \in R^n$, $c_0^{ik} \in R^n$. This problem arises as an auxiliary one when solve a class of resource allocation problems [6, 10]. In [10] it is shown that this problem can be replaced by the following BPP:

to minimize

$$z = \sum_{i=1}^{k} \sum_{k=1}^{r_i} (c^{ik} x + c_0^{ik}) y_{ik} \tag{10}$$

on subject

$$Ax \le a, \quad x \ge 0; \tag{11}$$

$$\begin{cases} \sum_{k=1}^{r_i} y_{ik} = 1, \quad i = \overline{1,l}, \\ y_{ik} \ge 0, \ l = \overline{1, r_i}, \ i = \overline{1,l}. \end{cases} \tag{12}$$

In this paper we propose an approach for solving BPP (1)–(3) which takes into account the particularity of the mentioned above cases of problems, i.e. when the matrix B is either identity one or step-diagonal one. The general scheme of the proposed approach is based on parametric linear programming. Using duality principle for the considered problem we show that it can be reduced in polynomial time to a problem of determining the consistency of the system of linear inequalities with right-hand members that depend on parameters, admissible values of which are defined by another system of linear inequalities. Then a specification of the proposed approach for the mentioned above linear boolean and resource allocation problems are developed and new algorithms for solving these classes of problems are derived. Computational complexity aspects of the proposed approach are discussed and a class of problems for which polynomial-time algorithms exist is described.

2 Parametrical programming approach for BPP

Let L be the size of BPP (1)–(3) with integer coefficients of the matrices C, A, B and vectors a, b, c', c'', i.e. L is the length of the input dates of BPP (1)–(3) [4,6].

If BPP (1)–(3) has solution then it can be solved by varying the parameter $h \in [-2^L, 2^L]$ in the problem of determining the consistency of the system

$$\begin{cases} Ax \le a; \\ xCy + c'x + c''y \le h; \\ By \le b; \\ x \ge 0, \ y \ge 0. \end{cases} \tag{13}$$

In the following we will reduce the consistency problem for system (13) to the consistency problem for the system of linear inequalities with a right-hand member depending on parameters.

Theorem 1. *Let solution sets X and Y of systems (2) and (3) be nonempty. Then system (13) has no solution if and only if the following system of linear inequalities*

$$\begin{cases} -A^T u \le Cy + c'; \\ au < c''y - h; \\ u \ge 0 \end{cases} \tag{14}$$

is consistent with respect to u for every y satisfying (3).

Proof. $\Rightarrow$ Let us assume that system (13) has no solution. This means that for every $y \in Y$ the following system of linear inequalities

$$\begin{cases} Ax \le a, \\ x(Cy + c') \le h - c''y, \\ x \ge 0 \end{cases} \tag{15}$$

has no solution with respect to x. Then according to Theorem 2.14 from [2] the inconsistency of system (15) involves the solvability with respect to u and t of the following system of linear inequalities

$$\begin{cases} A^T u + (Cy + c')t \ge 0; \\ au + (h - c''y)t < 0; \\ u \ge 0, \ t \ge 0, \end{cases} \tag{16}$$

for every $y \in Y$. Note that for every fixed $y \in Y$ of system (16) for an arbitrary solution (u^*, t^*) the condition $t^* > 0$ holds. Indeed, if $t^* = 0$, then it means that the system

$$\begin{cases} A^T u \ge 0; \\ au < 0, \ u \ge 0, \end{cases}$$

has solution, what, according to Theorem 2.14 from [2], involves the inconsistency of system (2) that is contrary to the initial assumption. Consequently, $t^* > 0$. Since

$t > 0$ in (16) for every $y \in Y$, then, dividing each of inequalities of this system by t and denoting $z = (1/t)u$, we obtain the following system:

$$\begin{cases} -A^T z \le Cy + c'; \\ az < c''y - h; \\ z \ge 0, \end{cases}$$

which has solution with respect to z for every $y \in Y$.

$\Leftarrow$ Let system (14) have solution with respect to u for every $y \in Y$. Then the following system of linear inequalities

$$\begin{cases} A^T u + (Cy + c')t \ge 0; \\ au + (h - c''y)t < 0; \\ u \ge 0,\ t > 0, \end{cases}$$

is consistent with respect to u and t for every $y \in Y$. However this system is equivalent to system (16) as it was shown that for every solution (u, t) of system (16) the condition $t > 0$ holds. Again using Theorem 2.14 from [2], we obtain from the solvability of system (16) with respect to u and t for every $y \in Y$ that system (15) is inconsistent with respect to x for every $y \in Y$. This means that system (13) has no solution. □

Theorem 2. *The minimal value of the object function in BPP (1)–(3) is equal to the maximal value h^* of the parameter h in the system*

$$\begin{cases} -A^T u \le Cy + c'; \\ au \le c''y - h; \\ u \ge 0 \end{cases} \tag{17}$$

for which it is consistent with respect to u for every $y \in Y$. An arbitrary point $y^* \in Y$, for which system (14) with $h = h^*$ and $y = y^*$ has no solution with respect to u, corresponds to one of the optimal points for BPP (1)–(3).

Proof. Let h^* be a maximal value of parameter h, for which system (17) with $h = h^*$ has solution with respect to u for every $y \in Y$. Then system (14) with $h = h^*$ has solution with respect to u not for every $y \in Y$. From this on the basis of Theorem 1 it results that system (13) with $h = h^*$ is consistent. Using Theorem 1 we can see that if for every fixed $h < h^*$ system (14) has solution with respect to u for every $y \in Y$, then system (13) with $h < h^*$ has no solution. Consequently, the maximal value h^* of parameter h, for which system (17) has solution with respect to u for every $y \in Y$, is equal to the minimum value of the object function of BPP (1)–(3).

Now let us prove the second part of the theorem. Let $y^* \in R^m$ be an arbitrary point for which system (14) with $h = h^*$ and $y = y^*$ has no solution with respect to u. Then on the basis of the duality principle the following system

$$\begin{cases} Ax \le a; \\ x(Cy^* + c') \le h^* - c''y; \\ x \ge 0 \end{cases}$$

has solution with respect to x. So, system (13) with $h = h^*$ is consistent and point $y^* \in Y$ together with certain $x^* \in X$ represents the solution of BPP. □

Corollary 3. *Let* $\overline{Y}_h = \{y \in R^m \,|\, U_h(y) \neq 0\}$, *where* $U_h(y)$ *is the set of solutions of system (17) with respect to* u *for given* $y \in R^m$ *and fixed* h. *Assume that* y^0 *is an arbitrary basic solution of system (3) such that*

$$Z^0 = \min_{x \in X}(xCy^0 + c'x + c''y^0) > h^*.$$

Then

i) $y^0 \in int\, \overline{Y}_{h^*}$, *i.e.* y^0 *is an interior point of set* $\overline{Y}_{h^*}$;

ii) $Y \subset int\, \overline{Y}_h$ *if* $h < h^*$.

Note that in an analogous way the same mathematical tool for system (13) can be applied considering x as a vector of parameters. This allows us to replace the main problem by the problem of determining the consistency of the system

$$\begin{cases} -B^T v \leq C^T x + c''; \\ bv \leq c'x - h; \;\; v \geq 0 \end{cases} \tag{18}$$

with respect to v for every x satisfying (2). This means that for the linear parametric system the following duality principle holds (see [9]).

Theorem 4. *The system of linear inequalities (17) is consistent with respect to* u *for every* y *satisfying (3) if and only if the system of linear inequalities (18) is consistent with respect to* v *for every* x *satisfying (2).*

It is easy to observe that if Y is a bounded set then the consistency property in our auxiliary problem can be verified by checking the consistency of system (17) for every basic solution of system (3). This fact follows from the geometrical interpretation of the problem. Indeed, let $UY \subseteq R^{n+k}$ be a solution set of system (17) with respect to u and y. Then $\overline{Y}_h$ for given h represents the orthogonal projection on R^k of the set $UY \subseteq R^{n+k}$. Therefore $Y \subseteq \overline{Y}_h$ if and only if system (17) is consistent for every basic solution of (3). Of course such an approach for solving the auxiliary problem cannot be used for systems with big number of variables. The approach we propose allows us to avoid exhaustive search. Moreover, we can see that in the case of problems (4)–(8) and (9)–(12) our approach efficiently solves the auxiliary problem.

The results described above show that BPP (1)–(3) can be solved efficiently if there exists an efficient algorithm for solving the following problem: to determine the maximal value h^* of parameter h such that a basic solution y^* of system (3) belongs to $bd\, \overline{Y}_{h^*}$.

In the following we show how to verify the condition $Y \subset int\overline{Y}_h$ and propose an algorithm for solving BPP (1)–(3) in the case when (3) is determined by (8) or (12).

3 Some auxiliary results

In order to explain the main results we need some auxiliary results related to dependent inequalities of linear systems. An inequality

$$\sum_{j=1}^{m} s_j y_j \le s_0 \tag{19}$$

is called dependent [2] on the consistent system of linear inequalities

$$\sum_{j=1}^{m} d_{ij} y_j \le d_{i0},\ \ i = \overline{1,p}, \tag{20}$$

if for an arbitrary solution of system (20) condition (19) holds.

The well-known Minkowski-Farkas theorem [2, 3] gives the necessary and sufficient condition of dependency (19) on (20) in the case of consistent system (20). We will extend this theorem for inconsistent systems and will use it in general form. In order to formulate this result we need the following definition.

Definition 1. Assume that system (20) is inconsistent. Inequality (19) is called dependent on system (20) if there exists a consistent subsystem

$$\sum_{j=1}^{m} d_{i_k j} y_j \le d_{i_k 0},\ \ i = \overline{1,r}, \tag{21}$$

of system (20) such that inequality (19) is dependent on (21).

Theorem 5. *Let be given system (20) with rank $r \le m$ $(m < p)$. Inequality (19) is dependent on system of linear inequalities (20) if and only if there exist numbers $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_p$ such that*

$$\begin{cases} s_j = \sum_{i=1}^{p} d_{ij}\lambda_i,\ \ j = \overline{1,m}; \\ s_0 = \sum_{i=1}^{p} d_{i0}\lambda_i + \lambda_0; \\ \lambda_j \ge 0,\ \ j = \overline{1,m}, \end{cases} \tag{22}$$

where no more than r components among $\lambda_1, \lambda_2, \dots, \lambda_p$ are nonzero.

Proof (Sketch). Necessary condition follows from [2] (Theorem 2.2). Indeed, if (19) is dependent on (20) then there exists nodal solution (21) such that (19) is dependent on (21) and necessary condition holds. Sufficient condition in the case of inconsistent system (21) can be proved in the following way. Assume that system (22) has solution $\lambda_0, \lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_r}, 0, 0, \dots, 0$, where $r \le m$. Then system (21), corresponding to $\lambda_{i_k} > 0$, $k = \overline{1,r}$, has a solution. This means that inequality (19) is dependent on consistent subsystem of linear inequalities (21). □

4 The main results

We consider the problem from Section 2 and describe an algorithm for checking if $Y \subset int\overline{Y}_h$ in the case when Y is determined by system (20), which satisfies the following conditions:

a) system (20) has rank m $(m < p)$ and Y is a bounded set with $intY \neq \emptyset$;

b) system (20) does not contain dependent inequalities;

c) if an arbitrary subsystem

$$\sum_{j=1}^{m} d_{i_k j} y_j \leq d_{i_k 0}, \;\; k = \overline{1,m}; \tag{23}$$

of system (20) has rank m then the solution of the system of linear inequalities

$$\sum_{j=1}^{m} d_{i_k j} y_j = d_{i_k 0}, \;\; k = \overline{1,m}; \tag{24}$$

is a solution of system (20), i.e. system (20) contains all possible nodal solutions.

It is easy to observe that system (3) when B is an identity matrix and B is a step-diagonal matrix represents a particular case of system (20) with properties a)–c). Therefore the results we describe below can be referred to problems (6)–(8) and (9)–(12).

In order to guarantee $int\overline{Y}_h \neq \emptyset$ we will fix $h \in [-2^L, N)$, where $N = \min[h^0, 2^L]$, h^0 is the optimal value of the object function in the linear programming problem: to maximize h on subject (17), i.e. to maximize h on the set of solutions of the following system

$$\begin{cases} -\sum_{j=1}^{q} a_{ji} u_j - \sum_{j=1}^{m} c_{ij} y_j \leq c'_i, \;\; i = \overline{1,n}; \\ \sum_{j=1}^{q} a_{j0} u_j - \sum_{j=1}^{m} c''_j y_j \leq -h; \\ u_j \geq 0, \;\; j = \overline{1,q}. \end{cases} \tag{25}$$

Theorem 6. *Let be given set Y determined by system of linear inequalities (20) satisfying conditions a)–c). In addition assume that $h \in [-2^L, N)$ and set X of solutions of system (2) is bounded with $intX \neq \emptyset$. Then $Y \not\subset int\overline{Y}_h$ if and only if the following system of linear inequalities*

$$\begin{cases} \sum_{i=1}^{n} a_{ji} \lambda_i \leq a_{j0}, \;\; j = \overline{1,q}; \\ \sum_{i=1}^{p} d_{ij} \mu_i + \sum_{i=1}^{n} c_{ij} \lambda_i = -c''_j, \;\; j = \overline{1,m}; \\ -\sum_{i=1}^{p} d_{i0} \mu_i + \sum_{i=1}^{n} c'_i \lambda_i \leq h; \\ \mu_i \geq 0, \;\; i = \overline{1,p}; \;\; \lambda_i \geq 0, \;\; i = \overline{1,n}. \end{cases} \tag{26}$$

has such a solution that $\sum_{i=1}^{p} d_{ij}\mu_i \neq 0$ *at least for an index* $j \in \{1, 2, \dots, m\}$ *and no more than* m *components among* $\mu_1, \mu_2, \dots, \mu_p$ *are nonzero.*

Proof. $\Rightarrow$ Assume that system (20) satisfies conditions a)–c) and $Y \not\subset int\overline{Y}_h$ for given $h \in [-2^L, N)$. Then $int\overline{Y}_h \neq \emptyset$ and there exists a nodal solution $y^0 = (y_1^0, y_2^0, \dots, y_m^0)$ of system (20) for which $y^0 \notin int\overline{Y}_h$, i.e. there exists subsystem (23) of system (20) such that for $y = y^0$ condition (24) holds and $y^0 \notin int\overline{Y}_h$. Note that an arbitrary nodal solution y^0 of system (20) can be regarded as a common vertex of two symmetrical cones one of which Y^0 is determined by system (23) and another one $\overline{Y}^0$ is determined by the following symmetric system

$$\sum_{j=1}^{m} d_{i_k j} y_j \geq d_{i_k 0}, \ \ k = \overline{1, m}, \tag{27}$$

which is a subsystem of the following inconsistent system

$$\sum_{j=1}^{m} d_{ij} y_j \geq d_{i0}, \ \ i = \overline{1, p}. \tag{28}$$

Based on properties a)-c) of system (20) we can show that there exists a nodal solution y^0 which determines the cone $\overline{Y}^0$ such that $\overline{Y}^0 \cap int\overline{Y}_h = \emptyset$.

This means that there exists a separating hyperplane $\sum_{j=1}^{m} s_j y_j = s_0$ [5] such that $\sum_{j=1}^{n} s_j y_j < s_0$ for $(y_1, y_2, \dots, y_n) \in int\overline{Y}_h$ and

$$-\sum_{j=1}^{m} s_j y_j \leq -s_0 \tag{29}$$

for $(y_1, y_2, \dots, y_n) \in \overline{Y}^0$. So, the inequality

$$\sum_{j=1}^{n} s_j y_j \leq s_0 \tag{30}$$

is dependent on system (25) with respect to variables $\mu_1, \mu_2, \dots, \mu_p, y_1, y_2, \dots, y_p$ and inequality (29) is dependent on system (27). If (29) is dependent on (27) then (29) is dependent on inconsistent system (28). Thus on the basis of Theorem 5, we obtain that the following systems

$$\begin{cases} 0 = -\sum_{i=1}^{n} a_{ji}\lambda_i + a_{j0}\lambda_0 - \lambda_{n+j}, \ \ j = \overline{1, q}; \\ s_j = -\sum_{i=1}^{n} c_{ij}\lambda_i - c''_j \lambda_0, \ \ j = \overline{1, q}; \\ s_0 = \sum_{i=1}^{n} c'_i \lambda_i - h\lambda_0 + \lambda_{n+q+1}; \\ \lambda_i \geq 0, \ \ i = \overline{0, n+q+1}; \end{cases} \tag{31}$$

$$\begin{cases} -s_j = -\sum_{i=1}^{p} d_{ij}\mu_i; \\ -s_0 = -\sum_{i=1}^{n} d_{i0}\mu_i + \mu_0; \\ \mu_i \geq 0, \;\; i = \overline{0,p}; \end{cases} \tag{32}$$

have solutions and no more than m components among $\mu_1, \mu_2, \ldots, \mu_p$ are nonzero.

Taking into account that we are seeking for a basic solution where $s_j \neq 0$ at least for an index $j \in \{1, 2, \ldots, m\}$ we obtain that $\lambda_0 \neq 0$. Therefore if we consider $\lambda = 1$ and introduce (32) in (31) then system (26) has a solution for which $\sum_{i=1}^{p} d_{ij}\mu_i \neq 0$ at least for an index $j \in \{1, 2, \ldots, m\}$ and no more than m components among $\mu_1, \mu_2, \ldots, \mu_p$ are nonzero.

$\Leftarrow$ Assume that problem (26) has solution with the properties mentioned in the theorem. This involves that systems (31), (32) have such a solution that $s_j \neq 0$ at least for an index $j \in \{1, 2, \ldots, m\}$ and there exist inequalities (29), (30) that (29) is dependent on (28) and (30) is dependent on a consistent subsystem (27) of inconsistent system (28). This means that there exists a nodal solution $y^0 = (y_1^0, y_2^0, \ldots, y_n^0)$ of system (20) for which $y^0 \notin int\overline{Y}_h$. □

Theorem 7. *Let h^* be the minimal value of parameter h for which system (26) has solution $\mu_1^*, \mu_2^*, \ldots, \mu_p^*, \lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$ such that $\sum_{i=1}^{p} d_{ij}\mu_i^* \neq 0$ at least for an index $j \in \{1, 2, \ldots, m\}$ and no more than m components among $\mu_1^*, \mu_2^*, \ldots, \mu_p^*$ are nonzero. Then h^* is equal to the optimal value of the object function in the following BPP: to minimize (1) on subject (2) and (20) with properties a)-c). An arbitrary solution $y^* = (y_1^*, y_2^*, \ldots, y_m^*)$ of the system of linear inequalities*

$$\begin{cases} \sum_{j=1}^{m} d_{ij}y_j \leq d_{i0}; \;\; i = \overline{1,p}; \\ \sum_{j=1}^{m} s_j^* y_j = s_0^*, \end{cases} \tag{33}$$

with $s_j^* = \sum_{i=1}^{p} d_{ij}\mu_i^*$, $j = \overline{0,m}$, corresponds to a solution of BPP (1), (2), (20).

Proof. Let h^* be the quantity which satisfies the condition of the theorem. Then for an arbitrary $h < h^*$ system (26) has no solution with the properties from Theorem 6. This means that $Y \subset int\overline{Y}_h$ for every $h < h^*$. So, h^* is the maximal value of parameter h for which system (17) is consistent with respect to u for every $y \in Y$. According to Theorems 1 and 2, the point y^* is a point for which system (14) with $h = h^*$ has no solution. Therefore y^* corresponds to a solution of BPP (1), (2), (20). Taking into account that equation $\sum_{j=1}^{m} s_j^* y_j = s_0^*$ determines a supporting plane for Y then a solution of system (32) is a solution of BPP (1), (2), (20). □

Now let us show how to find the solution of system (26) with the properties from Theorem 6.

Theorem 8. *Let be given system of linear inequalities (26) with fixed $h \in [-2^L, N)$ and consider the following $2m$ linear programming problems:*

$$\text{to maximize } f_j = \sum_{i=1}^{p} d_{ij}\mu_i \text{ on subject (26)}, \ \ j = \overline{1,m}; \tag{34}$$

$$\text{to minimize } f_j = \sum_{i=1}^{p} d_{ij}\mu_i \text{ on subject (26)}, \ \ j = \overline{1,m}. \tag{35}$$

Assume that $\overline{f}_1, \overline{f}_2, \ldots, \overline{f}_m$ represent the corresponding optimal values of object functions of problems (34) and $\overline{\overline{f}}_1, \overline{\overline{f}}_2, \ldots, \overline{\overline{f}}_m$ represent the corresponding optimal values of object functions of problems (35). Then system (26) has a solution with the property from Theorem 6 if and only if

1) at least for an index $j \in \{1, 2, \ldots, m\}$ either $\overline{f}_j \neq 0$ or $\overline{\overline{f}}_j \neq 0$;

2) the corresponding basic solution for which 1) holds satisfies the condition that no more than m components among $\mu_1^*, \mu_2^*, \ldots, \mu_p^*$ are nonzero.

Proof. The sufficient condition of the theorem is evident. Let us prove the necessary one. Assume that system (26) has solution $\mu_1^*,\mu_2^*,\ldots,\mu_p^*,\lambda_1^*,\lambda_2^*,\ldots,\lambda_n^*$ which satisfies conditions of Theorem 6. Then it is easy to observe that

$$\overline{f}_{j_0} \geq \sum_{i=1}^{p} d_{ij_0}\mu_i^* \text{ if } \sum_{i=1}^{p} d_{ij_0}\mu_i^* > 0$$

and

$$\overline{\overline{f}}_{j_0} \leq \sum_{i=1}^{p} d_{ij_0}\mu_i^* \text{ if } \sum_{i=1}^{p} d_{ij_0}\mu_i^* < 0.$$

□

Corollary 9. *For given $h \in [-2^L, N)$ a solution of system (25) with the properties from Theorem 6 can be found in polynomial time.*

Based on results described above we can propose the following algorithm.

Algorithm for solving BPP (1), (2), (20) with conditions a)–c)

We replace BPP (1), (2), (20) by system (25), where $h \in [-2^L, N)$. Then using the method of interval bisection after $2L+2$ steps we find $[h_{k-1}, h_k]$ with $\varepsilon = h_k - h_{k-1} < 2^{-2L-2}$ (see [7, 8]), where for $h = h_k$ system (25) has a solution with the property from Theorem 6 and for $h = h_{k-1}$ system (25) does not have such a solution. Based on results from [7, 8] we can find the exact solution h^* in polynomial time by using a special approximate procedure. Note that for problem (6)–(8) it is sufficient to find h_k with precision $\varepsilon \in [0, 1/2)$, because h^* is integer and therefore it can be found from h_k by simple roundoff procedure.

If h^* and a solution $\mu_1^*, \mu_2^*, \ldots, \mu_p^*, \lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$ of system (26) which satisfies conditions of Theorem 6 are known, then find $s_j^* = \sum_{i=1}^{p} d_{ij}\mu_i$, $j = \overline{0,m}$. After that solve system (33) and find the solution $y^* \in Y$. Then fix $y = y^*$ in (1) and solve the linear programming problem: to minimize $z = xCy^* + c'x + c''y^*$ on subject (2). In such a way we find (x^*, y^*).

The proposed algorithm can be used for a large class of integer programming problems and some new results related to computational complexity of the considered problem can be obtained on the basis of such approach.

References

[1] Altman M. *Bilinear Programming.* Bul. Acad. Polan. Sci., Ser. Sci. Math. Astron. et Phis, 1968, **16(9)**, p. 741–746.

[2] Chernicov S.N. *Linear Inequalities.* Moscow, Nauka, 1968 (in Russian)

[3] Farkas J. *Uber die Theorie der einfachen Ungleichungen.* J. reine and angew. Math., 1901, **124**, p. 1–24.

[4] Garey M. Johnson D. *Computers and Intractability: A Guide to the Theory of NP-Complectness.* Freeman, 1979.

[5] Karmanov V. *Mathematical Programming.* Moscow, Nauka, 1990 (in Russian).

[6] Karp R. *Reducibility among combinatorial problems.* In Miler R. and Thatcher J. editors, Complexity of Computer Computations, 1972, p. 83–103.

[7] Khachian L.G. *Polynomial time algorithm in linear programming.* USSR, Computational Mathematics and Mathematical Physics, 1980, **20**, p. 51–58.

[8] Khachian L.G. *On exact solution of the system of linear inequalities and linear programming problem.* USSR, Computational Mathematics and Mathematical Physics, 1982, **22**, p. 999–1002.

[9] Lozovanu D. *Duality Principle for Systems of Linear Inequalities with a Right-Hand Member that Depends on Parameters.* Izv. AN SSSR, Ser. Techn. Cyb., 1987, **6**, p. 3–13 (in Russian).

[10] Lozovanu D. *Extremal-Combinatorial Problems and Algorithms for their Solving.* Chisinau, Stiintsa, 1991 (in Russian).

[11] Lozovanu D. *Parametrical approach for studying and solving bilinear programming problem.* Proceedings of the International Workshop on Global Optimization, San Jose, Almeria, Spain, September 18-22th, 2005, p. 165–170.

Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
Academiei str. 5, MD-2028 Chişinău
Moldova
E-mail: *lozovanu@math.md*

Received September 20, 2007

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 3(55), 2007, Pages 102–110
ISSN 1024–7696

Computation of inertial manifolds in biological models. FitzHugh-Nagumo model

Cristina Nartea

Abstract. Inertial manifolds are related to the large time behaviour of dynamical systems. An algorithm, based on the Lyapunov-Perron method, is implemented here and used to construct a sequence of approximate inertial manifolds for a biological model. The hypotheses of the Jolly, Rosa, Temam's algorithm are verified for the FitzHugh-Nagumo model in the case of real eigenvalues. This algorithm is used for the construction of approximate inertial manifolds.

Mathematics subject classification: 92C20.
Keywords and phrases: Inertial manifolds, dynamical systems, Lyapunov-Perron method, biological model, FitzHugh-Nagumo model.

1 Introduction

The purpose of this paper is to study the approximate inertial manifolds for FitzHugh-Nagumo model, in the case of real eigenvalues using an algorithm developed by Jolly, Rosa and Temam in [5,6].

Let us consider the abstract evolution equation

$$\frac{du}{dt} + Au = f(u), \tag{1}$$

with the initial condition $u(0) = u_0$. Using the associated semigroup $\{S(t)\}_{t\geq 0}$, where $S(t) : u_0 \to u(t)$, $u(\cdot)$ is the solution of (1), with $u(0) = u_0$, the definition of inertial manifolds is given below.

Definition 1. [8]. An **inertial manifold** $\mathcal{M}$ is a finite-dimensional Lipschitz manifold, positively invariant (i.e. $S(t)\mathcal{M} \subset \mathcal{M}$, $t \geq 0$) and which exponentially attracts all orbits of (1).

Any inertial manifold contains the global attractor; and it is easier to describe then the attractor.

An *approximate inertial manifold* (a.i.m.) is a smooth finite dimensional manifold of the phase space which attracts all orbits to a thin neighborhood of it in a finite time uniformly with respect to the initial conditions from a given bounded set. This neighborhood contains the global attractor. The a.i.m.s are useful when an inertial manifold is not known to exist or its exact representation is not known, or when the dimension of the inertial manifold is too high and we want an approximation by a lower finite dimensional system. The algorithm we use in this paper keeps constant the dimension of the a.i.m.s.

2 The algorithm

In [5] and [6] was developed an algorithm for the computation of inertial manifolds. The assumptions presented below guarantee the existence of an inertial manifold and also the convergence of the algorithm.

Consider the equation (1), $u(0) = u_0$, where A is a linear operator, $u \in E$ and E is a Banach space.

A1. The nonlinear term f is globally Lipschitz continuous from E into another Banach space F, $E \subset F \subset \mathcal{E}$, the injections being continuous, each space dense in the following one, and $\mathcal{E}$ is a Banach space. It follows that

$$|f(u)|_F \leq M_0 + M_1 |u|_E,$$

for $M_0 \geq 0$.

A2. The linear operator $-A$ generates a strongly continuous semigroup $\{e^{-tA}\}_{t\geq 0}$ of bounded operators on $\mathcal{E}$ such that $e^{-tA}F \subset E$ for all $t > 0$.

A3. There exist two sequences of numbers $\{\lambda_n\}_{n=n_0}^{n_1}, \{\Lambda_n\}_{n=n_0}^{n_1}$, $n_0 \in \mathbb{N}$, $n_1 \in \mathbb{N} \cup \infty$ such that $0 < \lambda_n \leq \Lambda_n$, for all $n_0 \leq n \leq n_1$, and a sequence of finite-dimensional projectors $\{P_n\}_{n=n_0}^{n_1}$ such that $P_n\mathcal{E}$ is invariant under e^{-tA} for $t \geq 0$, and $\{e^{-tA}|P_n\mathcal{E}\}_{t\geq 0}$ can be extended to a strongly continuous semigroup $\{e^{-tA}P_n\}_{t\in\mathbb{R}}$ of bounded operators on $P_n\mathcal{E}$ with

$$\|e^{-tA}P_n\|_{\mathcal{L}(E)} \leq K_1 e^{-\lambda_n t},\ t \leq 0,$$

$$\|e^{-tA}P_n\|_{\mathcal{L}(F,E)} \leq K_1 \lambda_n^{\alpha} e^{-\lambda_n t},\ t \leq 0,$$

$Q_n\mathcal{E}$ is positively invariant under e^{-tA} for $t \geq 0$, with

$$\|e^{-tA}Q_n\|_{\mathcal{L}(E)} \leq K_2 e^{-\Lambda_n t},\ t \geq 0,$$

$$\|e^{-tA}Q_n\|_{\mathcal{L}(F,E)} \leq K_2 (t^{-\alpha} + \Lambda_n^{\alpha}) e^{-\Lambda_n t},\ t > 0,$$

where $K_1, K_2 \geq 1$ and $0 \leq \alpha < 1$.

A4. The equation (1) has a continuous semiflow $\{S(t)\}_{t\geq 0}$ in E.

A5. There exists $K_3 \geq 0$ independent of n such that $\|AP_n\|_{\mathcal{L}(E)} \leq K_3 \lambda_n$.

A6. A is invertible.

A7. The spectral gap condition

$$\Lambda_n - \lambda_n > 3M_1 K_1 K_2 [\lambda_n^{\alpha} + (1 + \gamma_\alpha)\Lambda_n^{\alpha}],$$

holds for some $n \in \mathbb{N}$, where $\gamma_\alpha = \begin{cases} \int\limits_0^\infty e^{-r} r^{-\alpha} dr, & \text{if } 0 < \alpha < 1, \\ 0, & \text{if } \alpha = 0. \end{cases}$

3 An alternative formulation of the FitzHugh-Nagumo Model

The FitzHugh-Nagumo system [1], modelling the electrical potential in the nodal system of the heart, reads

$$\begin{cases} \dot{x} = c(x + y - x^3/3), \\ \dot{y} = -(x - a + by)/c. \end{cases} \tag{2}$$

To its solution the initial condition $x(0) = x_0$, $y(0) = y_0$ is imposed, where x, y represent the electrical potential of the cell membrane and the excitability, respectively, a, b are real parameters depending on the number of channels of the cell membrane which are open for the ions of K^+ and Ca^{++} and $c > 0$ is the relaxation parameter.

In [2, 3] the global bifurcation diagram provides the qualitative responses of the model for all values of the parameters.

In order to apply to the FitzHugh-Nagumo model the numerical algorithm, this model must be reformulated in an appropriate way. This is done in the present section.

With the notation

$$A = \begin{pmatrix} -c & -c \\ 1/c & b/c \end{pmatrix}, \quad \mathbf{f}(x, y) = \begin{pmatrix} -cx^3/3 \\ a/c \end{pmatrix}. \tag{3}$$

system (2) can be written as

$$\dot{\mathbf{x}} + A\mathbf{x} = \mathbf{f}(\mathbf{x}), \tag{4}$$

where $\mathbf{x} = (x, y)$.

The eigenvalues of A are

$$\lambda_1 = \frac{b - c^2 - \sqrt{(c^2 + b)^2 - 4c^2}}{2c}, \quad \lambda_2 = \frac{b - c^2 + \sqrt{(c^2 + b)^2 - 4c^2}}{2c}$$

and the corresponding eigenvectors, read $v_1 = (1, -\dfrac{c + \lambda_1}{c})$, $v_2 = (1, -\dfrac{c + \lambda_2}{c})$.

We perform the following change of variables

$$\mathbf{x} = T\mathbf{u}, \tag{5}$$

where $\mathbf{u} = (u_1, u_2)$ and T contains the eigenvectors of A, i.e.

$$T = \begin{pmatrix} 1 & 1 \\ -\dfrac{c + \lambda_1}{c} & -\dfrac{c + \lambda_2}{c} \end{pmatrix}. \tag{6}$$

Then, equation (4) becomes

$$T\dot{\mathbf{u}} + AT\mathbf{u} = \mathbf{f}(\mathbf{Tu}).$$

Multiplying the last equation by T^{-1}, we obtain

$$\dot{\mathbf{u}} + T^{-1}AT\mathbf{u} = T^{-1}\mathbf{f}(T\mathbf{u}),$$

Denoting $B = T^{-1}AT$ and $\mathbf{g}(\mathbf{u}) = T^{-1}\mathbf{f}(T\mathbf{u})$, we obtain the modified FitzHugh-Nagumo system, which will be studied further in this paper, namely

$$\dot{\mathbf{u}} + B\mathbf{u} = \mathbf{g}(\mathbf{u}), \tag{7}$$

where B is the diagonal matrix

$$\begin{pmatrix} \zeta & 0 \\ 0 & \eta \end{pmatrix} \tag{8}$$

with

$$\zeta = -\frac{c^4 + 2bc^2 + c^2\sqrt{(c^2+b)^2 - 4c^2} - 4c^2 + b^2 - b\sqrt{(c^2+b)^2-4c^2}}{2c\sqrt{(c^2+b)^2-4c^2}},$$

$$\eta = \frac{c^4 + 2bc^2 - c^2\sqrt{(c^2+b)^2 - 4c^2} - 4c^2 + b^2 + b\sqrt{(c^2+b)^2-4c^2}}{2c\sqrt{(c^2+b)^2-4c^2}},$$

and

$$\mathbf{g}(\mathbf{u}) = \begin{pmatrix} -\dfrac{(c^2+b+\sqrt{(c^2+b)^2-4c^2})(u_1+u_2)^3 c}{2\sqrt{(c^2+b)^2-4c^2}} + \dfrac{ca}{\sqrt{(c^2+b)^2-4c^2}} \\ \\ \dfrac{(c^2+b-\sqrt{(c^2+b)^2-4c^2})(u_1+u_2)^3 c}{2\sqrt{(c^2+b)^2-4c^2}} - \dfrac{ca}{\sqrt{(c^2+b)^2-4c^2}} \end{pmatrix}. \tag{9}$$

4 Checking the hypotheses of the algorithm for the modified FitzHugh-Nagumo model

We deal only with the case of real eigenvalues, i.e. $b \in (-\infty, -c^2 - 2c] \cup [-c^2 + 2c, +\infty)$, because for complex eigenvalues we cannot choose λ_n and Λ_n to satisfy the conditions A3 and A7 of the numerical algorithm.

We consider $E = F = \mathcal{E} = \mathbb{R}^2$.

Assumption A1. The first assumption is that the nonlinear term $\mathbf{g}$ is globally Lipschitz. In order to have this condition fulfilled, we shall further use the prepared equation, as in like [6]. First we verify the Lipschitz condition for $\mathbf{g}$ restricted to the disk of radius r and then we construct the prepared equation, inside the ball of radius ρ the flow of the initial one, being the same with that of the prepared one.

First we compute the Lipschitz constant for each component of $\mathbf{g} = (g_1, g_2)$, and then for $\mathbf{g}$. Let $\mathbf{u} = (u_1, u_2), \mathbf{v} = (v_1, v_2)$ be in the disk of radius r, i.e. $u_1^2 + u_2^2 \leq r^2$ and $v_1^2 + v_2^2 \leq r^2$. We use the norm $\|\mathbf{u}\| = \max\{|u_1|, |u_2|\}$.

$$|g_1(u_1,u_2) - g_1(v_1,v_2)| = \left| \frac{c(c^2+b+\sqrt{(c^2+b)^2-4c^2})[-(u_1+u_2)^3 + (v_1+v_2)^3]}{2\sqrt{(c^2+b)^2-4c^2}} \right| =$$

$$= c\left|\frac{c^2+b+\sqrt{(c^2+b)^2-4c^2}}{2\sqrt{(c^2+b)^2-4c^2}}\right|\cdot|v_1+v_2-u_1-u_2|\cdot|v_1^2+v_2^2-v_1v_2-u_1^2-u_2^2+u_1u_2|.$$

Using $|v_1+v_2-u_1-u_2| \leq |v_1-u_1|+|v_2-u_2| \leq 2\max\{|v_1-u_1|,|v_2-u_2|\} = \|(u_1,u_2)-(v_1,v_2)\|$ and $|u_1|,|u_2|,|v_1|,|v_2| \leq r$, we obtain

$$|g_1(u_1,u_2)-g_1(v_1,v_2)| \leq c\left|\frac{c^2+b+\sqrt{(c^2+b)^2-4c^2}}{2\sqrt{(c^2+b)^2-4c^2}}\right|\cdot 2\|(u_1,u_2)-(v_1,v_2)\|\cdot 6r^2.$$

Hence,

$$|g_1(u_1,u_2)-g_1(v_1,v_2)| \leq c\left|\frac{c^2+b+\sqrt{(c^2+b)^2-4c^2}}{\sqrt{(c^2+b)^2-4c^2}}\right|\cdot 6r^2\|(u_1,u_2)-(v_1,v_2)\|$$

and

$$|g_2(u_1,u_2)-g_2(v_1,v_2)| \leq c\left|\frac{c^2+b-\sqrt{(c^2+b)^2-4c^2}}{\sqrt{(c^2+b)^2-4c^2}}\right|\cdot 6r^2\|(u_1,u_2)-(v_1,v_2)\|.$$

We conclude that

$$\|\mathbf{g}(\mathbf{u})-\mathbf{g}(\mathbf{v})\| \leq M_r\|\mathbf{u}-\mathbf{v}\|, \tag{10}$$

where

$$M_r = \frac{6cr^2}{\sqrt{(c^2+b)^2-4c^2}}\max\{|c^2+b+\sqrt{(c^2+b)^2-4c^2}|,|c^2+b-\sqrt{(c^2+b)^2-4c^2}|\}. \tag{11}$$

Now we determine M such that $\|\mathbf{g}(\mathbf{u})\| \leq M$, for $\mathbf{u}$ inside the disk of radius r. We have

$$\|\mathbf{g}(\mathbf{u})\| = \max\left\{\left|-\frac{(c^2+b+\sqrt{(c^2+b)^2-4c^2})(u_1+u_2)^3c}{2\sqrt{(c^2+b)^2-4c^2}}+\frac{ca}{\sqrt{(c^2+b)^2-4c^2}}\right|,\right.$$
$$\left.\left|\frac{(c^2+b-\sqrt{(c^2+b)^2-4c^2})(u_1+u_2)^3c}{2\sqrt{(c^2+b)^2-4c^2}}-\frac{ca}{\sqrt{(c^2+b)^2-4c^2}}\right|\right\} \leq$$
$$\leq \frac{\max\{|c^2+b+\sqrt{(c^2+b)^2-4c^2}|,|c^2+b-\sqrt{(c^2+b)^2-4c^2}|\}}{2\sqrt{(c^2+b)^2-4c^2}}c|u_1+u_2|^3+$$
$$+\frac{c|a|}{\sqrt{(c^2+b)^2-4c^2}}.$$

Since $|u_1|,|u_2| \leq r$, we have $|u_1+u_2| \leq 2r$ and $|u_1+u_2|^3 \leq 8r^3$. Thus,

$$\|\mathbf{g}(\mathbf{u})\| \leq \frac{4cr^3\max\{|c^2+b+\sqrt{(c^2+b)^2-4c^2}|,|c^2+b-\sqrt{(c^2+b)^2-4c^2}|\}+c|a|}{\sqrt{(c^2+b)^2-4c^2}}. \tag{12}$$

The prepared equation is

$$\frac{d\mathbf{u}}{dt} + B\mathbf{u} = \mathbf{g}_\rho(\mathbf{u}), \tag{13}$$

where $\mathbf{g}_\rho(\mathbf{u}) = \chi_\rho(r)\mathbf{g}(\mathbf{u})$, $\chi_\rho(r) = \chi\left(\frac{r^2}{\rho^2}\right)$, $\chi \in \mathcal{C}^1(\mathbb{R}_+)$, $\chi/_{[0,1]} = 1$, $\chi/_{[2,\infty)} = 0$, $0 \le \chi(s) \le 1, \forall s \in [1,2]$. Thus, the nonlinear term, $\mathbf{g}_\rho(\mathbf{u})$ is zero outside the ball of radius $\rho\sqrt{2}$. For $\chi(s) = 2(s-1)^3 - 3(s-1)^2 + 1$, $s \in [1,2]$, $\chi'(s) = 6(s^2-3s+2)$, hence $\chi'(s) \in \left[-\frac{3}{2}, 0\right]$, i.e. $\chi'(s) \le \frac{3}{2}$. For $s \in \mathbb{R}\backslash[1,2], \chi'(s) = 0 \le \frac{3}{2}$.

Let us compute the Lipschitz constant for $\mathbf{g}_\rho$. For $u_1^2 + u_2^2 \le r_1^2$ and $v_1^2 + v_2^2 \le r_2^2$, we have

$$\|\mathbf{g}_\rho(\mathbf{u}) - \mathbf{g}_\rho(\mathbf{v})\| = \|\chi_\rho(r_1)\mathbf{g}(\mathbf{u}) - \chi_\rho(r_2)\mathbf{g}(\mathbf{v})\| = \|\chi\left(\frac{r_1^2}{\rho^2}\right)\mathbf{g}(\mathbf{u}) - \chi\left(\frac{r_2^2}{\rho^2}\right)\mathbf{g}(\mathbf{v})\| =$$

$$= \|\chi\left(\frac{r_1^2}{\rho^2}\right)\mathbf{g}(\mathbf{u}) - \chi\left(\frac{r_2^2}{\rho^2}\right)\mathbf{g}(\mathbf{u}) + \chi\left(\frac{r_2^2}{\rho^2}\right)\mathbf{g}(\mathbf{u}) - \chi\left(\frac{r_2^2}{\rho^2}\right)\mathbf{g}(\mathbf{v})\| \le$$

$$\le |\chi\left(\frac{r_1^2}{\rho^2}\right) - \chi\left(\frac{r_2^2}{\rho^2}\right)| \cdot \|\mathbf{g}(\mathbf{u})\| + |\chi\left(\frac{r_2^2}{\rho^2}\right)| \cdot \|\mathbf{g}(\mathbf{u}) - \mathbf{g}(\mathbf{v})\| \le$$

$$\le |\chi'(\xi)| \cdot \left|\frac{r_1^2 - r_2^2}{\rho^2}\right| \cdot \|\mathbf{g}(\mathbf{u})\| + \|\mathbf{g}(\mathbf{u}) - \mathbf{g}(\mathbf{v})\|.$$

We have used the Lagrange Theorem, with ξ between $\frac{r_1^2}{\rho^2}$ and $\frac{r_2^2}{\rho^2}$, and $\left|\chi\left(\frac{r_2^2}{\rho^2}\right)\right| \le 1$. Since $|\chi'(\xi)| \le \frac{3}{2}$, using (10) we obtain

$$\|\mathbf{g}_\rho(\mathbf{u}) - \mathbf{g}_\rho(\mathbf{v})\| \le \frac{3}{2\rho^2}|r_1 + r_2| \cdot |r_1 - r_2| \cdot \|\mathbf{g}(\mathbf{u})\| + M_r\|\mathbf{u} - \mathbf{v}\|,$$

with M_r defined in (11).

If $r_{1,2}^2 > 2\rho^2$, then $\|\mathbf{g}_\rho(\mathbf{u}) - \mathbf{g}_\rho(\mathbf{v})\| = 0$. If $r_{1,2}^2 \le 2\rho^2$, then $|r_1 - r_2| \le \sqrt{2}\|\mathbf{u} - \mathbf{v}\|$ and thus, $\|\mathbf{g}_\rho(\mathbf{u}) - \mathbf{g}_\rho(\mathbf{v})\| \le \frac{3}{2\rho^2}2\rho\sqrt{2} \cdot \sqrt{2}\|\mathbf{u} - \mathbf{v}\| \cdot \|\mathbf{g}(\mathbf{u})\| + M_r\|\mathbf{u} - \mathbf{v}\|$. Using $r_{1,2}^2 \le 2\rho^2$ in (12) and (11), we obtain

$$\|\mathbf{g}_\rho(\mathbf{u}) - \mathbf{g}_\rho(\mathbf{v})\| \le M_\rho\|\mathbf{u} - \mathbf{v}\| \tag{14}$$

where

$$M_\rho = \frac{\max\{|c^2 + b + \sqrt{(c^2+b)^2 - 4c^2}|, |c^2 + b - \sqrt{(c^2+b)^2 - 4c^2}|\}}{\sqrt{(c^2+b)^2 - 4c^2}} \times$$

$$\times (48\sqrt{2} + 12)c\rho^2 + \frac{6c|a|}{\rho\sqrt{(c^2+b)^2 - 4c^2}} \tag{15}$$

Assumption A3. We choose the following projectors

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We have $\|e^{-tB}P\| = e^{-\lambda_1 t}$ and $\|e^{-tB}Q\| = e^{-\lambda_2 t}$. We have to choose $0 < \lambda_n \leq \Lambda_n$ to satisfy the conditions A3.

I. The case $0 < \lambda_1 \leq \lambda_2$. We have $\|e^{-tB}P\| = e^{-\lambda_1 t} \leq 1e^{-\lambda_1 t}, \quad \forall t \leq 0$, $\|e^{-tB}Q\| = e^{-\lambda_2 t} \leq 1e^{-\lambda_2 t}, \quad \forall t \geq 0$. So, we can choose $\lambda_n = \lambda_1$, $\Lambda_n = \lambda_2$, $K_1 = 1$, $K_2 = 1$ and $\alpha = 0$.

II. The case $\lambda_1 \leq 0 < \lambda_2$. $\|e^{-tB}P\| = e^{-\lambda_1 t} \leq e^0 < 1e^{-10^{-1}t}, \quad \forall t \leq 0$, $\|e^{-tB}Q\| = e^{-\lambda_2 t} \leq 1e^{-\lambda_2 t}, \quad \forall t \geq 0$. Consequently, for $\lambda_n = 10^{-1}$, $\Lambda_n = \lambda_2$, $K_1 = 1$, $K_2 = 1$ and $\alpha = 0$, we have A3 satisfied if $\lambda_2 \geq 10^{-1}$.

III. The case $\lambda_1 < \lambda_2 \leq 0$. In this case we can not have the conditions A3 satisfied. This would imply that $e^{-\lambda_2 t} \leq K_2 e^{-\Lambda_n t}$ for all $t \geq 0$, i.e. $\Lambda_n \leq \lambda_2 < 0$, which is impossible. Thus, in this case, we can not apply this algorithm.

Assumption A5. $\|BP\| = |\lambda_1|$. In the first case, $\lambda_1 > 0$, hence $\|BP\| = \lambda_1$, $\lambda_n = \lambda_1$, and $K_3 = 1$. In the second case $\lambda_1 < 0$ and we must have $\|BP\| = -\lambda_1 \leq K_3\lambda_n$, where $\lambda_n = \dfrac{1}{10}$.

In conclusion, there exists $K_3 \geq 0$ independent of n such that $\|BP\| \leq K_3\lambda_n$, for λ_n defined as above.

Assumption A7 (Spectral Gap Condition). We must have $\Lambda_n - \lambda_n > 3M_\rho K_1 K_2[\lambda_n^\alpha + (1+\gamma_\alpha)\Lambda_n^\alpha]$. For $\alpha = 0$, we have $\gamma_\alpha = 0$, the condition reads then

$$\Lambda_n - \lambda_n > 6M_\rho, \tag{16}$$

with M_ρ defined in (15).

5 The approximate inertial manifolds for the prepared equation

Using the Jolly, Rosa, Temam's algorithm (see [5],[6]), we have implemented a program, using Scilab software (see [10]), for the construction of approximate inertial manifolds.

The approximate inertial manifolds are the collections of trajectories given by $\mathcal{M}_j = \text{graph}\Phi_j$, where $\Phi_j : P\mathbb{R}^2 \to Q\mathbb{R}^2$, $\Phi_j(p_0) = Q\varphi^j(p_0)(0)$.

For the following choice of parameters, we have all conditions satisfied: $a = 0.01, b = 5, c = 1$; we also choose $\rho = 1/20$. The eigenvalues are $\lambda_1 = 2 - 2\sqrt{2}$ and $\lambda_2 = 2 + 2\sqrt{2}$, i.e. the second case. We take $\lambda_n = 10^{-1}$, $\Lambda_n = \lambda_2$ and then, the spectral gap condition becomes $\frac{19}{10} + 2\sqrt{2} > 2.68$, which is satisfied.

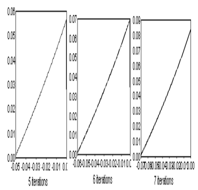

Fig. 1

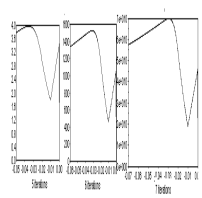

Fig. 2

The graphical representations of $Q\varphi^j$ vs time, for different numbers of iterations, for the initial conditions $u_0 = 1, v_0 = 1$ are shown in Fig. 1. For the same choice of parameters, but for $u_0 = 5, v_0 = 3$ we have the graphics in Fig. 2.

For $a = 0.01, b = 0.9, c = 0.1$, we are situated in the first case, real positive eigenvalues, $\lambda_n = \lambda_1 = 0.011, \Lambda_n = \lambda_2 = 8.89$. Choosing $\rho = 1/10$, the spectral gap condition becomes $8.88 > 1.038$, which is satisfied. For $u_0 = 5, v_0 = 3$ we have Fig. 3.

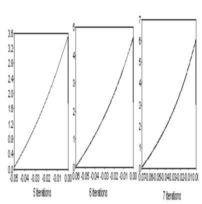

Fig. 3

References

[1] FitzHugh R. *Impulses and physiological states in theoretical models of nerve membrane.* Biophysical J., 1961, **1**, p. 445–446.

[2] Georgescu A., Rocoreanu C., Giurgieanu N. *Global bifurcations in FitzHugh-Nagumo model.* Trends in Mathematics: Bifurcations, Symmetry and Patterns, Birkhäuser Verlag Basel, Switzerland, 2003, p. 197–202.

[3] Rocoreanu C., Georgescu A., Giurgieanu N. *The FitzHugh-Nagumo model. Bifurcation and Dynamics.* Kluwer, Dordrecht, 2000.

[4] Jolly M.S., Rosa R. *Computation of non-smooth local centre manifolds.* IMA J. of Numerical Analysis, 2005, **25**, p. 698–725.

[5] Jolly M.S., Rosa R., Temam R. *Accurate computations on inertial manifold.* SIAM J. Sci. Comput., 2000, **22**, N 6, p. 2216–2238.

[6] Rosa R. *Approximate inertial manifolds of exponential order.* Discrete and Continuous Dynamical Systems, 1995, **1**, p. 421–448.

[7] Rosa R., Temam R. *Inertial manifolds and normal hyperbolicity.* ACTA Applicandae Mathematicae, 1996, **45**, p. 1–50.

[8] Temam R. *Infinite-dimensional dynamical systems in mechanics and physics.* Applied Mathematical Sciences, 1997, **68**, Springer, Berlin.

[9] Tu P. *Dynamical systems. An introduction with applications in economics and biology.* Springer, Berlin, 1994.

[10] www.scilab.org/

University of Pitesti
Romania
E-mail: *cristina.nartea@k.ro*

Received September 28, 2006