Contents

1 Motivation for multicriteria strategies in distributed system resource management 1
   Costel Aldea

2 Some new properties of Lie triple systems 7
   Ilie Burdujan

3 Some problems of the theory of compactifications of topological spaces 25
   Laurențiu I. Calmuțchi, Mitrofan M. Choban

4 New iterative chain classes of pseudo-boolean functions 53
   Vadim Cebotari, Mefodie Rață

5 Applications to Lie algebras 57
   Camelia Ciobanu

6 Perturbed bifurcation in a demand-offer model 65
   Elena Codeci

7 Inverse problems of transitive closure 71
   Adrian Deaconu

8 Further study of a microconvection model using the direct method 77
   Ioana Dragomirescu, Adelina Georgescu

9 An optimization method for free surface steady flows 87
   Tiberiu Ioana, Titus Petrila

10 Approximate inertial manifolds for an advection-diffusion problem 91
    Anca-Veronica Ion
1. Finite state machine testing from an OR state refinement design
   *Florentin Ipate, Tudor Bălănescu*

2. Sharpening of critical spectral parameter by pseudoperturbation method
   *Boris V. Loginov*, *Olga V. Makeeva**, *Elena V. Foliadova***

3. Existence results for a class of nonlinear Volterra integrodifferential problem
   *Rodica Luca–Tudorache*

4. Computer realization of the branching equation construction
   *Oleg V. Makeev*

5. On a sparse matrix memorizing method
   *Vlad Monescu, Silviu Dumitrescu, Bogdana Pop*

6. Regenerative process with a rare event
   *Bogdan-Gh. Munteanu*

7. Stability of limit cycles in a calcium oscillations dynamics model
   *Ileana Rodica Nicola*

8. Estimation of Hausdorff and packing linear measure and dimensions for some plane sets of points
   *Antonio Nuică*

9. Error handling in software systems: modeling and testing with Finite State Machines
   *Radu Oprișa*

10. Using stochastic Petri nets in performance analyze of distributed systems
    *Norocel Petrușche*

11. Solving method for multiple objective fuzzy constraints mixed variables optimization
    *Bogdana Pop, Silviu Dumitrescu, Vlad Monescu*

12. On the applicability of the Lindstedt-Poincaré method in the two dimensional case
    *Nicolae - Dora Stănescu*
<table>
<thead>
<tr>
<th>23</th>
<th>The generalized algorithm for solving the fractional multiobjective transportation problem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alexandra I. Tkacenko</td>
</tr>
<tr>
<td>24</td>
<td>LISC – A MATLAB package for linear differential problems</td>
</tr>
<tr>
<td></td>
<td>Damian Trif</td>
</tr>
<tr>
<td>25</td>
<td>Optimal control and stochastic uniform observability</td>
</tr>
<tr>
<td></td>
<td>Viorica Mariela Ungureanu</td>
</tr>
<tr>
<td>26</td>
<td>Mathematical models of a pipeline - pressure sensor mechanical system</td>
</tr>
<tr>
<td></td>
<td>Petr A. Velmisov, Yuliya V. Pokladova</td>
</tr>
<tr>
<td>27</td>
<td>Extending object-oriented databases for fuzzy information modeling</td>
</tr>
<tr>
<td></td>
<td>Cristina-Maria Vladarean</td>
</tr>
</tbody>
</table>
MOTIVATION FOR MULTICRITERIA STRATEGIES IN DISTRIBUTED SYSTEM RESOURCE MANAGEMENT PROBLEMS

Costel Aldea

University "Transilvania" of Brasov
costel@info.unitbv.ro

Abstract In the resource management systems the following aspects must be considered: the virtual organization policies and the application requirements and, in some cases, the user preferences. According to these multicriteria strategies for distributed resource management should be formulated and considered. Some available multicriteria optimization techniques and methods are discussed. The dynamic behaviour, uncertainty and incomplete information are important aspects of the distributed system nature that gives the core of the resource management process. Some aspects of the resource management process are: complete information about the accessible resource, applications requirements and user preferences, association of the tasks with the resource in the best possible ways, supporting the changes in the distributed system with respect to distributed application control.

1. INTRODUCTION

In order to describe the resource management process in a distributed system it is important to know the participants to this process. There are also many aspects in modelling their preferences in order to obtain a good association between the tasks and the resources. Nevertheless, one must observe other techniques for satisfying the requirements given by the dynamic behaviour of the tasks related to the resources as well as for the uncertainty and incomplete information.

Some basic definition and notions related to multicriteria approach are introduced. The aim is to motivate the importance of multicriteria strategies in distributed systems resource management.

The performance and the high-throughput of the resource management implies the following comparison between two main resource management strategies: application level scheduling and job scheduling. It is obvious that the combination between those two strategies can bring together all the benefits of the both strategies using multicriteria strategies approach and the support of artificial intelligence techniques. An analysis of possible criteria that can be used in distributed system resource management process is presented.
and how the preference models can be exploited in order to assign resources to the tasks in the most appropriate way is shown.

2. MULTICRITERIA APPROACH

A formulation for the multicriteria decision problem is made and some basic definition and various ways of modelling decisions maker’s preferences are described.

Multicriteria decision problem. If there are several different ways to achieve a goal, a decision problem needs to be solved. More then that, to reach the goal, optimized choices of the best method must be used. These different ways are decision actions. They are also called the compromise solutions or schedules.

In general the main goal of the multicriteria decision making process is to build a global model of the decision makers’ preferences and exploit them in order to find the solution that is accepted by a decision maker with respect to the values on all criteria. In order to accomplish this, preferential information concerning the importance of particular criteria must be obtained from a decisionmaker.

If these preferences are not available, then the only objective information is the dominance relation in the set of decision actions.

Basic definitions. For the choice of the compromise solution the concepts of non-dominated and Pareto optimal are used. There are two important spaces to consider: decision variable space and the criteria space. The former contains possible solutions along with values of their attributes. Let us denote by $D$ the set of solutions. The set $D$ can be easily mapped into its image in the criteria space, which creates the sets of points denoted by $C(D)$. The functions $\phi(x)$ represent values of criteria and their number is denoted by $k$.

**Definition 2.1 (Pareto dominance).** The point $z \in C(D)$ dominates $z' \in C(D)$, denoted as $z \succ z'$, if and only if $(\forall) j \in \{1, \cdots, k\}, z_j \geq z'_j \land (\exists) i \in \{1, \cdots, k\}: z_i > z'_i$ ($z'$ is partially less than $z$). Thus, one point dominates another if it is not worse with respect to all criteria and it is better for at least one of them.

**Definition 2.2 (Pareto optimality).** The point $z' \in C(D)$ is said to be non-dominated with respect to the set $C(D)$ if there is no $z \in C(D)$ that dominates $z'$. A solution is $x$ Pareto optimal (efficient) if its image in the criteria space is non-dominated.

**Definition 2.3 (Pareto-optimal set).** The set of $P^*$ of all Pareto-optimal solutions is called the Pareto optimal set. Thus, it is defined as

$$P^* = \left\{ x \in D \mid \neg (\exists) x' \in D : z' \succ z \right\}, \text{ where } z' \text{ and } z \in C(D)$$

**Definition 2.4 (Pareto Front).** For a given Pareto optimal set $P^*$, Pareto front ($PF^*$) is defined as $PF^* = \left\{ z = \{f_1 = z_1, \cdots, f_k(x) = z_k\} \mid x \in P^* \right\}$. A
Pareto front is also called a non-dominated set because it contains all non-dominated points.

Preference models. A preference model defines a preference structure in the set of decision actions. Preference models can aggregate evaluations criteria into:
- functions (e.g. weighted sums);
- statements (e.g. binary relation: action $a$ is at least as good as action $b$);
- logical statements (e.g. decision rules: if condition on criteria then decision, where if condition on criteria is the conditional part of the rule and then decision is the decision part of the rule).

3. MOTIVATION FOR MULTIPLE CRITERIA

Multicriteria approaches focus on a compromise solution (in this case compromise schedule). In this way one can increase the level of satisfaction of many stakeholders of the decision making process and try to combine various points of view, rather than provide solutions that are very good from only one specific perspective as is currently common in other distributed system resource management approaches (grids). Such specific perspective result in different, often contradictory, criteria (along with preferences) and make the process of mapping jobs to resources difficult or even impossible. Consequently all the different perspectives and preferences must be somehow aggregated to please to all participants of the distributed system resource management process.

Various stakeholders and their preferences. Distributed systems scheduling and resource management potentially involves the interaction of many human players (although possibly indirectly). These players can be divided into three classes:
- end users making use of distributed system applications and portals;
- resource administrators and owners;
- virtual organization (VO) administrators and VO policy makers.

One goal of the resource management process is to automate the scheduling and resource management process in order to maximize stakeholders participations in the entire process. This is why the final decision is often delegated to such systems. By decision maker notion one means a scheduler and all human players listed above are stakeholders of a decision making process. This model assumes a single artificial decision maker in a VO. However, in real distributed systems, multiple stakeholders (agents) may often act according to certain strategies, and the decision may be made by means of negotiations between these agents. Such an approach requires distributed decision making models with multiagent interactions [1], which will not be discussed here.

One refers to the assignment of resources to tasks as a solution or a schedule. Since one considers many contradictory criteria and objective functions,
finding the optimal solution is not possible. The solution that is satisfactory from all the stakeholders points of view and that takes into consideration all criteria is called the **compromise solution**.

The need for formulation of distributed systems resource management as a multicriteria problem follows from the characteristics of the distributed system environment itself. Such an environment has to meet requirements of different groups of stakeholders listed at the beginning of this section. Thus, various points of view and policies must be taken into consideration, and results need to address different criteria for the evaluation of schedules. Different stakeholders have different, often contradictory, preferences that must be somehow aggregated to please all stakeholders. For instance, administrators require robust and reliable work of the set resources controlled by them, while end users often want to run their applications on every available resource. Additionally, site administrators often want to maintain a stable load of their machines and thus load-balancing criteria become important. Resource owners want to achieve maximal throughput (to maximize work done in a certain amount of time), while end users expect good performance of their application or, sometimes, good throughput of their set of tasks. Finally, VO administrators and policy makers are interested in maximizing the whole performance of the VO in the way that satisfies both end users and administrators.

Different stakeholders are not the only reason for multiple criteria. Users may evaluate schedules simultaneously with respect to multiple criteria. For example, one set of end users may want their applications to complete as soon as possible, whereas another one try to pay the minimum cost for the resources to be used. Furthermore, the stakeholders may have different preferences even inside one group. For example, a user may consider time of execution more important than the cost (which does not mean that the cost is ignored).

**Job Scheduling** involves the use of one central scheduler that is responsible for assigning the distributed system resources to applications across multiple administrative domains. There are three main levels: a set of applications, a central scheduler, and the distributed system resources. In principle, this approach is in common use and can be found today in many commercial and public-domain job-scheduling systems. A detailed analysis and their comparative study can be found in [2].

Note that the dynamic behaviour seen in distributed systems environments is most often the result of competing jobs executed on the resources. By having a control over all the applications, as happens in the job scheduling approach, a central scheduler is able to focus on factors that cause variable, dynamic behavior, whereas application-level schedulers have to cope with this effects.

**Application-Level Scheduling.** With application-level scheduling, applications make scheduling decisions themselves, adapting to the availability of
Motivation for multicriteria strategies in distributed system resource management

resources by using additional mechanisms and optimizing their own performance. That is, applications try to control their behavior on the distributed system resources independently. Note that all resource requirements are integrated within applications.

Such an assumption causes many ambiguities in terms of the classic scheduling definitions. When one considers classic scheduling problems, the main goal is to effectively assign all the available task requirements to available resources so that they are satisfied in terms of particular objectives. In general, this definition fits more the job scheduling approach, where job scheduling of systems take care of all the available applications and map them to resources in a suitable way. In both cases application-level scheduling differs from the job-scheduling approach in the way applications compete for the resources. In application-level scheduling, an application schedule have no knowledge of the other applications, and in particular of any scheduling decisions made by another application. Hence, if the number of applications increases (which is a common situation in large and widely distributed systems like a Grid), one observe worse scheduling decisions from the system perspective, especially in terms of its overall throughput. Moreover, since each application must have its own scheduling mechanisms, information processes may be unnecessarily duplicated (because of the lack of knowledge about other applications’ scheduling mechanisms).

Hence, a natural question arises: why does one need the application level scheduling approach? In practice, job scheduling systems do not know about many specific internal application mechanisms, properties, and behaviors. Obviously, many of them can be represented in the form of before mentioned hard constraints, as applications requirements, but still it is often impossible to specify all of them in advance. More and more, end users are interested in new Grid-aware application scenarios. New scenarios assume that dynamic changes in application behavior can happen during both launch-time and runtime. Consequently, Grid-aware applications adapt on-the-fly to the dynamically changing Grid environment if a pure job scheduling approach is used. The adaptation techniques, which in fact vary with classes of applications, include different application performance models, and the ability to checkpoint and migrate.

Consequently, job scheduling-systems are efficient from high-throughput perspective (throughput based criteria) where as application-level schedulers miss this important criterion, even though internal scheduling mechanisms can significantly improve particular application performance (application performance based criteria).

**Hard and soft constraints.** By extending present job-scheduling strategies (especially to the war application requirements are expressed), one is able to deal with both throughput and application-centric needs. As noted in the
previous section, a central scheduler receives many requests from different end users to execute and control their applications. Requests are represented in the form of hard constraints, a set of applications and resources requirements that must be fulfilled in order to run an application or begin any other action on particular resources. These constraints are usually specified by a resource description language such as RSL in the Globus Toolkit, ClassAds in Condor, or JDL in DataGrid. Simply stated, a resource description language allows the definition of various hard constraints, such as the operating system type, environment variables, memory and disk size, and number and performance of processors, and generally varies with classes of applications on the distributed system. Many extensions to the resource description languages are needed to represent soft constraints, the criteria for resource utilization, deadlines, response time, and so forth needed when many stakeholders are involved. Unfortunately, soft constraints are rarely expressed in the resource description language semantics, and consequently they are omitted in many resource management strategies.

The main difference between hard and soft constraints is that hard constraints must be obeyed, whereas soft constraints should be taken into consideration in order to satisfy all the stakeholders as far as possible. In other words, one is able to consider soft constraints if and only if all hard constraints are met. Yet, soft constraints are actually criteria in the light of decision making process, and they are tightly related to the preferences that the stakeholders of the process want to consider.

We have indicated that resource management in a distributed environment requires a multicriteria approach. However, proposition of a method copying with multiple criteria during the selection of the compromise schedule does not solve the resource management problem entirely.

References

SOME NEW PROPERTIES OF LIE TRIPLE SYSTEMS

Ilie Burdujan

Dept. of Math., Univ. of Agricultural Sciences and Veterinary Medicine
"Ion Ionescu de la Brad" Iași
burdujan.ilie@yahoo.com

Abstract

The categories $\mathcal{LTS}$ of Lie triple systems (briefly, Ltss) and $\mathcal{LTALG}$ of LT-algebras are considered. There exists a covariant functor $\mathcal{LT}: \mathcal{LTALG} \to \mathcal{LTS}$ that assigns to each LT-algebra $A$ a specific Ltss denoted by $\mathcal{LT}(A)$. $\mathcal{LT}$ has a left adjoint functor, i.e., there exists a functor $\mathcal{U}: \mathcal{LTS} \to \mathcal{LTALG}$ having the property $\text{Hom}_{\mathcal{LTS}}(T, \mathcal{LT}(A)) \cong \text{Hom}_{\mathcal{LTALG}}(\mathcal{U}(T), A)$. The existence of $\mathcal{U}$ gives rise to the construction of the universal enveloping LT-algebra. For any Ltss $T$ a universal enveloping algebra $\mathcal{U}(T)$ was constructed [4] as being the LT-algebra obtained by factorizing the nonassociative tensor algebra of $T$ by an appropriate ideal. The usual results concerning such a universal enveloping were proved in a similar way as the corresponding results on the universal enveloping algebra of a Lie algebra. A similar result as that stated in Poincaré-Birkhoff-Witt Theorem is proved, namely: the cosets of 1 and the standard monomials for a basis of Ltss $T$ form a basis for the universal enveloping $\mathcal{U}(T)$.

Keywords: Lie triple system, universal enveloping algebra.

2000 MSC: 17A40, 17B35

1. INTRODUCTION

Lie triple systems (briefly, Ltss) are subspaces of any Lie algebra $L([.,.])$ which are closed under the ternary composition $[x, y, z] = [[x, y], z]$. They were first noted by É. Cartan in his study on totally geodesic submanifolds [5]. More exactly, if $\mathfrak{g}$ is the Lie algebra of the Lie group $G$, then Ltss in $\mathfrak{g}$ are connected with the totally geodesic subspaces of $G$ in the same way that Lie subalgebras of $\mathfrak{g}$ are related to analytic subgroups of $G$. The Ltss arose also in the studies on symmetric spaces [5], [20], [22]. They were also used in differential geometry by P. I. Kovaljov in the study of certain manifolds endowed with affine connections [17], [18]. It must be also remarked that they arose naturally in the investigation by H. Freudenthal (1954) on the geometries of exceptional simple Lie groups [8]. Recently, N. Kamiya and S. Okubo [16] connect Ltss with the study of the Yang-Baxter equation.

From the algebraic point of view, Ltss were studied by N. Jacobson [14], [15], W. G. Lister [19], K. Yamaguti [25], [26], [27], B. Harris [9], N. Hopkins
Ilie Burdujan
[11], [12], [13], T. Hodge [10] et all. Actually, they were introduced by N. Jacobson [14] as being the abstract algebraic structures describing the subspaces of an associative algebra that are closed relative to the ternary operations \([x, y], z\), where \([x, y] = xy - yx\). A crucial moment in the development of Lts structural theory was the result, proved in 1951 by N. Jacobson [14], stating that every Lts \((T, [\cdot, \cdot, \cdot])\) can be embedded into a Lie algebra \((L, [\cdot, \cdot])\) such that its ternary operation is a superposition of Lie brackets of \(L\) (i.e., \([x, y, z] = [(x, y), z], \forall x, y, z \in T\)). This embedding suggests that Lts are the tangent algebras of particular homogeneous spaces.

Another important class of Lts is connected with a class of commutative algebras satisfying an identity of degree 4, the so called LT-algebras [4]. This connection is yielded by means of the ternary operation \([x, y, z] = x \cdot (y \cdot z) - y \cdot (x \cdot z)\). Moreover, such a commutative algebra can be associated with every nonassociative algebra, that allows us to give an embedding result for Lts similar to Jacobson’s embedding result. It was proved [4] that a universal enveloping LT-algebra can be associated with every Lts. More exactly, for any Lts \(T\), the universal enveloping LT-algebra \(\mathcal{U}(T)\) is a nonassociative algebra having the property that any Lts homomorphism of \(T\) into the Lts associated with a LT-algebra \(A\) extends to an algebra isomorphism of \(\mathcal{U}(T)\) into \(A\). The algebra \(\mathcal{U}(T)\) is a quotient of the nonassociative tensor algebra \(\mathcal{T}T\) by a suitable two sided ideal and it is a filtered algebra.

In this paper we prove, for \(\mathcal{U}(T)\), a similar result to that stated in the Poincaré-Birkhoff-Witt theorem for universal enveloping of a Lie algebra, namely: the cosets of 1 and the standard monomials for a basis of Lts \(T\) form a basis for the universal enveloping \(\mathcal{U}(T)\).

2. PRELIMINARIES

Definition 2.1 A Lie triple system (briefly, Lts) is a vector space \(T\) over the field \(K\), with a ternary composition \([\cdot, \cdot, \cdot] : T \times T \times T \to T, (a, b, c) \to [a, b, c]\), which is three-linear and satisfies the following axioms:

\[
\begin{align*}
\text{(Lts.1)} & \quad [x, x, y] = 0, \\
\text{(Lts.2)} & \quad [x, y, z] + [y, z, x] + [z, x, y] = 0, \\
\text{(Lts.3)} & \quad [x, y, [u, v, w]] = [[x, y, u], v, w] + [u, [x, y, v], w] + [u, v, [x, y, w]],
\end{align*}
\]

\(\forall x, y, z, u, v, w \in T\).

We consider, for every \(x, y \in T\), the vector space endomorphism \(D_{(x,y)} : T \to T\) defined by \(D_{(x,y)}(z) = [x, y, z]\); (Lts. 3) assures us that \(D_{(x,y)}\) is a Lts-derivation which is called an inner derivation of \(T\). Moreover, \(\mathcal{D} = \text{Span}_K \{D_{(x,y)} | x, y \in T\}\) has a Lie algebra structure; \(\mathcal{D}\) is the so-called inner derivation algebra and it is also denoted by \(\text{InnDer}(T)\).

A Lts is called Abelian if \([x, y, z] = 0\) for any triple \((x, y, z)\). Of course,
every $K$-module $T$ is an Abelian Lts by defining the Lts-bracket as $[x, y, z] = 0, \forall x, y, z \in T$.

Let $\mathcal{LTS}_K$ denote the category of Lie triple systems over $K$ with the Lts homomorphisms as morphisms.

If $L$ is a Lie algebra with product $[a, b]$, then the ternary composition $[[a, b], c]$ satisfies the above identities. In 1951 N. Jacobson showed [15] that any Lts may be considered as a subspace of a Lie algebra $L$ in such a way that $[[a, b], c]$; this Lie algebra is called the standard enveloping Lie algebra of $T$, and it is denoted by $L_s(T)$ or merely $L_s$; the natural embedding map of $T$ into $L_s(T)$ is 1-1.

Further, a universal associative algebra (with identity) $\mathbb{U}(T)$ with the same property relative to the homomorphisms of $T$ into a Lie algebra extends to a homomorphism of $L_u(T)$. Further, a universal associative algebra of $L_u(T)$, with all other products vanishing.

The existence of a new kind of universal enveloping algebra for Lts was proved in [4]. This is neither a Lie algebra nor an associative one, but it is a commutative algebra satisfying a weak associative law, the so-called LT-algebra. Recall the definition of the LT-algebra.

**Definition 2.2** Any commutative algebra $A(\cdot)$ (with or without identity element) satisfying the identity

$$2((y \cdot x) \cdot x) \cdot x + y \cdot (x^2 \cdot x) = 3(y \cdot x^2) \cdot x, \quad \forall x, y \in A \quad (1)$$

is called a LT-algebra.

By linearizing (2.1) we get its equivalent form

$$f(x, y, z, u) = 0 \quad (2)$$

where

$$f(x, y, z, u) = x \cdot (y \cdot (z \cdot u) - y \cdot (x \cdot z \cdot u)) - (x \cdot (y \cdot z)) \cdot u + + (y \cdot (x \cdot z)) \cdot u - z \cdot (x \cdot (y \cdot u)) + z \cdot (y \cdot (x \cdot u)). \quad (3)$$

As examples we consider: 1° LT-algebras obtained by factorizing a commutative algebra $A(\cdot)$ by the ideal generated by $\{f(x, y, z, w); x, y, z, w \in A\}$ [4], 2° Jordan algebras satisfying the conditions from Corollary of Theorem 2 [23] are LT-algebras.

Let us denote by $\mathcal{LJAG}_K$ the category of LT-algebras over $K$ with the algebra homomorphisms as morphisms.
Every LT-algebra $A(\cdot)$ becomes a Lts relative to the composition
\[ [x, y, z] = x \cdot (y \cdot z) - y \cdot (x \cdot z), \quad \forall x, y, z \in A; \] (4)
this Lts is denoted by $\mathcal{LT}(A)$. Consequently, there exists a covariant functor $\mathcal{LT}: \mathcal{LTS}_{KT} \to \mathcal{LTS}_{K}$, carrying every LT-algebra $A(\cdot)$ in $\mathcal{LT}(A)$. It must be remarked that every inner derivation for $\mathcal{LT}(A)$ is a derivation for $A(\cdot)$.

Moreover, the following identities hold in any LT-algebra
\[ x_1 \cdot (x_2 \cdot (a_1 \cdot (a_2 \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots))) = \quad (5) \]
\[ = x_2 \cdot (x_1 \cdot (a_1 \cdot (a_2 \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots)) + \]
\[ + D_{(x_1,x_2)}(a_1 \cdot (a_2 \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots)), \quad m \geq 2, \]
\[ a_1 \cdot (a_2 \cdot \ldots \cdot (a_{i-1} \cdot (a_{i+1} \cdot (a_{i+2} \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots))) = \quad (6) \]
\[ = a_1 \cdot (a_2 \cdot \ldots \cdot (a_{i-1} \cdot (a_{i+1} \cdot (a_{i+2} \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots))) + \]
\[ a_1 \cdot (a_2 \cdot \ldots \cdot (a_{i-1} \cdot (D_{(a_{i+1})}(a_{i+2} \cdot \ldots \cdot (a_{m-1} \cdot a_m) \ldots))). \]

Notations. For the later use we introduce the notations
\[ [x, y, z] = x \otimes (y \otimes z) - y \otimes (x \otimes z). \]
\[ f \otimes (x, y, z, w) = x \otimes (y \otimes (z \otimes w)) - y \otimes (x \otimes (z \otimes w)) + \]
\[ + z \otimes (y \otimes (x \otimes w)) - (x \otimes (y \otimes z)) \otimes w + (y \otimes (x \otimes z)) \otimes w. \]

They are obtained by substituting everywhere in the expressions (2.4), (2.3) of $[x, y, z], f(x, y, z, w)$, respectively, the binary operation symbol “” by “$\otimes$”. In [4] the existence of a left adjoint for $\mathcal{LT}$ was proved; this is equivalent with the construction of a universal enveloping algebra $\mathcal{U}(T)$ for any Lts $T$. $\mathcal{U}(T)$ is obtained by a factorization of the nonassociative tensor algebra $\mathcal{T}\{T\}$ by an apropriate ideal. Recall that
\[ \mathcal{T}\{T\} = T^{\otimes 0} \oplus T^{\otimes 1} \oplus T^{\otimes 2} \oplus \ldots \oplus T^{\otimes n} \oplus \ldots \]
where, $T^{\otimes 0} = K1, T^{\otimes 1} = T$ and, for $n \geq 2$,
\[ T^{\otimes n} = \sum_{i=1}^{n-1} T^{\otimes i} \otimes T^{\otimes n-i}, \]
the product of $v \in T^{\otimes i}$ and $w \in T^{\otimes j}$ is defined as $v \cdot w = v \otimes w$. The algebra $\mathcal{T}\{T\}$ has the following universal property: for any $K$-algebra $A$ and any $K$-linear map $f : T \to A$ there exists a unique algebra homomorphism $f_0 : \mathcal{T}\{T\} \to A$ extending $f$. In other words, the functor $\mathcal{T}$ is left adjoint to
Let $X$ be an ordered set.

Consider the ordered set $(v_1, \ldots, v_n)$ realized in pairs delimited by a fixed selection $s$ of parenthesis, and $[f(v_1) \cdot f(v_2) \cdot \ldots \cdot f(v_n)]_s$ is the element obtained in $A$ by composing the elements $f(v_i)$ and preserving the selection $s$ of parenthesis.

**Definition 2.3** Let $T$ be a Lts over $K$. The pair $(U, i)$, where $U$ is a LT-algebra over $K$ and $i$ is a Lts-homomorphism from $A$ to $L(T(U))$ (i.e., it is a linear map such that

$$i([x, y, z]) = i(x) \cdot [i(y) \cdot i(z)] - i(y) \cdot [i(x) \cdot i(z)], \quad \forall \ x, y, z \in T,$$

(7)

is called a **universal enveloping LT-algebra** of $T$ if the following property holds: for any LT-algebra $B$ and any Lts-homomorphism $j$ from $T$ to $L(T(B))$, there exists a unique homomorphisms of LT-algebras $\Phi : U \rightarrow B$ such that $\Phi \circ i = j$.

The uniqueness of the universal enveloping LT-algebra $(U, i)$ of $T$ is easily provable in a standard way. Further, the LT-algebra $U = U(T)$ is generated by $i(T)$. In order to prove the existence of a suitable pair $(U, i)$ we consider $T \{ T \}$ and factorize it by the two sided ideal $J$ generated by $\{ x \otimes y - y \otimes x, x \otimes (y \otimes z) - y \otimes (x \otimes z) - [x, y, z], f^\otimes(x, y, z, w) \}$, where $T$ is considered naturally imbedded in $T \{ T \}$. We put $U = U(T) = T \{ T \} / J$, denote by $\pi : T(T) \rightarrow U(T)$ the canonical homomorphism, and define $i : T \rightarrow U(T)$ to be the restriction of $\pi$ to $T$. Notice that $J \subset \otimes_{i=1}^\infty T^\otimes(T)$, so $\pi$ maps $T$ isomorphically into $T \subset T \{ T \}$, and consequently, $i$ is a 1-1 map. Then, $(U(T), i)$ is a universal enveloping algebra of $T$. In [4] the following theorem was proved.

**Theorem 2.1** 1. Let $T_1$ and $T_2$ be two Lts, $(U_1, i_1)$, $(U_2, i_2)$-their corresponding universal enveloping LT-algebras, and $\alpha : T_1 \rightarrow T_2$ be a Lts-homomorphism. Then, there exists a unique LT-homomorphism $\alpha' : U_1 \rightarrow U_2$ such that $\alpha' \circ i_1 = i_2 \circ \alpha$.

2. Let $B$ be a two-sided ideal of $T$ and let $R$ be the two-sided ideal of $U$ generated by $i(B)$. If $a \in T$, then $j : a + B \rightarrow i(a) + R$ is a Lts-homomorphism of $T/B$ and $L(T(\mathcal{V}))$, where $\mathcal{V} = U/R$; further $(\mathcal{V}, j)$ is the universal enveloping LT-algebra for $T/B$.

3. If $D$ is a derivation for the Lts $T$, then there exists a uniquely defined derivation $D'$ for the LT-algebra $U$ such that $D' \circ i = i \circ D$.

### 3. STANDARD POLYNOMIALS

In what follows, we consider the Lts $T$ having the basis $X = \{ x_j \}_{j \in J}$ where $J$ is an ordered set (possible, a finite one). In particular, we can consider $X = \{ x_1, x_2, \ldots, x_n, \ldots \}$ or $X = \{ x_1, x_2, \ldots, x_n \}$. Let us consider the set
Let \( + u \) hold for all \( \in X \). Then, the following identities hold:

\[
\begin{align*}
& (i) \quad u_i \otimes u_j = u_j \otimes u_i (\text{mod} J), \\
& (ii) \quad u_i \otimes (u_j \otimes u_k) = u_j \otimes (u_i \otimes u_k) + [u_i, u_j, u_k] (\text{mod} J), \\
& (iii) \quad u_i \otimes (u_j \otimes (u_k \otimes u_l)) = u_j \otimes (u_i \otimes (u_k \otimes u_l)) + u_k \otimes [u_i, u_j, u_l] + \\
& \quad \quad + [u_i, u_j, u_k] \otimes u_l (\text{mod} J), \\
& (iv) \quad u_j \otimes (u_j \otimes (u_j \otimes (u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots))) = \\
& \quad \quad = u_j \otimes (u_j \otimes (u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots)) + \\
& \quad \quad + D_{(u_j, u_j)}^{(u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots)}(u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots) (\text{mod} J), m \geq 5,
\end{align*}
\]

hold for all \( u_i, u_j, u_k, u_l, u_m \in T \); here \( D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)} \) and \( D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)} \) are defined, respectively, by

\[
D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)}([a_1, a_2, \ldots, a_n]_s) =
\]

\[
= \sum_{i=1}^{n} [a_1, a_2, \ldots, a_{i-1}, [x, y, a_i]_{s}, a_{i+1}, \ldots, a_n]_s,
\]

4. **THE ANALOG OF POINCARÉ-BIRKHOFF-WITT THEOREM**

Let \( T \) be a \( K \)-Lts and \( \mathfrak{u}(T) \) be its universal enveloping. Also, let \( X = \{x_1, x_2, \ldots, x_n, \ldots\} \) be a basis for \( T \) and \( t_{jkm}^{ij} \in K \) be its structure constants. Firstly, we shall prove the following result.

**Lemma 4.1** Let \( T \) be a Lts and \( \mathcal{T}(T) \) be its nonassociative tensor algebra. Then, the following identities hold for any \( u_i, u_j, u_k, u_l \in T \):

\[
\begin{align*}
& (i) \quad u_i \otimes u_j = u_j \otimes u_i (\text{mod} J), \\
& (ii) \quad u_i \otimes (u_j \otimes u_k) = u_j \otimes (u_i \otimes u_k) + [u_i, u_j, u_k] (\text{mod} J), \\
& (iii) \quad u_i \otimes (u_j \otimes (u_k \otimes u_l)) = u_j \otimes (u_i \otimes (u_k \otimes u_l)) + u_k \otimes [u_i, u_j, u_l] + \\
& \quad \quad + [u_i, u_j, u_k] \otimes u_l (\text{mod} J), \\
& (iv) \quad u_j \otimes (u_j \otimes (u_j \otimes (u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots))) = \\
& \quad \quad = u_j \otimes (u_j \otimes (u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots)) + \\
& \quad \quad + D_{(u_j, u_j)}^{(u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots)}(u_j \otimes \ldots \otimes (u_j \otimes u_j) \ldots) (\text{mod} J), m \geq 5,
\end{align*}
\]

where \( D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)}(x \otimes \ldots \otimes (x \otimes y) \ldots) \) and \( D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)}(x \otimes \ldots \otimes (x \otimes y) \ldots) \) are defined, respectively, by

\[
D_{(x,y)}^{(x \otimes \ldots \otimes (x \otimes y) \ldots)}([a_1, a_2, \ldots, a_n]_s) =
\]

\[
= \sum_{i=1}^{n} [a_1, a_2, \ldots, a_{i-1}, [x, y, a_i]_{s}, a_{i+1}, \ldots, a_n]_s.
\]
= \sum_{i=1}^{n} [a_1, a_2, ..., a_{i-1}, [x, y, a_i], a_{i+1}, ..., a_n].

**Proof.** The proof must certainly take account of the generators for \( J \).
Obviously, we have \( u_i \otimes u_j = u_j \otimes u_i + (u_i \otimes u_j - u_j \otimes u_i) = u_j \otimes u_i (\text{mod} J) \).
From the definition of \([u_i, u_j, u_k]^{\circ}\), we get
\[
(u_i \otimes (u_j \otimes u_k)) = u_j \otimes (u_i \otimes u_k) + [u_i, u_j, u_k]^{\circ} = u_j \otimes (u_i \otimes u_k) + ([u_i, u_j, u_k]^{\circ} - [u_i, u_j, u_k]) + [u_i, u_j, u_k] = u_j \otimes (u_i \otimes u_k) + [u_i, u_j, u_k] (\text{mod} J).
\]
In order to prove formula (iii) we make use of the identity
\[
x_1 \cdot (x_2 \cdot (x_3 \cdot x_4)) = (9)
\]
which is satisfied for every binary algebra \( A(\cdot) \) with \([x, y, z] \) and \( f(x, y, z, w) \) defined by (2.4) and (2.3), respectively. Replacing everywhere \( \cdot \) by \( \otimes \) we get
\[
u_i \otimes (u_j \otimes (u_k \otimes u_\ell)) = u_j \otimes (u_i \otimes (u_k \otimes u_\ell)) + u_k \otimes [u_i, u_j, u_\ell]^{\circ} + [u_i, u_j, u_k]^{\circ} \otimes u_\ell + f^{\circ}(u_i, u_j, u_k, u_\ell) = u_j \otimes (u_i \otimes (u_k \otimes u_\ell)) + u_k \otimes (u_i, u_j, u_\ell) + [u_i, u_j, u_k] \otimes u_\ell + f^{\circ}(u_i, u_j, u_k, u_\ell) = u_j \otimes (u_i \otimes (u_k \otimes u_\ell)) + u_k \otimes [u_i, u_j, u_\ell] + [u_i, u_j, u_k] \otimes u_\ell (\text{mod} J).
\]
We shall prove now, recurrently, formula (iv). Formula (iii) shows us that (iv) is valid in the case \( m = 4 \). Suppose that (iv) is valid for certain \( m \). We put it in the form
\[
u_{j_1} \otimes (u_{j_2} \otimes (u_{j_3} \otimes ... \otimes (u_{j_{m-1}} \otimes u_{j_m}))) = u_{j_2} \otimes (u_{j_1} \otimes (u_{j_3} \otimes ... \otimes (u_{j_{m-1}} \otimes u_{j_m}))) + \sum_{i=3}^{m-1} u_{j_3} \otimes (u_{j_4} \otimes ... \otimes (u_{j_{i-1}} \otimes (u_{j_1}, u_{j_2}, u_{j_3}) \otimes (u_{j_{i+1}} \otimes ... \otimes (u_{j_{m-1}} \otimes u_{j_m})))) + u_{j_3} \otimes (u_{j_4} \otimes ... \otimes (u_{j_{m-1}} \otimes [u_{j_1}, u_{j_2}, u_{j_m}])) (\text{mod} J), \quad m \geq 5.
\]
Replacing in this formula \( u_{j_m} \) by \( u_{j_m} \otimes u_{j_{m+1}} \) and taking into account that \([x, y, z]^{\circ} \) is congruent mod \( J \) with \([x, y, z] \), the previous formula can be written in the form
\[
u_{j_1} \otimes (u_{j_2} \otimes (u_{j_3} \otimes ... \otimes (u_{j_{m-1}} \otimes (u_{j_m} \otimes u_{j_{m+1}})))) = \]
The following formula holds:

\[ u_{j_2} \otimes (u_{j_1} \otimes (u_{j_3} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \otimes u_{j_{m+1}}))) \ldots )) + \]

\[ + \sum_{i=3}^{m-1} u_{j_3} \otimes \ldots \otimes (u_{j_{i-1}} \otimes ((u_{j_1} \otimes u_{j_2}, u_{j_1}) \otimes (u_{j_{i+1}} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \otimes u_{j_{m+1}}))) \ldots )) + \]

\[ + u_{j_3} \otimes (u_{j_4} \otimes \ldots \otimes (u_{j_{m-1}} \otimes [u_{j_1}, u_{j_2}, u_{j_m} \otimes u_{j_{m+1}}]) \otimes \ldots )) (mod J). \]

To conclude, it is enough to remark the readily provable identity

\[ [u_{j_1}, u_{j_2}, u_{j_m} \otimes u_{j_{m+1}}] \otimes = [u_{j_1}, u_{j_2}, u_{j_m}] \otimes u_{j_{m+1}} + \]

\[ + u_{j_m} \otimes [u_{j_1}, u_{j_2}, u_{j_{m+1}}] \otimes + f(u_{j_1}, u_{j_2}, u_{j_m}, u_{j_{m+1}}). \]

This completes the proof of Lemma.

Formulae (ii)-(iv) give the opportunity to make the following remarks:

1) since \([u_1, u_j, u_k] \in T\) it follows that it is a standard polynomial of degree 1,

2) since \([u_i, u_j, u_k] \in T\) it follows that \(u_k \otimes [u_i, u_j, u_\ell] + [u_i, u_j, u_k] \otimes u_\ell\) is congruent \(mod J\) with a standard polynomial of degree 2,

3) the element

\[ D_{(u_{j_1}, u_{j_2})} (u_{j_3} \otimes \ldots \otimes (u_{j_{m-1}} \otimes u_{j_m})) \]

is congruent \(mod J\) with a standard polynomial of degree \(m - 2\).

Then, we can restate Lemma 4.1 in the form.

**Lemma 4.2 (Switching Lemma)** Let \(T\) be Lts and \(T\{T\}\) be its nonassociative tensor algebra. Then, the following identities hold:

1') \(u_i \otimes u_j = u_j \otimes u_i (mod J)\),

2') \(u_i \otimes (u_j \otimes u_k) = u_j \otimes (u_i \otimes u_k) + P_1^\otimes (x_1, \ldots , x_n, \ldots ) (mod J)\),

3') \(u_i \otimes (u_j \otimes (u_k \otimes u_\ell)) = u_j \otimes (u_i \otimes (u_k \otimes u_\ell)) + P_2^\otimes (x_1, \ldots , x_n, \ldots ) (mod J)\),

4') \(u_{j_1} \otimes (u_{j_2} \otimes (u_{j_3} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \ldots ))) = \]

\[ = u_{j_2} \otimes (u_{j_1} \otimes (u_{j_3} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \ldots ))) + P_{m-2}^\otimes (x_1, \ldots , x_n, \ldots ) (mod J), \quad m \geq 5, \]

where \(P_k^\otimes (x_1, \ldots , x_n, \ldots ) (k = 1, 2, \ldots )\) are left standard polynomials of degree \(k\) (i.e. they are finite linear combinations of left standard monomials of \(S[X]\)).

It must be remarked that \(P_k^\otimes (x_1, \ldots , x_n, \ldots )\) can depend on those elements from \(X\) which occur in the linear combinations expressing the elements defining the analyzed monomial as well as those introduced by the expressions of \([u_i, u_j, u_k]\) in the basis \(X\).

**Corollary 4.1** The following formula holds

\[ u_{j_1} \otimes (u_{j_2} \otimes \ldots \otimes (u_{j_{i-1}} \otimes (u_{j_i} \otimes (u_{j_{i+1}} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \ldots ))) = (11) \]

\[ = u_{j_1} \otimes (u_{j_2} \otimes \ldots \otimes (u_{j_{i-1}} \otimes (u_{j_{i+1}} \otimes (u_{j_i} \otimes (u_{j_{i+2}} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m} \ldots ))) + \]

...
Every element of \( P_{m-i-1}(x_1, \ldots, x_n, ...) \) (mod \( J \)) depends on those elements from \( X \) which occur in connection of \( u_{j_1}, u_{j_{i+1}}, \ldots, u_{j_m} \).

**Proof.** The assertion follows from Lemma 4.2 applied to \( u_{j_1} \otimes (u_{j_{i+1}} \otimes \ldots \otimes (u_{j_{m-1}} \otimes (u_{j_m}) \ldots) \) and taking into account that \( u_{j_1} \otimes (u_{j_2} \otimes \ldots \otimes (u_{j_{i+1}} \otimes (u_{j_{i+1}}) \ldots) \) is not involved in the computations.

In what follows, we need to consider the subspace \( \mathcal{T}_l(T) \) of \( \mathcal{T}(T) \), named the left tensor algebra of \( T \). It is defined as

\[
\mathcal{T}_l(T) = T^{\otimes 0} \otimes T^{\otimes 1} \otimes T^{\otimes 2} \otimes \ldots \otimes T^{\otimes n} \otimes \ldots
\]

where \( T^{\otimes 0} = K1, T^{\otimes 1} = T, \) and \( T^{\otimes n} = T \otimes T^{\otimes n-1} \) for \( n \geq 2 \).

**Lemma 4.3** Every element of \( \mathcal{T}_l(T) \) is congruent mod \( J \) to a \( K \)-linear combination of 1 and left standard monomials.

**Proof.** The assertion is clear for \( K1 \) and monomials of degree 1. Let \( X = \{x_1, x_2, \ldots, x_n, \ldots\} \) be a basis for \( T \). It is sufficient to apply the results of Lemma 4.2 to monomials in \( X \). For any left monomial of degree \( n \geq 2 \), we shall prove the following formula

\[
x_{j_1} \otimes (x_{j_2} \otimes (x_{j_3} \otimes \ldots \otimes (x_{j_{n-1}} \otimes (x_{j_n}) \ldots) = \sum_{1 \leq i_1 < i_2 < \ldots < i_n \leq n} P_{n-2}^{\otimes}(x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_n} \otimes (x_{i_n})) \tag{12}
\]

where \( (i_1, i_2, \ldots, i_n) \) is the set, obtained by putting in natural order the elements of the set \( (j_1, j_2, \ldots, j_n) \), and \( P_{n-2}^{\otimes}(x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_n} \otimes (x_{i_n})) \) is a left standard polynomial of degree \( n-2 \), i.e. a (finite) linear combination of elements from \( S_l[X] \). If the set \( (j_1, j_2, \ldots, j_n) \) has \( k \) inversions, then we get (4.12) after \( k \) times of using Lemma 4.2. Of course, \( P_{n-2}^{\otimes}(x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_n} \otimes (x_{i_n})) \) can depend (not necessarily) on \( x_{j_1}, x_{j_2}, \ldots, x_{j_n} \) as well as on those \( x_j \) occurring through the usual formulae defining the structure constants \( \{x_{j_1}, x_{j_2}, x_{j_3}, x_{j_4} = \sum_{1 \leq i_1 < i_2 < \ldots < i_n \leq n} t'_{j_1j_2j_3j_4} x_{i_n} \) of the Lts \( T \).

**Lemma 4.4** Every element of \( \mathcal{T}(T) \) is congruent mod \( J \) to a \( K \)-linear combination of 1 and left standard monomials.

**Proof.** Since every monomial of degree \( n \) in \( X \) is uniquely obtained as some tensorial products of elements from \( T_l(T) \), the proof is the result of using Lemma 4.4 for every element of \( T_l(T) \) which is factor of the analyzed monomial.

We prove now that the classes mod \( J \) of 1 and standard monomials are linearly independent and, consequently, they represent a basis in \( \mathcal{U}(T) \). To
this aim we consider the vector spaces \( B_n \) spanned by all standard monomials of degree \( n \) and the vector space \( B = K1 \oplus B_1 \oplus ... \oplus B_n \oplus ... \). Their linear independence follows from the following statement.

**Lemma 4.5** There exists a linear mapping \( \sigma \) between \( T \{ T \} \) and \( B \) such that

\[
\sigma(1) = 1, \tag{13}
\]

\[
\sigma([x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_n}]_s) = \tag{14}
\]

\[
= [x_{j_1} \cdot x_{j_2} \cdot ... \cdot x_{j_n}]_s \in S[X], \quad \forall [x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_n}]_s \in S[X]
\]

\[
\sigma([x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{i-1}} \otimes [x_{k_1}, x_{k_2}, x_{k_3}] \otimes x_{j_{i+1}} \otimes ... \otimes x_{j_{n-2}}]_s) = \tag{15}
\]

\[
= \sigma([x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{i-1}} \otimes [x_{k_1}, x_{k_2}, x_{k_3}] \otimes x_{j_{i+1}} \otimes ... \otimes x_{j_{n-2}}]_s),
\]

\[
i \in \{1, 2, ..., n\}, \quad n \geq 3
\]

\[
\sigma([x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{i-1}} \otimes f_{s}(x_{k_1}, x_{k_2}, x_{k_3}, x_{k_4}) \otimes x_{j_{i+1}} \otimes ... \otimes x_{j_{n-3}}]_s) = \tag{16}
\]

\[
= 0, \quad i \in \{1, 2, ..., n\}, \quad n \geq 4
\]

**Proof.** This definition is connected with the number of the inversions that are necessary to put the set of label in the natural order. We define it such that to take as the value on a monomial the corresponding standard polynomial. Actually, the existence of \( \sigma \) is proved if the standard polynomial corresponding to a monomial is unique. Certainly, it is the case of monomials of degree 2. That is why we set

\[
\sigma(1) = 1, \tag{17}
\]

\[
\sigma(x_j) = x_j, \forall x_j \in X. \tag{18}
\]

Taking into account Lemma 4.1 (1) we define

\[
\sigma(x_i \otimes x_j) = x_i \cdot x_j, \quad if \quad i \leq j, \tag{19}
\]

\[
\sigma(x_i \otimes x_j) = x_j \cdot x_i, \quad if \quad i > j.
\]

A monomial of degree 3 can have 0, 1, 2 or 3 inversions. According to these cases we define

\[
\sigma(x_i \otimes (x_j \otimes x_k)) = x_i \cdot (x_j \cdot x_k), \quad if \quad i < j < k \tag{20}
\]

\[
\sigma(x_i \otimes (x_j \otimes x_k)) = x_i \cdot (x_k \cdot x_j), \quad if \quad i < k < j
\]

\[
\sigma(x_i \otimes (x_j \otimes x_k)) = x_j \cdot (x_i \cdot x_k) + [x_i, x_j, x_k], \quad if \quad j < i < k
\]

\[
\sigma(x_i \otimes (x_j \otimes x_k)) = x_j \cdot (x_k \cdot x_i) + [x_i, x_j, x_k], \quad if \quad j < k < i
\]

\[
\sigma(x_i \otimes (x_j \otimes x_k)) = x_k \cdot (x_i \cdot x_j) + [x_i, x_k, x_j], \quad if \quad k < i < j
\]
The monomial has no inversion iff it is a standard monomial. In this case we necessarily define $\sigma$ by (4.201). If the labels of a monomial have only one inversion, then, by Lemma 4.1, (1), (2), the monomial is congruent $\mod J$ with a unique standard monomial; this supports the definitions (4.202) and (4.203). The formulae (4.204) and (4.205) are also correct because, even if the monomials present exactly two inversions, the order of making acting these inversions is necessarily unique. Finally, (4.206) occurs by applying Lemma 4.1 (i) and (ii) to realize the inversions $(j, k), (i, k)$ and $(j, i)$. If we should make the three inversions in the order $(i, j), (i, k)$ and $(j, k)$, then we obtain

$$\sigma(x_i \otimes (x_j \otimes x_k)) = x_k \cdot (x_j \cdot x_i) + [x_i, x_k, x_j], \text{ if } k < j < i$$

By subtracting the two expressions for $\sigma(x_j \otimes x_k)$ we get

$$[x_i, x_k, x_j] - [x_j, x_k, x_i] - [x_i, x_j, x_k] = 0,$$

i.e., (4.206) is well-defined. Moreover, as it can easily be checked, the formulae (20) imply

$$\sigma([x_i, x_j, x_k] \otimes - [x_i, x_j, x_k]) = 0.$$

As $f \otimes (x_1, x_2, x_3, x_4)$ is a generator for $J$, we put

$$\sigma(f \otimes (x_1, x_2, x_3, x_4)) = 0, \forall x_1, x_2, x_3, x_4 \in X. \quad (21)$$

Then, we shall prove the assertion of Lemma by a complete induction on the degree $n > 3$ of the monomials and on the number of inversions of labels occurring in these monomials; the number of inversions of labels occurring in a monomial is called its index. We denote by $T^{\otimes n, j}$ the subspace of $T^{\otimes n}$ spanned by the monomials of degree $n$ and index $\leq j$. In order to ascertain the affiliation to a subspace $T^{\otimes n, j}$, we associate with the monomial $x_{j_1} \otimes (x_{j_2} \otimes (x_{j_3} \otimes \ldots \otimes (x_{j_{n-1}} \otimes (x_{j_n})))),$ with $n > 3$, a permutation

$$\tau : \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

where $i_1$ is the ordinal number giving the position of $j_1$ in the totally ordered set $\{j_1, j_2, \ldots, j_n\}$ etc. It is well-known that every permutation $\tau$ is the composition of a finite number of transpositions. In their turn, each transposition is the product of inversions.

Unfortunately, a representation of any permutation as a product of all its inversions is unique up to their ordering. Certainly, there exists a natural bijection between the inversions in the labels of monomial $x_{j_1} \otimes (x_{j_2} \otimes (x_{j_3} \otimes \ldots \otimes (x_{j_{n-1}} \otimes (x_{j_n}))))$.
Suppose that a linear mapping \( \sigma \) satisfying to conditions stated in Lemma 4.2 has already been defined for \( T^0 \oplus T_1 \oplus \ldots \oplus T^{n-1} \) for the monomials in \( T_j \{T \} \), only. In order to extend linearly to \( T^0 \oplus T^1_\ell \oplus \ldots \oplus T^{n-1}_\ell \oplus T^{n,0}_\ell \) we put

\[
\sigma(x_{j_1} \otimes (x_{j_2} \otimes (x_{j_3} \otimes \ldots \otimes (x_{j_{n-1}} \otimes (x_{j_n})) \ldots) = x_{j_1} \cdot (x_{j_2} \cdot (x_{j_3} \cdot \ldots (x_{j_{n-1}} \cdot (x_{j_n})) \ldots),
\]

if \( j_1 \leq j_2 \leq \ldots \leq j_n \), i.e. we have solved the problem in the case when the left monomials of degree \( n \) have no inversion. Consider now the left monomials having a single inversion. The most representative such monomial is \( x_2 \otimes (x_1 \otimes (x_3 \otimes \ldots \otimes (x_{n-1} \otimes x_n)) \ldots). \) Then, by making use of the identity (2.5) and the induction hypothesis, we give the following definition

\[
\sigma(x_2 \otimes (x_1 \otimes (x_3 \otimes \ldots \otimes (x_{n-1} \otimes x_n))) = \\
= x_1 \cdot (x_2 \cdot (x_3 \cdot \ldots (x_{n-1} \cdot x_n))) + \\
+ \sigma(D(x_2,x_1)(x_3 \otimes \ldots (x_{n-1} \otimes x_n))).
\]

Of course, this definition is correct. Moreover, the identity (2.6) suggests us to use the following definition for any monomial of degree \( n \) and index 1

\[
\sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{h-1}} \otimes (x_{j_{i+1}} \otimes (x_{j_{i+2}} \otimes \ldots \otimes (x_{n-1} \otimes x_n)))) \ldots)) = \\
= \sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{h-1}} \otimes (x_{j_{i+1}} \otimes (x_{j_{i+2}} \otimes \ldots \otimes (x_{n-1} \otimes x_n)))) \ldots)) + \\
+ \sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (D^{j_{i+1}}_i,x_{j_{i+1}})(x_{j_{i+2}} \otimes \ldots \otimes (x_{n-1} \otimes x_n))))).
\]

Assume that \( \sigma \) has already been defined for \( T^0 \oplus T^1 \oplus \ldots \oplus T^{n-1} \oplus T^{n,i-1} \) satisfying the assertion of Lemma for monomials belonging to this space and let \( x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{n-1} \otimes x_n)) \) be a left monomial of the index \( i \geq 1 \). Suppose \( j_k > j_{k+1} \). Then we set

\[
\sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{k-1}} \otimes (x_{j_k} \otimes (x_{j_{k+1}} \otimes \ldots \otimes (x_{n-1} \otimes x_n)))) = (22) \\
= \sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{k-1}} \otimes (x_{j_k} \otimes (x_{j_{k+1}} \otimes \ldots \otimes (x_{n-1} \otimes x_n)))))) + \\
+ \sigma(x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{k-1}} \otimes (D^{j_{k+1}}_{j_{k+1}})(x_{j_{k+2}} \otimes \ldots \otimes (x_{n-1} \otimes x_n))))).
\]

This definition makes a sense since the terms on the right-hand side are in \( T^0 \oplus T^1 \oplus \ldots \oplus T^{n-1} \oplus T^{n,3-1} \). First we show that this definition is independent of the choice of the pair \( (j_k,j_{k+1}), j_k > j_{k+1} \). Let \( (j_\ell,j_{\ell+1}) \) be another pair with \( j_\ell > j_{\ell+1} \). Essentially there are two cases: I. \( \ell > k + 1 \), II. \( \ell = k + 1 \). Since the computations do not involve the part \( x_{j_1} \otimes (x_{j_2} \otimes \ldots \otimes (x_{j_{k-1}} of the
monomial and taking into account the existence of a bijection between the in-
versions in the labels of a monomial and those of the associated permutation,
it follows that it is sufficient to consider the following two types of monomials

\[ I. \quad x_2 \otimes (x_1(x_3 \cdots (x_{i-1} \otimes (x_{i+1} \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots)))\cdots))) \]

\[ II. \quad x_3 \otimes (x_2 \otimes (x_1 \otimes (x_4 \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))) \]

I. The associated permutation is \( \tau = (1, 2)(i, i + 1) \). Then we get the first
evaluation \( E_1 \) by starting with \((1, 2)\), namely

\[ E_1 = \]
\[ = \sigma(x_2 \otimes (x_1 \otimes (x_3 \otimes \cdots \otimes (x_{i-1} \otimes (x_{i+1} \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) = \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sigma(x_3 \otimes \cdots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_{i+1}) \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sigma((x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots)) + \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sigma(x_3 \otimes \cdots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_{i+1}) \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

If we start with \((i + 1, i)\), we get the second evaluation \( E_2 \), namely

\[ E_2 = \]
\[ = \sigma(x_2 \otimes (x_1 \otimes (x_3 \otimes \cdots \otimes (x_{i-1} \otimes (x_{i+1} \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) = \]

\[ + \sigma(x_2 \otimes (x_1 \otimes (x_3 \otimes \cdots \otimes (x_{i-1} \otimes (x_{i+1} \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sigma(x_3 \otimes \cdots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_{i+1}) \otimes (x_i \otimes (x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots)))) + \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]

\[ + \sum_{k=3}^{i-1} \sigma(x_3 \otimes (x_4 \otimes \cdots \otimes (x_{k-1} \otimes (D_{(x_2,x_1)}(x_k) \otimes (x_{k+1} \otimes \cdots \otimes (x_{i-2} \otimes x_{i-1})\otimes (x_{i+1} \otimes (x_i \otimes x_{i+2} \otimes \cdots \otimes (x_{n-1} \otimes x_n))\cdots))))) + \]
Firstly, we use the order of transformations following the pairs 
\((3,2)\) and get the first form
\[ E_1 = \sigma(x_3 \otimes (x_4 \otimes \ldots \otimes x_{i-1} \otimes (D_{(x_2,x_1)}^\circ (x_{i+2} \otimes \ldots \otimes x_{n-1} \otimes x_n)) \ldots)). \]

By substracting \( E_1 - E_2 \) it follows
\[ E_1 - E_2 = \]
\[ = \sigma((x_3 \otimes \ldots \otimes x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_i) \otimes (x_{i+2} \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) - 1) \]
\[ = \sigma((x_3 \otimes \ldots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_i) \otimes (x_{i+2} \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) + 1) \]
\[ = \sigma((x_3 \otimes \ldots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_i) \otimes (x_{i+2} \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) - 1) \]
\[ = \sigma((x_3 \otimes \ldots \otimes (x_{i-2} \otimes (x_{i-1} \otimes (D_{(x_2,x_1)}(x_i) \otimes (x_{i+2} \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) + 1) \]

By using (Lts.5) and (Ls.6) it follows that the two evaluations coincide.

II. Let us consider the case \( \sigma(x_3 \otimes (x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n)) \ldots)). \)

Firstly, we use the order of transformations following the pairs 
\((3,2)\), \((3,1)\) and 
\((2,1)\) and get the first form \( E_1 \) of evaluation, namely
\[ E_1 = \sigma(x_3 \otimes (x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) = \]
\[ = \sigma(x_3 \otimes (x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(D_{(x_3,x_2)}(x_1)(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(x_1 \otimes (D_{(x_3,x_2)}(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) = \]
\[ = \sigma(x_2 \otimes (x_1 \otimes (x_3 \otimes (x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(D_{(x_3,x_2)}(x_1)(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(x_1 \otimes (D_{(x_3,x_2)}(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) = \]
\[ = \sigma(x_1 \otimes (x_2 \otimes (x_3 \otimes (x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) + \]
\[ + \sigma(D_{(x_3,x_2)}(x_3)(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(x_3 \otimes (D_{(x_3,x_2)}(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(x_2 \otimes (D_{(x_3,x_2)}(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(D_{(x_3,x_2)}(x_1)(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)) + \]
\[ + \sigma(x_1 \otimes (D_{(x_3,x_2)}(x_4 \otimes \ldots \otimes (x_{n-1} \otimes x_n) \ldots)))) = \]

If we use the order of transformations following the pairs 
\((2,1)\), \((3,1)\) and 
\((3,2)\) we get the second form \( E_2 \) of evaluation, namely
The cosets of $1$ and the standard monomials

**Theorem 4.1**

Birkhoff-Witt Theorem. Using Lemmas 4.1 and 4.2 we can prove the following analogue of Poincaré-the conditions stated by Lemma. This completes the proof of Lemma. By combination of elements of the form $T_1$ and the cosets of the standard monomials. Lemma 4.2 gives a linear mapping $\mathbb{B}$ into $\mathbb{A}$, $i \in \{1, n\}$, $n \geq 1$, $s \in \{0, 1\}$, $E_2 = $

$$
E_2 = \\
= \sigma(x_3 \otimes (x_2 \otimes (x_1 \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...)))) + \\
+ \sigma(x_3 \otimes (D(x_2,x_1)(x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
= \sigma(x_1 \otimes (x_3 \otimes (x_2 \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...)))) + \\
+ \sigma(D(x_3,x_1)(x_2 \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
= \sigma(x_1 \otimes (x_2 \otimes (x_3 \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...)))) + \\
+ \sigma(x_1 \otimes (D(x_2,x_3)(x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
= \sigma(x_1 \otimes (D(x_2,x_3)(x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
+ \sigma(D(x_3,x_2)(x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
+ \sigma(x_2 \otimes (D(x_3,x_1)((x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
+ \sigma(x_3 \otimes (D(x_2,x_1)((x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...)))) = 0.
$$

Subtracting $E_1 - E_2$ one gets

$$
E_1 - E_2 = \\
= \sigma([x_2, x_1, x_3] \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
+ \sigma([x_3, x_2, x_1] \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) + \\
+ \sigma([x_1, x_3, x_2] \otimes (x_4 \otimes ... \otimes (x_{n-1} \otimes x_n)...))) = 0.
$$

By using (Lts.3) and (Ls.4) it follows that the two evaluations are coincident. Consequently, the definition (19) is independent of the choice of the pair $(i + 1, i)$. We now apply this definition for defining $\sigma$ for the monomials of degree $n$ and index $i$. Since every monomial of degree $n$ is obtained uniquely as a tensorial product (with a specific ordering for parenthesis) $[w_1 \otimes w_2 \otimes ... \otimes w_p]_s$ and $s \leq n - 1$, it follows that $\sigma$ is naturally extended to the space $T^{\otimes n,i}$. Consequently, we get a mapping on $T^{\otimes 0} \oplus T^{\otimes 1} \oplus ... \oplus T^{\otimes n-1} \oplus T^{\otimes n,i}$ satisfying the conditions stated by Lemma. This completes the proof of Lemma. By using Lemmas 4.1 and 4.2 we can prove the following analogue of Poincaré-Birkhoff-Witt Theorem.

**Theorem 4.1** The cosets of $1$ and the standard monomials mod$J$ form a basis for $U(T)/J$.

**Proof.** Lemma 4.1 shows that every coset is a linear combination of $1 + J$ and the cosets of the standard monomials. Lemma 4.2 gives a linear mapping of $\mathcal{J}T$ into $B$ satisfying the stated conditions. Every element of $J$ is a linear combination of elements of the form

$$
x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{n-1}} \otimes x_{j_n} \otimes x_{j_{n+1}} \otimes ... \otimes x_{j_{n+2}} - \\
- x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{n-1}} \otimes x_{j_{n+1}} \otimes x_{j_{n+2}} \otimes ... \otimes x_{j_n}, \\
x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{n-1}} \otimes [x_{k_1}, x_{k_2}, x_{k_3}] \otimes x_{j_{n+1}} \otimes ... \otimes x_{j_{n+2}} - \\
- x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{n-1}} \otimes [x_{k_1}, x_{k_2}, x_{k_3}] \otimes x_{j_{n+1}} \otimes ... \otimes x_{j_{n+2}} - \\
i \in \{1, 2, ..., n\}, \ n \geq 3, \\
x_{j_1} \otimes x_{j_2} \otimes ... \otimes x_{j_{n-1}} \otimes f^\otimes(x_{k_1}, x_{k_2}, x_{k_3}, x_{k_4}) \otimes x_{j_{n+1}} \otimes ... \otimes x_{j_{n+3}} - \\
i \in \{1, 2, ..., n\}, \ n \geq 4.
$$
Since \( \sigma \) maps these elements into 0, \( \sigma(J) = 0 \) and so \( \sigma \) induces a linear mapping of \( \mathfrak{U}(T) \) into \( B \). This mapping carries the cosets of 1 and standard monomials \([x_{j_1} \otimes x_{j_2} \otimes \ldots \otimes x_{j_n}]_s \in S[X] \), into 1 and \([x_{j_1} \cdot x_{j_2} \cdot \ldots \cdot x_{j_n}]_s \in S[X] \), respectively. Since these images are linearly independent in \( B \), we have the linear independence in \( \mathfrak{U}(T) \) of the cosets of 1 and the standard monomials.

**Corollary 4.2** The mapping \( i \) of \( T \) in \( \mathfrak{U}(T) \) is 1-1 and \( K_1 \cap i(T) = \{0\} \).

**Proof.** If \( X = \{x_1, x_2, \ldots, x_n, \ldots\} \) is a basis for \( T \) over \( K \), then 1 = 1 + \( J \) and the cosets \( i(x_j) = x_j + J \) are linearly independent. This implies both statements.

**References**


Some new properties of Lie triple systems


SOME PROBLEMS OF THE THEORY OF COMPACTIFICATIONS OF TOPOLOGICAL SPACES

Laurenţiu I. Calmuţchi, Mitrofan M. Choban
Tiraspol State University, Republic of Moldova

Abstract
If $bX$ and $cX$ are two compactifications of a space $X$, then a mapping $\varphi : bX \to cX$ is canonical if $\varphi$ is continuous and $\varphi(x) = \varphi^{-1}(x) = x$ for any $x \in X$. The Wallman compactification $wX$ may be constructed for an arbitrary $T_0$-space $X$. A compactification $cX$ of a space $X$ is a $\rho$-compactification if there exits a canonical mapping $\varphi : \omega X \to cX$ onto $cX$. If $\varphi$ is closed, then we say that $cX$ is an $\omega\alpha$-compactification.

Keywords: compactification, g-compactification, $\omega\alpha$-proximity, perfect compactification, WS-compactification.


1. INTRODUCTION

Compactness is one of the most important notions and the problems connected with special embeddings of spaces in compact spaces were studied in many works [7, 9, 10, 13, 30, 40].

The general problems of the theory of compactifications are the following.

Problem 1. To find the methods of construction and study of the compactifications with concrete properties of a given space.

Problem 2. To construct, for every space $X$ from a given class $K$ of spaces and a class $M$ of continuous mappings, a class $L(X)$ of compactifications of $X$ such that:
- in the class $L(X)$ there exists a maximal element $m_L X$ for any $X \in K$;
- if $X, Y \in K$ and $f : X \to Y$ is a mapping from $M$, then there exists a continuous extension $g : m_L X \to m_L Y$ of the mapping $f$.

One of the solutions of the Problem 2 for a class $K$ of completely regular spaces is the class $L(X)$ of Hausdorff compactifications. The class of Hausdorff compactifications was profoundly studied (see [3, 8, 9, 11, 12, 13, 15, 16, 23, 30, 34, 35, 38, 40]).

Since the class of all $T_1$-compactifications of a given infinite $T_1$-space is not a set, there exist many obstacles for solution of these problems in the classes of $T_1$-spaces and $T_0$-spaces.
Significant moments in the general theory of $T_1$-compactifications of $T_1$-spaces were:

- the construction by L. Wallman of the Wallman compactification $wX$ of a $T_1$-space $X$ (see [1, 7, 13, 41]);
- the Shanin’s generalization of the Wallman method of constructions of $T_1$-compactifications (see [7, 10, 17, 20, 21, 22, 24, 27, 28, 29, 32, 33, 36, 37]);
- the construction of the theory of $\omega\alpha$-compactifications and of $\omega\alpha$-proximities (see [25, 26]);
- the existence of a Hausdorff compactification $bX$ for some discrete space $X$ which is not of the Wallman-Shanin type (see [33, 39]).

Various methods of construction of extensions of spaces were proposed in [3, 5, 7, 10, 16, 18, 20].

The aim of the present work is to investigate the class of $T_0$-compactifications of $T_0$-spaces.

In Section 1 we discuss the general notions of compactifications and $g$-compactifications.

Sections 2-6 are devoted to the investigation of the Wallman-Shanin method of construction of $T_0$-compactifications of $T_0$-spaces.

In Section 7 we study the Hausdorff $g$-compactifications of spaces.

In Section 8 the notion of $\rho$-compactification of a $T_0$-space is introduced.

In Section 9–10 the $\omega\alpha$-compactifications of $T_0$-spaces are studied.

In Section 11 we consider the locally compact-small spaces (see [19]).

In Section 12 we define the notion of the $\omega\alpha$-proximity for $T_0$-spaces.

In Sections 13 and 14 the notions of perfect and punctiform compactifications are studied for $T_0$-spaces.

We use the terminology from [13].

By $cl_A X$ or $cl A$ it is denoted the closure of a set $A$ in a space $X$. By $|A|$ we denote the cardinality of a set $A$. The weight of a space $X$ is denoted by $w(X)$. If $X$ is a space, then $c(X) = \{x \in X : \text{the set } \{x\} \text{ is closed in } X\}$ and $z(X) = X \setminus c(X)$.

By $exp(X)$ we denote the cardinality of a set of all subsets of a set $X$.

Every space is considered to be a non-empty $T_0$-space. Let $N = \{1, 2, \ldots\}$. If $\tau$ is a cardinal number, then we put $exp(\tau) = 2^\tau$.

1. COMPACTIFICATIONS OF SPACES

1.1 Definition. A $g$-compactification of a space $X$ is called a pair $(Y, f)$, where $Y$ is a compact space, $f : X \to Y$ is a continuous mapping and the set $f(X)$ is dense in $Y$. If $f$ is an embedding of $X$ into $Y$, then $(Y, f)$ is a compactification of $X$.

If $(Y, f)$ is a compactification of a space $X$, then we identify $x \in X$ with $f(x) \in Y$ and we consider that $X$ is a dense subspace of $Y$, i.e $X = f(X) \subseteq Y$. 
Denote by $GC(X)$ the class of all $g$-compactifications and by $C(X)$ the class of all compactifications of a space $X$.

Let $(Y, f)$ and $(Z, g)$ be $g$-compactifications of a space $X$. We consider that $(Z, g) \leq (Y, f)$ if there exists a continuous mapping $\varphi : Y \to Z$ such that $g = \varphi \circ f$. If $\varphi$ is a homeomorphism of $Y$ onto $Z$, then we consider that $(Y, f) = (Z, g)$. We identify the equal $g$-compactifications. If $(Z, g) \leq (Y, f)$ and $(Y, f) \leq (Z, g)$, then the $g$-compactifications $(Y, f)$ and $(Z, g)$ are called equivalent and we denote $(Y, f) \sim (Z, g)$.

For every space $X$ and every $g$-compactification $(Y, f)$ of $X$ the class $\{(Z, g) \in GC(X) : (Z, g) \sim (Y, f)\}$ is not a set. In particular, $GC(X)$ and $C(Y)$ are not sets (see [7], Proposition 1.1.7).

1.2 Definition. (see [21] for $T_1$-compactifications). A $g$-compactification $(Y, f)$ of a space $X$ is called:

- a weakly correct $g$-compactification if $\{\text{cl}_Y f(A) : A \subseteq X\}$ is a closed base of $Y$;

- a correct $g$-compactification if it is weakly correct and for every $y \in Y \setminus f(X)$ the set $\{y\}$ is closed in $Y$;

- a strongly correct $g$-compactification if it is correct and if $\Phi$ is a closed subset of $Y$, $F$ is a closed subset of a subspace $f(X)$ and $y \in \Phi \cap (\text{cl}_Y F \setminus F)$, then there exists a filter $\eta$ of closed subsets of $f(X)$ such that $F \in \eta$ and $y \in \{\text{cl}_Y H : H \in \eta\} \subseteq \Phi$.

Denote by $wKGC(X)$ the class of all weakly correct $g$-compactifications, by $KGC(X)$ the class of all correct $g$-compactifications and by $sKGC(X)$ the class of all strongly $g$-compactifications of a space $X$.

1.3. Proposition. The class $wKGC(X)$ of all weakly correct $g$-compactifications of a space $X$ is a set.

Proof. If $(Y, f) \in wKGC(X)$, then $w(Y) \leq \exp(X)$ and $|Y| \leq \exp(\exp(X))$. The proof is complete.

1.4. Proposition. If $(Y, f)$ and $(Z, g)$ are equivalent correct $g$-compactifications of a space $X$, then $(Y, f) = (Z, g)$.

Proof. Let $\varphi : Y \to Z$ and $\psi : Z \to Y$ be continuous mappings for which $g = \varphi \circ f$ and $f = \psi \circ g$. We put $X_1 = f(X)$ and $h(y) = \psi(\varphi(y))$ for any $y \in Y$. It is obvious that $y = h(y)$ for any $y \in X_1$. Suppose that $y \in Y$ and $y_1 = h(y) \neq y$. It is clear that $y \in Y \setminus X_1$ and $\{y\}$ is a closed subset of $Y$. There exists a closed subset $F$ of $X_1$ such that $\{y, y_1\} \cap \text{cl}_Y F = \{y\}$. Since $h$ is a continuous mapping, $h(\text{cl}_Y F) \subseteq \text{cl}_Y h(F)$. By construction, $y_1 \in h(\text{cl}_Y F)$ and $F = h(F)$, a contradiction. Thus $h(y) = y$ for any $y \in Y$ and $\psi = \varphi^{-1}$. The proof is complete.

1.5. Operation of correction. Let $(Y, f)$ be a $g$-compactification of a space $X$. On $Y$ we consider the topology $T'$ generated by the closed base $\{\cap \{\text{cl}_Y f(A) : i \leq n\} : A_1, A_2, \ldots, A_n \subseteq X, n \in N\}$. If $T$ is the initial topology on $Y$, then $T' \subseteq T$. For any $y \in Y$ we put $i(y) = \cap \{U \setminus V : U, V \in T', y \in$
There exists a mapping $\varphi : Y \to Z$, where $\varphi^{-1}(\varphi(y)) = i(y)$ for any $y \in Y$. On $Z$ we consider the topology $\{H \subseteq Z : \varphi^{-1}(H) \in T'\}$. Then $Z$ is a $T_0$-space and $(Z, g = \varphi \circ f)$ is a $g$-compactification of $X$. By construction, $\varphi \mid f(X) : f(X) \to \varphi(f(X))$ is a homeomorphism and $\{c_{LY}(A) : A \subseteq X\}$ is a closed base of the space $Z$. Thus $(Z, g)$ is a weakly correct compactification of $X$. If $(Y, f)$ is a compactification of $X$, then $(Z, g)$ is a compactification of $X$ too. The $g$-compactification $(Z, g)$ is called the correction of the $g$-compactification $(Y, f)$. If $(S, h)$ is a correct $g$-compactification of $X$ and $(S, h) \subseteq (Y, f)$, then $(S, h) \subseteq (Z, g)$. Thus $(Z, g)$ is the maximal weakly correct $g$-compactification from the set $\{(S, h) \in KGC(X) : (S, h) \subseteq (Y, f)\}$.

1.6 Definition. We say that a $g$-compactification $(Y, f)$ of a space $X$:

- is weakly incompressible if a compact subset $\Phi$ of $Y$ such that $Y \neq \Phi$ and $f(X) \subseteq \Phi$ does not exist;
- is incompressible if for every subset $A$ of $f(X)$ a compact subset $\Phi$ of $Y$ such that $\Phi \neq \text{cl}_Y f(A)$ and $f(X) \cap \text{cl}_Y f(A) \subseteq \Phi \subseteq \text{cl}_Y f(A)$ does not exist.

1.7 Example. Let $D$ be an infinite discrete space, $A$ and $B$ be two infinite subsets of $D$ for which $A \cap B = \emptyset$ and $D = A \cup B$. Let $a, b$ be two distinct points, $D \cap \{a, b\} = \emptyset$ and $cD = D \cup \{a, b\}$. On $cD$ we consider the topology generated by the open base $\{\{x\} : x \in D\} \cup \{(D \setminus F) \cup \{a\} : F$ is a finite subset of $D \cup \{(B \setminus F) \cup \{b\} : F$ is a finite subset of $D\}$. The space $cD$ is a compactification of $D$ with the properties:

- $cD$ is not weakly correct;
- $cD$ is not incompressible;
- $cD$ is not weakly incompressible.

1.8 Example. Let $A, B, C$ be infinite subsets of a discrete space $D$ and $A \cap B = A \cap C = B \cap C = \emptyset$, $D = A \cup B \cup C$. On $bD = D \cup \{a, b\}$, where $a \neq b$ and $D \cap \{a, b\} = \emptyset$, consider the topology generated by the open base $\{\{x\} : x \in D\} \cup \{\{a\} \cup ((A \cup B) \setminus F) : F$ is a finite subset of $D\} \cup \{\{b\} \cup ((B \cup C) \setminus F) : F$ is a finite subset of $D\}$. The space $bD$ is a compactification of $D$ with the proprieties:

- if the set $L \subseteq B$ is infinite, then the sets $L \cup \{a\}$, $L \cup \{b\}$ are compact and $\text{cl}_{bD} L = L \cup \{a, b\}$;
- $bD$ is not weakly correct;
- $bD$ is not incompressible;
- $bD$ is weakly incompressible.

1.9 Example. Let $Y = [0, 1]$ with the topology $T = \{X, \emptyset\} \cup \{[0, x) : 0 < x < 1\}$. Fix $a \in Y$ and set $X = Y \setminus \{a\}$. Then $Y$ is a weakly correct compactification of $X$. The compactification $Y$ is incompressible if and only if $a = 1$. For $a < 1$ the space $X$ is compact and $Y$ is not a correct compactification. The compactification $Y$ is strongly correct if and only if $a = 1$. 

$U \setminus V$. }
2. **WS-COMPACTIFICATIONS**

A family \( \mathcal{L} \) of subsets of a space \( X \) is called a ring of subsets of \( X \) if \( X, \emptyset \in \mathcal{L} \) and \( A \cap B, A \cup B \in \mathcal{L} \) for all \( A, B \in \mathcal{L} \).

If \( \mathcal{L} \) is a ring of subsets of a space \( X \) and any set \( A \in \mathcal{L} \) is closed in \( X \), then we say that \( \mathcal{L} \) is a closed ring of \( X \).

If a ring is a closed base of a space, then \( \mathcal{L} \) is called a basic ring of \( X \).

Fix a ring \( \mathcal{L} \) of closed subsets of \( X \). For any \( x \in X \) we put \( \xi(x, \mathcal{L}) = \{ F \in \mathcal{L} : x \in F \} \). A family \( \xi \) of subsets of \( X \) is called an \( \mathcal{L} \)-filter if: \( \xi \subseteq \mathcal{L}; \emptyset \notin \xi; A \cap B \in \xi \) for any \( A, B \in \xi \); if \( A \subseteq B, A \in \xi \) and \( B \in \mathcal{L} \), then \( B \in \xi \). A maximal \( \mathcal{L} \)-filter is called an \( \mathcal{L} \)-ultrafilter. A filter \( \xi \) is a free \( \mathcal{L} \)-filter if \( \cap \xi = \emptyset \).

We put \( \omega_\mathcal{L} X = \{ \xi(x, \mathcal{L}) : x \in X \} \cup \{ \xi \subseteq \mathcal{L} : \xi \) is a free \( \mathcal{L} \)-ultrafilter \}.  

Consider the mapping \( \omega_\mathcal{L} : X \to \omega_\mathcal{L} X \), where \( \omega_\mathcal{L}(x) = \xi(x, \mathcal{L}) \) for any \( x \in X \). For every \( A \in \mathcal{L} \) we put \( c_\mathcal{L}(A) = \{ \xi \in \omega_\mathcal{L} X : A \in \xi \} \).

**2.1. Lemma.** For every ring \( \mathcal{L} \) of closed subsets of a space \( X \) we have:
1. \( c_\mathcal{L}(\emptyset) = \emptyset \) and \( c_\mathcal{L}(X) = X \);
2. \( c_\mathcal{L}(A \cup B) = c_\mathcal{L}(A) \cup c_\mathcal{L}(B) \) for all \( A, B \in \mathcal{L} \);
3. \( c_\mathcal{L}(A \cap B) = c_\mathcal{L}(A) \cap c_\mathcal{L}(B) \) for all \( A, B \in \mathcal{L} \);
4. \( c(\mathcal{L}) = \{ c_\mathcal{L}(A) : A \in \mathcal{L} \} \) is a closed base for some \( T_0 \)-topology on \( \omega_\mathcal{L} X \);
5. \( c(\mathcal{L}) \) is a ring of subsets;
6. \( \omega_\mathcal{L}^{-1}(c_\mathcal{L}(A)) = A \) for any \( A \in \mathcal{L} \).

**Proof.** The assertions 1 and 2 are obvious. If \( \xi \in \omega_\mathcal{L} X \) and \( A, B \in \mathcal{L} \), then \( A \cup B \in \xi \) if and only if \( \xi \cap \{ A, B \} \neq \emptyset \). Thus the assertion 3 is true. Hence \( c(\mathcal{L}) \) is a ring of subsets of \( \omega_\mathcal{L} X \). In particular, \( c(\mathcal{L}) \) is a closed base of some \( T_0 \)-topology on \( \omega_\mathcal{L} X \). The assertions 4 and 5 are proved. The assertion 6 is obvious. The proof is complete.

On \( \omega_\mathcal{L} X \) we consider the topology generated by the closed base \( c(\mathcal{L}) \).

**2.2. Theorem.** Let \( \mathcal{L} \) be a closed ring of subsets of a space \( X \). Then \( (\omega_\mathcal{L} X, \omega_\mathcal{L}) \) is a correct weakly incompressible \( g \)-compactification of the space \( X \). Moreover, if \( Y = \omega_\mathcal{L} X \), then:
1. \( \text{cl}_Y \omega_\mathcal{L}(A) = c_\mathcal{L}(A) \) for any \( A \in \mathcal{L} \);
2. if \( \xi \in \omega_\mathcal{L} X \setminus \omega_\mathcal{L}(X) \), then the set \( \{ \xi \} \) is closed in \( \omega_\mathcal{L} X \);
3. if \( A \in \mathcal{L}, \emptyset \) is a compact subset of \( Y \) and \( \omega_\mathcal{L}(A) \subseteq \emptyset \), then \( \text{cl}_Y \omega_\mathcal{L}(A) \subseteq \emptyset \).

**Proof.** It is well-known that \( \omega_\mathcal{L} X \) is a compact space. Fix \( A \in \mathcal{L} \). By definition of the topology on \( \omega_\mathcal{L} X \), we have \( \text{cl}_Y \omega_\mathcal{L}(A) \subseteq c_\mathcal{L}(A) \). Suppose that \( \xi \in c_\mathcal{L}(A) \setminus \text{cl}_Y \omega_\mathcal{L}(A) \). Then \( \xi \) is a free \( \mathcal{L} \)-ultrafilter, \( A \in \xi, \omega_\mathcal{L}(H) \cap \text{cl}_Y \omega_\mathcal{L}(A) \subseteq \omega_\mathcal{L}(H \cap A) \setminus \emptyset \) for any \( H \in \xi \) and \( \cap \{ c_\mathcal{L}(H) : H \in \xi \} = \{ \xi \} \). Thus \( \cap \{ c_\mathcal{L}(A) \setminus \text{cl}_Y \omega_\mathcal{L}(A) \cap \text{cl}_Y \omega_\mathcal{L}(H) : H \in \xi \} = \emptyset \), a contradiction with the condition of compactness of the set \( \text{cl}_Y \omega_\mathcal{L}(A) \). The assertions 1 and 2 are proved. The proof of the assertion 3 is similar with the proof of the assertion 1. The proof is complete.
2.3. **Corollary.** If $\mathcal{L}$ is a closed ring base of a space $X$, then $\omega_{\mathcal{L}}X$ is a correct weakly incompressible compactification of the space $X$.

2.4. **Definition.** If $\mathcal{L}$ is a ring of closed subsets of $X$, then $(\omega\mathcal{L}X, \omega_{\mathcal{L}})$ is called the Wallman-Shanin $g$-compactification or the $WS$-$g$-compactification generated by the closed ring $\mathcal{L}$.

2.5. **Definition.** A $g$-compactification $(Y, f)$ of a space $X$ is called a $WS$-$g$-compactification of $X$ if $(Y, f) = (\omega\mathcal{L}X, \omega_{\mathcal{L}})$ for some closed ring $\mathcal{L}$ of $X$.

2.6. **Definition.** If $\mathcal{L}$ is the ring of all closed subsets of $X$, then $(\omega X, \omega_{\mathcal{L}})$ is called the Wallman compactification of the space $X$.

2.7. **Corollary.** The Wallman compactification $\omega X$ of a space $X$ is correct, incompressible and the set $\{\xi\}$ is closed in $\omega X$ for any point $\xi \in \omega X \setminus X$.

2.8. **Remark.** The compactifications $bD$ and $cD$ of the discrete space $D$ from Examples 1.7 and 1.8 are not $WS$-compactifications. If $a = 1$ in the $WS$-compactification $(Y, f)$ of a space $X$, then $Y = \omega X$.

2.9. **Definition.** A mapping $g : X \to Y$ of a space $X$ into a space $Y$ is **weakly perfect** if $g$ is a closed continuous mapping and the set $g^{-1}(y)$ is compact provided $y \in z(Y)$.

Thus, every continuous closed mapping $g : X \to Y$ of a $T_0$-space $X$ into a $T_1$-space $Y$ is weakly perfect.

2.10. **Theorem.** Let $g : X \to Y$ be a weakly perfect mapping of a space $X$ into a space $Y$. Then:

1. there exists a weakly perfect mapping $\varphi : \omega X \to \omega Y$ such that $g = \varphi \restriction X;
2. if $\Psi : \omega X \to \omega Y$ is a closed continuous mapping and $g = \Psi \restriction X$, then $\Psi = \varphi$.

**Proof.** We consider that $Y = g(X)$. If $\xi$ is an ultrafilter of closed subsets of $X$, then $g(\xi) = \{g(H) : H \in \xi\}$ is an ultrafilter of closed subsets of $Y$. In particular, if $\xi \in \omega X \setminus X$, then $g(\xi) \in \omega Y$ and $g(\xi)$ is an ultrafilter of closed subsets of $Y$ and the set $\{g(\xi)\}$ is closed in $\omega X$.

We put $\varphi(x) = g(x)$ for $x \in X$ and $\varphi(\xi) = g(\xi)$ for $\xi \in \omega X \setminus X$. If $y \in z(Y)$ then the set $g^{-1}(y)$ is compact and $\xi \notin g^{-1}(y)$ for any ultrafilter $\xi \in \omega X$. Thus $\varphi^{-1}(y) = g^{-1}(y)$. By construction, $\varphi^{-1}(cl_{\omega Y} F) = cl_{\omega X} g^{-1}(F)$ for any closed subset $F$ of $Y$. Thus $\varphi$ is a continuous mapping and the set $\varphi^{-1}(y)$ is compact for any $y \in \omega X$. If $F$ is a closed subset of $X$, then $\varphi(cl_{\omega X} F) = cl_{\omega Y} g(F)$. Thus the mapping $\varphi$ is closed. The assertion 1 is proved. The assertion 2 is obvious. The proof is complete.

2.11. **Corollary.** Let $(Y, f)$ be a $g$-compactification of a space $X$, $Z = f(X), \varphi : \omega X \to Y$ be a closed continuous mapping and $f = \varphi \restriction X : X \to Z$ be a weakly perfect mapping.

1. Then there exists a unique weakly perfect mapping $\Psi : \omega X \to \omega Z$ and a continuous closed mapping $h : \omega Z \to Y$ such that $f = \Psi \restriction X = h \circ \Psi$ and $h(z) = h^{-1}(z) = z$ for any $z \in Z$. 


2. \((Y, f)\) is a correct \(g\)-compactification of the space \(X\) and a correct compactification of the space \(Z\).

3. \((Y, f)\) is an incompressible \(g\)-compactification of the space \(X\).

4. If \(F\) is a closed subset of \(Z\) and \(y \in \text{cl}_Y F \setminus F\), then there exists an ultrafilter \(\eta\) of closed subsets of \(Z\) such that \(F \in \eta\) and \(\cap \{\text{cl}_Y H : H \in \eta\} = \{y\}\).

3. CONSTRUCTION OF ALL STRONGLY CORRECT \(G\)-COMPACTIFICATIONS WITH A FINITE REMAINDER

3.1. Theorem. Let \((Y, f)\) be a strongly correct \(g\)-compactification of a space \(X\), \(Y \setminus f(X) \subseteq c(Y)\) and the set \(Y \setminus f(X)\) is finite. Then:

1. \((Y, f)\) is a weakly incompressible \(g\)-compactification of \(X\).

2. \((Y, f)\) is a \(WS\)-\(g\)-compactification of the space \(X\).

Proof. Let \(Z = f(X)\). The assertions are obvious if \(Z = Y\). Suppose that \(Y \setminus Z\) is a non-empty finite set. Let \(Y \setminus Z = \{y_1, y_2, \ldots, y_n\}\), where the points \(y_1, y_2, \ldots, y_n\) are distinct. If \(R = \{y_1, y_2, \ldots, y_n\}\), then every subset \(L \subset R\) is closed in \(Y\). Since the sets \(L \subset R\) are closed and \(\{\text{cl}_Y A : A \subset Z\}\) is a closed base of \(Y\), then there exist the closed subsets \(F_1, F_2, \ldots, F_n\) of \(Z\) such that:

1. \(\text{cl}_Y F_i = F_i \cup \{y_i\}\) for any \(i \leq n\);

2. \(F_i \cap F_j = \emptyset\) for all \(1 \leq i < j \leq n\).

Let \(L_0 = \{H \subset Z : H\) is closed in \(Y\}\). It is obvious that \(\emptyset \in L_0\) and \(A \cup B, A \cap B \in L_0\) for any \(A, B \in L_0\).

If \(\Phi\) is a closed compact subset of \(Z\) and \(\Phi \notin L_0\), then the set \(\Phi\) is not closed in \(Y\) and the \(g\)-compactification \(Y\) is not incompressible. Suppose that \(\Phi\) is a compact subset of \(Y\) and \(Z \subset \Phi \neq Y\). We may assume that \(y_1 \notin \Phi\). Since \(\{y_1\}\) is a closed subset of \(Y\) and \(y_1 \in \text{cl}_Y F_1\), there exists a filter \(\eta\) of closed subsets of \(Z\) such that \(F_1 \in \eta\) and \(\cap \{\text{cl}_Y H : H \in \eta\} = \{y_1\}\). Then \(H \cap \Phi = \emptyset\) for some \(H \in \eta\) and \(H \subset Z \setminus \Phi\), a contradiction. Thus \(Y\) is a weakly incompressible \(g\)-compactification.

For every \(i \leq n\) there exists a free ultrafilter \(\eta_i\) of closed subsets of \(Z\) such that \(F_i \in \eta_i\). In this case \(\cap \{\text{cl}_Y H : H \in \eta_i\} = \{y_i\}\) for all \(i \leq n\). Let \(L_n = \eta_n\). For any \(i \leq n - 1\) we put \(L_i = \{H_0 \cup H_1 \cup \ldots \cup H_i : H_0 \in L_0\}\) and \(H_j \in \{L \in \eta_j : L \subset F_j\}\) for \(1 \leq j \leq i\). Let \(L^* = L_0 \cup L_1 \cup \ldots \cup L_n\). Then \(L^*\) is a closed base and a ring of subsets of the space \(Z\). In this case \(\omega_L Z = Y\).

If \(M = \{f^{-1}(H) : H \in L\}\), then \(\omega_M X = \omega_L Z\). The proof is complete.

3.2. Definition. Let \(L\) be a ring of closed subsets of \(X\). The family \(L_0 = L \setminus \{H \in L : H \in \xi\) for some free \(L\)-filter \(\xi\}\) is called the nucleus of the ring \(L\).

3.3. Corollary. If \(L\) is a ring of closed subsets of a space \(X\), then \(L_0 = \{H \in L : \omega_L(H)\) is closed in \(\omega_L X\}\).
3.4. Definition. Let $Ł$ be a ring of closed subsets of $X$ and $n \in \mathbb{N}$. We say that the rank $r(Ł) \leq n$ if there exists $F_1, F_2, \ldots, F_n \in Ł$ such that $ξ \cap \{F_1, F_2, \ldots, F_n\} \neq \emptyset$ for any free $Ł$-ultrafilter $ξ$.

We put $|A| = |A|$ if the set $A$ is finite and $|A| = \infty$ if the set $A$ is infinite.

3.5. Corollary. $r(Ł) = |ω_ŁX \setminus ω_Ł(X)|$ for any ring $Ł$ of closed subsets of a space $X$.

3.6. Construction. Let $Ł_0, ξ_1, ξ_2, \ldots, ξ_n$ be families of closed subsets of $X$ and $F_1, F_2, \ldots, F_n$ be closed subsets of $X$ such that:

1. $F_i \cap F_j = \emptyset$ for $1 \leq i < j \leq n$;
2. $F_i \in ξ_i$ for any $i \leq n$;
3. $\cap ξ_i = \emptyset$ for any $i \leq n$;
4. if $A, B \in Ł_0$, then $A \cup B, A \cap B \in Ł_0$;
5. if $A, B \in ξ_i$ and $1 \leq i \leq n$, then $A \cup B, A \cap B \in ξ_i$ and $\emptyset \notin ξ_i$;
6. if $1 \leq i \leq n$ and $A \in Ł_0$, then $A \cap H = \emptyset$ for some $H \in ξ_i$.

We put $Ł_n = \{X\} \cup \{A_0 \cup (A_1 \cap F_1) \cup \ldots \cup (A_{n-1} \cap F_{n-1}) \cup A_n, A_0 \in Ł_0, A_1 \in ξ_1, \ldots, A_{n-1} \in ξ_{n-1}, A_n \in ξ_n\}$ and $Ł_i = \{A_0 \cup (A_1 \cap F_1) \cup \ldots \cup (A_i \cap F_i) : A_0 \in Ł_0, A_1 \in ξ_1, \ldots, A_i \in ξ_i\}$ for any $1 \leq i \leq n - 1$. Let $Ł = Ł_0 \cup Ł_1 \cup \ldots \cup Ł_n$.

Then $Ł$ is a ring of closed subsets of $X$, $Ł_0$ is the nucleus of $Ł$ and $r(Ł) = n$. The pair $(ω_ŁX, ω_Ł)$ is a $WŚ$-g-compactsification and $|ω_ŁX \setminus ω_Ł(X)| = n$.

Every $WŚ$-g-compactsification with the finite remainder may be obtained in this way.

3.7. Remark. The one-point correct $T_1$-compactsifications of $T_1$-spaces were constructed by P. C. Osmatescu [28].

3.8. Construction. Let $Ł_0$ be a family of compact closed subsets of a space $X$ such that $A \cup B, A \cap F \in Ł_0$ for all $A, B \in Ł_0$ and any closed subset $F$ of $X$. Fix a free ultrafilter $ξ$ of closed subsets of $X$. Then $Ł_0 \cup ξ = Ł$ is a base ring of closed subsets of $X$ and $ω_ŁX$ is a one-point compactsification of the space $X$. We say that $ω_ŁX$ is generated by the nucleus $Ł_0$.

If $aX$ and $bX$ are two one-point $WŚ$-compactsifications of the space $X$, then $Ł_1 = \{F \subseteq X : F$ is closed in $aX\}$ is the nucleus of $aX$, $Ł_2 = \{F \subseteq X : F$ is closed in $bX\}$ is the nucleus of $bX$ and $bX = aX$ if and only if $Ł_1 = Ł_2$.

If $aX$ is a one-point compactsification of $X$ and every compact closed subset of $X$ is closed in $aX$, then $aX$ is called the one-point Alexandroff compactsification of the space $X$.

4. EXISTENCE OF A $WŚ$-COMPACTIFICATION WHICH IS NOT A CLOSED IMAGE OF THE WALLMAN COMPACTIFICATION

Let $X$ be a $T_0$-space and $Φ = \{a_n : n \in \mathbb{N}\}$ be an infinite set of $X$ such that every subset $L \subseteq Φ$ is closed in $X$. We put $B = \{b_n = a_{3n+1} : n \in \mathbb{N}\}$, $C = \{c_n = a_{3n+2} : n \in \mathbb{N}\}$ and $A = Φ \setminus (A \cup B)$. There exists the free
ultrafilters $\lambda, \mu, \eta$ of closed subsets of $X$ such that $A \subseteq \lambda$, $B \subseteq \mu$, $C \subseteq \eta$. Fix a family $\mathcal{L}_0$ of closed compact subsets of $X$ such that $A \cup B, A \cap F \in \mathcal{L}_0$ for all $A, B \in \mathcal{L}_0$ and a closed subset $F$ of $X$. We put $\mathcal{L}_2 = \eta, \mathcal{L}_1 = \{H_0 \cup (H_1 \cap A) \cup (H_2 \cap B) : H_0 \in \mathcal{L}_0, H_1 \subseteq \lambda$ and $H_2 \subseteq B\}$ and $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_2$. Then $\mathcal{L}$ is a base ring of closed subsets of $X$. Suppose that $\varphi : \omega X \rightarrow \omega L X$ is a mapping of $\omega X$ onto $\omega L X$ and $\varphi^{-1}(x) = x$ for any $x \in X$.

**Case 1.** $X$ is not an open subset of $\omega X$. In this case the mapping $\varphi$ is not continuous: $X$ is an open subset of $\omega L X$ and $\varphi^{-1}(X) = X$.

**Case 2.** There exists a closed compact subset $F$ of $X$ such that $F \notin \mathcal{L}_0$. In this case $F$ is not closed in $\omega L X$ and $\varphi(F) = F$. Thus in this case the mapping $\varphi$ is not closed.

**Case 3.** $X$ is a normal locally compact space and for $X$ a Hausdorff compactification does not exist, $bX$ for which $bX \setminus X$ is a finite set and $|bX \setminus X| \geq 2$ does not exist. Since $|\omega L X \setminus X| = 2$, then $\varphi$ is not a closed mapping. In particular, if $\omega X \setminus X$ is a connected space, then $\varphi$ is not a closed mapping.

5. **Existence of Compactifications of the Given Weight**

5.1. **Theorem.** For a space $X$ and a cardinal number $\tau$ the following assertions are equivalent:

1. $w(X) \leq \tau$;
2. there exists a compactification $bX$ of $X$ such that $w(bX) \leq \tau$;
3. there exists a $\omega L$-compactification $\omega L X$ such that $w(\omega L X) \leq \tau$.

**Proof.** The implications $2 \rightarrow 3 \rightarrow 1$ are obvious. If $w(X) \leq \tau$, then there exists an open base $\mathcal{B}$ of $X$ such that $|\mathcal{B}| \leq \tau$. Let $\mathcal{L}$ be the minimal family of closed subsets of $X$ for which $\{X \setminus U : U \in \mathcal{B}\} \subseteq \mathcal{L}$. Then $|\mathcal{L}| \leq \tau$ and $\omega L X$ is the desired compactification. The proof is complete.

5.2. **Definition.** A $\omega$-network on $X$ is a family $\mathcal{B}$ of closed subsets such that for every point $x \in X$ and any neighbourhood $U$ of $X$ there exists an element $H \in B$ for which $x \in H \subseteq U$.

The $\omega$-network weight of a space $X$ is defined as the smallest cardinal number $|\mathcal{B}|$, where $\mathcal{B}$ is a $\omega$-network of $X$ and this cardinal number is denoted by $\text{cow}(X)$.

The cardinal number $\text{cow}(X)$ is defined if and only if $X$ is a $T_1$-space. If for $X$ the cardinal number $\text{cow}(X)$ is not defined, then we put $\text{cow}(X) = \infty$.

5.3. **Theorem.** For a space $X$ and a ring $\mathcal{L}$ of closed subsets of $X$ the following assertions are equivalent:

1. $\omega L X$ is a $T_1$-compactification of the space $X$;
2. $\mathcal{L}$ is a closed base and a $\omega$-network of the space $X$;
3. for every $x \in X$ the family $\xi(x, \mathcal{L})$ is an $\mathcal{L}$-ultrafilter;
4. $\mathcal{L}$ is a closed base of $X$ and $c(\mathcal{L})$ is a $\omega$-network of $\omega L X$. 

Proof. 1. The ring $L$ is a closed base of $X$ if and only if $\omega_L X$ is a compactification of $X$. In this case we assume that $X \subseteq \omega_L X$.

2. $\omega_L X$ is a $T_1$-compactification of $X$ if and only if $L$ is a closed base and $c(L)$ is a $\alpha$-network of $\omega_L X$. The implications $4 \rightarrow 1 \rightarrow 4 \rightarrow 2$ are proved.

3. If $\omega_L X$ is a $T_1$-compactification of $X$, then every family $\xi(x, L)$ is an $L$-ultrafilter. A point $\xi \in \omega_L X$ is closed in $\omega_L X$ if and only if $\xi$ is an $L$-ultrafilter.

The implication $1 \rightarrow 3$ is proved.

4. Let $L$ be a closed base and a $\alpha$-network of $X$. Fix $x \in X$. If $F \in L$ and $x \notin F$, then there exists $H \in L$ such that $x \in H \subseteq X \setminus F$. Thus $F \notin \xi(x, L)$ and $\xi(x, L)$ is an $L$-ultrafilter. The implication $2 \rightarrow 3$ is proved. The proof is complete.

5.4. Theorem. (A.V. Arhangel’skii [4]). For a $T_1$-space $X$ and an infinite cardinal number $\tau$ the following assertions are equivalent:

1. there exists a $T_1$-compactification $bX$ of $X$ such that $w(bX) \leq \tau$;
2. there exists a $WS$-compactification $bX$ of $X$ such that $w(bX) \leq \tau$;
3. $w(X) + c\omega(X) \leq \tau$.

Proof. The implications $1 \rightarrow 3 \rightarrow 1$ are proved by A.V. Arhangel’skii [4]. The implication $2 \rightarrow 1$ is obvious. If $L_1$ is a closed base of $X$, $L_2$ is a $\alpha$-network of $X$, $|L_1| + |L_2| \leq \tau$ and $L$ is the minimal ring of closed subsets of $X$ such that $L_1 \cup L_2 \subseteq L$, then $|L| \leq \tau$, $\omega_L X$ is a $T_1$-compactification and $w(\omega_L X) \leq |L| \leq \tau$. The proof is complete.

5.5. Example. (see A. V. Arhangel’skii [4]). Let $X = [0, 1]$, $Q$ be the set of rational numbers of $X$ and $U(x, \varepsilon) = \{x\} \cup \{y \in X \setminus Q : x - \varepsilon < y < x + \varepsilon\}$ for every $x \in X$ and $\varepsilon > 0$. On $X$ we consider the topology generated by the open base $\{U(x, \varepsilon) : x \in X \text{ and } \varepsilon > 0\}$. It is obvious that the set $X \setminus Q$ is open in $X$ and $X$ is a Hausdorff $H$-closed space with a countable base. The $\alpha$-weight of $X$ is not countable. Thus for $X$ a $T_1$-compactification of countable weight does not exist.

6. THE MAXIMALITY OF A RING OF CLOSED SETS

6.1. Theorem. For a $g$-compactification $(Y, f)$ of a space $X$ and a ring $L$ of closed subsets of $X$ the following assertions are equivalent:

1. $(Y, f) = (\omega_L X, \omega_L)$;
2. a family $b(L) = \{b(H) = cl_Y f(H) : H \in L\}$ is a closed base of the space $Y$, $b(F) \cap f(Y) = f(F)$, $F = f^{-1}(f(F))$ and $b(F \cap H) = b(F) \cap b(H)$ for all $F, H \in L$.

Proof. The implication $1 \rightarrow 2$ follows from Lemma 2.1.

Let $Z = f(X)$, $b(L)$ be a closed base of $Y$, $b(F) \cap Z = f(F)$ and $b(F \cap H) = b(F) \cap b(H)$ for all $F, H \in L$. For every $x \in X$ we put $\varphi(\xi(x, L)) = f(x)$ and $\varphi(\xi) = \cap\{b(H) : H \in \xi\}$ for any $\xi \in \omega_L X \setminus \omega_L(X)$. The mapping $\varphi$ is one-to-
one and \( \varphi(c_L(H)) = b(H) \) for any \( H \in L \). Thus \( \varphi \) is a homeomorphism. The proof is complete.

6.2. Definition. A ring \( L \) of closed subsets of \( X \) is called a maximal ring if \( (\omega_L X, \omega_L) \neq (\omega_M X, \omega_M) \) for every ring \( M \) of closed subsets of \( X \) for which \( L \subseteq M \) and \( L \neq M \).

From Theorem 6.1 it follows

6.3. Corollary. For a ring \( L \) of closed subsets of a space \( X \) the following assertions are equivalent:

1. the ring \( L \) is maximal;
2. for every closed subset \( \Phi \) of \( X \), for which \( \Phi \notin L \), there exists \( F \in L \) such that \( cl_Y(\omega_L(F \cap \Phi)) \neq cl_Y(\omega_L F \cap cl_Y(\omega_L \Phi)) \).

6.4. Corollary. Let \((L_\alpha : \alpha \in A)\) be a family of rings of closed subsets of \( X \), \( L = \cup\{ L_\alpha : \alpha \in A \} \) and for every \( \alpha, \beta \in A \), \( F \in L_\alpha \) and \( \Phi \in L_\beta \) there exists \( \mu \in A \) such that \( F, \Phi \in L_\mu \). If \( (\omega_{L_\alpha} X, \omega_{L_\alpha}) = (\omega_{L_\beta} X, \omega_{L_\beta}) \) for all \( \alpha, \beta \in A \), then \( L \) is a ring of closed subsets of \( X \) and \( (\omega_L X, \omega_L) = (\omega_{L_\alpha} X, \omega_{L_\alpha}) \) for any \( \alpha \in A \).

6.5. Corollary. For every ring \( L \) of closed subsets of \( X \) there exists a maximal ring \( M \) of closed subsets of \( X \) such that \( L \subseteq M \) and \( (\omega_L X, \omega_L) = (\omega_M X, \omega_M) \).

6.6. Example. Let \( X \) be a space and \( \xi, \mu \) be two distinct free ultrafilters of closed subsets of \( X \). We put \( \lambda = \{ F \subseteq X : F \text{ is a compact closed subset of } X \} \), \( L = \lambda \cup \xi \) and \( M = \mu \cup \lambda \). Then \( L, M \) are maximal rings of closed subsets of \( X \), \( L \neq M \) and \( (\omega_L X, \omega_L) = (\omega_M X, \omega_M) \) is the one-point Alexandroff compactification \( aX \) of the space \( X \).

6.7. Remark. From Corollary 6.3 it follows that the proof of Proposition 4.10 from [19] is not correct. By virtue of Example from [39] it follows that the Proposition 4.10 and Theorem 4.11 from [19] are not true.

7. HAUSDORFF G-COMPACTIFICATIONS

Let \( HGC(X) \) be the family of all Hausdorff \( g \)-compactifications of a space \( X \) and \( HC(X) = C(X) \cap HGC(X) \). A space \( X \) is completely regular if and only if \( HC(X) \neq \emptyset \).

It is obvious that every Hausdorff \( g \)-compactification is strongly correct and incompressible.

7.1. Definition. A class \( B \) of \( g \)-compactifications of a space \( X \) is called:

- a complete upper semilattice if \( B \) is a non-empty set and for every non-empty set \( M \subseteq B \) there exists a \( g \)-compactification \( (\vee M) \cap B \);
- a complete lattice if \( B \) is a complete upper semilattice and \( (\wedge B) \cap B \neq \emptyset \).

If \( B \) is a complete upper semilattice of \( g \)-compactifications of a space \( X \), then in \( B \) there exists some maximal element \( (\beta_B X, \beta_B) \).
It is obvious that $HGC(X)$ is a complete lattice of $g$-compactifications of a space $X$ and the maximal Hausdorff $g$-compactification $(\beta X, \beta_X)$ is unique. The $g$-compactification $(\beta X, \beta_X)$ is called the Stone–Čech $g$-compactification of the space $X$.

7.2. Theorem. (see [13], Theorem 3.6.21 for $T_1$-spaces). If $f : X \to Y$ is a continuous mapping of a space $X$ into a Hausdorff compact space $Y$, then there exists a unique perfect mapping $\varphi : \omega X \to Y$ such that $f = \varphi \mid X$. If $f(X)$ is dense in $Y$, then $(Y, f) \in HGC(X)$.

Proof. Suppose that the subspace $Z = f(X)$ is dense in $Y$. For $x \in X$ we put $\varphi(x) = f(x)$. If $\xi \in \omega X \setminus X$, then there exists a unique point $y \in Y$ such that $y \in \cap \{cl_Y f(H) : H \in \xi\}$ and we put $\varphi(\xi) = y$. By construction, $\varphi : \omega X \to Y$ is a single-valued mapping. Fix a closed subset $F$ of $Z$. Let $\Phi = f^{-1}(F)$. If $x \in cl_{\omega X} \Phi$, then $\varphi(x) \in cl_Y F$. Suppose that $y \in cl_Y F$ and $\eta = \{cl_Z (U \cap Z) : U$ is open in $Y$ and $y \in U\}$. Then $\{y\} = \cap \{cl_Y (H \cap F) : H \in \eta\}$. There exists an ultrafilter $\xi$ of closed subsets of $X$ such that $\Phi \in \xi$ and $f^{-1}(H) \in \xi$ for any $H \in \eta$. Then $\xi \in cl_{\omega X} \Phi$ and $y = \varphi(\xi)$. Thus $cl_{\omega X} \Phi = \varphi^{-1}(cl_Y F)$ and the mapping $\varphi$ is continuous. Since $Y$ is a Hausdorff space, then $\varphi$ is a perfect mapping. The proof is complete.

7.3. Corollary. There exists a unique continuous mapping $\varphi : \omega X \to \beta X$ such that $\beta_X = \varphi \mid X$.

7.4. Corollary. For every continuous mapping $\varphi : X \to Y$ there exists a unique continuous mapping $\beta f : \beta X \to \beta Y$ such that $\beta f \circ \beta_X = \beta_Y \circ f$. In particular, the mapping $\beta f$ is perfect.

8. RELAXED COMPACTIFICATIONS

8.1. Definition. A $g$-compactification $(Y, f)$ is called a relaxed $g$-compactification or a $\rho$- $g$-compactification of a space $X$ if there exists a continuous mapping $\varphi : \omega X \to Y$ of $\omega X$ onto $Y$ such that $f = \varphi \mid X$.

Let $\rho GC(X)$ be the class of all $\rho$- $g$-compactifications of a space $X$ and $\rho C(X) = C(X) \cap \rho GC(X)$.

It is obvious that $\rho GC(X)$ is a set and $\rho C(X) \neq \emptyset$. Moreover, $HGC(X) \subseteq \rho GC(X)$. The compactification $\omega X$ is the maximal element of the classes $\rho GC(X)$ and $\rho C(X)$.

We say that the $g$-compactification $(Y, f)$ is a $\rho^*$- $g$-compactification if $(Y, f) \in \rho GC(X)$ and for every two continuous mappings $\varphi, \psi : \omega X \to Y$ for which $\varphi \mid X = \psi \mid X = f$ we have $\varphi = \psi$. Let $\rho^* GC(X)$ be the class of all $\rho^*$- $g$-compactifications of $X$ and $\rho^* C(X) = C(X) \cap \rho^* GC(X)$.

8.2. Example. Let $D$ be a discrete space and $cD = D \cup \{a, b\}$ be the compactification constructed in Example 1.7. We put $\varphi(x) = \psi(x) = x$ for any $x \in D$, $\varphi^{-1}(a) = cl_{\omega D} A \setminus A$, $\varphi^{-1}(b) = cl_{\omega D} B \setminus B$, $\psi^{-1}(b) = \emptyset$ and
\[ \psi^{-1}(a) = \omega D \setminus D. \] The mappings \( \varphi, \psi \) are continuous and \( \varphi \neq \psi \). Thus \( e \in \rho C(X) \setminus \rho^*C(X) \).

8.3. Proposition. Let \( \{(Y_\alpha, f_\alpha) : \alpha \in A\} \) be a non-empty set of \( g \)-compactifications of a space \( X \), \( \{\varphi_\alpha : \omega X \to Y_\alpha : \alpha \in A\} \) be a family of continuous mappings, \( f_\alpha = \varphi_\alpha \mid X \) for any \( \alpha \in A \), \( \varphi(x) = (\varphi_\alpha(x) : \alpha \in A) \in \cap\{Y_\alpha : \alpha \in A\} \) for every \( x \in \omega X \), \( Y = \varphi(\omega X) \) and \( f = \varphi \mid X \). Then \( (Y, f) \) is a \( \rho \)-\( g \)-compactification of the space \( X \) and \( (Y_\alpha, f_\alpha) \leq (Y, f) \) for any \( \alpha \in A \). If \( (Y_\alpha, f_\alpha) \in \rho^*GC(X) \) for every \( \alpha \in A \), then \( (Y, f) = \lor\{(Y_\alpha, f_\alpha) : \alpha \in A\} \) and \( (Y, f) \in \rho^*GC(X) \).

Proof. It is obvious that \( (Y, f) \in \rho GC(X) \) and \( (Y_\alpha, f_\alpha) \leq (Y, f) \) for every \( \alpha \in A \).

Suppose that \( (Y_\alpha, f_\alpha) \in \rho^*GC(X) \) for every \( \alpha \in A \). Let \( (Z, g) \in \rho GC(X) \) and \( (Y_\alpha, f_\alpha) \in (Z, g) \) for any \( \alpha \in A \).

Fix a continuous mapping \( h : \omega X \to Z \), where \( g = h \mid X \). For every \( \alpha \in A \) fix a continuous mapping \( h_\alpha : Z \to Y_\alpha \), where \( f_\alpha(x) = h_\alpha(h(\omega)) \) for every \( x \in X \).

Since \( (Y_\alpha, f_\alpha) \in \rho^*GC(X) \), \( \varphi_\alpha = h_\alpha \circ h \) for any \( \alpha \in A \). Thus \( \psi(z) = (h_\alpha(z) : \alpha \in A) = (\varphi_\alpha(h^{-1}(z)) : \alpha \in A) \) is a continuous mapping of \( Z \) into \( Y \) and \( f = g \circ \psi \). The proof is complete.

8.4. Corollary. For every space \( X \):

- The set \( \rho^*GC(X) \) is a complete lattice with a maximal element \( (\omega X, \omega X) \) and a minimal element \( (mX, mX) \), where \( mX \) is the singleton space;
- The set \( \rho^*C(X) \) is a complete upper semilattice with the maximal element \( (\omega X, \omega X) \).

8.5. Theorem. Let \( (Y, f) \) be a \( g \)-compactification and \( \varphi : \omega X \to Y \) be a continuous closed mapping for which \( f = \varphi \mid X \). Then:

1. \( (Y, f) \in \rho^*GC(X) \);
2. if \( F \) is a closed subset of the space \( Z = f(X) \) and \( y \in \text{cl}_{Y}F \setminus F \), then there exists an ultrafilter \( \xi \) of closed subsets of \( Z \) such that \( F \in \xi \) and \( \cap\{\text{cl}_{Y}H : H \in \xi\} = \{y\} \).

Proof. Let \( F \) be a closed subset of \( Z \) and \( y \in \text{cl}_{Y}F \setminus F \). Then \( y \in Y \setminus Z \), \( \varphi^{-1}(y) \) is a compact closed subset of \( \omega X \) and \( \varphi^{-1}(y) \subseteq \omega X \setminus X \). It is obvious that \( \text{cl}_{\omega X}f^{-1}(F) \cap \varphi^{-1}(y) \neq \varnothing \). Let \( \eta \in \varphi^{-1}(y) \cap \text{cl}_{\omega X}f^{-1}(F) \). Then \( \eta \) is an ultrafilter of closed subsets of \( F \) and \( F^{-1}(F) \in \eta \). Hence \( \xi = f(\eta) \) is the desired ultrafilter. The assertion 2 is proved.

Suppose that \( \psi : \omega X \to Y \) is a continuous mapping, \( \psi \mid X = f \) and \( \varphi \neq \psi \). There exists a point \( \eta \in \omega X \setminus X \) such that \( \varphi(\eta) \neq \psi(\eta) \). Since \( \varphi \) is a closed mapping, the set \( \{\varphi(\eta)\} \) is closed and \( \{\psi(\eta)\} = \cap\{\text{cl}_{Y}f(H) : H \in \eta\} \). There exists \( F \in \eta \) such that \( \psi(\eta) \notin \text{cl}_{Y}f(F) \). Thus \( \eta \in \text{cl}_{\omega X}F \) and \( \psi(\eta) \notin \text{cl}_{Y}f(F) = \text{cl}_{Y}(F) \), a contradiction. Thus \( (Y, f) \in \rho^*GC(X) \). The proof is complete.
8.6. **Example.** Let \( X \) be a locally compact non-compact \( T_1 \)-space and \( bX = X \cup \{b\} \) be the one-point compactification of \( X \) with the nucleus \( \mathcal{L}_0 = \{F \subseteq X : F \) is a finite set\}. Then \( bX \in \rho^*C(X) \). If the space \( X \) is not discrete, then does not exist a closed mapping \( \varphi : \omega X \to bX \) such that \( \varphi(x) = x \) for any \( x \in X \) does not exist.

8.7. **Remark.** \( HGC(X) \subseteq \rho^*GC(X) \) for any space \( X \).

9. **\( \omega \alpha \)-COMPACTIFICATIONS**

9.1. **Definition.** A \( g \)-compactification \((Y, f)\) of a space \( X \) is called:
- an \( \omega \alpha \)-compactification of the space \( X \) if there exists a continuous closed mapping \( \varphi : \omega X \to Y \) such that \( f = \varphi \mid X \), the mapping \( f : X \to f(X) \) is closed and the set \( f^{-1}(y) \) is compact for any point \( y \in z(f(X)) \);
- an \( \alpha \)-compactification if there exists a closed continuous mapping \( \varphi : \omega X \to Y \) such that \( f = \varphi \mid X \).

9.2. **Lemma.** Let \((Y, f)\) be a \( g \)-compactification of a space, \( f : X \to f(X) \) be a closed mapping, \( \varphi : \omega X \to Y \) be a closed continuous mapping and \( f = \varphi \mid X \). The following assertions are equivalent:
1. the mapping \( \varphi \) is perfect;
2. \( \varphi^{-1}(y) \) is a compact subset of \( Y \) for any \( y \in z(Y) \);
3. if \( Z = f(X) \), then the subset \( f^{-1}(y) \) is compact for any \( y \in z(Z) \);
4. \( \varphi^{-1}(y) = f^{-1}(y) \) and the subset \( f^{-1}(y) \) is compact for any \( y \in z(Z) \);

**Proof.** By definition, \( c(Y) = \{y \in Y : \) the set \( \{y\} \) is closed in \( Y \} \) and \( z(Y) = Y \setminus c(Y) \). The implications 1 \( \rightarrow \) 2 and 4 \( \rightarrow \) 3 are obvious.

If \( \xi \in \omega X \setminus X \), then the set \( \{\varphi(\xi)\} \) is closed in \( Y \) and \( \varphi(\xi) \in c(Y) \). Thus \( \varphi^{-1}(y) = f^{-1}(y) \) for any \( y \in z(Y) \). Since \( z(Z) \subseteq z(Y) \), then the implications 2 \( \rightarrow \) 3 \( \rightarrow \) 4 are proved. If \( y \in c(Y) \), then \( \varphi^{-1}(y) \) is a closed compact subset of \( \omega X \). The implication 4 \( \rightarrow \) 1 is proved. The proof is complete.

Let \( \Omega AC(X) \) be the set of all \( \omega \alpha \)-\( g \)-compactifications of a space \( X \) and \( \Omega AC(X) = C(X) \cap \Omega AGC(X) \).

From Theorem 8.5 it follows

9.3. **Corollary.** For every space \( X \):
1. \( \Omega AC(X) \subseteq \rho^*GC(X) \);
2. \( \Omega AC(X) \subseteq \rho^*C(X) \);
3. \( \Omega AC(X) \) and \( \Omega AGC(X) \) are sets of \( g \)-compactifications with the maximal element \( (\omega X, \omega X) \) and \( (mX, mX) \) is a minimal element of the set \( \Omega AGC(X) \).

9.4. **Proposition.** For a space \( X \) the following assertions are equivalent:
1. \( X \) is a \( T_1 \)-space;
2. \( \omega X \) is a \( T_1 \)-space;
3. some compactification of \( X \) is a \( T_1 \)-space;
4. every \( \omega \alpha \)-compactification of \( X \) is a \( T_1 \)-space;
5. every \( \omega \alpha \)-\( g \)-compactification of \( X \) is a \( T_1 \)-space.
Proof. Every subspace of a $T_1$-space is a $T_1$-space and the closed image of the $T_1$-space is a $T_1$-space. If $Y$ is a $\omega\alpha$-compactification of $X$, then $z(Y) = z(X)$. The proof is complete.

9.5. Theorem. Let $(Y, f)$ be a $g$-compactification of a space $X$, $\varphi : \omega X \to Y$ be a continuous mapping, $f = \varphi | X$ be a closed mapping, $c(f(X)) \subseteq c(Y)$ and the set $f^{-1}(y)$ be compact for any $y \in z(f(X))$. The following assertions are equivalent:

1. $(Y, f)$ is a $\omega\alpha$-$g$-compactification of the space $X$;
2. the mapping $\varphi$ is closed;
3. $(Y, f)$ is a strongly correct and incompressible $g$-compactification;
4. $(Y, f)$ is an incompressible $g$-compactification.

Proof. The implication $2 \to 1 \to 2$ and $3 \to 4$ are obvious. The implication $1 \to 3$ follows from Corollary 2.11. Suppose now that $(Y, f)$ is an incompressible $g$-compactification. Since $f(\omega X)$ is a compact subset and $Z = f(X) \subseteq f(\omega X)$, $f(\omega X) = Y$. Since $\varphi$ is continuous, $\varphi^{-1}(y)$ is a closed compact subset for any $y \in i(Y)$. Let $F$ be a closed subset of $X$. Then $\Phi = f(F) = \varphi(F)$ is a closed subset of $Z$. The set $\varphi(c\omega_X F)$ is compact and $\Phi \subseteq \varphi(c\omega_X F) \subseteq cl\Phi$. Since $(Y, f)$ is incompressible, $cl\Phi = \varphi(c\omega_X F)$ and $\varphi(c\omega_X F)$ is a closed subset of $Y$. Let $\xi \in \omega X \cdot X$. Then $\xi$ is a ultrafilter of closed subsets of $X$ and $\eta = f(\xi)$ is an ultrafilter of closed subsets of $Z$.

If $y \in \cap\eta$, then $\{y\} = \cap\eta$ and $y \in c(Z)$. Thus $\varphi^{-1}(y) = f^{-1}(y)$ is a compact subset of $\omega X$ for any $y \in z(Z)$. Suppose that $y \in Y \setminus Z$, $z \in Y$, $z \neq y$ and $z \in cl_Y \{y\}$. In this case $L = Y \setminus \{y\}$ is a compact subset of $Y$ and $X \subseteq L$, a contradiction. Thus, $\{y\}$ is a closed subset for any $y \in Y \setminus Z$. Since $c(Z) \subseteq c(Y)$, then $z(Y) = z(Z)$ and $\varphi$ is a compact mapping. If $\Phi$ is a closed subset of $\omega X$, then $\varphi(\Phi) = \cap\{\varphi(c\omega_X F) : F \text{ is a closed subset of } X \text{ and } \Phi \subseteq c\omega_X F\}$ is a closed subset of $Y$. Thus, $\varphi$ is a perfect mapping. The implication $4 \to 2$ is proved. The proof is complete.

9.6. Example. Let $X$ be a non-compact space, $x_0 \in X$ and $\{x_0\} = \cap\{c\omega_X U : U \text{ is open in } X \text{ and } x_0 \in U\}$. Let $(Y, f)$ be a $\omega\alpha$-compactification of $X$ and $y_0 \in Y \setminus X$, where $f(X) = X$. There exists a perfect mapping $\varphi : X \to Y$ such that $\varphi^{-1}(x) = \varphi(x) = f(x) = \{x\}$ for any $x \in X$. On $Y$ we consider a new topology $T' = \{U \subseteq Y \setminus \{y_0\} : U \text{ is open in } Y\} \cup \{U \cup V : U, V \text{ are open in } Y \text{ and } x_0 \in U\}$. Denote by $Z$ the set $Y$ with the topology $T'$.

The mapping $\varphi : \omega X \to Y$ is continuous, the set $\varphi^{-1}(x) = \{x\}$ is compact for any $x \in X$ and $\varphi(\{x_0\}) = f(\{x_0\})$ is not closed in $Z : cl_Z \{x_0\} = \{x_0, y_0\}$. Thus $(Z, f)$ is not a $\omega\alpha$-compactification of $X$. By construction, $x_0 \in c(X) \cap z(Z)$. The compactification $(Z, f)$ is incompressible. Thus the requirement $c(f(X)) \subseteq c(Y)$ is essential in the conditions of Theorem 9.4.

9.7. Example. Let $X$ be a non-compact space, $U$ be an open subset of $X$, $c\omega_X U \neq U$, $V$ be an open subset of $\omega X$ and $V \cap X = U$. Consider the mapping $\varphi : \omega X \to Y$, where $\varphi^{-1}(\varphi(x)) = \{x\}$ for any $x \in V$ and $\varphi^{-1}(\varphi(x)) = V$ for
any \( x \in V \). In \( Y \) we consider the topology \{\( H \subseteq Y : \varphi^{-1}(H) \) is open in \( \omega X \)\}. Let \( f = \varphi \mid X \) and \( z_0 = \varphi(V) \). Then \{\( z_0 \)\} is an open subset of \( Y \) and \((Y,f)\) is a \( \rho\)-\( g\)-compactification of \( X \). The mappings \( f \) and \( \varphi \) are continuous. Suppose now that if \( L \subseteq U \) and \( L \) is closed in \( X \), then \( L = \emptyset \). In this case \( V = U \) and the mappings \( \varphi : \omega X \to Y \) and \( f : X \to f(X) \) are closed. If the set \( V \) is not compact, then \((Y,f)\) is not a \( \omega\)-\( g\)-compactification of \( X \).

9.8. Example. Let \( X = [0,1) \) with the topology \( T = \{[0,x) : 0 \leq x \leq 1 \} \). Let \( U = [0,2^{-1}) \). Then \( \omega X = [0,1] \) with the topology \{\( \omega X \) \cup \( T \)\}. In this case the \( g\)-compactification \((Y,f)\) is not a \( \omega\)-\( g\)-compactification of \( X \), as in Example 9.6.

9.9. Example. Let \( X = [0,1] \) with the topology \{\( [0,x) : 0 \leq x \leq 1 \)\}. Then \( \omega X = X \cup \{1\} \) with the topology \{\( \omega X \) \cup \{\( [0,x) : 0 \leq x \leq 1 \)\} \} and \( X \) is open in \( \omega X \). Consider the space \( Y = \{0,1\} \) with the topology \{\( \emptyset,\{0\}, Y \)\}. Let \( \varphi(x) = f(x) = 0 \) for any \( x \in X \) and \( \varphi(1) = 1 \). Then \( \varphi : \omega X \to Y \) is a continuous closed mapping and \( f : X \to f(X) = \{0\} \) is a closed mapping. Thus \((Y,f)\) is an \( \alpha\)-compactification and a \( \rho\)-compactification of the space \( X \). Since the mapping \( \psi : \omega X \to Y \), where \( \psi(x) = 0 \) for any \( x \in \omega X \), is continuous and \( f = \psi \mid X \), then \((Y,f)\) is not a \( \rho^*\)-compactification and an \( \omega\)-\( \alpha\)-compactification.

10. INTRINSICAL CHARACTERISTICS OF \( \omega\)-\( G\)-COMPACTIFICATIONS

10.1. Definition. Let \( Z \) be a subspace of a space \( Y \). A set \( H \subseteq Y \) is called a \emph{quasi-clopen} in \( Y \) relatively to \( Z \) subset or a \( Z\)-clopen subset of \( Y \) if:

1. \( H \cap Z \) is an open subset of \( Z \);
2. if \( F \) is a closed subset of \( Z \) and \( F \subseteq H \), then \( \text{cyl} F \subseteq H \).

10.2. Definition. Let \( Z \) be a subspace of a space \( Y \) and \( L \subseteq H \subseteq Y \). A set \( H \) is a \( Z\)-\emph{clopen neighbourhood of} \( L \) in \( Y \) if \( H \) is a \( Z\)-clopen subset of \( Y \) and there exists an open subset \( U \) of \( Y \) such that \( L \subseteq U \subseteq H \).

10.3. Definition. Let \( Z \) be a subspace of a space \( Y \). A family \( B \) of subsets of \( Y \) is a \emph{\( Z\)-clopen basis of} \( Y \) if for every open subset \( V \) of \( Y \) and any point \( y \in V \) there exists a \( Z\)-clopen neighbourhood \( H \in B \) of \( y \) in \( Y \), such that \( y \in H \subseteq V \).

10.4. Lemma. \( Z \) is a subspace of a space \( Y \) and \( B \) is a \( Z\)-clopen basis of \( Y \), then the set \( Z \) is dense in \( Y \).

\textbf{Proof}. It is obvious.

10.5. Theorem. Let \((Y,f)\) be a \( \omega\)-\( g\)-compactification of a space \( X \), \( Z = f(X) \) and \( B_Y \) be the set of all \( Z\)-clopen subsets of \( Y \).

1. \( B_Y \) is a \( Z\)-clopen basis of \( Y \);
2. if \( A, B \in B_Y \), then \( A \cup B, A \cap B \in B_Y \).

\textbf{Proof}. It is obvious that \( A \cap B \in B_Y \) for all \( A, B \in B_Y \).
Fix a perfect mapping $\varphi : \omega X \to Y$ such that $f = \varphi | X$ and $f^{-1}(y) = \varphi^{-1}(y)$ for any $y \in \omega(Z)$.

For any open subset $U$ of $X$ we put $< U > = \omega X \setminus cl_{\omega X}(X \setminus U) = \cup \{ V : V$ is open in $\omega X$ and $X \cap V \subseteq \cup \}$. The set $< U >$ is $X$-clopen in $\omega X$. Let $H$ be a $X$-clopen subset of $\omega X$ and $U = H \cap X$. The set $U$ is open in $X$.

Suppose that $F$ is a closed subset of $X$ and $\xi \in cl_{\omega X} F \setminus F$. Then $\xi$ is an ultralimit of closed subsets of $X$ and $F \in \xi$. If $F \subseteq U$, then $X \setminus U \notin \xi$ and $\xi \in \omega X \setminus cl_{\omega X}(X \setminus U) = < U >$. Thus $cl_{\omega X} F \subseteq < U >$. Therefore a set $H \subseteq \omega X$ is a $X$-clopen subset of $\omega X$ if and only if the set $U = H \cap X$ is open in $X$ and $< U > \subseteq H$. Denote by $\mathcal{B}_{\omega X}$ the set of all $X$-clopen subsets of $\omega X$.

By construction, $< U > : U \text{ open in } X \} \text{ is a Z-clopen base of } \omega X$. Thus $\mathcal{B}_{\omega X}$ is an $X$-clopen base of $\omega X$.

Since $< U \cup V > = < U > \cup < V >$ for all open subsets $U, V$ of $X$, then $A \cup B \in \mathcal{B}_{\omega X}$ for all $A, B \in \mathcal{B}_{\omega X}$. It is obvious that $A \cap B \in \mathcal{B}_{\omega X}$ for $A, B \in \mathcal{B}_{\omega X}$.

Since the mapping $\varphi$ is closed, $\varphi^{-1}(\mathcal{B}_Y) \subseteq \mathcal{B}_{\omega X}$.

Let $H \subseteq Y$ and $L = \varphi^{-1}(H)$ be an $X$-clopen subset of $\omega X$. Let $F \subseteq Z$ be a closed subset of $Z$ and $F \subseteq L \cap Z$. The set $\Phi = f^{-1}(F)$ is closed in $X$ and the set $U = X \cap L = f^{-1}(X \cap Z)$ is open in $X$. Since $\Phi \subseteq U, cl_{\omega X} \Phi \subseteq L$. Therefore $cl_Y \Phi = \varphi(cl_{\omega X} \Phi) \subseteq \varphi(L) = H$ and $H$ is a $Z$-clopen subset of $Y$.

Therefore $A \cup B \in \mathcal{B}_Y$ for any $A, B \in \mathcal{B}_Y$.

Let $V$ be an open subset of $Y$ and $y_0 \in V$. The set $\varphi^{-1}(y_0)$ is compact. There exists an open subset $U$ of $X$ such that $\varphi^{-1}(y_0) \subseteq < U > \subseteq \varphi^{-1}(V)$. Let $H = \{ y \in Y : \varphi^{-1}(y) \subseteq < U > \}$ and $L = H \cup (\varphi(< U >) \cap (Y \setminus Z))$. The set $H$ is open in $Y$, $y_0 \in H$ and $L \cap Z = H \cap Z$. Let $F$ be a closed subset of $Z$ and $F \subseteq L$. Then $f^{-1}F \subseteq U$ and $cl_Y F \setminus F \subseteq \varphi(cl_{\omega X} \Phi f^{-1}(F)) \cap (Y \setminus Z) \subseteq (\varphi(< U >) \cap (Y \setminus Z)) \subseteq L$.

Since $\varphi^{-1}(H) \subseteq < U >$, then $\varphi^{-1}(H) \subseteq < X \cap \varphi^{-1}(H) > \subseteq U$. By construction, $\varphi^{-1}(y_0) \subseteq < X \cap \varphi^{-1}(H) >$. Thus there exists a closed subset $\Phi$ of $X$ such that $\Phi \subseteq \varphi^{-1}(H)$ and $\varphi^{-1}(y_0) \cap cl_{\omega X} \Phi \neq \emptyset$. Then $f(\Phi) \subseteq H \cap Z$ and $y_0 \in cl_X f(\Phi) \subseteq L$.

Therefore $L \in \mathcal{B}_Y$ and $y_0 \in H \subseteq L$. The proof is complete.

10.6. Proposition. Let $Z$ be a subspace of a compact space $Y$, $Y \setminus Z \subseteq c(Y)$ and $\mathcal{B}$ be a Z-clopen base of the space $Y$ such that $A \cup B \in \mathcal{B}$ for every $A, B \in \mathcal{B}$.

1. if $H \in \mathcal{B}$, then $F = Z \setminus H$ is a closed subset of $Z$ and $cl_Y F \cup H = Y$;
2. if $F$ is a closed subset of $Z$ and $y \in cl_Y F \setminus F$, then there exists an ultrafilter $\xi$ of closed subsets of $Z$ such that $F \in \xi$ and $\{ y \} = \bigcap (cl_Y H : H \in \xi)$.

Proof. For every $H \in \mathcal{B}$ we put $\omega(H) = \cup \{ V : V$ is open in $Y$ and $V \subseteq H \}$. Then $H \cap Z = Z \cap \omega(H)$ and if $\Phi \subseteq Z \cap H$ is a closed subset of $Z$, then $cl_Y \Phi \subseteq H$. 

Fix $H \in \mathcal{B}$. Let $F = Z \setminus H$. Suppose that $Y \setminus (cl_Y F \cup H) \neq \emptyset$. Fix $y \in Y \setminus (cl_Y F \cup H)$. Since $cl_Y F$ is a compact set, there exists $A \in \mathcal{B}$ such that $cl_Y F \subseteq \omega(A) \subseteq A \subseteq Y \setminus \{y\}$. Then $A \cup H = B \in \mathcal{B}$, $Z \subseteq \omega(B) \subseteq B \subseteq Y \setminus \{y\}$ and $cl_Y Z \subseteq B$, a contradiction. Therefore $H \subseteq Y \setminus cl_Y F$. The assertion 1 is proved.

Let $F$ be a closed subset of $Z$ and $y \in cl_Y F \setminus F$. Since $y \in Y \setminus Z$, the set $\{y\}$ is closed in $Y$. Let $\{H_\alpha : \alpha \in A\} = \{H \in \mathcal{B} : y \notin H\}$. For every $\alpha, \beta \in A$ there exists $\lambda = \lambda(\alpha, \beta) \in A$ such that $H_\lambda = H_\alpha \cup \lambda \beta$. Let $F_\alpha = F \setminus H_\alpha$ for any $\alpha \in A$. We affirm that $F_\alpha \neq \emptyset$. If $F_\alpha = \emptyset$, then $F \subseteq Z \cap H_\alpha \subseteq H_\alpha$ and $cl_Y F \subseteq H_\alpha$, a contradiction. Thus $F_\alpha \neq \emptyset$ for any $\alpha \in A$. We may consider that $H_\alpha = \emptyset$ and $F_\alpha = F$ for some $\alpha \in A$.

Suppose that $\alpha \in A$ and $y \notin cl_Y F_\alpha$. Then, for some $\beta \in \mathcal{B}$, we have $cl_Y F_\beta \subseteq \omega(H_\beta) \subseteq H_\beta \subseteq Y \setminus \{y\}$. If $\lambda = \lambda(\alpha, \beta) \in A$, then $F \subseteq H_\lambda \cap Z$ and $cl_Y F \subseteq H_\lambda \subseteq Y \setminus \{y\}$, a contradiction. Thus $y \notin cl_Y F_\alpha$ for any $\alpha \in A$. If $x \in Y \setminus \{y\}$, then there exists $\alpha \in A$ such that $x \in \omega(H_\alpha) \subseteq H_\alpha$ and $x \notin cl_Y F_\alpha \subseteq Y \setminus \omega(H_\alpha)$. Therefore $\cap\{cl_Y F_\alpha : \alpha \in A\} = \{y\}$. Denote by $\xi$ the ultrafilter of closed subsets of $Z$ for which $\{F_\alpha : \alpha \in A\} \subseteq \xi$. Since $\emptyset \neq \cap\{cl_Y H : H \in \xi\} \subseteq \cap\{cl_Y F_\alpha : \alpha \in A\} = \{y\}$, the proof is complete.

10.7. Theorem. (First intrinsic characteristic, P. C. Osmatescu [25] for $T_1$-compactifications). Let $(Y, f)$ be a g-compactification of a space $X$, $f : X \to f(X)$ be a closed mapping, $z(f(X)) = z(Y)$ and the set $f^{-1}(y)$ be compact for any $y \in z(f(X))$. The following assertions are equivalent:

1. $(Y, f)$ is an $\omega$-g-compactification of the space $X$;
2. if $Z = f(X)$, then there exists a $Z$-clopen base $\mathcal{B}$ of the space $Y$ such that $A \cup B \in \mathcal{B}$ for any $A, B \in \mathcal{B}$.

Proof. The implication 1 $\Rightarrow$ 2 follows from Theorem 10.5.

Let $B$ be a $Z$-clopen base of $Y$ and $A \cup B \in \mathcal{B}$ for all $A, B \in \mathcal{B}$.

For any $x \in X$ we put $\varphi(x) = f(x)$.

Let $\xi \in \omega X \setminus X$. Then $f(\xi)$ is an ultrafilter of closed subsets of $Z$. We put $\varphi(\xi) = \cap\{cl_Y f(H) : H \in \xi\}$. By construction, $\varphi(\xi)$ is a closed compact subset of $Y$. Suppose that $y_1 \in \varphi(\xi)$ and $y_2 \in Y$. If $Z \cap \varphi(\xi) \neq \emptyset$, then $|\varphi(\xi) \cap Z| \leq 1$. Hence, we may suppose that $y_2 \in Y \setminus Z$. Since the set $\{y_2\}$ is closed in $Y$, then there exist $H \in \mathcal{B}$ and an open subset $U$ of $Y$ such that $y_1 \in U \subseteq H \subseteq Y \setminus \{y_2\}$. From Proposition 10.6 it follows that $y_2 \in cl_Y (Z \setminus H) \subseteq Y \setminus U \subseteq Y \setminus \{y_1\}$. Moreover, if $y_2 \in cl_Y F$ for every $F \in f(\xi)$, then $F \setminus H \neq \emptyset$ for every $F \in f(\xi)$. In this case $F \setminus H \in f(\xi)$ and $y_1 \notin \varphi(\xi)$, a contradiction. Thus $\varphi(\xi)$ is a singleton set for any $\xi \in \omega X \setminus X$. Hence $\varphi : \omega X \to Y$ is a single-valued mapping.

By construction, $\varphi(cl_{\omega X} F) \subseteq cl_Y f(F)$ for any closed subset $F$ of $X$.

Let $F$ be a closed subset of $X$ and $y \in cl_Y f(F)$. If $y \in f(F)$, then $y \in \varphi(cl_{\omega X} F)$. Suppose that $y \in cl_Y f(F) \setminus Z$. Then there exists an ultrafilter $\eta$ of closed subsets of $Z$ such that $f(F) \in \eta$ and $\{y\} = \cap\{cl_Y H : H \in \eta\}$.
Since \( f \) is a closed mapping, there exists an ultrafilter \( \xi \) of closed subsets of \( X \) such that \( \eta = f(\xi) \) and \( F \in \xi \). Then \( \varphi(\xi) = y \) and \( \xi \in cl_{\omega X} F \). Hence \( \varphi(cl_{\omega X} F) = cl_Y f(F) \) for every closed subset \( F \) of \( X \).

Let \( x \in \omega X, y = \varphi(x), W \) be an open subset of \( Y \) and \( y \in W \). There exist \( H \in \mathcal{B} \) and an open subset \( V \) of \( Y \) such that \( H \subseteq Y = V \subseteq \omega X \) and \( y \in V \subseteq H \subseteq W \). Let \( U = f^{-1}(V) \) and \( U' = \omega X \setminus cl_{\omega X} (X \setminus U) \). If \( x \notin U' \), then \( x \in cl_{\omega X} (X \setminus U) \) and \( y \notin cl_Y (Z \setminus V) \subseteq Y \setminus V \), a contradiction. Thus \( x \in U' \). By construction, \( \varphi(U' \cap X) = f(U' \cap X) = V \setminus Z \subseteq V \). Let \( x \in (\omega X \setminus X) \cap U' \). Then \( x \) is an ultrafilter of closed subsets of \( X \), \( \eta = f(x) \) is an ultrafilter of closed subsets of \( Y \) and \( y' = \cap \{ cl_H : H \in \eta \} \). By construction, \( Z \setminus V \notin \eta \) and \( y' \notin (Z \setminus V) \). Since \( Y = H \cup cl_Y (Z \setminus V) \), then \( y' \in H \). Thus \( \varphi(U') \subseteq H \subseteq W \). Hence \( \varphi \) is a continuous mapping. In this case \( \varphi^{-1}(y) \) is a non-empty closed compact subset of \( \omega X \) for any \( y \in c(Y) \). If \( y \in z(Z) = z(Y) \), then \( \varphi^{-1}(y) = f^{-1}(y) \) is a compact subset of \( X \).

Let \( F \) be a closed subset of \( \omega X \) and \( \{ H_\alpha : \alpha \in A \} = \{ H \subseteq X : F \subseteq cl_{\omega X} H, F \) is a closed subset of \( X \} \). For every \( \alpha, \beta \in A \) we have \( H_\alpha \subseteq H_\alpha \cap H_\beta \) for some \( \lambda \in A \). Since \( \varphi \) is a compact mapping and \( \varphi(cl_{\omega X} H_\alpha) \) is a closed subset of \( Y \), then \( \varphi(F) = \cap \{ \varphi(cl_{\omega X} H_\alpha) : \alpha \in A \} \) is a closed subset of \( Y \). Thus \( \varphi \) is a perfect mapping. The implication 2 \( \rightarrow 1 \) and the theorem are proved.

10.8. **Theorem.** Let \( (Y, f) \) be an \( \omega_\alpha \)-\( g \)-compactification of a space \( X \) and \( Z = f(X) \). There exists the perfect mappings \( \varphi : \omega X \rightarrow Y \), \( \psi : \omega Z \rightarrow Y \) and \( h : \omega X \rightarrow \omega Z \) such that:

1. \( f = h \mid X = \varphi \mid X \);
2. \( \varphi = \psi \circ h \);
3. if \( g = \psi \mid Z \), then \( (Y, g) \) is an \( \omega_\alpha \)-\( g \)-compactification of the space \( Z \);
4. \((\omega Z, f)\) is an \( \omega_\alpha \)-\( g \)-compactification of the space \( X \).

**Proof.** Since \( (Y, f) \) is an \( \omega_\alpha \)-\( g \)-compactification, there exists a perfect mapping \( \varphi : \omega X \rightarrow Y \) such that \( f = \varphi \mid X \).

The family \( \mathcal{B} = \{ \omega Z \setminus cl_{\omega Z} (Z \setminus U) : U \) is an open subset of \( Z \} \) is a Z-clopen base of the space \( \omega Z \). By virtue of Theorem 10.7, there exists a perfect mapping \( h : \omega X \rightarrow \omega Z \) such that \( f = h \mid X \). The assertion 4 is proved. Now we put \( \psi(z) = \varphi(h^{-1}(z)) \) for any \( z \in \omega Z \). The mappings \( \varphi, \psi, h \) are constructed. The proof is complete.

10.9. **Definition.** A \( g \)-compactification \((Y, f)\) of a space \( X \) is a \( g \)-compactification with the properties \( \Omega \alpha \) if:

\( \Omega \alpha_1. \) the mapping \( f : X \rightarrow f(X) \) is closed and \( f^{-1}(y) \) is a compact set for any \( y \in z(f(X)) \);

\( \Omega \alpha_2. \) if \( C, D \) are closed subsets of \( Y \) and \( C \cap D = \emptyset \), then there exists a closed subset \( F \) of \( f(X) \) such that \( C \subseteq cl_Y F \subseteq Y \setminus D \) and \( C \cap cl_Y H = \emptyset \) provided \( H \) is a closed subset of \( f(X) \) and \( F \cap H = \emptyset \);
\(\Omega\alpha3.\) if \(y \in Y \setminus f(X),\) \(F\) is a closed subset of \(f(X)\) and \(y \in \text{cl}_Y F,\) then there exists a filter \(\eta\) of closed subsets of \(f(X)\) such that \(F \in \gamma\) and \(\{y\} = \cap \{\text{cl}_Y H : H \in \eta\};\)

\(\Omega\alpha4.\) the family \(\{\text{cl}_Y A : A \subseteq f(X)\}\) is a closed base of the space \(Y.\)

**10.10. Lemma.** If \((Y, f)\) is a \(g\)-compactification of a space \(X\) with Property \(\Omega\alpha3,\) then \(z(Y) = z(f(X)).\)

**Proof.** One easily can verify that \(Y \setminus f(X) \subseteq c(Y),\) i.e. the set \(\{y\}\) is closed in \(Y\) for every \(y \in Y \setminus f(X)\). Let \(y \in f(X),\) the set \(\{y\}\) be closed in \(f(X)\) and \(y' \in \text{cl}_Y \{y\}.\) By assumptions, \(f(X) \cap \text{cl}_Y \{y\} = \{y\}.\) If \(F = f(X) \cap \text{cl}_Y \{y\} = \{y\}, y' \neq y\) and \(y' \in \text{cl}_Y F,\) then \(y' \in f(X) \setminus X\) and there exists a filter \(\eta\) of closed subsets of \(f(X)\) such that \(F \in \eta,\) \(\eta,\) and \(\cap \{\text{cl}_Y H : H \in \eta\} = \{y'\}.\) For some \(H \in \eta\) we have \(y \notin H.\) Thus \(\varnothing = F \cap H \in \eta,\) a contradiction. The proof is complete.

**10.11. Lemma.** If \((Y, f)\) is a \(T_1\)-\(g\)-compactification of a space \(X\) with Property \(\Omega\alpha2,\) then \((Y, f)\) is a \(g\)-compactification with Property \(\Omega\alpha4.\)

**Proof.** Let \(C\) be a closed subset of \(Y\) and \(y \notin C.\) Since the set \(D = \{y\}\) is closed in \(Y,\) then there exists a closed subset \(F\) of \(f(X)\) such that \(C \subseteq \text{cl}_Y F \subseteq Y \setminus \{y\}.\) The proof is complete.

**10.12. Theorem.** (Second intrinsic characteristic). For a \(g\)-compactification \((Y, f)\) of a space \(X\) the following assertions are equivalent:

1. \((Y, f)\) is an \(\omega\alpha-g\)-compactification of \(X;\)
2. \((Y, f)\) is a \(g\)-compactification of \(X\) with Properties \(\Omega\alpha.\)

**Proof.** Let \(Z = f(X), \varphi : \omega X \rightarrow Y\) be a perfect mapping of \(X\) onto \(Y\) and \(f = \varphi|X : X \rightarrow Y\) be a closed mapping for which the set \(f^{-1}(y)\) is compact provided \(y \in z(Z).\)

Let \(C\) and \(D\) be two closed subsets of \(Y\) and \(C \cap D = \varnothing.\) There exists a closed subset \(F_1\) of \(X\) such that \(\varphi^{-1}(C) \subseteq \text{cl}_{\omega X} F_1 \subseteq \omega X \setminus \varphi^{-1}(D).\) Let \(F = \varphi(F_1).\) If \(H\) is closed in \(Z\) and \(H \cap F = \varnothing,\) then \(C \cap \text{cl}_Y H = \varnothing.\)

Thus \((Y, \alpha)\) is a \(g\)-compactification with Property \(\Omega\alpha2.\) From Theorem 8.5 it follows that \((Y, f)\) is a \(g\)-compactification with Property \(\Omega\alpha3.\) It is obvious that \(\{\text{cl}_Y A : A \subseteq Z\}\) is a closed base of \(Y.\) The implication \(1 \rightarrow 2\) is proved.

Let \((Y, f)\) be a \(g\)-compactification with Properties \(\Omega\alpha.\)

If \(x \in X\) we put \(\varphi(x) = f(x).\)

If \(x \in X\) we put \(\varphi(x) = f(x).\)

If \(x \in X\) we put \(\varphi(x) = f(x).\)

If \(x \in X\) we put \(\varphi(x) = f(x).\)

\(\Omega\alpha3.\) it follows that \(\varphi(\text{cl}_{\omega X} F) = \text{cl}_Y f(F)\) for any closed subset of \(X.\) Thus \(\varphi^{-1}(\text{cl}_Y \Phi) = \text{cl}_{\omega X} f^{-1}(\Phi)\) for any closed subset \(\Phi\) of \(Z.\) By that fact and Property \(\Omega\alpha4.\) it follows that \(\varphi\) is a continuous
compact mapping. Since \(\{cl_\omega X : A \subseteq X\}\) is a closed base of \(\omega X\), \(\varphi\) is a compact mapping and the set \(\varphi(cl_\omega X)\) is closed in \(Y\), the mapping \(\varphi\) is perfect. The implication \(2 \rightarrow 1\) and Theorem are proved.

10.13. Definition. A \(g\)-compactification \((Y, f)\) of a space \(X\) is a \(g\)-

compactification of \(X\) with Properties \(\omega\) if:

\(\omega\alpha\). the mapping \(f : X \rightarrow f(X)\) is closed and the set \(f^{-1}(y)\) is compact

for any \(y \in z(f(X))\);

\(\omega\beta\). if \(C\) and \(D\) are closed subsets of \(f(X)\) with \(cl_Y C \cap cl_Y D = \emptyset\), then

there exists a closed subset \(B\) of \(f(X)\) such that \(C \subseteq B\), \(cl_Y B \cap cl_Y D = \emptyset\)

and \(cl_Y F \cap cl_Y C = \emptyset\) provided the set \(F\) is closed in \(f(X)\) and \(F \cap B = \emptyset\);

\(\omega\gamma\). if \(F\) is a closed subset of \(f(X)\) and \(y \in F \cap (Y \setminus f(X))\), then there

exists a filter \(\eta\) of closed subsets of \(f(X)\) such that \(\{y\} = \cap\{cl_H H : H \in \eta\}\);

\(\omega\delta\). \(Y \setminus f(X) \subseteq c(Y)\).

10.14. Theorem. (Third intrinsic characteristic). For a \(g\)-compactification

\((Y, f)\) of a space \(X\) the following assertions are equivalent:

1. \((Y, f)\) is an \(\omega\alpha\)-\(g\)-compactification of \(X\);

2. \((Y, f)\) is a \(g\)-compactification with Properties \(\omega\).

Proof. The implication \(1 \rightarrow 2\) is obvious. From Properties \(\omega\beta\) and \(\omega\gamma\) follow the Properties \(\omega\beta\) and \(\omega\delta\). Theorem 10.2 complete the proof.

10.15. Corollary. Let \(\mathcal{L}\) be a ring of closed subsets of a space \(X\). The

following assertions are equivalent:

1. \((\omega_\mathcal{L} X, \omega_\mathcal{L})\) is an \(\omega\alpha\)-\(g\)-compactification of \(X\).

2. \(\mathcal{L}\) is a ring with the following properties:

2.1. the mapping \(\omega_\mathcal{L} : X \rightarrow \omega_\mathcal{L}(X)\) is continuous, closed and the set \(\omega_\mathcal{L}^{-1}(y)\)

is compact for any \(y \in z(Z)\), where \(Z = f(X)\);

2.2. if \(A, B \in \mathcal{L}\) and \(A \cap B = \emptyset\), then there exists \(F \in \mathcal{L}\) such that

\(A \subseteq F \subseteq X \setminus B\) and if \(H\) is a closed subset of \(\omega_\mathcal{L}(X)\) such that \(F \cap \omega_\mathcal{L}^{-1}(H) = \emptyset\),

then \(cl_Y \omega_\mathcal{L}(A) \cap cl_Y H = \emptyset\), where \(Y = \omega_\mathcal{L} X\);

2.3. if \(F\) is a closed subset of \(Z\) and \(y \in cl_Y F \setminus Z\), then there exists a filter

\(\eta\) of closed subsets of \(Z\) such that \(F \in \eta\) and \(\cap\{cl_H H : H \in \eta\} = \{y\}\).

Proof. Implication \(1 \rightarrow 2\) follows from Theorem 10.14. It is obvious that

\((\omega_\mathcal{L} X, \omega_\mathcal{L})\) is a \(g\)-compactification with Properties \(\omega\alpha\), \(\omega\beta\), \(\omega\gamma\) and \(\omega\delta\).

Let \(C\) and \(D\) be closed subsets of \(Y\) and \(C \cap D = \emptyset\). From Lemma 2.1 it

follows that there exist \(C', D' \in \mathcal{L}\) such that \(C' \cap D' = \emptyset\), \(C \subseteq cl_Y \omega_\mathcal{L}(C')\), \(D \subseteq cl_Y \omega_\mathcal{L}(D')\). From Condition 2.2 there exists a closed subset \(F\) of \(Z\) such that

\(C \subseteq \omega_\mathcal{L}(C') \subseteq F \subseteq Z \setminus D'\) and \(cl_Y \omega_\mathcal{L}(C') \cap cl_Y H = \emptyset\) provided \(H\) is a closed

subset of \(Z\) and \(H \cap F = \emptyset\). Thus \((\omega_\mathcal{L} X, \omega_\mathcal{L})\) is a \(g\)-compactification with

Properties \(\omega\). The proof is complete.
10.16. Example. Let $X$ be a non-compact space $a \in X$, $b \in \omega X \setminus X$ and $C \notin \omega X$. Suppose that $\{a\}$ is a closed subset of $X$. We put $Y = Z = \omega X \cup \{c\}$. On $Y$ and $Z$ we consider the following topologies:

- $\omega X$ is an open subspace of the spaces $Y$ and $Z$;
- $\{\{c\} \cup U : U$ is open in $\omega X$ and $a \in U\}$ is the base of the space $Y$ at a point $c$;
- $\{\{c\} \cup (U \setminus \{b\}) : U$ is open in $\omega X$ and $b \in U\}$ is the base of the space $Z$ at a point $c$.

Thus $Y$ and $Z$ are compactifications of the space $X$. It is obvious that:

- $Y$ is a compactification with Properties $\Omega \alpha_1, \Omega \alpha_4, \omega \alpha_1, \omega \alpha_2, \omega \alpha_4$ and without Properties $\Omega \alpha_3, \omega \alpha_3$;
- $Z$ is a compactification with Properties $\Omega \alpha_1, \omega \alpha_1, \omega \alpha_2$ and without other properties;
- the compactifications $\omega X, Y, Z$ are equivalent.

11. LOCALLY COMPACT SPACES

A set $B$ of a space $X$ is compact-small if every closed subset $F \subseteq B$ of the space $X$ is compact (see [19]).

A space $X$ is called:

- locally compact if for every point $x \in X$ there exists an open subset $U$ of $X$ such that $x \in U$ and the set $cl_X U$ is compact;
- weakly locally compact if for every point $x \in X$ there exists an open subset $U$ and a compact subset $F$ such that $x \in U \subseteq F$;
- locally compact-small if for every point $x \in X$ there exists an open compact-small subset $U$ of $X$ such that $x \in U$.

For every non-compact space $X$ it is determined the one-point Alexandroff compactification $aX$, where $aX = X \cup \{a\}$, $a \notin X$, $X$ is an open subspace of $aX$ and $\{aX \setminus F : F$ is a compact closed subset of $X\}$ is a base of the space $aX$ at a point $a$.

For a compact space $X$ we put $aX = X$.

11.1. Theorem. A space $X$ is a locally compact-small space if and only if $aX$ is an $\omega \alpha$-compactification of $X$.

Proof. If $U$ is an open compact-small set, then $U$ is open in $\omega X$. Let $\varphi : \omega X \rightarrow aX$ be the mapping for which $\varphi^{-1}(a) = \omega X \setminus X$ and $\varphi(x) = x$ for any $x \in X$. If $X$ is a locally compact-small space, then $X$ is open in $\omega X$ and $aX$. Thus $\varphi$ is a perfect mapping and $aX$ is an $\omega \alpha$-compactification. If $\varphi$ is continuous, then $X$ is open in $\omega X$. For every point $x \in X$ there exists a closed subset $F$ of $X$ such that $x \notin F$ and $\omega X \setminus X \subseteq cl_{\omega X} F$. Then $X \setminus F$ is an open compact-small subset. The proof is complete.

11.2. Theorem. ([19] for $T_1$-spaces). For a space $X$ the following assertions are equivalent:
1. \( X \) is a locally compact-small space;
2. \( aX \) is a \( \omega \alpha \)-compactification of \( X \);
3. \( X \) is an open subspace of \( \omega X \);
4. \( aX \) is a \( \rho \)-compactification of \( X \);
5. \( X \) is open in some \( \omega \alpha \)-compactification;
6. \( X \) is open in every \( \omega \alpha \)-compactification.

12. PROXIMITIES ON SPACES

Let \( X \) be a set, \( \delta \) be a relation on a family of all subsets of \( X \). We shall write \( \delta(A,B) = 0 \) if the subsets \( A, B \subseteq X \) are \( \delta \)-related, otherwise we shall write \( \delta(A,B) = 1 \).

We put \( \delta(\{x\}, A) = \delta(x,A), \delta(A) = \{x \in X : \delta(x,A) = 0\} \) and \( [A]_{\delta} = \cap\{B \subseteq X : A \subseteq B = \delta(B)\} \). We put \( \delta(x,y) = \delta(\{x\}, \{y\}) \).

12.1. Definition. A relation \( \delta \) on the family of all subsets of a set \( X \) is a proximity on the set \( X \) if \( \delta \) satisfies the following conditions:

A1. \( \delta(x,y) + \delta(y,x) = 0 \) if and only if \( x = y \);
A2. \( \delta(A,B \cup C) = \min\{\delta(A,B), \delta(A,C)\} \);
A3. \( \delta(A \cup B, C) = \min\{\delta(A,C), \delta(B,C)\} \);
A4. \( \delta(\emptyset, X) = \delta(X, \emptyset) = 1 \);
A5. \( \delta(A,B) = \delta(A, [B]_{\delta}) \).

Conditions (A1)–(A5) imply the following properties of proximities:

B1. if \( \delta(A,B) = 0 \), \( A \subseteq A_1 \) and \( B \subseteq B_1 \), then \( \delta(A_1, B_1) = 0 \);
B2. if \( A \cap B \neq \emptyset \), then \( \delta(A,B) = 0 \);
B3. \( \delta(\emptyset, A) = \delta(A, \emptyset) = 1 \) for any \( A \subseteq X \);
B4. \( T(\delta) = \{X \setminus [A]_{\delta} : A \subseteq X\} \) is a \( T_0 \)-topology on the set \( X \).

We say that \( T(\delta) \) is the topology generated by the proximity \( \delta \).

For every proximity \( \delta \) on \( X \) we put \( \hat{\delta}(A,B) = \min\{\delta(A,B), \delta(B,A)\} \). The following properties are obvious:

B5. \( \hat{\delta}(A,B) = \hat{\delta}(B,A) \);
B6. \( \hat{\delta} \) is a proximity on \( X \) if and only if \( T(\hat{\delta}) \) is a \( T_1 \)-topology on \( X \);
B7. if \( \hat{\delta} \) is a proximity on \( X \), then \( T(\hat{\delta}) = T(\delta) \).

12.2. Definition. We say that the proximity \( \delta \) on a set \( X \) is an \( \omega \alpha \)-proximity on \( X \) if

A6. if \( \delta(A,B) = 1 \), then there exists a subset \( C \subseteq X \) such that \( A \subseteq C, \hat{\delta}(A, X \setminus C) = 1 \) and \( \hat{\delta}(X \setminus C \cap [D]_{\delta}, B) = 0 \) provided \( \hat{\delta}(D,B) = 0 \).

If \( \delta_1, \delta_2 \) are proximities on \( X \) and \( \delta_1(A,B) \leq \delta_1(A,B) \) for all subsets \( A, B \) of \( X \), then we put \( \delta_1 \leq \delta_2 \).

12.3. Remark. The notion of the symmetric \( \omega \alpha \)-proximity was introduced by A. V. Arhanghel’skii (see [26]). For a symmetric proximity \( \delta \) Condition A5 follows from Conditions A1, A2, A3, A4 and A6 (see[26], p.194).
12.4. **Theorem.** Let $X$ be a compact space with the topology $T$ and $\delta_X^m(A,B) = 1$ if and only if $A \cap \text{cl}_X B = \emptyset$. Then:
1. $\delta_X^m$ is a $\omega$-proximity on $X$ and $T = T(\delta_X^m)$;
2. if $\delta$ is a proximity on $X$ and $T = T(\delta)$, then $\delta_X^m \leq \delta \leq \delta_X^m$.

**Proof.** It is obvious.

12.5. **Remark.** Let $\delta$ and $\hat{\delta}$ be proximities on $X$. Then $\delta$ is an $\omega$-proximity if and only if $\hat{\delta}$ is an $\omega$-proximity.

12.6. **Theorem.** Let $Y$ be a compactification of a space $X$ and $\delta_Y(A,B) = \delta_Y^m(A,B)$ for any subsets $A, B$ of $X$. Then:
1. $\delta_Y$ is a proximity on $X$ which generated the topology of the space $X$;
2. if $Y$ is an $\omega$-compactification, then $\delta_Y$ is an $\omega$-proximity.

**Proof.** The assertion1 it is obvious.

By construction, $\hat{\delta}_Y(A,B) = 1$ if and only if $\text{cl}_Y A \cap \text{cl}_Y B = \emptyset$. Fix $A, B \subseteq X$. Suppose that $\hat{\delta}(A,B) = 0$. By virtue of Theorem 10.7, there exist an open subset $U$ of $Y$ and an $X$-clopen subset $C$ of $Y$ such that $\text{cl}_Y A \subseteq U \subseteq C \subseteq Y \setminus \text{cl}_Y B$. Let $D$ be a closed subset of $X$ and $\delta(D,B) = 0$. We put $H = D \cap (X \setminus C)$. Fix $y \in \text{cl}_Y D \cap \text{cl}_Y B$. There exists an ultrafilter $\xi$ of closed subsets of $X$ such that $D \in \xi$ and $\{y\} = \cap\{\text{cl}_YE : E \in \xi\}$. By construction, $U \cap X = C \cap X$ is an open subset of $X$ and $\hat{\delta}(A,X \setminus C) = 1$. If $X \setminus C \notin \xi$, then $E \cap (X \setminus C) = \emptyset$ and $y \in \text{cl}_Y E \subseteq C \subseteq Y \setminus \text{cl}_Y B$ for some $E \in \xi$, a contradiction. Thus $X \setminus C \in \xi$, $F = (X \setminus C) \cap D \in \xi$ and $\delta(D \cap (X \setminus C), B) = \delta(F,B) = 0$. The proof is complete.

12.7. **Definition.** The proximities $\delta_1, \delta_2$ on $X$ are called **equivalent** if $\hat{\delta}_1 = \hat{\delta}_2$.

If $\hat{\delta}$ is a proximity on $X$, then the proximities $\delta, \hat{\delta}$ are equivalent.

12.8. **Corollary.** If $Y, Z$ are two distinct $\omega$-compactifications of a space $X$, then the $\omega$-proximities $\delta_Y, \delta_Z$ on $X$ are not equivalent.

12.9. **Theorem.** For every $\omega$-proximity $\delta$ on a space $X$ there exists a unique $\omega$-compactification $Y$ such that $\delta$ and $\delta_Y$ are equivalent.

**Proof.** The uniqueness of the $\omega$-compactification $Y$ follows from Corollary 12.8.

Fix an $\omega$-proximity $\delta$ on a space $X$. Two ultrafilters $\xi, \mu \in \omega X$ are called $\delta$-**equivalent** if $\delta(A,B) = 0$ for all $A \in \xi$ and $B \in \mu$. From Condition A6 it follows that the $\delta$-equivalence of ultrafilters on $X$ is an equivalence relation. Thus, there exist a set $Y$ and a mapping $\varphi : \omega X \rightarrow Y$ of $\omega X$ onto $Y$ such that $\varphi^{-1}(\varphi(x)) = x$ for any $x \in X$ and $\varphi^{-1}(\varphi(\xi)) = \{\mu \in \omega X \setminus X : \xi \text{ and } \mu \text{ are } \delta\text{-equivalent ultrafilters on } X \}$ for every $\xi \in \omega X \setminus X$. Thus $\varphi^{-1}(y)$ is a class of $\delta$-equivalent ultrafilters for any $y \in Y \setminus X$, where $X = \varphi(x)$. If $F$ is a closed subset of $X$, then $c(F) = F \cup \{y \in Y \setminus X : F \in \xi \text{ for some } \xi \in \varphi^{-1}(y)\}$. By construction, $\varphi^{-1}(c(F)) = c_{\omega X}F$. On $Y$ consider the topology generated by the closed base $\{c(F) : F \text{ is a closed subset of } X\}$. Thus $\varphi$ is a continuous
mapping and $\varphi(\text{cl}_{\omega}X F) = c(F)$ for every closed subset $F$ of $X$. If $y \in Y \setminus X$ and $\xi \in \varphi^{-1}(y)$, then $\{y\} = \cap\{c(F) : F \in \xi\}$ is a closed subset of $Y$. Thus the mapping $\varphi$ is compact. Hence $\varphi$ is a perfect mapping and $Y$ is an $\omega\alpha$-compactification of $X$. The proof is complete.

13. PERFECT COMPACTIFICATIONS

13.1. Definition. Let $Y$ be a compactification of a space $X$ and $s_Y(A) = \cup\{V \subseteq Y : V$ is open in $Y$ and $V \cap X \subseteq A\}$. We say that $Y$ is a perfect compactification of $X$ if $\text{cl}_Y(s_Y(U)) \setminus s_Y(U) = \text{cl}_Y(\text{cl}_X U \setminus U)$ for every open subset $U$ of $X$.

A subset $B$ of a space $X$ is connected in $X$ if there do not exist two closed subsets $F$ and $H$ of $X$ such that $F \cap H = \emptyset$, $B \subseteq F \cap H$, $B \cap F \neq \emptyset$, $B \cap H \neq \emptyset$.

13.2. Theorem. Let $Y$ be an $\omega\alpha$-compactification of a space $X$. The following assertions are equivalent:

1. $Y$ is a perfect compactification of $X$;

2. there exists a closed continuous mapping $\varphi : \omega X \rightarrow Y$ such that $\varphi(x) = x$ for every $x \in X$ and the set $\varphi^{-1}(y)$ is connected in $\omega X$ for any $y \in Y$.

Proof. For $T_1$-spaces the theorem is proved in [26]. The case of $T_0$-spaces is similar.

We mention, that the case of completely regular spaces was studied by E. G. Sklearenco [34].

13.3. Remark. If $X$ is a space, then:

1. $\omega X$ is a perfect compactification;

2. if $X$ is a completely regular space, then the Stone-Čech compactification $\beta X$ of $X$ is perfect;

3. if $X$ is a rimcompact space (i.e. the family $\{U : U$ is open in $X$ and $\text{cl}_X U \setminus U$ is compact $\}$ is a base of $X$), then the Freudenthal-Morita compactification $\gamma X$ of $X$ (see[6]) is perfect and $\gamma X \leq Y$ for every perfect $\omega\alpha$-compactification $Y$ of $X$.

14. PUNCTIFORM COMPACTIFICATIONS

14.1. Definition. A compactification $Y$ of a space $X$ is called punctiform if $|B| \leq 1$ for every connected in $Y$ subset $B \subseteq Y \setminus X$.

14.2. Theorem. Let $f : X \rightarrow Y$ be a perfect mapping of a space $X$ onto a space $Y$, $bX$ be a perfect $\omega\alpha$-compactification of $X$ and $bY$ be a punctiform $\omega\alpha$-compactification of $Y$. Then there exists a perfect mapping $\varphi : bX \rightarrow bY$ such that $f = \varphi | X$.

Proof. There exist two perfect mappings $g : \omega X \rightarrow bX$ and $h : \omega Y \rightarrow bY$ such that $g^{-1}(x) = x$ and $h^{-1}(y) = y$ for all $x \in X$ and $y \in Y$. Since $f$ is perfect, there exists a perfect mapping $\psi : \omega X \rightarrow \omega Y$ such that $\psi^{-1}(y) =
The mapping \( \varphi(x) = h(\psi(g^{-1}(x))) \), \( x \in bX \) is the desired one. The proof is complete.

The case of separable metric spaces was examined by Duda, the case of completely regular spaces – by E. G. Skliarenco and the case of \( T_1 \)-spaces – by P. C. Osmatescu (see [26, 34]).

**14.3. Corollary.** If \( bX \) is a perfect \( \omega\alpha \)-compactification of \( X \) and \( cX \) is a punctiform \( \omega\alpha \)-compactification of \( X \), then \( cX \leq bX \).

**References**

Some problems of the theory of compactifications of topological spaces

NEW ITERATIVE CHAIN CLASSES OF PSEUDO-BOOLEAN FUNCTIONS

Vadim Cebotari, Mefodie Rață
Institute of Mathematics and Computer Science of A.S.R.M., Chișinău
cebotari@math.md, ratsa@math.md

Abstract We continue the investigations founded in the book [8]. The basic idea consists of new approach to the problem of the description of closed classes of the functions of simplest non-classical logic. The approach consists of the considerations of the notion of chain classes as the most fundamental. We describe a new infinite sequence of iterative chain classes of pseudo-boolean functions, and other 8 the same classes.

Keywords: iterative pseudo-boolean function, iterative chain class, First Jaśkowski’s Matrix, expressibility, pre-complete system, conserving of predicate (relation).

2000 MSC: 03B20, 03B50, 03B55

E. Post [1, 2, 3] described all closed (with respect to superpositions) classes of boolean functions. Each of these classes possesses a finite basis, and they in all constitutes an algebraic countable lattice. Otherwise is the situation with closed classes of k-valued logic when $k = 3, 4, \ldots$. Namely, Yu. Yanov and A. Muchnik [3] proved that there exist classes of the functions of 3-valued general logic with countable bases, and since they constitute in all a continuous set, there are classes in general without basis.

M. Rață [5, 8] proved that all mentioned new phenomena take place even in the simplest not classical logic. That is the logic of First Jaśkowski’s Matrix [6, 9]. We denote this logic as well as the class of its functions by the letter $J$. The formulas of the $J$ logic are constructed from variables $x, y, z, \ldots$, possibly indexed, by means of logical operations $\&$, $\lor$, $\supset$ and $\neg$, and round brackets. We denote the formulas $(x \& \neg x)$ (constant 0), $(x \supset x)$ (constant 1), $(x \lor \neg x)$ (ternondatia of $x$) and $((x \supset y) \& (y \supset x))$ (equivalence of $x$ and $y$) by symbols 0, 1, $\bot x$ and $x \sim y$, respectively. The logic $J$ is intermediate between classical and intuitionistic logics. It can be defined as the logic of 3-element pseudo-boolean algebra $< \{0, \tau, 1\}; \& , \lor , \supset , \neg >$ by means of the relations: $0 < \tau < 1$, $x \& y = \min(x, y)$, $x \lor y = \max(x, y)$, $x \supset y = 1$ if $x \leq y$, and $x \supset y = y$ if $x > y$, and $\neg x = x \supset 0$.

Assertion 1 [8, 9]. Any function of 3-valued general logic belongs to the $J$ logic if and only if it can be expressed by means of some formula.
Following A. Kuznetsov [7], we say that the function \( f(x_1, \ldots, x_n) \) conserves the predicate (relation) \( R(x_1, \ldots, x_m) \) on the set \( M \) if, for every elements \( \alpha_{ij} \in M \quad (i = 1, \ldots, m; j = 1, \ldots, n) \), the truth expressions

\[
R(\alpha_{11}, \ldots, \alpha_{m1}), \ldots, R(\alpha_{1n}, \ldots, \alpha_{mn})
\]

imply the fact that the expression

\[
R(f(\alpha_{11}, \ldots, \alpha_{1n}), \ldots, f(\alpha_{m1}, \ldots, \alpha_{mn}))
\]

is also true. Often it is convenient to use, instead of predicate \( R \), the corresponding matrix

\[
(\alpha_{ij}) \quad (i = 1, \ldots, m; j = 1, \ldots, l),
\]

such that \( R \) is true on m-element set \( \langle \alpha_{1j}, \ldots, \alpha_{mj} \rangle \) if and only if the last one constitutes some column of matrix.

**Assertion 2** [8, 9]. Any function of 3-valued general logic belongs to the logic \( J \) if and only if it conserves both predicates \( x \neq \tau \) and \( \neg x = \neg y \).

We say that, the formula \( F \) is expressible in the logic \( L \) by means of a \( \Sigma \) system of formulas when \( F \) may be obtained from the formulas belonging to \( \Sigma \) and from variables by means of a finite number of the applications of weak substitution rule and replacement by equivalent in \( L \). A system of \( \Sigma \) formulas (in the language of logic \( L \)) is called a complete in \( L \) if all the formulas are expressible in \( L \) by means of \( \Sigma \). The system \( \Sigma \) of functions of class \( K \) is said to be pre-complete in \( K \) if, for every function \( f \in K \setminus \Sigma \), the system \( \Sigma \cup \{ f \} \) is complete in \( K \).

It is known [8, 9] that the class \( J \) contains 10 pre-complete classes denoted by the symbols \( J_0, J_1, \ldots, J_9 \) and defined as the subclasses of functions of \( J \) conserving, respectively, the predicates: \( x = 0, x = 1, x \leq y \) and \( (x \sim y) = (z \sim u) \) on \( \{0, 1\} \), \( (x \neq \tau) \& ((x \& y) \neq \tau), (x \neq \tau) \& ((x \lor y) \neq \tau) \), \( (x \leq y) \& (\neg x = \neg y), ((x \lor y) \neq \tau) \& (\neg x = \neg y) \) and \( (((x \sim y) \& \neg \neg y) = z) \& (\neg x = \neg y = \neg z) \) on \( \{0, \tau, 1\} \).

We say that, a system \( \Sigma \) of functions of class \( K \) is a chain in \( K \) if all closed classes of \( K \) to which \( \Sigma \) is included, are comparable with respect to inclusion (i.e. they constitute a chain). It is clear that every of classes \( J_0, J_1, \ldots, J_9 \) is a chain in \( J \). In the monograph [8] all chain classes containing the function \( (x \& \neg x) \) (constant 0) are described .

Let us introduce in analysis, for any \( \mu = 1, 2, \ldots \), the predicates

\[
R^1_1, R^2_1, \ldots, R^\mu_1, \ldots
\]

(1)

over the set \( \{0, \tau, 1\} \), where

\[
R^\mu(x_1, \ldots, x_n) == (x_1 \neq \tau) \& \ldots \& (x_\mu \neq \tau) \& ((x_1 \lor \cdots \lor x_\mu) = 1).
\]
Remark that $R_1^1$ defines the class $J_1$ not containing the constant 0. It is necessary to introduce the following 8 matrices:

\[
M_{1.1} = (\tau), M_{1.3} = \begin{pmatrix} 0 & 1 \\ \tau & 1 \end{pmatrix}, M_{1.4} = \begin{pmatrix} 0 & 1 & 1 \\ \tau & \tau & 1 \end{pmatrix},
\]

\[
M_{1.5} = \begin{pmatrix} 0 & 0 & 1 \\ \tau & 1 & 1 \end{pmatrix}, M_{1.6} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & \tau & 1 \end{pmatrix}, M_{1.7} = \begin{pmatrix} \tau & \tau & 1 \\ \tau & 1 & 1 \end{pmatrix},
\]

\[
M_{1.8} = \begin{pmatrix} \tau & 1 & 1 \\ 1 & \tau & 1 \end{pmatrix}, M_{1.9} = \begin{pmatrix} \tau & \tau & 1 \\ \tau & 1 & 1 \\ 1 & \tau & 1 \end{pmatrix}.
\]

We denote the classes of pseudo-boolean functions conserving the introduced predicates of line (1.1) or matrix $M_{1.1}$ by symbols $J_1^1$, $J_1^2$, ..., $J_1^9$, respectively.

**Theorem 1** (criterion of completeness in $J_1$). **In order that a system $\Sigma$ of functions of $J_1$ be complete in $J_1$ it is necessary and sufficient that $\Sigma$ be not included in any of the classes $J_0 \cap J_1$, $J_3 \cap J_1$, $J_4 \cap J_1$, $J_{1.1}$, $J_{1.3}$, $J_{1.4}$, ..., $J_{1.9}$.**

(2)

**Consequence 1.** The classes of line (2) and only they are pre-complete in the class $J_1$.

**Consequence 2.** The classes $J_1.1$, $J_1.3$, $J_1.4$, ..., $J_1.9$ and only they are iterative chains in the class $J_1$.

Let us consider the function

\[ h_\mu^\ast = \land_{i=1}^{\mu+1} (x_1 \lor \cdots \lor x_{i-1} \lor x_{i+1} \lor \cdots \lor x_{\mu+1}), \quad \mu = 1, 2, \ldots, \]

and its particular cases

\[ h_1^\ast = x_1 \land x_2, \quad h_2^\ast = (x_1 \land x_2) \lor (x_1 \land x_3) \lor (x_2 \land x_3). \]

**Lemma 1.** $h_\mu^\ast \in J_\mu^\mu \setminus J_\mu^{\mu+1}$, $\mu = 1, 2, \ldots$.

This Lemma is dual to Lemma 2.1 of monograph [8].

**Consequence 3.** The following strong inclusions

\[ J_\mu^{\mu+1} \subset J_\mu^\mu, \quad \mu = 1, 2, \ldots \]

hold.

**Theorem 2.** The classes $J_1^1$, $J_1^2$, ... constitute an infinite decreasing sequence of iterative chain classes of pseudo-boolean functions.
References


APPLICATIONS TO LIE ALGEBRAS
Camelia Ciobanu
"Mircea cel Batran" Naval Academy, Constanta

Abstract
In order to find the Jacobson radical of finite-dimensional associative algebra, polynomial time algorithms are described. An efficient algorithm for computing the solvable and nilpotent radicals of Lie algebra is presented.

Keywords: Lie algebra, Jacobson radical

1. INTRODUCTION

Consider some basic algorithmic problems related to finite dimensional associative algebras. Our starting point is the structure theory of these algebras and we touch upon some applications of the associative decomposition algorithms, including efficient algorithms for calculating the radicals of Lie algebras.

2. BASIC DEFINITIONS AND THEOREMS

A linear space \( L \) over the field \( k \) is an algebra over \( k \) if it is equipped with a binary, \( k \)-bilinear operation (called multiplication). Denote by \( xy \) the product of \( x, y \in L \) . Multiplication is assumed to be associative, i.e. \( x(yz) = (xy)z \) for every \( x, y, z \in L \). Throughout we assume that \( \dim_k L = n < \infty \). We say that \( L \) is a commutative algebra if \( xy = yx \) for every \( x, y \in L \). A \( k \)-subspace \( S \) of \( L \) is a subalgebra of \( L \) if \( S \) is closed under multiplication: if \( x, y \in S \) then \( xy \in S \). A \( k \) subspace \( I \) of \( L \) is a left ideal of \( L \) if \( gx \in L \) whenever \( x \in I \) and \( y \in L \). A right ideal is defined similarly. An \( k \)-subspace \( I \) of \( L \) is an ideal of \( L \) if \( I \) is both left and right ideal of \( L \). If \( I \) is an ideal in \( L \), then we can form the factor algebra \( L/I \). The notions of homomorphism and \( L \)-module are used in the standard way. The algebra \( L \) is simple if it has no ideals except (0) and \( L \), and \( LL \neq 0 \), where \( LL \) is the algebra generated by products \( ab \) with \( a, b \in L \). We say that \( L \) is the direct sum of its ideals \( \text{L}_1, ..., \text{L}_s \) (written as \( \bigoplus L_i \)) if \( L \) is the direct sum of these linear subspaces.

Theorem 2.1 (Representation Theorem). Let \( L \) be an algebra over the field \( k \) and suppose that \( \dim_k L = n \). Then \( L \) is isomorphic to a subalgebra of \( M_{n+1}(k) \) - the algebra of all \( n+1 \) by \( n+1 \) matrices over \( k \).
An element \( x \in L \) is called nilpotent if \( x^p = 0 \) for some positive integer \( p \). An element \( x \) is strongly nilpotent if \( xy \) is nilpotent for every \( y \in L \). The Jacobson radical \( \text{Rad}(L) \) of \( L \) is the set of strongly nilpotent elements of \( L \). It is not difficult to see that \( \text{Rad}(L) \) is an ideal of \( L \) and \( L/\text{Rad}(L) \) has no nonzero strongly nilpotent elements. It can be shown that \( \text{Rad}(L) \) is a nilpotent ideal: there exists a positive integer \( p \) such that \( x_1x_2...x_p = 0 \), for all \( x_1, x_2, ..., x_p \in \text{Rad}(L) \). An algebra \( L \) is semisimple if \( |L| \geq 2 \) and \( \text{Rad}(L) = (0) \).

A characterization of semisimple algebra is included in the next theorem.

**Theorem 2.2** (Wedderburn’s Theorem). If \( L \) is a finite-dimensional semisimple associative algebra over the field \( k \), then \( L \) is expressible as a direct sum \( L = L_1 \oplus L_2 \oplus ... \oplus L_s \), where the \( L_i \) are exactly the minimal nonzero ideals of \( L \). Moreover, \( L_i \) is isomorphic to matrix algebra \( M_{n_i}(k_i) \) where \( k_i \) is a possibly noncommutative extension field of \( k(1 \leq i \leq s) \).

In the next section we present algorithms in order to compute the Jacobson radical. The interesting case here is when \( k \) (and consequently \( L \)) is finite. We explain some basic methods of finding the structural ingredients of algebras in a computationally efficient way. These methods are applied in the last section, leading to the computation of the (solvable) radical and the nilradical of Lie algebras.

Recall some basic facts about Lie algebras. Detailed exposition can be found in Jacobson [10] and Humphreys [9]. A linear space \( G \) over the field \( k \) is a **Lie algebra**, if \( G \) is equipped with a \( k \)-bilinear binary operation \( [\cdot, \cdot] \) such that \( [x, x] = 0 \) for every \( x \in G \) and \( [[x, y], z] + [[y, z], x] + [[z, x], y] = 0 \) for every \( x, y, z \in G \) (the Jacobi identity). Just like in the associative case, we have the familiar notions of subalgebra, ideal, factor algebra and homomorphism for Lie algebras. The derived series of \( G \) is the collection \( G^{(i)} \) of ideals in \( G \) defined as \( G^{(0)} = G \) and \( G^{(i+1)} = [G^{(i)}, G^{(i)}] \) for \( i > 0 \). A Lie algebra is called **solvable** if the derived series reaches \( (0) \) in finitely many steps: \( G^{(n)} = 0 \) for some natural number \( n \). Here we consider finite dimensional Lie algebra only.

In this case \( G \) has an unique maximal solvable ideal, denoted by \( R(G) \), the radical of \( G \). The **descending central series of** \( G \) is the sequence \( G^j \) of ideals of \( G \), where \( G^0 = G \) and \( G^{i+1} = [G, G^i] \) for \( i \geq 0 \). A Lie algebra \( G \) is nilpotent if \( G^n = (0) \) for some natural number \( n \). If \( \dim_k G < \infty \), then \( G \) has an unique maximal nilpotent ideal \( N(G) \) the **nilradical of** \( G \).

**Example** Let \( L \) be an associative algebra over \( k \). For two elements \( a, b \in L \), the additive commutator is \( [a, b] = ab - ba \). It is easy to check that this operation satisfies the identities of a Lie-bracket. As a consequence, if a \( k \)-subspace \( S \) of \( L \) is closed with respect to the operation \( [\cdot, \cdot] \), then \( S \) can be considered as a Lie algebra. Particularly important are the Lie subalgebras of this form which are obtained from \( G = M_n(k) \). They are called **linear Lie algebras**.
There is a straightforward analogue of the regular representation for a Lie algebra $G$. For an $x \in G$, let $ad(x) : G \to G$ be the linear map that maps $y \in G$ to $[x, y]$. The map $x \to ad(x)$ is a Lie algebra homomorphism from $G$ to linear Lie algebra $gl(G)$ of all linear transformations of the $k$-space $G$. Unfortunately, this map is far from being faithful (if $G$ is simple, then this map is faithful). We just remark here that, according to a deep theorem of Ado and Iwasuwa [10], every finite-dimensional Lie algebra is actually isomorphic to a linear Lie algebra.

We are interested in exact computations, such that $k$ is either a finite field or an algebraic number field. We specify now the input of the algorithmic problems addressed. In order to obtain sufficient general results, we consider an algebra to be given as a collection of structure constants.

If $L$ is an algebra over the field $k$ and $e_1, e_2, ..., e_n$ is a basis of the $k$-space $L$, then multiplication is completely described if we express the product $e_i e_j$ as linear combinations of the basis elements $e_i e_j = \gamma_{ijk} e_1 + ... + \gamma_{ijn} e_n$. The coefficients $\gamma_{ijk} \in k$ are called structure constants. When an algebra is given as input, we assume that it is represented as an array of structure constants. Substructures (such as subalgebras, ideals, subspaces) can then be represented by bases whose elements are linear combinations of basis elements of the ambient structure (algebra).

In our cases $k$ can be viewed as an algebra over its prime field $P$; therefore $k$ can also be represented with structure constants from $P$ (If $k$ is finite, then $P = k_p$ for some prime $p$, if $k$ is number field, then $P = Q$). In these cases $k$ is usually specified by giving the (monic) minimal polynomial $f$ of a single generating element $\alpha$ over the prime field $P$. This is a special case of the representation with structure constants. The coefficients of $f$ give the structure constants with respect to $P$-basis $1, \alpha, \alpha^1, ..., \alpha^{n-1}$ of $k$ where $n = \dim_p k$.

Another important way to represent an algebra is in the form of a matrix algebra. In these cases we are given a collection of matrices which generate the algebra. The algorithms described in this paper are applicable in this setting as well. From such a matrix representation one can efficiently find a basis of the algebra and then calculate structure constants with respect to this basis.

We would like to consider algorithms which have a theoretical guarantee for their efficiency. From the perspective of computer science these are the polynomial time algorithms. An algorithm runs in polynomial time if on inputs of length $n$ the computation requires at most $n^c$ bit operations. Here $c > 0$ is a constant independent of $n$, and $n$ is a positive integer.
3. COMPUTING THE RADICAL

Suppose that \( L \) is a finite-dimensional algebra over the field \( k \), given as a collection of structure constants. Our objective is to find a basis of \( \text{Rad}(L) \), the radical of \( L \), in time polynomial in the input size.

If \( \text{char} \ k = 0 \), then the problem is equivalent to solving a system of linear equations over the ground field as follows from the characterization of the radical by Dickson

**Theorem 3.1** Let \( L \) be a finite dimensional algebra of matrices over a field \( k \), and \( \text{char} \ k = 0 \). Then \( \text{Rad}(L) = \{ x \in L/\text{Tr}(yx) = 0 \text{ for every } y \in L \} \).

In fact, if \( e_1, e_2, \ldots, e_n \) is a linear basis of \( L \) over \( k \), then to find \( \text{Rad}(L) \), it suffices to solve the linear system \( \text{Tr}(e_i x) = 0, i = 1, \ldots, n \), where \( x \) is an ,,unknown'' element of \( L \).

We now turn to the case where \( L \) (and hence \( k = k_q \)) is finite. We assume that \( p \) is a prime, \( q \) is a power of \( p \), \( k = k_q \) and that \( L \) is a subalgebra of \( M_n(k) \). The statement of Dickson’s Theorem is no longer valid in positive characteristic. There is, however, a more subtle, and still useful, description of the radical in this case. We explain this in the sequel. Define the natural characteristic. There is, however, a more subtle, and still useful, description of the radical in this case. We explain this in the sequel. Define the natural characteristic.

**Theorem 3.2.** Let \( L \leq M_n(k) \) be an algebra of matrices over the finite field \( k \) of characteristic \( p \). Put \( l = \lceil \log_p n \rceil \) and let \( R_0, R_1, \ldots, R_{l+1} \) be as defined above. Then:

1. \( R_0, R_1, \ldots, R_{l+1} \) are ideals of \( L \);
2. \( R_{l+1} = \text{Rad}(L) \);
3. for every \( j \in \{0, \ldots, l\} \) the function \( T_j \) is \( p^j \)-semilinear on \( R_j \), i.e. \( T_j(\alpha a + \beta b) = \alpha^{p^j} T_j(a) + \beta^{p^j} T_j(b) \) for every \( \alpha, \beta \in k \) and \( a, b \in R_j \).
Property 3 implies that we can obtain a basis of $R_{j+1}$ from a basis $R_j$ by solving a system of linear equations over $k$. Indeed, set $a_0 = I$, and let $a_1, ..., a_s$ be a basis of $L$ over $k$. Suppose that $\{b_1, b_2, ..., b_r\}$ is a basis of $R_j$ over $k$, and we are looking for a basis of $R_{j+1}$. Semilinearity implies that an element $a \in R_j$, $a = \sum_{i=1}^{r} \lambda_i b_i$ is in $R_{j+1}$ iff $\sum_{i=1}^{r} T_j (a_i b_i) \lambda_i^{p^j} = 0$, $(t = 0, ..., s)$.

The inverse of the automorphism $\lambda \to \lambda^{p^s}$ of the finite field $k = k_q$ can be computed efficiently, hence the above system can be translated into

$$\sum_{i=1}^{r} T_j (a_i b_i) \lambda_i^{p^j} = 0 \ (t = 0, ..., s).$$

The latter is a system of linear equations in the variables $\lambda_1, \lambda_2, ..., \lambda_r$. Thus, we start with $R_0 = L$ and then in turn proceed to compute $R_1, ..., R_{l+1}$.

From a basis of $R_t$ we obtain a basis of $R_{t+1}$ by solving a system of linear equations over $k$. The number of equations and the number of variables is at most $n^2$, hence the system can be solved in time $(n + \log q)^{O(l)}$. We obtain a basis of $\text{Rad}(L)$ in $l + 1 = 0 (\log n)$ such rounds; therefore the overall cost of the computation is $(n + \log q)^{O(l)}$ bit operations. Below we give a formal description of the algorithm.

Radical $(L) :=$

$$A := \{I\} \cup \text{basis of } L;$$

$$B := \text{basis of } L;$$

for $j$ from 1 to $\lfloor \log_p n \rfloor + 1$ do

if $B \neq \emptyset$ then

$$C := \left( T_j (ab)^{\frac{1}{p^j}} \right) a \in A, b \in B$$

$$\Lambda := \text{a basis of } \text{Ker}C :$$

$$B := \{\lambda_1 b_1 + ... + \lambda_r b_r | (\lambda_1, ..., \lambda_r) \in \Lambda\}$$

fi

od

return $B$.

The proof of Theorem 3.2 follows immediately from the next sequence of lemmas. The statement of the first lemma can be considered as a special case of the theorem, where the underlying module is simple.

**Lemma 3.3.** Let $S$ be a simple algebra over the finite field $k$ and $U$ be a simple $S$-module. Then there exists an element $a \in S$ with $\text{Tr}_U(a) = 1$, where $\text{Tr}_U(a) = 1$ stands for the ordinary trace of the action of $a$ on $U$.

Below we show that semilinearity and other useful properties hold for the trace functions $T_j$ on certain ideals.

**Lemma 3.4.** Let $L \leq M_n(k)$ be a matrix algebra over the field $k$ of characteristic $p$. Assume that $L \neq \text{Rad}(L)$. Let $(0) = U_0 < U_1 < ... < U_r = U$ be
a composition series of the $L$-module $U = k^n$. Let $I_1, I_2, \ldots, I_t$ be the minimal elements of the set of ideals of $L$ properly containing $\text{Rad}(L)$. For every index $i \in \{1, 2, \ldots, t\}$ fix a simple $L$-module $V_i$ that belongs to the ideal $I_i$ and denote the multiplicity of $V_i$ in the composition series by $m_i$. Put $l = \lfloor \log_p n \rfloor$ and define the ideals $R'_0, R'_1, \ldots, R'_{l+1}$ as

$$R_j = \text{Rad}(L) + \sum_{p^j/m_i} I_i.$$  

Then:

i) $R'_{l+1} = \text{Rad}(L)$;

ii) for every $j \in \{0, \ldots, l\}$ the function $T_j$ is $p^j$-semilinear on $R'_j$ (in the sense of Theorem 3.2);

iii) $T_j(ab) = T_j(ba)$ for every $j \in \{0, \ldots, l\}, b \in L$ and $a \in R'_j$;

iv) $T_j$ is identically zero on $R'_{j+1}$ ($j = 0, \ldots, l$);

v) $T_j$ is not identically zero on ideals $J_i$ so that the multiplicity $m_i$ is divisible by $p^j$ but not by $p^{j+1}$ ($j = 0, 1, \ldots l$).

Lemma 3.4 provides a tool to inductively check that the subsets $R_j$ coincide with the ideal $R'_j$ defined in that lemma.

\textbf{Lemma 3.5.} Keeping the notation of Lemma 3.4 for each $j \in \{0, \ldots, l\}$ we have $R'_{j+1} = \{a \in R'_j | T_j(ab) = 0$ for every $b \in \{I\} \cup L\}$.

\textbf{Remark 3.6.} The approach presented here is a simplified and specialized version of a result from [4] where arbitrary fields of characteristic $p$ are allowed. In the general case the characterization of the ideals $R_j$ is slightly more complicated than formula $(*)$.

4. COMPUTATIONS IN LIE ALGEBRAS

Just like associative algebras, Lie algebras can be conveniently described by structure constants. As we saw, if $G$ is a Lie algebra over a field $k$ and $e_1, e_2, \ldots, e_n$ is a basis of $G$, then the bracket is described if we have the products $e_i e_j$ as linear combinations of the basis elements: $[e_i, e_j] = \gamma_{ij} e_1 + \ldots + \gamma_{ijn} e_n$, $\gamma_{ij} \in k$ are called structure constants.

Let us now outline algorithms for computing the nilpotent and the solvable radical of a Lie algebra. These problems can be reduced to associative radical computations. First we consider the nilradical. We need a theorem of Jacobson [10].

Let $G$ be a finite dimensional Lie algebra over an arbitrary field $k$. 
Theorem 4.1 (Jacobson’s Theorem). Let \( L \) be the associative (matrix-) algebra generated by the linear transformations \( \text{ad}(x), x \in L \), i.e., the image \( \text{ad}(L) \) of the adjoint representation of \( L \). Then an element \( x \in L \) is in the nilradical \( \mathcal{N}(L) \) if and only if \( \text{ad}(x) \in \text{Rad}(L) \).

This result offers a reasonable way to computing \( \mathcal{N}(L) \) if the ground field \( k \) is a finite field or an algebraic number field. Indeed, we can compute first a basis of \( L \), and then compute \( \text{Rad}(L) \) with the algorithms of the previous section. We compute the intersection of the \( k \)-subspaces \( \text{ad}(L) \) and \( \text{Rad}(L) \) by solving a system of linear equations. By Jacobson’s Theorem, the inverse image in \( L \) of the intersection \( \text{ad}(L) \cap \text{Rad}(L) \) is \( \mathcal{N}(L) \). A formal description of our method reads as follows.

Nilradical \((L) := \)
\[
L := \text{associative algebra generated by } \text{ad}(L) \\
\text{return } \text{ad}^{-1}\text{Rad}(L).
\]

Corollary 4.3. Let \( G \) be a finite-dimensional Lie algebra over the field \( k \), where \( k \) is either a finite field or an algebraic number field. Suppose that \( G \) is given as a collection of structure constants. Then the nilradical \( \mathcal{N}(G) \) can be computed in time polynomial in the input size.

Next we address the problem of computing the solvable radical \( R(G) \). For finite \( k \) the problem of computing \( R(G) \) can be reduced efficiently to the problem of computing \( \mathcal{N}(G) \).

First remark that \( \mathcal{N}(G) \subseteq R(G) \) and if \( \mathcal{N}(G) = (0) \), then \( R(G) = (0) \) because starting with the next to the last element of the derived series of \( R(G) \) they are Abelian, hence nilpotent ideals of \( G \). With these we define the sequence \( G_i \) of Lie algebras as follows: let \( G_0 = G \); if \( \mathcal{N}(G_i) \neq (0) \), then let \( G_{i+1} = G_i/\mathcal{N}(G_i) \); if \( \mathcal{N}(G_i) = (0) \), then \( G_{i+1} \) is not defined. This sequence of Lie algebras has no more than \( \dim G + 1 \) elements. From Corollary 4.3. it follows that all algebras \( G_i \) can be computed in polynomial time over finite \( k \). Let \( G_j \) be the last algebra of the sequence. We then have \( G_j \cong G/\text{R}(G) \). Moreover, we can construct a basis for \( \text{R}(G) \) by keeping track of the preimages of the ideals we factored out during the computation of the sequence \( G_0, G_1, ..., G_j \).

It is instructive to view this computation in terms of ideals of \( G \). For \( i > 0 \) let \( J_i \) denote the kernel of the composition of the natural maps \( G_0 \to G_1 \to ... \to G_i \). We then have \( J_1 \subset J_2 \subset ... \subset J_i, \mathcal{N}(G/J_i) = J_{i+1}/J_i \) for \( 0 < i < j \) and \( J_j = R(G) \). From a basis of \( J_i \) a basis of \( J_{i+1} \) is obtained by a single call of the nilradical - algorithm with the algebra \( G/J_i \) as input. As a result, we obtain elements \( h_1, h_2, ..., h_k \in G \) such that \( h_1 + J_i, ..., h_k + J_i \) form a basis of \( J_{i+1}/J_i \). Now the elements \( h_i \) together with a basis of \( J_i \) will constitute a basis of \( J_{i+1} \). In such \( j \) rounds we obtain a basis of \( R(G) \). Below we give a formal description of the algorithm:

\[ \text{Solvable Radical}(G) := \]
\[ S := (0) : \]
\[ loop \]
\[ S := Nilradical (G/S) ; \]
\[ \phi := \text{natural map } G \rightarrow G/S ; \]
\[ S := \phi^{-1}(S) ; \]
\[ until \ S = (0) ; \]
\[ return S . \]

**Corollary 4.4.** Let \( G \) be a finite-dimensional Lie algebra over \( k_q \), given by structure constants. Then (a basis of) the solvable radical \( R(G) \) can be computed in time polynomial in \( \dim_k G \) and \( \log q \).

Variants of the radical algorithms discussed here are implemented by Wilhelm de Graaf in a general library of Lie algebras algorithm called ELIAS (for Eindhoven Lie Algebra System), which is built into the computer algebra systems GAP4 and MAGMA.

**Acknowledgements**

This paper was supported by the CNCSIS grant 1075/2005

**References**


PERTURBED BIFURCATION IN A DEMAND-OFFER MODEL

Elena Codeci
College 'V. Voda', Curtea de Arges
codeci.jili@yahoo.com

Abstract For a demand-offer model with cubic demand function the universal unfoldings
and the transient manifolds are determined. The sections in the static bifurcation diagram and the universal unfoldings of the manifold $S$ are characterized.
The cases are enumerated where these universal unfoldings cannot be found,
the respective germs having a greater than three or infinite codimension.

The mathematic model associated with the economical demand-offer model
with a cubic demand function is the Cauchy problem for the system of ordinary
differential equations(s.e.d.o.)

$$\begin{align*}
\dot{x} &= \alpha x^3 + cx^2 - y, \\
\dot{y} &= \mu x - y - \gamma,
\end{align*}$$

with $\alpha, c, \mu, \gamma$ - the four parameters and $x, y$ - the state functions. In the
following we use the notation, definitions from [GS] (for perturbed bifurcation)
and [GMO] (for general theory of bifurcation and nonlinear dynamics). The
method of normal form is that presented in [AP] and [Kus].

1. DETERMINATION OF THE EQUILIBRIUM MANIFOLD AND S MANIFOLDS

Proposition 1.1 The bifurcation equation for (1) is

$$\alpha x^3 + cx^2 - \mu x + \gamma = 0.$$
Proposition 1.2 For the equilibrium points $P_1 \left( \frac{9\alpha \gamma + \mu c}{6\mu \alpha + 2c^2}, \mu \left( \frac{9\alpha \gamma + \mu c}{6\mu \alpha + 2c^2} \right) - \gamma \right)$ of s.e.d.o. (1), the linearized s.e.d.o. has at the origin a saddle-node bifurcation, and the $S$ manifold is

$$S = \left\{ (\alpha, c, \mu, \gamma) \mid 4c^3 \gamma - \mu^2c^2 + 18\alpha \mu \gamma c + 27\alpha^2 \gamma^2 - 4\alpha \mu^3 = 0 \right\}.$$ (3)

The normal form for the linearized s.e.d.o. around the origin, in the phase space, for a parameter $(\alpha, c, \mu, \gamma) \in S$ fixed, is

$$\begin{pmatrix} \dot{n}_2 \\ \dot{n}_1 \end{pmatrix} = \begin{pmatrix} \mu - 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} n_2 \\ n_1 \end{pmatrix} + \begin{pmatrix} -2a_1 \mu n_1 n_2 \\ -\mu a_1 n_1^2 \end{pmatrix} + O(n^3),$$ (4)

with $a_1 = \frac{3\alpha x_0 + c}{\mu - 1}$. The graphical representation of $S$, for a fixed value of the parameter $\mu$, is drawn in fig.2a.

Expression (1.2) follows by eliminating $x_0$ between (1.1) and the condition that the equilibrium point $(x_0, y_0)$ has a null eigenvalue.

Fig.2b, reveals the presence of a saddle-node.
2. PERTURBED BIFURCATION IN THE STATIC BIFURCATION DIAGRAM

Case $\alpha \neq 0$.

**Proposition 2.1** The static bifurcation diagram is an equation $G(x) = 0$, where the $G$ function is

$$G(x, a, b, d) = x^3 + ax^2 + bx + d. \tag{5}$$

The germ $G$, for $x$ selected as a state function, $b$ as a control parameter, $a$ and $d$ as small parameters, is the universal unfolding for the fork at the origin.

**Proposition 2.2** For $d$ selected as control parameter, $a$ and $b$ as small parameters, then $G$ in (5) is a two parameter unfolding of a hysteresis point situated at the origin.

Let $E_{x,\lambda}$ be the set of germs in $x$ and $\lambda$ variables.

**Proposition 2.3** The germ $g_1(x, a) = x^3 + ax^2$ has an infinite codimension, and, generally, any germ $g(x, \lambda) = x^2q(x, \lambda)$, with $q(x, \lambda) \in E_{x,\lambda}$ has an infinite codimension.

Case $\alpha = 0$, $c \neq 0$. For this case, the unfolding of the static bifurcation diagram for the s.e.d.o. (1) is

$$G(x, \gamma, \mu, c) = cx^2 - \mu x + \gamma, \text{ i.e. } \dot{x} = cx^2 - \mu x + \gamma. \tag{6}$$

**Proposition 2.4** The unfolding (6) factorizes with the universal unfolding of the limit point.

2.1. DETERMINATION OF THE UNIVERSAL UNFOLDINGS FOR $S$, CORRESPONDING TO THE NON-HYPERBOLIC EQUILIBRIA $P_1$

Case $\gamma \neq 0$.

**Proposition 2.5** For the bifurcation equation (3), assuming $c$ as a state function, $\alpha$ as a control parameter and $\gamma \neq 0$ as fixed, then the origin $(c, \alpha) = (0, 0)$ is a bifurcation winged-cusp point, if $\mu = 0$. In addition, the function from (3) represents a map strongly equivalent with the normal form of the winged cusp.
In fig. 3 we give sections in the transient manifolds of the universal unfoldings $H(x, \lambda, \alpha, \beta, \gamma) = x^3 + \lambda^2 + \alpha + \beta \lambda + \gamma x \lambda$ of the winged cusp, with plane $\gamma = \text{const.}$, together with all the persistent bifurcations.

Fig. 3 Sections in the transient manifolds of the universal unfoldings $H(x, \lambda, \alpha, \beta, \gamma) = x^3 + \lambda^2 + \alpha + \beta \lambda + \gamma x \lambda$ of the winged cusp, with planes $\gamma = \text{const.}$ and the persistent unfoldings.

Proposition 2.6 For the bifurcation equation (3), taking $\mu = 0$, $c$ as state function, $\gamma \neq 0$ as a control parameter, i.e. for the bifurcation problem $G_2(c, \alpha, \gamma) = G_1(c, \alpha, \gamma, 0)/\gamma = 0$, it follows that the origin $(c, \gamma) = (0, 0)$ is a bifurcation hysteresis point.

Proposition 2.7 For the bifurcation equation (3), with $\mu = 0$, choosing $c$ as a state function, $\gamma$ as control parameter, and $\alpha \neq 0$ as fixed, then $G_1(c, \alpha, \gamma, 0)$ is an infinite codimension bifurcation germ.

case $\gamma = 0$. The bifurcation problem (3) becomes

$$G_1(c, \alpha, 0, \mu) = -4\alpha \mu^3 - \mu^2 c^2 = 0.$$  (7)
Proposition 2.8 In (7), for \( \mu \) considered as a state function, \( \alpha \neq 0 \) as fixed, \( c \) as a control parameter, we have that at the origin an infinite codimension bifurcation occurs.

Proposition 2.9 The bifurcation problem (7), for \( \mu \) taken as a state function, \( \alpha \neq 0 \) as fixed, \( c \) as a control parameter, has at the origin a fork bifurcation.

Proposition 2.10 In (7), for \( c \) taken as a state function, \( \mu \) as a control parameter, \( \alpha \neq 0 \) as fixed, then \( G_1(c, \alpha, 0, \mu) \) has an infinite codimension.

Proposition 2.11 For (7) with \( \mu \neq 0 \) fixed, \( c \) chosen as a state function, \( \alpha \) as a control parameter, we have at the origin \( (c, \alpha) = (0, 0) \) a zero codimension limit point.

Case \( \alpha \neq 0 \) fixed.

Proposition 2.12 The bifurcation problem (3), with \( \mu \) chosen as a state function, \( \gamma \) as a control parameter, \( \alpha \neq 0 \) as fixed, has at the origin \( (\mu, \gamma) = (0, 0) \), a cusp bifurcation, for \( c = 0 \).

Indeed for \( c = 0 \), from (3) we obtain the unfolding

\[ G_1(\mu, \gamma, \alpha, 0) = -4\alpha\mu^3 + 27\alpha^2\gamma^2. \]  

Proposition 2.13 i) If in (8) we take \( \mu \) as a state function, \( \alpha \) as a control parameter and \( \gamma \) as fixed, we obtain that \( G_1(\mu, \alpha, \gamma, 0) \) is a germ in \( (\mu, \alpha) \) with an infinite codimension.

ii) If in (8) we take \( \gamma \) as a state function, \( \alpha \neq 0 \) as fixed and \( \mu \) as the control parameter, then we obtain that \( G_1(\gamma, \mu, \alpha, 0) \) is a zero codimension germ, strongly equivalent with the limit point germ, \( g(x, \lambda) = x^2 + \lambda \).

iii) If in (8) we take \( \gamma \) as a state function, we fixe \( \mu \neq 0 \), and \( \alpha \) is the control parameter, then \( G_1(\gamma, \alpha, \mu, 0) \) is a germ in \( E_{\gamma, \alpha} \) with an infinite codimension.

iv) Taking in (8), \( \alpha \) as a state function, \( \gamma \neq 0 \) as fixed and \( \mu \) as a control parameter, we obtain that \( G_1(\alpha, \mu, \gamma, 0) = 0 \) is equivalent with the bifurcation problem \( g(x, \lambda) = x^2 - \lambda^2 = 0 \), i.e. it is a simple bifurcation.

References


**INVERSE PROBLEMS OF TRANSITIVE CLOSURE**

Adrian Deaconu

*Univ. “Transilvania” of Brașov, Faculty of Mathematics and Informatics*

Deaconu@info.unitbv.ro

**Abstract**

A graph $G'$ starting from a self-transitive closure graph $G$ with as least as possible arcs so that the transitive closure of $G'$ is $G$ is found. Two algorithms are presented. The number of optimal solutions is computed.

**Keywords:** graphs, permutations

1. **INVERSE PROBLEMS OF TRANSITIVE CLOSURE**

Let $G = (V, E)$ be an oriented graph.

**Definition 1.** The oriented graph $G^* = (V, E^*)$ is called the *transitive closure* of the graph $G$, where $E^* = \{(x, y) |$ there is a directed path from $x$ to $y\}$.

**Definition 2.** The oriented graph $G = (V, E)$ is a *self-transitive closure* iff its transitive closure does not differ from $G$, i.e. $G^* = G$.

Let us consider a self-transitive closure graph $G = (V, E)$. The inverse problem of transitive closure (ITC) is to find $G^{-1} = (V, E^{-1})$, where $E^{-1}$ has as least arcs as possible from $G$ and the transitive closure of $G^{-1}$ is $G$, i.e. $(G^{-1})^* = G$. Let us present other two definitions, which make the difference between the quality of the solutions of the inverse transitive closure problem.

**Definition 3.** $G^{-1} = (V, E^{-1})$ is called a *solution* for the inverse problem of transitive closure, if and only if $(G^{-1})^* = G$ and the elimination of any arc $(x, y)$ from $E^{-1}$ leads to a graph $G' = (V, E')$, where $E' = E^{-1} - \{(x, y)\}$, for which the transitive closure is not $G$, i.e. $((G')^* \neq G$.

**Definition 4.** $G^{-1} = (V, E^{-1})$ is called an *optimal solution* for the inverse problem of transitive closure, if and only if any graph $G' = (V, E')$ so that $(G')^* = G$ has at least the number of arcs of $G^{-1}$, i.e. $|E'| \geq |E^{-1}|$.

Any optimal solution satisfies the inverse problem of transitive closure.

2. **ALGORITHMS FOR THE ITC PROBLEM**

The first idea that appears for solving the inverse problem is to find criteria for elimination of arcs from $E$ so that the resulted graph $G' = (V, E')$ (\(E' = \))
Theorem 1. If an arc \((x, y)\) is eliminated from the graph \(G\) and in the resulted graph \(G' = (V, E')\) \((E' = E - \{(x, y)\})\) there is still a directed path from \(x\) to \(y\), then \((G')^*_r = G^*_r\).

Proof. Let \((x, y)\) \(\in E\) be an arc so that in the graph \(G' = (V, E' = E - \{(x, y)\})\) there is a directed path \(P'\) from \(x\) to \(y\) (1). Suppose that \((G')^*_r \neq G^*_r\) \(\Rightarrow \exists (u, v) \in E^*_r\) and \((u, v) \notin (E')^*_r\) iff there is a directed path \(P\) from \(u\) to \(v\) in \(G^*_r\) and there is no directed path from \(u\) to \(v\) in \((G')^*_r\).

There are two situations:
1. if the arc \((x, y)\) \(\notin P\), then all arcs from \(P\) are in \(E\) and \(E'\), so \(P\) is a directed path in \(G'\) (contradiction);
2. if the arc \((x, y)\) \(\in P\), then \(P'' = (u, ..., x, ..., y, ..., v)\), where all arcs are in \(E'\). So, there is a directed path \(P''\) in \(G'\) from \(u\) to \(v\) (contradiction).

In both cases we obtained contradiction, hence the assumption \((G')^*_r \neq G^*_r\) is false. So, \((G')^*_r = G^*_r\) and the theorem is proved.

Starting with Theorem 1 and remarking that the arcs \((x, x), x \in V\) can be also eliminated from \(G\), we can easily write an algorithm for finding a solution of the inverse problem of transitive closure.

\begin{enumerate}
  \item \(E^{-1} := E \setminus \{(x, x) | x \in V\}\);
  \item \(\text{for } (x, y) \in E^{-1} \text{ do}
    \begin{align*}
      &\text{if there is a directed path from } x \text{ to } y \text{ in } G'=(V, E'),
      \text{ where } E' = E^{-1} \setminus \{(x, y)\} \text{ then go;}
    \end{align*}
  \text{end for;}
\end{enumerate}

Fig. 1. Algorithm 1 (elimination of arcs).

Theorem 2. The algorithm 1 finds a solution \(G^{-1} = (V, E^{-1})\) of the inverse problem of transitive closure in a complexity of \(O(m^2)\), where \(m\) is the number of directed arcs in \(G = (V, E)\), i.e. \(m = |E|\).

Proof. Obviously \(G^{-1}\) is a solution, because every directed arc was tested for elimination, such that as a result of application of the algorithm 1 there is no other arc in \(E^{-1}\) which can be eliminated any more. The test if there is a directed path from a node \(x\) to a node \(y\) can be done in a complexity of \(O(m)\), using a search algorithm (BFS or DFS). There are \(m\) tests, so the complexity of the algorithm 1 is \(O(m^2)\).

The initial graph \(G\) is a self-transitive closure graph and due to that, usually it has many arcs, it is a so-called dense graph. So, algorithm 1 is slow, often it runs in a complexity of \(O(n^4)\), because in a dense graph \(m\) is closed to
Inverse problems of transitive closure

There is another problem with the algorithm 1. It finds a solution of the inverse problem, which is not necessarily optimal. Here is an example to illustrate this case.

Let us apply algorithm 1 to the graph in the fig. 2.

\[ G^{-1} \]

First the algorithm eliminates the arcs \((1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\). Suppose that in the second step the arcs \((1, 2), (1, 3), (1, 4), (2, 1), (2, 3), (2, 4), (4, 1), (4, 2), (4, 3)\) are eliminated. Finally, it is obtained the graph \(G^{-1}\) from the fig. 3. At the end of the algorithm, the graph \(G^{-1}\) has 7 directed arcs. This is a solution of the inverse problem, but it is not optimal, because an optimal solution has only 5 directed (fig. 4).

\[ G \]

Let us present a faster algorithm and prove that the found solution is optimal. The idea of this algorithm starts with the remark that the strongly connected components of any solution \(G^{-1}\) of the inverse problem are the same with the strongly connected components of the initial graph \(G\).

**Definition 5.** \(K \subseteq V\) is a strongly connected component of the graph \(G = (V, E)\) if for each nodes \(u, v \in K\) there is a directed path in \(G\) from \(u\) to \(v\).

**Definition 6.** The graph \(G^c = (V^c, E^c)\) is called the condensed graph of \(G = (V, E)\) if it has as nodes all the strongly connected components of \(G\) and the directed arcs of \(G^c\) are the connections between the strongly connected components of \(G\), i.e. \(V^c = \{K | K\ is\ a\ strongly\ connected\ component\ of\ G\}\) and \(E^c = \{(K_1, K_2) | K_1, K_2 \in V^c\ and \ \exists u \in K_1, v \in K_2 : (u, v) \in E\}\).

Using as few directed arcs from \(E\) as possible for each strongly connected component so that it remains strongly connected and using the directed arcs of the condensed graph, then an optimal solution for the inverse problem of the transitive closure is obtained.
Theorem 3. In a self-transitive closure graph \( G \) there is a directed path from \( x \) to \( y \) (\( x, y \in V \)), iff there is a directed path from \( x \) to \( y \) in a solution of the inverse problem denoted by \( G^{-1} \).

Proof. 
(" <= ") If a directed path \( P \) is from \( G^{-1} \), as \( E^{-1} \subseteq E \), it exists in \( G \). 
(" => ") Let \( P = (x = u_1, u_2, ..., u_r = y) \) be a directed path in \( G \). For any directed arc \((u_i, u_{i+1})(i \in 1, 2, ..., r-1)\) from \( P \), there is a directed path \( P_i \) in \( G^{-1} \), from \( u_i \) to \( u_{i+1} \). We replace every arc \((u_i, u_{i+1})\) with the path \( P_i \) in \( P \). It is obtained the path \( P' = (Pu_1, Pu_2, ..., Pu_{r-1}) \) from \( x \) to \( y \).

Theorem 4. \( K \subseteq V \) is a strongly connected component of a self-transitive closure graph \( G = (V, E) \), iff \( K \) is a strongly connected component of any solution \( G^{-1} \) of the ITC problem.

Proof. Directly from Theorem 3.

Theorem 5. If \( K \subseteq V \) is a strongly connected component of a self-transitive closure graph \( G = (V, E) \), then the graph \( G_K = (K, E_K) \) is complete, where \( E_K = \{(x, y) \in E \mid x, y \in K\} \), i.e. \( E_K = \{(x, y) \mid x, y \in K\} \).

Proof. Let \( x \) and \( y \) be two arbitrary nodes from \( K \). Since \( K \) is a strongly connected component, then there is a directed path from the node \( x \in K \) to node \( y \in K \) and as \( G = G^* \) it follows that \((x, y) \in E_K \), whence the theorem.

Theorem 6. Let \( G^c \) be the condensed graph of the self-transitive closure graph \( G = G^* \). Any directed arc \((K_i, K_j)(i \neq j)\) of the condensed graph must be in the set of arcs \( E^{-1} \) of any solution \( G^{-1} = (V, E^{-1}) \) of the ITC problem, i.e. \( \exists x \in K_i \) and \( \exists y \in K_j \) so that \((x, y) \in E^{-1} \).

Proof. Let \((G^{-1})^c = \{V', (E^{-1})^c\}\) be the condensed graph of \( G^{-1} \). Using the Theorem 4, it is obviously that \(((G^{-1})^c)^* = (G^c)^*\) (2). Let \((K_i, K_j)\) be an arbitrary chosen directed arc of the condensed graph \( G^c \). Then \((K_i, K_j) \in (E^c)^*\) (3).

Suppose that for any \( x \in K_i \) and for any \( y \in K_j : (x, y) \notin E^{-1} \). It follows that \((K_i, K_j) \notin (E^{-1})^c \), therefore \((K_i, K_j) \notin ((E^{-1})^c)^*\) (4). From (2), (3) and (4), a contradiction follows.

1. \( \mathbb{E}^k := \emptyset \);
2. Find the condensed graph \( G^c=(V^c, \mathbb{E}^c) \), where \( V^c = \{K_i \mid i \in \{1, 2, ..., p\}\} \), \( \mathbb{E}^c = \{k_{ij} \mid i \in \{1, 2, ..., p\}\} \); 
3. \( \text{for } i:=1 \text{ to } p \text{ do}
   \text{if } s_i > 1 \text{ then}
   \text{for } j:=1 \text{ to } s_i \text{ do}
     \mathbb{E}^{-1} := \mathbb{E}^c \uplus \{ (k_{ij}, k_{ij+1}) \}; \text{ (where } \bar{k}_{ij} = \bar{k}_{ij} \}
   \text{end for;}
   \text{end if;}
\end{for;}
4. \( \mathbb{E}^{-1} := \mathbb{E}^{-1} \uplus \{ (s_i, 1) \mid s_i \in \mathbb{E}^c \} \);

Fig. 5. Algorithm 2 (using the condensed graph).
Theorem 7. The algorithm 2 finds an optimal solution \( G^{-1} = (V, E^{-1}) \) of the inverse transitive closure problem.

Proof. In the step 3 of the algorithm, in the set \( E^{-1} \) we introduce directed arcs that form directed elementary cycles of each strongly connected component (which has the minimum number of arcs of all directed cycles). In the step 4, in the set \( E^{-1} \) we introduce directed arcs that connects the strongly connected components of the condensed graph. The algorithm finds an optimal solution for the inverse problem of transitive closure (\( (G^{-1})^* = G \) and there is no other solution with a less number of arcs) - see Theorems 4, 6.

Theorem 8. The complexity of algorithm 2 is \( O(\min\{m+n, n \cdot p\}) \), where \( m \) is the number of directed arcs, \( n \) is the number of nodes of the self-transitive closure graph \( G = (V, E) \), i.e. \( m = |E| \), \( n = |V| \) and \( p \) is the number of strongly connected components of \( G \).

Proof. The complexity of the step 1 of the algorithm is \( O(1) \). The complexity of the step 2 is \( O(m+n) \) using the depth first search algorithm (DFS). The complexity of the step 3 is \( O(n) \), because \( s_1 + s_2 + ... + s_p = n \). The complexity of the step 4 is \( O(n) \) (there are at most n-1 arcs found in step 2 that connect the strongly connected components, because the condensed graph is a forest of trees). So, the complexity of the whole algorithm is \( O(m+n) \). Using the fact that in the self-transitive closure graph \( G \) any strongly connected component is complete (Theorem 5), the condensed graph can be also found in step 2 in a complexity of \( O(n \cdot p) \) with the algorithm presented in fig. 6.

Being a self-transitive closure graph, the graph \( G \) has many arcs (it is "dense"). In most of the cases \( m \) is larger than \( n \cdot p \). So, it is better to use the algorithm from fig. 6 to find the condensed graph.

3. THE NUMBER OF OPTIMAL SOLUTIONS

Here we compute the number of optimal solutions for the ITC problem.

Theorem 9. The number of optimal solutions of the inverse transitive closure problem is \( \prod_{i=1}^p (s_i-1)! \prod_{1 \leq i < j \leq p} N_{i,j}, \) where \( N_{i,j} = s_i \cdot s_j \), if \( \exists (K_i, K_j) \in E^c \) or \( \exists (K_j, K_i) \in E^c \) and \( N_{i,j} = 1 \), otherwise.

Proof. All optimal solutions are generated from the condensed graph \( G^c = (V, E^c) \). In order to obtain different optimal solutions, the nodes from every strongly connected component must be permuted (noncircular). The noncircular permutations of \( k \) order are obtained by holding one position and permuting the other \( k-1 \). So, there are \( (k-1)! \) noncircular permutations of \( k \) order (1).

If the directed arcs which connect the strongly connected components links different nodes from components, different optimal solutions are obtained. Here \( N_{i,j} \) is the number of possible different links between the strongly connected components \( K_i \) and \( K_j \), if there is an arc \( (K_i, K_j) \) or \( (K_j, K_i) \) that
connects the components in the condensed graph (2). So, using (1) and (2), the theorem is proved.

\begin{verbatim}
r := 0;
p := 1;
select y \in V;
k_{0,1} := \{y\};
s_{0,1} := 1;
for \forall x \in V \setminus \{y\} do
  for i := 1 to p do
    if \((x, k_{i-1,1}) \in E\) and \((k_{i-1,1}, x) \in E\) then
      s_{i,1} := s_{i,1} + 1
      k_{i,1} := k_{i-1,1}
    else
      if \((x, k_{i-1,1}) \in E\) then
        p := p + 1;
        k_{p,1} := x;
        r := r + 1;
        s_{p,1} := s_{p,1};
        s_{r,1} := s_{r,1};
      end if;
      if \((k_{i-1,1}, x) \in E\) then
        p := p + 1;
        s_{p,1} := s_{p,1};
        k_{p,1} := x;
        r := r + 1;
        s_{r,1} := s_{r,1};
      end if;
  end for;
end for;
\end{verbatim}

Fig. 6. Algorithm for finding the strongly connected components of a self-transitive closure graph.

**Theorem 10.** If the graph \(G\) is (weakly) connected (there is a path between every two nodes of \(G\)), then the number of optimal solutions of the inverse problem of transitive closure is \(\prod_{i=1}^{p} s_i! \cdot s_i^{p-2}\).

**Proof.** If the graph \(G\) is connected then, between every two components \(K_i\) and \(K_j\) there is an arc \((K_i, K_j)\) or \((K_j, K_i)\) that connects them. This implies that

\[ \prod_{1 \leq i < j \leq p} N_{i,j} = \prod_{1 \leq i < j \leq p} s_i \cdot s_j = \prod_{i=1}^{p} s_i^{p-1}. \]

It follows (by Theorem 9) that the number of optimal solutions of the ITC problem is

\[ \prod_{i=1}^{p} (s_i - 1)! \cdot \prod_{1 \leq i < j \leq p} N_{i,j} = \prod_{i=1}^{p} (s_i - 1)! \cdot \prod_{i=1}^{p} s_i^{p-1} = \prod_{i=1}^{p} s_i! \cdot s_i^{p-2}. \]
FURTHER STUDY OF A MICROCONVECTION MODEL USING THE DIRECT METHOD

Ioana Dragomirescu, Adelina Georgescu

Univ. "Politehnica" of Timisoara, Univ. of Pitesti

Abstract

The analytical study, partially treated in [1] by using the direct method, of the linear model of natural convection under microgravity conditions for a binary liquid layer in the presence of the Soret effect is completed. Due to the large number of parameters involved, a lot of particular cases are encountered. The closed form expressions of the bifurcation points of the characteristic manifolds are found. The false secular points, occurring in direct formal applications of numerical methods, are detected. In the parameter space, they correspond to bifurcation set of the characteristic manifold.

Keywords: thermal convection, microgravity

2000 MSC: 76E06

1. EIGENVALUE PROBLEM. CHARACTERISTIC EQUATION

Many processes performed under microgravity conditions favour a better understanding of the chemical reactions and kinetic and transport processes involved. In this respect, the researchers from the Max-Planck Insitute of Extraterrestrial Physics proved that complex plasmas under microgravity conditions can self-organize spontaneously to a crystalline-like state (plasma crystal). On Earth, such crystals cannot be obtained since the gravity squeezes the crystals together. The self-organization is modelled by bifurcation emerged as a result of the loss of stability. The new patterns (crystals) correspond to the bifurcating solutions of the equations governing those processes. Combustion, convection, material science, are other domains in which the gravity effects are studied.

In this paper we are concerned with a class of microconvection models with strong Soret effect in an infinite horizontal binary liquid layer, bounded by two impermeable rigid walls on which the normal heat flux is specified. This system is placed in a uniform gravity field with the acceleration $g$.

In [1] the linear model of natural convection under microgravity conditions for a binary liquid layer in the presence of Soret effect was treated analytically using the direct method based on the characteristic equation. There it was proved that the characteristic equation associated with the two-point problem
for the governing equations has no solution of third algebraic multiplicity order except for a few particular cases. In these cases the secular equation was obtained. By simultaneously solving the characteristic and the secular equation, and keeping some parameters constant, the neutral manifolds were analytically determined.

When the characteristic equation has multiple roots the straightforward application of numerical method can lead to false secular points. The false secular manifolds were analytically detected in [1] for third order multiple characteristic values. Herein we complete the study with the analytical investigation of the case of double roots.

The two-point problem for the nondimensional equations linearized about the mechanical equilibrium governing the thermal convection [2] is

\[
\begin{align*}
(D^2 - a^2)^2 \Psi' + iaG'(T' + C') &= 0, \\
-\varepsilon (ia\Psi' + ST' + LeDC') &= (D^2 - a^2)T', \\
\varepsilon \sigma (ia\Psi' + ST' + LeDC') &= Le(D^2 - a^2)(C' - \sigma T'),
\end{align*}
\]

\[D\Psi' = ia(ST' + LeC'), \quad \Psi' = DT' = DC' = 0, \quad \text{at } z = 0 \text{ and } 1. \tag{1}
\]

where \(D = \frac{d}{dz}\). Here \(G' = \frac{G}{1 + \varepsilon(T_0 + C_0)^2}[2]\) and is assumed to be constant. \(0.5 \leq G' \leq 1\). The values of the physically admissible parameters are: the Lewis number, \(Le\), \(0 < Le \leq 0.5\), the Boussinesq parameter, \(\varepsilon\), \(0.005 \leq \varepsilon \leq 0.05\) and the separation ratio, \(\sigma\), \(1 \leq \sigma \leq 1.5\). Only small wave numbers \(10^{-2} \leq a \leq 10^{-1}\) are considered. These values are chosen so that the problem has physical meaning.

In [1] we rewrote the two-point problem (1)-(2) in terms of \(T'\) only in the form

\[
Le(D^2 - a^2)^3 T' + \varepsilon Le(1 - \sigma)D(D^2 - a^2)^2 T' + a^2 \varepsilon G'[Le + \sigma(Le - 1)]T' = 0, \tag{3}
\]

\[DT' = (D^2 - a^2)T' = D^3T' = 0 \quad \text{at } z = 0 \text{ and } 1. \tag{4}
\]

Denote \(a_1 = \varepsilon(1 - \sigma), \quad a_2 = a^2 \varepsilon G'[1 + \sigma(1 - 1/Le)]\) and take into account that the Lewis number is a nonnull physical parameter. Then equation (3) becomes

\[
(D^2 - a^2)^3 T' + a_1 D(D^2 - a^2)^2 T' + a_2 T' = 0. \tag{5}
\]

With the notation \(\mu = \lambda/a, \quad 6b = a_1/a, \quad d = (a_2 - a^6)/a^6\), the characteristic equation associated with the two-point problem (5)-(4) achieves the simplified form [1]

\[
P \equiv \mu^6 + 6b\mu^5 - 3\mu^4 - 12b\mu^3 + 3\mu^2 + 6b\mu + d = 0, \tag{6}
\]

where the cases \(b = 0\) and/or \(d = -1\) are not considered in the sequel because they have been treated in [1].
2. FALSE SECULAR MANIFOLDS

The characteristic equation (6) reads in the equivalent form

\[(\mu^2 - 1)^3 + 6b\mu(\mu^2 - 1)^2 + d + 1 = 0.\] (7)

If (7) has a double root, then it is a root for its first derivative too, i.e.

\[(\mu^2 - 1)(\mu(\mu^2 - 1) + 5b(\mu^2 - 1) + 4b] = 0.\] (8)

The case \(\mu = \pm 1\), i.e. \(d = -1\), equivalent to \(a_2 = 0\), was treated in [1], so we assume that \(a_2 \neq 0\). Therefore instead of (8) we study the equation

\[P_1 \equiv \mu^3 + 5b\mu^2 - \mu - b = 0.\] (9)

Further we apply a variant of the Euclid algorithm to (7) and (9). Thus, from (9) we find \(\mu^2 - 1 = -\frac{4b}{\mu + 5b}\), \(\mu \neq -5b\) which replaced in (7) yields

\[P_2 \equiv \mu^3 u + \mu^2 (96b^3 + 15bu) + \mu (480b^4 + 75bu) + (125b^3u - 64b^3) = 0.\] (10)

The assumption \(\mu \neq -5b\) holds because if \(\mu = -5b\), then (9) implies \(b = 0\) and (6) yields \(d = 0\). This case was studied in [1] and it was shown that it corresponds to false secular points.

The common roots of (7) and (8) are the common roots of (9) and (10). At first, we investigate the possibility that all roots of the equations (9) and (10) are the same, which leads to

\[
\frac{1}{d + 1} = \frac{5b}{96b^3 + 15b(d + 1)} = \frac{-1}{480b^4 + 75b^2(d + 1)} = \frac{-1}{61b^3 + 125b^3d}. \tag{11}
\]

The first two equalities from (11) read \(d + 1 = -\frac{48b^2}{5} = \frac{-480b^4}{75b^2 + 1}\), i.e. \(25b^2 + 1 = 0\). Since \(b\) is a real physical parameter the last situation is not possible, so the two equations cannot be equivalent. Thus, we must look for two or one common root.

In order to simplify the implied cumbersome computations by reducing the degree of equations (9) and (10) we perform the linear combination \((10) - (d + 1)(9)\) leading to \(Q_1 \equiv \mu^2(96b^3 + 10bd + 10b) + \mu(480b^4 + 75b^2d + 75b^2 + d + 1) + (61b^3 + 125b^3 + db + b) = 0.\) (12)

and the linear combination \(\mu(12) - (9)(96b^3 + 10bd + 10b)\) implying

\[Q_2 \equiv \mu^2(25b^2 + 1)(d + 1) + \mu(11db + 11b + 157b^3 + 125db^3) + (10db^2 + 10b^2 + 96b^4) = 0.\] (13)
Equations (12) and (13) can have two common solutions, only one common solution or none. Each of these cases must be treated separately.

Consider first the case of double roots of (12), (13); they correspond to two pairs of equal solutions of (6).

**Proposition 1.** In the five-dimensional \((a, \varepsilon, \sigma, G', Le)\) parameter space the two hypersurfaces

\[
\begin{align*}
H1 : a\varepsilon(1 - \sigma) &= \sqrt{-3 + 8\sqrt{6}} \frac{375}{a^4} \\
H2 : \frac{\varepsilon G'[1 + \sigma(1 - 1/Le)]}{a^4} &= \frac{16(29 + 6\sqrt{6})}{625}
\end{align*}
\]

defined by

\[
b^* = \pm \sqrt{-3 + 8\sqrt{6}} \frac{375}{a^4} \quad \text{and} \quad d^* = \frac{-161 + 96\sqrt{6}}{625}
\]

are false secular manifolds.

**Proof.** For these values of \(b\) and \(d\) in Appendix it is shown that the characteristic equation has two double roots \(\mu_1 = \mu_3\) and \(\mu_2 = \mu_4\) because (12) and (13) have two common roots. The corresponding expressions of \(\mu_{1,3}\) are given also in Appendix by (21). Since we know the algebraic multiplicity for each root of the characteristic equation (in fact, we know the roots themselves), we can write the general solution \(T'\) of the equation. Replacing it into the boundary conditions we obtain the secular equation

\[
\begin{bmatrix}
\lambda_1 & \lambda_2 & 1 & 1 & \lambda_5 & \lambda_6 \\
\lambda_1 e^{\lambda_1} & \lambda_2 e^{\lambda_2} & (\lambda_1 + 1)e^{\lambda_1} & (\lambda_2 + 1)e^{\lambda_2} & \lambda_5 e^{\lambda_5} & \lambda_6 e^{\lambda_6} \\
r_1 & r_2 & 2\lambda_1 & 2\lambda_2 & r_5 & r_6 \\
r_1 e^{\lambda_1} & r_2 e^{\lambda_2} & (2\lambda_1 + r_1)e^{\lambda_1} & (2\lambda_2 + r_2)e^{\lambda_2} & r_5 e^{\lambda_5} & r_6 e^{\lambda_6} \\
\lambda_1^3 & \lambda_2^3 & 3\lambda_1^2 & 3\lambda_2^2 & \lambda_3^3 & \lambda_4^3 \\
\lambda_1^2 e^{\lambda_1} & \lambda_2^2 e^{\lambda_2} & (3\lambda_1^2 + \lambda_2^2)e^{\lambda_1} & (3\lambda_2^2 + \lambda_2^2)e^{\lambda_2} & \lambda_3^2 e^{\lambda_3} & \lambda_4^2 e^{\lambda_4}
\end{bmatrix}
= 0 \quad (16)
\]

where \(\lambda_i = \mu_i a, r_i = \lambda_i^2 - a^2, i = 1, 6\). This equation (16) contains a single variable \(a\) and has only the physically unrealistic root \(a = 0\). On the other hand, for \(a = 0\) we have the case \(\lambda_{1,6} = 0\) for which (16) is not entitled to be considered as a secular equation.

**Fig. 1.** The characteristic equation for \(b = b^* < 0\) and \(d = d^*\). (a) \(b = -\sqrt{-3 + 8\sqrt{6}} \frac{375}{a^4}\) \(d = -\frac{-161 + 96\sqrt{6}}{625}\).
**Proposition 2.** The characteristic equation has double roots when the real parameters $b$ and $d$ satisfy the relation $m = 0$, where the expression of $m$ in terms of $b$ and $d$ is given by (24).

**Proof.** In Appendix, the relation $m = 0$ connecting the parameters $b$ and $d$ is a third degree equation in $d$, the coefficients of which depend on $b$. Therefore its solutions $d_1$, $d_2$, and $d_3$ can be expressed in a simple form by means of the cubic Cardano's formulas.

For the physical admissible values of the parameters $a$, $Le$, $\sigma$, $\varepsilon$, $G'$, the numerical evaluations show that $d < -1$. This is why, in the following we are interested in this case. However, for the sake of completeness, theoretically we treat also the case $d > -1$.

Remark that the equation (26) is invariated by a symmetry with respect to the $Od$-axis, i.e. $(\mu, b, d) = (-\mu, -b, d)$. This is why, the study can be restricted to the half-space $\mu > 0$. In addition, the form (7) of the characteristic equation shows that if $|\mu| < 1$, then $d < -1$ (fig. 2).

Let $Hs$ be the set of points of the space $(b, d, \mu)$ corresponding to $d_1$, $d_2$, and $d_3$. They are situated on three curves $d_i = d_i(b)$, $i = 1, 2, 3$.

**Proposition 3.** The points on $Hs$ are not secular.

**Proof.** As can be seen in the Appendix, the discriminant $\Delta$ of the equation $m = 0$ can be equal to zero, only for $b = 0$, when $d^*_1 = 0$, $d^{**}_{2,3} = -1$, or for $b = b^*$, when $d_1 = d_3 = d_*$.

For all other values of $b$ and $d$ we have $d_1 \geq 0$, $d_3 \geq -1$, $d_2 \leq -1$. The points $(0, 0)$, $(0, -1)$, $(b^*, d^*)$ are bifurcation points for the set of the roots of $m = 0$ (fig. 2).

![Fig. 2. The curves $d_i(b)$](image)

For $(b, d) = (0, -1)$ the characteristic equation has three double roots [1], for $(b^*, d^*)$ and $(b, -1)$, $(b \neq 0, b^*)$ it has two double roots while for all other bifurcation points, i.e. for $(b, d) = (b, d_i(b))$, $i = 1, 2, 3$ $(b \neq 0, b^*)$, only one double root corresponds.

Fig. 3 shows that the sheets of the characteristic roots coalesce along the curves $d_i = d_i(b)$ in the $(b, d)$-plane.

If $Le$ assumes the physically admissible values, i.e. $0 < Le \leq 0.5$, and taking into account that $a$, $\varepsilon$, $G'$ are positive parameters and the expression $1 - \sigma$ is
negative, then it follows that \( a_2 \equiv a^2 \varepsilon G'[1 + \sigma(1 - 1/Le)] \leq a^2 \varepsilon G'(1 - \sigma) < 0 \), implying \( d < -1 \). In this case, only the root \( d_2 \) of equation \( m = 0 \) exists.

For the points \((b, d_2(b))\), with \( b \neq 0 \), the general form of the solution of (5) has the form

\[
T'(z) = (A + Bz)e^{\lambda_1 z} + \sum_{i=3}^{6} A_i e^{\lambda_i z},
\]

and, taking into account the boundary conditions, the secular equation reads

\[
\begin{vmatrix}
\lambda_1 & 1 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\
\lambda_1 e^{\lambda_1} & (\lambda_1 + 1) e^{\lambda_1} & \lambda_3 e^{\lambda_1} & \lambda_4 e^{\lambda_1} & \lambda_5 e^{\lambda_1} & \lambda_6 e^{\lambda_1} \\
r_1 & 2r_1 & r_3 & r_4 & r_5 & r_6 \\
r_1 e^{\lambda_1} & (r_1 + 2\lambda_1) e^{\lambda_1} & r_3 e^{\lambda_1} & r_4 e^{\lambda_1} & r_5 e^{\lambda_1} & r_6 e^{\lambda_1} \\
\lambda_1^3 & 3\lambda_1^3 & \lambda_3^3 & \lambda_3^3 & \lambda_3^3 & \lambda_3^3 \\
(\lambda_1^3 + 3\lambda_1^3) e^{\lambda_1} & \lambda_3^3 e^{\lambda_1} & \lambda_3^3 e^{\lambda_1} & \lambda_3^3 e^{\lambda_1} & \lambda_3^3 e^{\lambda_1} & \lambda_3^3 e^{\lambda_1}
\end{vmatrix} = 0. \tag{17}
\]

Solving simultaneously the characteristic and the secular equation, we get that, for \( d = d_2(b) \) and for the physically admissible values of the parameters \( a, Le, \sigma, \varepsilon \), our numerical computations indicate that the points on \( H_s \) are not secular.

3. APPENDIX

Two common roots of (12), (13). In this case the ratio of the corresponding coefficients of (12) and (13) must be equal, i.e.

\[
\frac{96b^3 + 10bu}{(25b^2 + 1)u} = \frac{480b^4 + (75b^2 + 1)u}{11b + 125b^3} = \frac{(125b^2 + 1)u - 64b^2}{96b^3 + 10bu}, \tag{18}
\]

where \( u = d + 1 \). Assume, that, the coefficients of (12) and (13) are nonnull. Then system in \( u \), (19), reads

\[
\begin{cases}
C_1 \equiv (-625b^4 + 10b^2 - 1)u^2 + 896b^4u + 3072b^6 = 0 \\
C_2 \equiv (-3125b^4 - 50b^2 - 1)u^2 + (3520b^4 + 64b^2)u + 9216b^6 = 0,
\end{cases} \tag{19}
\]
where the coefficients of \( u^2 \) are nonnull for every \( b \in \mathbb{R} \). Assume that equations (19) have two common roots. Then

\[
\begin{align*}
-625b^4 + 10b^2 - 1 &= \frac{896b^4}{3520b^4 + 64b^2} = \frac{3072b^6}{9216b^6} = \frac{1}{3},
\end{align*}
\]

implying the system in \( b \), \( 625b^4 + 40b^2 - 1 = 0 \), \( 13b^4 + b^2 = 0 \), which has no solution. Therefore in (19) the two equations cannot be equivalent.

A possible common root of the equations (19) must satisfy the second degree equation, obtained, for instance, by performing the operations \( 3(19)_1 - (19)_2 \), the roots of which are \( u = 0 \) and

\[
u = \frac{32b^2(13b^2 + 1)}{625b^5 + 40b^2 - 1}.
\]

The root \( u = 0 \), i.e. \( d = -1 \), is disregarded because we do not study this case. The second root must satisfy one arbitrary equation from (19). Imposing this condition, it follows

\[
(1 + 25b^2)(625b^4 - 10b^2 + 1)(375b^4 + 6b^2 - 1)(125b^4 + 22b^2 + 1) = 0.
\]

Therefore, we must have \( 375b^4 + 6b^2 - 1 = 0 \) leading to \( b^* \). Then (20) yields \( d^* \), whence (15).

Replacing the values (15) of \( b^* \) and \( d^* \) in equation (6) we obtain the following roots of the characteristic equation

\[
\begin{align*}
\mu_1 = \mu_3 &= \pm \frac{(6 + \sqrt{6})\sqrt{120\sqrt{6} - 45}}{150} + \frac{\sqrt{2}}{2}, \\
\mu_2 = \mu_4 &= \pm \frac{(6 + \sqrt{6})\sqrt{120\sqrt{6} - 45}}{150} - \frac{\sqrt{2}}{2}, \\
\mu_5 &= \frac{(3 + \sqrt{6})\sqrt{120\sqrt{6} - 45}}{75} + \frac{2i}{25} \sqrt{45\sqrt{6} - 95}, \\
\mu_6 &= \frac{(3 + \sqrt{6})\sqrt{120\sqrt{6} - 45}}{75} - \frac{2i}{25} \sqrt{45\sqrt{6} - 95},
\end{align*}
\]
the sign − corresponding to $b^* > 0$ and + to $b^* < 0$. In order to verify the role of the roots $\mu_1, \mu_2$ we apply the Euclid algorithm, i.e.

$$P(\mu + 5b)^3 = P_1[(\mu^2 - 1)^2(\mu + 5b)^2 + b(\mu^2 - 1)(6\mu^2 + 30b\mu - 4)(\mu + 5b) - 4b^2(6\mu^2 + 30b\mu - 4)] + P_2,$$

$$P_2 = (d + 1)P_1 + Q_1,$$

$$P_1 = ((6b^3 + 10bu)^{-1} \{ (\mu - \frac{25b^2 + 1}{96b^3 + 10bu})Q_1 - \frac{\mu C_1 - bC_2}{96b^3 + 10bu} \}$$

which shows that, for $b^{*2} = -\frac{3}{375} + 8\sqrt{6}$, we have $C_1 = C_2 = 0$, $\mu_{1,2}$ are the roots of $Q_1 = 0$, of $P_1$, $P_2$ and $P$.

So far, we assumed that none of the coefficients in (12) and (13) are equal to zero. Now we treat separately the cases in which we have double roots and at least one of these coefficients is null.

a) $96b^3 + 10bu = 0$. In this case in (19) we must have also $(125b^2 + 1)u - 64b^2 = 0$ which cannot be satisfied simultaneously with the assumed relation.

b) $480b^4 + u(1 + 75b^2) = 0$. Similarly, in order for (19) to be satisfied it is necessary to have simultaneously $32b^3 + (11b + 125b^3)u = 0$ implying $b^2 = -9 + 2\sqrt{39}, u = -\frac{16(-9 + 2\sqrt{39})}{125}$. However, for these expressions, (19) is not satisfied. As a conclusion, the equations (12), (13) do not have two common roots in the case b).

c) $u = \frac{64b^2}{1 + 125b^2}$. This case, with $u > 0$ cannot stand because (18) implies $u = -\frac{96b^2}{10} > 0$.

We conclude that the equation (7) has two pairs of double roots only for

$$b^2 = b^{*2} = -\frac{3}{375} + 3\sqrt{6}$$

and $u = u^* = \frac{16(29 + 6\sqrt{6})}{625}$.

**Only one common root of (12), (13).** If this root exists, it must satisfy, in particular, the first degree equation $(10b(d + 1) + 96b^3)(13) - (25b^2 + 1)(d + 1)(12)$, the root of which is $\mu = -\frac{b[d^2(1 + 50b^2 + 3125b^4) + d(2 + 36b^2 + 2730b^4) + (1 - 14b^2 - 395b^4 - 9216b^6)]}{d^2(625b^4 - 10b^2 + 1) + d(354b^4 - 20b^2 + 2) + (-3072b - 271b^4 - 10b^2 + 1)}$,
where we assume that $10b(d + 1) + 96b^3 \neq 0$ and the denominator is different from zero. Imposing to (23) to satisfy (12) and (13) we obtain

$$\begin{align*}
&\left\{ -256b^5(48b^2 + 5d + 5)^2 \cdot m = 0 \\
&-128b^4(d + 1)(25b^2 + 1)(48b^2 + 5d + 5) \cdot m = 0,
\end{align*}$$

where

$$m \equiv d^3 + d^2(2 - 3b^2 - 375b^4 - 3125b^6) + d(1 - 6b^2 - 366b^4 - 3050b^6) + (-3b^2 + 9b^4 + 1099b^6 + 9216b^8).$$

Owing to the fact that $b \neq 0$ and $d \neq -1$, it remains to discuss the situation $m = 0$. Therefore, consider the relation $m = 0$ as a third degree equation in $d$, namely $d^3 + \alpha d^2 + \beta d + \gamma = 0$. Reduce it to the canonical form $y^3 + py + q = 0$, by putting $y = d + \alpha / 3$, $p = \beta - \alpha^2 / 3$, $q = 2\alpha^3 / 27 - \alpha \beta / 3 + \gamma$, where

$$\alpha = 2 - 3b^2 - 375b^4 - 3125b^6, \quad \beta = 1 - 6b^2 - 366b^4 - 3050b^6, \quad \gamma = -3b^2 + 9b^4 + 1099b^6 + 9216b^8.$$ Let $D = -\frac{1024}{27}b^6(375b^4 + 6b^2 - 1)^2(125b^4 + 22b^2 + 1)^3$ be the associated discriminant. As expected, the case $D = 0$ is equivalent to the case of two roots of (12), (13), namely $375b^4 + 6b^2 - 1 = 0$, leading to (15). Since $D \leq 0$ we must investigate in addition, the case $D < 0$. In this case, all roots of the equation $m = 0$ are real and they read $y_1 = 2\sqrt[3]{r} \cos \frac{\phi}{3}$,

$$y_2 = 2\sqrt[3]{r} \cos \left( \frac{\phi}{3} + 120^\circ \right), \quad y_3 = 2\sqrt[3]{r} \cos \left( \frac{\phi}{3} + 240^\circ \right),$$

where $r = \sqrt{-p^3 / 27}$, $\cos \phi = -q / 2r$. The corresponding expressions $d_i = d_i(b)$ represent three curves in the parameter space $(b, d)$ (fig. 2).

Remark that for $d = -1$ and $b$ arbitrary, (6) has two double roots, namely $\mu_{1,2} = 1$ and $\mu_{3,4} = -1$. Consequently, for each point $(b, d)$ on these curves different from $(0, 0), (0, -1), (b^*, d^*), (b, -1)$ $(b \neq 0, b^*)$ the characteristic equation (6) has one double root only, i.e. $\lambda_1 = \lambda_2 = \mu \cdot a$, where $\mu$ is given by (22). Thus, for any other point $(b, d)$ of the parameter space, the characteristic equation has six mutually distinct roots, situated in the 4-dimensional space $(b, d, Re\mu, Im\mu)$.

If the denominator of (22) is zero, in order to have a possible root, the numerator must also be zero. This implies

$$\begin{align*}
&d^2(625b^4 - 10b^2 + 1) + d(354b^4 - 20b^2 + 2)(1 - 10b^2 - 271b^4 - 3072b^6) = 0 \\
&d^2(1 + 50b^2 + 3125b^4) + d(2 + 366b^2 + 2730b^4)(1 - 14b^2 - 395b^4 - 9216b^8) = 0.
\end{align*}$$

(24)

From (24) we obtain

$$d = \frac{20b^2 - 354b^4 + 2 \pm 64\sqrt{3}b^6 + 166b^8 + 1875b^{10}}{2(625b^4 - 10b^2 + 1)}$$

(25)
For \( d \) corresponding to the sign plus in front of the radical, the two equations (25) have a common root for \( b = \pm \frac{1}{75} \sqrt{-45 + 120\sqrt{6}} \), i.e. \( b = b^* \). This case was already discussed. For the second expression of \( d \) in (26), the equation (25) has no real root \( b \). Therefore, the two equations (25) have no common roots. Thus, the vanishing denominator in (23) can lead only to the case \((b^*,d^*)\).

If \( 10bu + 96b^3 = 0 \), then either \( b = 0 \) (which is not our concern), or \( 10u + 96b^2 = 0 \). In the second case, (18) cannot occur, therefore to this case no double root of (12) correspond.

4. CONCLUSIONS

The two-point eigenvalue problem governing the microconvection model presented here depends on five parameters. As a consequence the investigation of the bifurcation of the characteristic manifold is difficult. Partially it was performed in [1], while in this paper we deduced analytical expressions for all other bifurcation points of this manifolds. Numerically we found that all of them correspond to false secular points.

References

AN OPTIMIZATION METHOD FOR FREE SURFACE STEADY FLOWS

Tiberiu Ioana, Titus Petrila

"Babes-Bolyai" University, Cluj-Napoca

Abstract The free surface steady flow of an incompressible heavy fluid (inviscid and viscous) is studied. The main goal of this work is to develop an efficient approach based on an optimal shape design formulation that can be used for numerical calculations.

1. INTRODUCTION

We consider the two dimensional flow of an incompressible fluid freely moving along a rigid and impermeable bottom of arbitrary geometry. Apart from the influence of gravity, the fluid flow is free of any physical restrictions (the presence of some obstacles within or above the flow domain has been excluded).

It is assumed that the far field flow has a constant velocity (of unitary magnitude in the nondimensional form). We also consider, without loosing the generality, that the uniform stream has a unitary depth.

The flow is referred to a Cartesian coordinates system $O_{xy}$ with the $Ox$ axis directed along the unperturbed farfield free line, while $Oy$ is oriented along the direction of the outward normal to the unperturbed free surface.

We presume that the flow is steady, the free surface may be written using a 'height' function, and the flow domain is bounded.

2. PROBLEM FORMULATION

Consider a two-dimensional domain $D$ of boundary $\partial D = S \cup \Gamma$, where $S$ is an a priori unknown part while $\Gamma$ is the known part. The boundary $\Gamma$ contains the fixed bottom and two artificially boundaries (of inflow and outflow) on which appropriate boundary conditions preserving the initial behavior of the unbounded fluid flow are imposed. Assume a steady flow past a rigid and impermeable wall presenting a smooth bump located within the interval $[0, 1]$.

The Froude number associated with this flow is considered greater than unity $Fr \geq 1$ as well. Within this framework the waves cannot be generated by the fluid flow [1]. That is why the free surface may be defined by using a function with zero values at its two edges. Thus, we define the free surface by some Lipschitz continuous functions of the type $\alpha : [0, 1] \rightarrow [0, l]$,
\( l \in \mathbb{R}, \ 0 < l < 1, \alpha(0) = \alpha(1) = 0 \) and \(|\alpha(x_1) - \alpha(x_2)| \leq \beta(x_1 - x_2)\), for any \(x_1 \) and \(x_2 \) belonging to \((0, l)\). In this way, the free boundary becomes \(S(\alpha) = \{(x, y) \in [0, 1] \times [0, l] : y = \alpha(x)\}\). The Lipschitz continuity of \(\alpha\) functions is an essential property \([2]\). Under this condition the flow domain \(D(\alpha)\) may be extended beyond its boundary, that provides a mathematical support for using an optimization procedure in order to establish the \(a \ priori\) unknown free boundary and further on all the fluid flow properties. Moreover, the Lipschitz continuity condition prevents the occurrence of meaningless physical solutions produced by the optimization procedure \([6]\). Therefore, the family \(U_{ad}^p\) of the admissible functions which may define the boundary \(S(\alpha)\), is of the form

\[
U_{ad}^p = \{\alpha \in C^{0,1}[0,1] : 0 \leq \alpha(x) \leq l_0 < l, \quad |\alpha(x_1) - \alpha(x_2)| \leq \beta|x_1 - x_2|, \quad \forall x_1, x_2 \in [0, 1], \quad \alpha(0) = \alpha(1) = 0\}.
\]

Consequently, in the case of fluid flows with a Froude number above unity we have defined a set of conditions which have to be fulfilled by the function describing the free boundary such that the flow domain could be extended beyond its boundaries (for details of how to prove this assertion, see \([2]\)).

Let \(n\) be the number of boundary conditions which has to be imposed on the fixed boundary \(\Gamma\). Because the free boundary \(S\) is \(a \ priori\) unknown, usually the number of boundary conditions imposed here should be \(n + 1\).

Denote by \(a(u, p) = a^*(x, \alpha(x))\) a global representation of the first \(n\) conditions imposed on \(S\) while \(s(u, p) = s^*(x, \alpha(x))\) is the \((n + 1)th\) (last one). Let \(b(u, p) = b^*(x, y)\) be a compact representation for the boundary conditions imposed on \(\Gamma\).

Hence the envisaged problem may be finally formulated in the following terms: find \(\alpha(x), u(x, y)\) and \(\varphi(x, y)\) as the solutions of the boundary value problem (\(Re\) being the Reynolds number)

\[
\begin{align*}
\nabla \cdot u + \nabla \varphi - Re^{-1} \nabla \cdot \left(\nabla u + (\nabla u)^T\right) &= -Fr^{-1}j, \\
\n\nabla \cdot u &= 0, \\
s(u, p) &= s^*(x, \alpha(x)), \quad a(u, p) = a^*(x, \alpha(x)), \quad b(u, p) = b^*(x, y).
\end{align*}
\]

It is this final formulation that will be used in our algorithm.

### 3. VARIATIONAL FORMULATION

Let us proceed with a representation \(S_0\) of the free boundary \(S\), on which we assume that the conditions \(a(u, p) = a^*(x, \alpha(x))\) are fulfilled. The boundary \(S_0\) would be the right \(S\) if the condition \(s(u, p) = s^*(x, \alpha(x))\) is fulfilled as well.

In order to check whether the condition \(s(u, p) = s^*(x, \alpha(x))\) holds on the boundary \(S_0\) and in view of correcting its shape (if necessary), let us introduce
A positive functional defined on the set of all admissible boundaries. The surface we are looking for is that one which minimizes this functional. Denoting this functional by $J$, it can be written as

$$J(S) = \int_S (s(u, p) - s^*(x, \alpha(x)))^2 \, dS.$$ 

So, within the optimization framework the problem may be formulated as follows: find the solution $S$ of the problem $\min_S J(S)$, where $u(x, y)$ satisfy

$$u \cdot \nabla u + \nabla p - Re^{-1} \nabla \cdot \left( \nabla u + (\nabla u)^T \right) = -Fr^{-1}j,$$

$$\nabla \cdot u = 0,$$

$$a(u, p) = a^*(x, \alpha(x)),$$

$$b(u, p) = b^*(x, y).$$

4. THE INVISCID CASE

Let $\mathbf{n}$ and $\mathbf{t}$ denote the outward normal unit vector drawn at the free boundary $S$ and the tangent unit vector, respectively.

For the case of an inviscid incompressible potential fluid flow ($u = \nabla \phi$) the associated mathematical model reads

$$\Delta \phi = 0, \quad (x, y) \in D, \quad (1)$$

$$\mathbf{n} \cdot \nabla \phi = 0, \quad (x, y) \in \Gamma, \quad (2)$$

$$\mathbf{n} \cdot \nabla \phi = 0, \quad (x, y) \in S, \quad (3)$$

$$\frac{1}{2} - \left( \frac{1}{2} |\nabla \phi|^2 + Fr^{-2}y \right) = p_1, \quad (x, y) \in S,$$

where $Fr$ stands for the Froude number and $p_1$ is the atmospheric pressure.

By introducing an auxiliary function $a : \Gamma \to \mathbb{R}$, the unbounded flow domain $D$ may be restricted to a bounded domain such that the physical meaning of the initial problem be preserved. Keeping the same symbols for the quantities, the governing boundary value problem is (4.1), (4.3) and

$$p = p_1, \quad \text{on} S, \quad (4)$$

$$\mathbf{n} \cdot \nabla \phi = a, \quad \text{on} \Gamma. \quad (5)$$

Define the objective functional $J(S) = \int_S (p - p_1)^2 \, dS$. Then the problem is reduced to $\min_S J(S)$ such that (1), (3), (5) hold.

Using a classical variational technique, the boundary value problem can be associated with variational problem for the objective functional. So we get a new objective functional, denoted by $J^*$, of the form

$$J^*(S) = J(S) + \int_D \psi \Delta \phi \, dD.$$
Here $\psi$ is an auxiliary variable on the domain $D$ (a Lagrange multiplier). Then the optimization problem rereads as: find $S$ the solution of $\text{min}_{S} J^*(S)$.

Assume that the free boundary is perturbed along the direction of the normal $n$ with a small quantity $\epsilon > 0$, such that each point $(x, y)$ of the boundary $S$ will be mapped onto a point $(x_\epsilon, y_\epsilon)$ of $S_\epsilon$, which is related with its original position by $(x, y) \longrightarrow (x_\epsilon, y_\epsilon) = (x, y) - \epsilon n$ (this mapping is used, for instance, in [7]).

Within this framework we may use the Taylor series expansion for the potential $\phi$

$$\phi ((x, \alpha(x)) - \epsilon n) = \phi (x, \alpha(x)) - \epsilon n \cdot \nabla \phi (x, \alpha(x)) + O(\epsilon^2).$$

Then, on $S_\epsilon$, the velocity potential may written as $\phi_\epsilon = \phi - \epsilon n \cdot \nabla \phi + O(\epsilon^2)$.

As long as a gradient based approach is employed, in order to solve the above variational problem, the key element is the determination of the closed form of the gradient of the functional $J^*$. The main result of this paper, consists in just establishing the gradient expression under the form

$$\text{grad}_{n} J^* = - \int_{S} \left( p F r^{-2} n \cdot j + \psi n : \nabla \nabla \phi \right) dS,$$

where $\psi$ satisfies the problem

\[
\begin{align*}
\Delta \psi &= 0, & \text{in } D, \\
n \cdot \nabla \psi &= t \cdot \nabla (p t \cdot \nabla \phi), & \text{on } S, \\
n \cdot \nabla \psi &= 0, & \text{on } \Gamma.
\end{align*}
\]

References


Approximate Inertial Manifolds for an Advection-Diffusion Problem

Anca-Veronica Ion
University of Pitești

Abstract In the framework of the infinite-dimensional dynamical systems theory, a family of approximate inertial manifolds (a.i.m.s) for a problem modelling the Fickian diffusion of a substance into a Newtonian fluid is constructed. Estimates of the distance between these manifolds and the exact solution of the problem are given, proving that, at large times, the solution is kept in some very narrow neighbourhoods of the a.i.m.s.

1. Introduction

The concept of approximate inertial manifold (a.i.m.) arose in the framework of the theory of inertial manifolds. First defined in [3], inertial manifolds are finite-dimensional (at least) Lipschitz invariant manifolds, that attract exponentially all trajectories of an evolution equation.

For many evolution equations the existence of an inertial manifold is not yet proved, since the proof requires the existence of a certain large between two successive eigenvalues of the linear part (the so-called spectral gap) and this requirement is not fulfilled [7]. Even if an inertial manifold exists, this inertial manifold is difficult to construct. In this situation, approximations of the inertial manifolds were constructed, the a.i.m.s.[3]. An a.i.m. is a finite dimensional (at least) Lipschitz manifold having the property that every trajectory of the problem enters in a narrow neighbourhood of the a.i.m. at a certain time and remains there for ever. This property inspired some new methods of approximating the solutions of an evolution equation namely the nonlinear Galerkin and post-processed Galerkin methods. Thus, appart for the property of the a.i.m. of keeping the attractor of the problem in some narrow neighbourhood, these methods bring a new and very concrete motivation to study the a.i.m.s.

Among several methods of constructing an a.i.m. we choosed the algorithm given in [8] for the Navier-Stokes equation. After presenting in Sections 2-4 the problem and the results of existence of the solutions and dissipativity of the problem, in Sections 5, 6 we splitt the phase space in the direct sum of a finite-dimensional subspace and an infinite-dimensional one and give estimates of the norm of the projection of the solution on the infinite dimensional subspace.
We present the algorithm of [8] in Sections 7 and 8 and then use it in order to construct a family of a.i.m.s for our problem, in Sections 9 and 10. We also find the distance between the a.i.m.s and the exact solution.

2. THE PROBLEM, THE FUNCTIONAL FRAMEWORK

Consider the generalized problem governing the (Fickian) diffusion of a substance into a Newtonian incompressible fluid. It can be written as the following Cauchy problem for the evolution advection-diffusion equations

\[
\begin{align*}
\frac{\partial u}{\partial t} - \nu \Delta u + (u \cdot \nabla) u &= f(t), \\
\text{div} u &= 0, \\
\frac{\partial c}{\partial t} - D \Delta c + u \cdot \nabla c &= h(t), \\
u u(0) &= u_0, \\
c(0) &= c_0,
\end{align*}
\]

where \( u = u(t, x) \) is the fluid velocity, \( x \in (0, l) \times (0, l) = \Omega, u(., x) : [0, \infty) \to \mathbb{R}^2 \), \( c(., x) : [0, \infty) \to \mathbb{R} \), \( \nu \) is the kinematic viscosity, \( D \) is the diffusion coefficient and \( c = c(t, x) \) is the concentration of the substance that is diffused into the fluid; \( f = f(t, x) \) represent the body forces and \( h = h(t, x) \) is a production function for \( c \). These are given functions and they are supposed to depend analytically on time. We assume that the function \( h \) it may describe either a production or a consumption of the diffused substance and, as a consequence, it may have either positive or negative values.

We consider periodic boundary conditions. Hence it is assumed that, for every fixed \( t \), \( u(t, \cdot) \) belongs to the space \( \mathcal{H}_1 \overset{def}{=} \{ v \in [L^2_{\text{per}}(\Omega)]^2, \text{div} v = 0 \} \), with the scalar product \( \langle u, v \rangle = \int_\Omega (u_1v_1 + u_2v_2) \, dx \), (where \( u = (u_1, u_2), v = (v_1, v_2) \)) and \( | \cdot | \) the induced norm. For \( c \) we assume \( c(\cdot, \cdot) \in \mathcal{H}_2 \overset{def}{=} L^2_{\text{per}}(\Omega) \), where \( L^2_{\text{per}}(\Omega) \) is endowed with the standard scalar product and the induced norm denoted also by \( | \cdot | \). From the context it will be understood whether we talk about the norm in \( \mathcal{H}_1 \) or that in \( \mathcal{H}_2 \).

We assume that for every \( t \), \( f(t) \in \mathcal{H}_1 \), and \( h(t) \in \mathcal{H}_2 \). As far as the dependence on \( t \) is concerned the most realistic hypothesis is that of periodicity in time for the functions \( f(\cdot) \) and \( h(\cdot) \). Anyhow, we suppose that these functions are bounded: there is a number \( M_f > 0 \) such that \( |f(t)| \leq M_f \), and there is a number \( M_h > 0 \) such that \( |h(t)| \leq M_h \) for every \( t > 0 \).

We also use the spaces \( \mathcal{V}_1 \overset{def}{=} \{ u \in [H^1_{\text{per}}(\Omega)]^2 : \text{div} u = 0 \} \), with the scalar product \( \langle (u, v) \rangle = \sum_{i,j=1}^2 (\frac{\partial u_i}{\partial x_j}, \frac{\partial v_i}{\partial x_j}) \), and \( \mathcal{V}_2 \overset{def}{=} H^1_{\text{per}}(\Omega) \) with the
scalar product \(((c_1, c_2)) = \sum_{j=1}^{2} \left( \frac{\partial c_1}{\partial x_j}, \frac{\partial c_2}{\partial x_j} \right)\). The induced norms are denoted by \(\| \|\).

The operator \(A \overset{def}{=} -\Delta\) is defined on \(D(A) = \left\{ u \in \left[H^2_{\text{per}} (\Omega) \right] : \text{div} u = 0 \right\}\) and it is self-adjoint.

For the bilinear form \(B(u, v) = (u \cdot \nabla) v\) the following inequalities \([4], [6], [8]\)

\[
\begin{align*}
|B(u, v)| & \leq c_1 |u|^{\frac{3}{2}} |\Delta u|^\frac{1}{2} |v|, \quad (\forall) \ u \in D(A), \ v \in V_1, \quad (2.6) \\
|B(u, v)| & \leq c_2 |u| |v| \left[1 + \ln \left( \frac{|\Delta u|^2}{|u|^2} \right) \right]^\frac{1}{2}, \quad (\forall) \ u \in D(A), \ v \in V(2.7) \\
|B(u, v)| & \leq c_3 |u| |v| \left[1 + \ln \left( \frac{|\Delta v|^2}{|v|^2} \right) \right]^\frac{1}{2}, \quad (\forall) \ u \in V_1, \ v \in D(A)(2.8)
\end{align*}
\]

hold \([4], [6], [8]\). We remind the following properties of the three-linear form \(b(u, v, w) = (B(u, v), w)\) (valid for periodic boundary conditions \([5]\)):

\[
b(u, v, w) = -b(u, w, v), b(u, v, v) = 0,
\]

as well as the following inequalities \([6], [5]\)

\[
\begin{align*}
|b(u, v, w)| & \leq k |u|^{\frac{1}{2}} |u|^\frac{3}{2} |v| |w|^{\frac{1}{2}} |w|^\frac{1}{2}, \quad (\forall) \ u, v, w \in V_1, \quad (2.9) \\
|b(u, v, w)| & \leq k |u|^{\frac{3}{2}} |u|^\frac{1}{2} |v|^\frac{3}{2} |\Delta v|^\frac{1}{2} |w|, \quad (\forall) \ u \in V_1, \ v \in D(A), \ w \in \Omega(2.10)
\end{align*}
\]

The operator \(A \overset{def}{=} -\Delta\) is defined on \(D(A) = \left[H^2_{\text{per}} (\Omega) \right]\) and it is self-adjoint.

We define the bilinear form \(B(u, c) = u \nabla c\). The inequalities (2.6)-(2.8) have the immediate consequences

\[
\begin{align*}
|B(u, c)| & \leq c_1 |u|^{\frac{3}{2}} |\Delta u|^\frac{1}{2} |c|, \quad (\forall) \ u \in D(A), \ c \in V_2, \quad (2.11) \\
|B(u, c)| & \leq c_2 |u| |c| \left[1 + \ln \left( \frac{|\Delta u|^2}{|u|^2} \right) \right]^\frac{1}{2}, \quad (\forall) \ u \in D(A), \ c \in V(2.12) \\
|B(u, c)| & \leq c_3 |u| |c| \left[1 + \ln \left( \frac{|\Delta c|^2}{|c|^2} \right) \right]^\frac{1}{2}, \quad (\forall) \ u \in V_1, \ c \in D(A). (2.13)
\end{align*}
\]

For the periodic boundary conditions it is usual to consider the averages of the unknown functions on the periodicity cell, namely \([6] \) for \(u(t, x)\)

\[
u_m (t) = \frac{1}{l^2} \int_{\Omega} u(t, x) \, dx, \ c_m (t) = \frac{1}{l^2} \int_{\Omega} c(t, x) \, dx.
\]
These are solutions of the equations
\[
\frac{d}{dt} u_m(t) = f_m(t), \quad (2.14)
\]
\[
\frac{dc_m}{dt}(t) = h_m(t), \quad (2.15)
\]
where \( f_m, h_m \) are the averages of \( f \) respectively \( h \) over \( \Omega \), therefore we can assume that \( u_m \) and \( c_m \) are given.

The functions \( \tilde{u}(t, x) = u(t, x) - u_m(t) \), \( \tilde{c}(t, x) = c(t, x) - c_m(t) \), are solutions of the equations
\[
\frac{\partial \tilde{u}}{\partial t} - \nu \Delta \tilde{u} + (\tilde{u} \cdot \nabla) \tilde{u} + (u_m \cdot \nabla) \tilde{u} = f - f_m, \quad (2.16)
\]
\[
\frac{\partial \tilde{c}}{\partial t} - D \Delta \tilde{c} + \tilde{u} \cdot \nabla \tilde{c} + u_m \cdot \nabla \tilde{c} = h - h_m. \quad (2.17)
\]

Since \( u_m(t) \) is known, the study of the above equations is very similar to that of (2.1), (2.3). It is then usual [6] to study equations (2.1), (2.3) with the conditions \( u_m(t) = 0, c_m(t) = 0 \). In the following qualitative investigation we adopt these conditions, but we remind that, when a quantitative study is in view, the equations (2.16), (2.17) should be studied, together with (2.14), (2.15).

Thus, further, in the study of (2.1), (2.3), we use the spaces \( \mathcal{H}_1 = \{ u \in \mathcal{H}_1, \ u_m = 0 \} \), \( \mathcal{V}_1 = \{ u \in \mathcal{V}_1, \ u_m = 0 \} \), \( \mathcal{H}_2 = \{ c \in \mathcal{H}_2, \ c_m = 0 \} \), \( \mathcal{V}_2 = \{ c \in \mathcal{V}_2, \ c_m = 0 \} \).

The operator \( A \) is positive-definite on \( D(A) \cap \mathcal{H}_1 \) and has a compact inverse [4]. Similarly, the operator \( A \) is positive-definite on \( D(A) \cap \mathcal{H}_2 \) and has a compact inverse.

### 3. EXISTENCE OF THE SOLUTIONS

The flow of the incompressible viscous fluid in which the diffusion takes place is not affected by the substance that is diffused, whence the decoupling of the equations of the model (2.1)-(2.5).

Hence, for the problem (2.1), (2.4) we have the classical existence and uniqueness results for the equations Navier-Stokes in \( \mathbb{R}^2 \), with periodic boundary conditions.

**Theorem 1** [6].

a) If \( u_0 \in \mathcal{H}_1 \), \( f \) is analytical in time and for every \( t \), \( f(t, \cdot) \in \mathcal{H}_1 \), then the problem (2.1), (2.4) has an unique solution \( u \), analytical in time and such that for every \( t \), \( u(t, \cdot) \in \mathcal{V}_1 \).

b) If, in addition to the hypotheses in a), \( u_0 \in \mathcal{V}_1 \), \( f(t, \cdot) \in \mathcal{V}_1 \), then \( u(t, \cdot) \in D(A) \).
By using the Galerkin-Faedo method we can easily prove the following theorem.

**Theorem 2.** a) In the conditions a) of Theorem 1 and if \( h \) is analytical in time and for every \( t, h(t, \cdot) \in \mathcal{H}_2, c_0 \in \mathcal{H}_2 \), then there is a unique solution \( c \) of the problem (2.3)-(2.5), analytical in time and such that \( c(t, \cdot) \in \mathcal{H}_2 \).

b) In the conditions b) of Theorem 1 and if \( c_0 \in \mathcal{V}_2, h(t, \cdot) \in \mathcal{V}_2 \) then \( c(t, \cdot) \in D(A), \forall t > 0 \).

Since, by Theorems 1, 2 the existence and uniqueness of the solution \((u, c)\) is global with respect to time, it follows that the problem generates two semi-dynamical systems: \( \{S_1(t)\}_{t \geq 0}, S_1(t) : \mathcal{H}_1 \rightarrow \mathcal{H}_1, S_1(t) u_0 = u(t, \cdot, u_0), \)
where \( u(t, x, u_0) \) is the solution of (2.1), (2.2), (2.4), and \( \{S(t)\}_{t \geq 0}, S(t) : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathcal{H}_1 \times \mathcal{H}_2, S(t) (u_0, c_0) = (u(t, \cdot, u_0), c(t, \cdot, u_0, c_0)) \), where \( c(t, x, u_0, c_0) \) is the solution of (2.3), (2.5), with \( u \) solution of (2.1), (2.2), (2.4).

### 4. DISSIPATIVITY OF THE PROBLEM

It was proved [1], [10], that the semi-dynamical system generated by (2.1), (2.2), (2.4) is dissipative, in the sense that in \( \mathcal{H}_1 \) there is an absorbing ball for it. More precisely, there is a \( \rho_0 > 0 \) such that for every \( R > 0 \), there is a \( t_0(R) > 0 \) with the property that for every \( u_0 \in \mathcal{H}_1 \) with \( |u_0| \leq R \), we have \( |S_1(t) u_0| \leq \rho_0 \) for \( t > t_0(R) \). In addition, there is an absorbing ball in \( \mathcal{V}_1 \) for \( \{S_1(t)\} \), i.e. there is a \( \rho_1 > 0 \) such that for every \( R > 0 \), \( |u_0| \leq R \) implies \( |S_1(t) u_0| \leq \rho_1 \) for \( t > t_1(R) \).

The component \( c \) of the solution \((u, c)\) is also dissipative both in \( \mathcal{H}_2 \) and \( \mathcal{V}_2 \). More precisely, the following result holds.

**Theorem 3. a)** There is a \( \eta_0 > 0 \) with the property that for every \( R > 0 \) there is a \( t_{c0}(R) > 0 \) such that for \( |c_0| \leq R \)

\[ |c(t, \cdot, u_0, c_0)| \leq \eta_0, \ t \geq t_{c0}(R). \]

**b)** There is a \( \eta_1 > 0 \) with the property that for every \( R, R_c > 0 \) there is a \( t_{c1}(R, R_c) > 0 \) such that

\[ \|c(t, \cdot, u_0, c_0)\| \leq \eta_1 \]

for \( |u_0| \leq R, \ |c_0| \leq R_c, \ t \geq t_{c1}(R, R_c). \)

The proof of this theorem is contained in the proof of Theorem 2 and we do not insist on it.
5. SPLITTING OF SPACES, PROJECTION OF EQUATIONS

The eigenvalues of $A = -\Delta$ in $\mathcal{H}_1$ are

$$\lambda_{j_1,j_2} = \frac{4\pi^2}{l^2} (j_1^2 + j_2^2), \quad (j_1, j_2) \in \mathbb{Z}^2 \setminus \{(0,0)\}. \quad (5.18)$$

For $\mathcal{H}_1$ there is a total system $f$ whose elements are eigenfunctions of $A = -\Delta$. They have the form $[8] \frac{\sqrt{2}}{l j} \sin \left(2\pi \frac{lx}{l^2}\right), \; \frac{\sqrt{2}}{l j} \cos \left(2\pi \frac{lx}{l^2}\right), \text{ where } j = (j_1, j_2), \; \tilde{j} = (j_2, -j_1), \; |j| = (j_1^2 + j_2^2)^{1/2}, \; \Lambda := \frac{j_1 x_1 + j_2 x_2}{l}, \; (j_1, j_2) \in \mathbb{Z}^2 \setminus \{(0,0)\}.$

The eigenvalues of $A = -\Delta$ are also (5.18) and there is a total system for $\mathcal{H}_2$ formed of eigenfunctions of $A = -\Delta$, namely those having the form $\frac{\sqrt{2}}{l} \sin \left(2\pi \frac{lx}{l^2}\right), \; \frac{\sqrt{2}}{l} \cos \left(2\pi \frac{lx}{l^2}\right)$.

Due to the simetries of the functions sin and cos to every ordered pair of natural numbers $(j_1, j_2)$, for both operators $A$ and $A$, four eigenfunctions as above (with “+” or “-” between the two terms $\frac{j_1 x_1}{l}$ and $\frac{j_2 x_2}{l}$) correspond.

We fix a value $m \in \mathbb{N}$ and consider the set $\Gamma_m$ of eigenvalues $\lambda_{j_1,j_2}$ having $|j_1|, |j_2| \leq m$.

Denote $\Lambda := \frac{4\pi^2}{l^2}, \; \Lambda := \frac{4\pi^2}{l^2} (m + 1)^2,$ $\delta := \frac{\lambda}{\Lambda} = \frac{1}{(m+1)^2}$.

Let $w_1, \ldots, w_{m_1}$ (resp. $w_1, \ldots, w_{m_1}$) be the eigenfunctions of $A$ (resp. $A$) corresponding to $\lambda_{j_1,j_2} \in \Gamma_m$.

Let $L (u, v, \ldots)$ stand for the linear space spanned by $u, v, \ldots$.

Denote by $P$ the projection operator on $L (w_1, \ldots, w_{m_1})$, with $Q$ the projection operator on $L (w_{m_1+1}, \ldots)$, by $P$ the projection operator on $L (w_1, \ldots, w_{m_1})$, and by $Q$ the projection operator on $L (w_{m_1+1}, \ldots)$.

For the solution $u$ of the Navier-Stokes equation, $p = Pu, \; q = Qu$, and, similarly, for the diffusion component $c, \; c_p = Pc, \; c_q = Qc$.

By projecting the equations (2.1), (2.3) on the above introduced spaces, we have,

$$\frac{dp}{dt} - \nu \Delta p + PB(p + q) = Pf, \quad (5.19)$$

$$\frac{dq}{dt} - \nu \Delta q + QB(p + q) = Qf, \quad (5.20)$$

$$\frac{dc_p}{dt} - D \Delta c_p + PB(u, c) = Ph, \quad (5.21)$$

$$\frac{dc_q}{dt} - D \Delta c_q + QB(u, c) = Qh. \quad (5.22)$$
6. ESTIMATES FOR THE SMALL COMPONENT OF THE SOLUTION

It is already known that for the Navier Sokes equations the component $q(t)$ of the solution is small in the sense that it has the order of $\delta$ for $t$ large enough. In [8] it is proved that for every initial state $u_0$ with $|u_0| \leq R$ there is a moment $t_2(R)$ such that, for $t \geq t_2(R)$,

$$|q(t)| \leq K_0 L^{\frac{1}{2}} \delta, \quad \|q(t)\| \leq K_1 L^{\frac{1}{2}} \delta^{\frac{1}{2}},$$

$$|q'(t)| \leq K_2 L^{\frac{1}{2}} \delta, \quad |\Delta q(t)| \leq K_3 L^{\frac{1}{2}},$$

where, for the chosen projection spaces, we have $L = \ln(1 + 2m^2)$. Here and in the sequel the prime means time derivative.

Let us prove a similar result for $c_q$. We assume in the following that for a fixed $R$, we have $|u_0| \leq R$.

**Theorem 4.** There is a moment $t_{c2}(R)$ such that for every $t \geq t_{c2}(R)$,

$$|c_q(t)| \leq J_0 L^{\frac{1}{2}} \delta, \quad \|c_q(t)\| \leq J_1 L^{\frac{1}{2}} \delta^{\frac{1}{2}}, \quad |c_q'(t)| \leq J_2 L^{\frac{1}{2}} \delta, \quad |\Delta c_q(t)| \leq J_3 L^{\frac{1}{2}},$$

(6.23) (6.24)

hold, where $J_0, J_1, J_2, J_3$ are independent of $m$.

**Proof.** In the sequel we frequently use the embedding inequalities

$$\|c_q\| \geq \Lambda^{\frac{1}{2}} |c_q|, \quad |\Delta c_q| \geq \Lambda |c_q|,$$

easy to derive by considering the Fourier series of $c_q$ and the fact that $\Lambda$ is the least eigenvalue of $A$ on $L_2(w_{m+1}, \ldots)$.

Taking the scalar product of (5.22) by $c_q$ we have

$$\left( \frac{d c_q}{dt}, c_q \right) - D(\Delta c_q, c_q) + (u \nabla c, c_q) = (Qh, c_q),$$

such that, by using (2.12) and (2.13), we obtain

$$\frac{1}{2} \frac{d}{dt} |c_q|^2 + D\Lambda |c_q|^2 \leq |(Q (p \nabla c_p), c_q)| + |(Q (q \nabla c_p), c_q)| + |(Q h, c_q)| \leq CL^{1/2} \|p\| \|c_p\| |c_q| + CL^{1/2} \|q\| \|c_p\| |c_q| + |Q h| |c_q| \leq \frac{C^2 L}{D\Lambda} \rho^2 \eta^2 + \frac{D\Lambda}{4} |c_q|^2 + \frac{C^2 L}{D\Lambda} \eta^2 + \frac{M^2}{D\Lambda} + \frac{D\Lambda}{4} |c_q|^2 \quad \text{and (since } 1 < L \text{ for } m > 1) \text{ it follows}
\[ \frac{d}{dt} |c_q|^2 + D\Lambda |c_q|^2 \leq \delta \left( \frac{C^2 L}{D\lambda} \rho_1^2 \eta_1^2 + \frac{C^2 L}{D\lambda} \delta \eta_1^2 + \frac{M^2}{D\lambda} \right) = K_0 L \delta, \]

whence, by applying Gronwall Lemma, (6.23), where \( J_0 = \sqrt{K_0} \).

Taking the scalar product of (5.22) by \(-\Delta c_q\), we obtain

\[ \frac{1}{2} \frac{d}{dt} \|c_q\|^2 + D\Lambda \|c_q\|^2 \leq \|(u \nabla c, \Delta c_q)\| + \|(Qh, \Delta c_q)\| \]

\[ \leq \|(p \nabla c_p, \Delta c_q)\| + \|(p \nabla c_q, \Delta c_q)\| + \|(q \nabla c_q, \Delta c_q)\| + \|(Qh, \Delta c_q)\| \]

\[ \leq C L \left( \|p\| \|c_p\| + \|p\| \|c_q\| + \|q\| \|c_q\| \right) \| \Delta c_q \|

\[ + C |q|^2 \|A_q\| \|c_q\| \| \Delta c_q \| + |Qh| \| \Delta c_q \|. \]

The use of the inequality \( \|c_q\| \leq \|c\| \leq \eta_1 \) further leads to

\[ \frac{d}{dt} \|c_q\|^2 + D\Lambda \|c_q\|^2 \leq \frac{C L}{D} (\rho_1^2 + \delta) \eta_1^2 + \frac{1}{D} M^2, \]

and

\[ \frac{d}{dt} \|c_q(t)\|^2 \leq \|c_q(0)\|^2 e^{-D\Lambda t} \left( \frac{C L}{D} (\rho_1^2 + \delta) \eta_1^2 + \frac{1}{D^2} M^2 \right), \]

where \( C \) is a generic notation for a constant not depending on \( m \).

Since, usually, \( m \) is large, \( L \) is large too and we obtain (6.23), for \( t \) great enough.

Then, like for the Navier-Stokes equation [6], we can prove that the extension of the semigroup to the complex time variable is analytic in time in a neighbourhood of the time axis, and by using the Cauchy formula, we obtain (6.24).

Finally, by using (5.22) and the above estimates, the inequality (6.24) follows. \( \square \)

7. THE FAMILY OF A.I.M.S FOR THE NAVIER STOKES EQUATIONS

The family of a.i.m.s for the Navier Stokes problem, is defined in [8] as follows. The first one is the graph \( M_0 \) of the function \( \Phi_0 : P^\prime \kappa_1 \rightarrow Q^\prime \kappa_1 \), that satisfies the relation

\[ -\nu \Delta \Phi_0 (X) + QB(X) = Qf, \text{ where } X \in P^\prime \kappa_1. \]
We remark that this relation is similar to the equation \(-\nu \Delta q + QB(p) = Qf\), obtained from (5.20) by neglecting \(\frac{dq}{dt}\) in the presence of \(\Delta q\) and \(q\) in the presence of \(p\). Therefore, \(\Phi_0(X)\) has the following form

\[\Phi_0(X) = (-\nu \Delta)^{-1} (Qf - QB(X)).\]

The next a.i.m. defined in [8] is \(M_1\), the graph of the function \(\Phi_1: P^1 \rightarrow Q^1\), given by the solution of the problem

\[-\nu \Delta \Phi_1(X) + QB(X) + QB(X, \Phi_0(X)) + QB(\Phi_0(X), X) = Qf,\]

that is

\[\Phi_1(X) = - (\nu \Delta)^{-1} [Qf - QB(X) - QB(X, \Phi_0(X)) - QB(\Phi_0(X), X)].\]

For \(j \geq 2\), the a.i.m. \(M_j\) is defined as the graph of \(\Phi_j: P^j \rightarrow Q^j\), with \(\Phi_j(X)\) the solution of

\[-\nu \Delta \Phi_j(X) + QB(X) + QB(X, \Phi_{j-1}(X)) + QB(\Phi_{j-1}(X), X) + QB(\Phi_{j-2}(X), \Gamma_{u,j-2}(X)) = Qf,\]

where \(D\Phi_{j-2}(X)\) is the differential of \(\Phi_{j-2}(X)\), and

\[\Gamma_{u,j-2}(X) = \nu \Delta X - PB(X + \Phi_{j-2}(X)) + Pf.\]

Hence

\[\Phi_j(X) = - (\nu \Delta)^{-1} [Qf - QB(X) - QB(X, \Phi_{j-1}(X)) - \Phi_{j-2}(X)]\]

where \(D\Phi_{j-2}(X)\) is the differential of \(\Phi_{j-2}(X)\), and

\[\Gamma_{u, j-2}(X) = \nu \Delta X - PB(X + \Phi_{j-2}(X)) + Pf.\]

This construction is better understood with the definition of the so-called "induced trajectories".

### 8. INDUCED TRAJECTORIES FOR THE NAVIER-STOKES PROBLEM

In [8] a family of functions, \(\{k_{j,m}; j \in \mathbb{N}\}\), is defined by the equalities

\[-\nu \Delta k_{0,m} + QB(p) = Qf, \quad (8.25)\]

\[-\nu \Delta k_{1,m} + QB(p, k_{0,m}) + QB(k_{0,m}, p) = 0,\]
\(-\nu \Delta k_{jm} + QB(p, k_{j-1,m}) + QB(k_{j-1,m}, p) + \sum_{s=0}^{j-3} QB(k_{j-2,m}, k_s,m) + \sum_{r=0}^{j-3} QB(k_r,m, k_{j-2,m}) + QB(k_{j-2,m}, k_{j-2,m}) + k'_{j-2,m} = 0, \quad j \geq 2,\)

where \(p(t)\) is, as before, the projection through \(P\) of the solution \(u(t)\).

Then the functions \(q_j\) are defined \([8]\) through the relations
\[
q_j = k_{0,m} + \ldots + k_{j,m},
\]
and therefore, satisfy the equations
\[
-\nu \Delta q_0 + QB(p) = Qf,
-\nu \Delta q_1 + QB(p) + QB(p, q_0) + QB(q_0, p) = Qf,
-\nu \Delta q_j + q'_j - 2 + QB(p) + QB(p, q_j-1)
+ QB(q_j-1, p) + QB(q_{j-2}, q_{j-2}) = Qf, \quad j \geq 2.
\]

The connection between the definitions of the a.i.m.s and those of the functions \(q_j\) is obvious. Moreover, it is clear that for every \(t \geq 0, (p(t), q_j(t))\) lies on \(M_j\) for \(j \geq 0\).

The functions of time \(u_j(t) = p(t) + q_j(t)\) are called induced trajectories associated with the trajectory \(u(t) = p(t) + q(t)\).

The following inequalities
\[
|k_{jm}| \leq \kappa_j \delta^{1+j/2} L^{(j+1)/2}, \quad \|k_{jm}\| \leq \kappa_j \delta^{1+j/2} L^{(j+1)/2}, \quad (8.26)
\]
\[
|k'_{jm}| \leq \kappa_j \delta^{1+j/2} L^{(j+1)/2}, \quad (8.27)
\]
\[
|q_j| \leq \kappa_j \delta L^{1/2}, \quad \|q_j\| \leq \kappa_j \delta^{1/2} L^{1/2}, \quad |q'_j| \leq \kappa_j \delta L^{1/2}, \quad (8.28)
\]

are proved in \([8]\), where \(\kappa_j\) are independent of \(m\), but depend on \(j, \nu, |f|, \lambda\).

They will help us to estimate the distance between the trajectories of the Navier-Stokes problems and the a.i.m.s.

Indeed, since \(u_j(t) \in M_j\),
\[
dist_{\mathcal{G}_1}(u(t), M_j) \leq dist(u(t), u_j(t)) = |q(t) - q_j(t)|.
\]

By estimating this last norm and the similar one for \(V_1\), in \([8]\) and \([9]\) it is proved that
\[
dist_{\mathcal{G}_1}(u(t), M_j) \leq \kappa_j L^{\frac{j+1}{2}} \delta^{\frac{j+3}{2}},
\]
\[
dist_{V_1}(u(t), M_j) \leq \kappa'_j L^{\frac{j}{2}} \delta^{\frac{j+2}{2}}.
\]
9. "INDUCED TRAJECTORIES" FOR THE DIFFUSED SUBSTANCE

Following the above reasoning, we define the functions $h_{j,m} : \mathbb{R}^+ \to Q^H \mathcal{C}_2$, $j \in \mathbb{N}$, as solutions of the equations

\begin{align*}
- D \Delta h_{0,m} + QB (p, c_p) & = Qh, \quad (9.29) \\
- D \Delta h_{1,m} + QB (p, h_{0,m}) + QB (k_{0,m}, c_p) & = 0, \quad (9.30)
\end{align*}

\begin{align*}
h'_{j-2,m} - D \Delta h_{j,m} + QB (p, h_{j-1,m}) + QB (k_{j-1,m}, c_p) + \\
+ \sum_{r=0}^{j-3} QB (k_{r,m}, h_{j-2,m}) + \sum_{s=0}^{j-3} QB (k_{j-2,m}, h_{s,m}) + QB (k_{j-2,m}, h_{j-2,m}) & = 0, \quad (9.31)
\end{align*}

for $j \geq 2$.

In these definitions $p = p(t) = Pu(t)$ and $c_p = c_p(t) = Pc(t)$, where $(u(t), c(t))$ is the solution of the diffusion problem, and $k_{j,m}$ are given by (8.25).

We prove the following

**Theorem 5.** There are some constants $\kappa_j, \varsigma_j$ independent of $m$, but depending on $j$ and on $\nu, |f|, \lambda_1, D$, and there is a $t_{\mathcal{A}}(R)(\geq t_{\mathcal{A}}(R))$ such that for $m, j$, and every $t \geq t_{\mathcal{A}}(R)$, the estimates hold, $(\forall) j \geq 0$.

**Proof.** We look for an estimate for the norm of $h_{0,m}$. From (9.29), by using (2.12), we sucessively obtain

\begin{align*}
| h_{j,m} | & \leq \kappa_j L^{\frac{i}{2} + 1} | h_{j,m} | \leq \kappa_j L^{\frac{i}{2} + 1} \delta^{\frac{i}{2} + 1}, \quad (9.32) \\
| h'_{j,m} | & \leq \kappa_j L^{\frac{i}{2} + 1} | h'_{j,m} | \leq \kappa_j L^{\frac{i}{2} + 1} \delta^{\frac{i}{2}}, \quad (9.33)
\end{align*}

With $\kappa_0 = \frac{c \rho_1 \eta_1 + M_h D}{D \lambda}, \ k_0 = \frac{c \rho_1 \eta_1 + M_h D}{D \lambda^2}$, the relations (9.32) are proved for $j = 0$. 

\begin{align*}
| D \Delta h_{0,m} | & = | Q (p \nabla c_p) | + | Qh | \leq | p \nabla c_p | + | Qh | \\
& \leq c L^{\frac{1}{2}} \| p \| \| c_p \| + M_h,
\end{align*}
Since \(u\) and \(c\) are analytic in time, so are \(p\) and \(c_p\). Moreover, these last functions are the restrictions to the real axis of functions of a complex variable that are analytic in a neighbourhood of the real axis. It follows that \(h_{0,m}\) has the same property. Then, by using the same reasoning as for the Navier-Stokes equations [6], we prove the first relation of (9.33), for \(j = 0\).

From the definition of \(h_{1,m}\), with (2.12), (2.13), and by taking into account the estimates for \(h_{0,m}\), we successively deduce

\[
|D\Delta h_{1,m}| \leq c \|p\| \|h_{0,m}\| L^{\frac{3}{2}} + c \|k_{0,m}\| \|c_p\| L^{\frac{3}{2}}
\]

\[
\leq c \rho_1 \eta_0 \delta^{1/2} + c \rho_0 \delta^{1/2} L \eta_1 \leq (c \rho_1 \eta_0 + c \rho_0 \eta_1) L \delta^{1/2},
\]

\[
|h_{1,m}| \leq \frac{(c \rho_1 \eta_0 + c \rho_0 \eta_1) L \delta^{1/2}}{D\lambda},
\]

\[
||h_{1,m}|| \leq \frac{(c \rho_1 \eta_0 + c \rho_0 \eta_1) L \delta^{1/2}}{D\lambda^2},
\]

We find as for \(h_{0,m}\), that \(h_{1,m}\) is analytic in time and, then, \(h'_{1,m}\) is majorized by an expression of the order of \(D\delta^{1/2}\).

For \(k \geq 2\) we proceed by induction. We assume that the inequalities (9.32), (9.33) are true and that \(h_{j,m}\) is the restriction to the real axis of an analytic function of \(t\) for every \(j \leq k - 1\). From the definition of \(h_{k,m}\), we have

\[
|D\Delta h_{k,m}| \leq |h'_{k-2,m}| + |Q(p \nabla h_{k-1,m})| + |Q(k_{k-1,m} \nabla c_p)| + \sum_{r=0}^{k-3} |Q(k_{r,m} \nabla h_{k-2,m})| + \sum_{s=0}^{k-3} |Q(k_{s,m} \nabla h_{k-2,m})| + \sum_{s=0}^{k-3} |Q(k_{s,m} \nabla h_{k-2,m})|
\]

\[
\leq CL^{\frac{k}{k-1}} \delta^{\frac{k}{2}} + CL^{\frac{k}{k-1}} ||p|| ||h_{k-1,m}|| + CL^{\frac{k}{k-1}} ||k_{k-1,m}|| ||c_p|| +
\]

\[
+ CL^{\frac{k}{k-1}} \sum_{r=0}^{k-3} ||k_{r,m}|| ||h_{k-2,m}|| + CL^{\frac{k}{k-1}} \sum_{s=0}^{k-3} ||k_{s,m}|| ||h_{s,m}|| +
\]

\[
+ CL^{\frac{k}{k-1}} ||k_{k-2,m}|| ||h_{k-2,m}||,
\]

whence

\[
|D\Delta h_{k,m}| \leq CL^{\frac{k}{k-1}} \delta^{\frac{k}{2}} + CL^{\frac{k}{k-1}} \rho_1 \eta_{k-1} L^\frac{k}{2} \delta^{k/2} + CL^{\frac{k}{k-1}} \eta_1 \kappa_{k-1} \delta^{\frac{k}{2}} L^{\frac{k+1}{2}} +
\]

\[
+ \kappa_{k-2} L^\frac{k}{2} \delta^\frac{k-1}{2} \sum_{r=0}^{k-3} \kappa_r \eta^{\frac{k+1}{2}} L^{\frac{k+r}{2}} +
\]

\[
+ \kappa_{k-2} L^\frac{k}{2} \delta^\frac{k-1}{2} \sum_{s=0}^{k-2} \eta_{s} L^\frac{k+s+1}{2} \delta^\frac{k+s+1}{2} \leq C_k L^\frac{k}{2} \delta^{\frac{k}{2}}.
\]
Consequently,

$$|h_{k,m}| \leq \frac{C''_k}{D^k} L_1^{\frac{k+1}{2}} \delta_k^2 = \frac{C''_k}{D^k} L_1^{\frac{k+1}{2}} \delta_k^{k+1} = \gamma_k L_1^{\frac{k+1}{2}} \delta_k^{k+1},$$

where $\gamma_k = \frac{C''_k}{D^k}$, hence, the conclusion. The other inequalities are proved as for $j = 0$ and $j = 1$. □

We define now $c_{q_j} : \mathbb{R}^+ \to Q\mathcal{H}_2$, $j \in \mathbb{N}$ by

$$c_{q,j} = h_{0,m} + h_{1,m} + \ldots + h_{j,m}.$$  

As a consequence of their definition, these new functions satisfy the relations

$$-D\Delta c_{q,0} + QB(p, c_p) = Qh, \quad (9.34)$$

$$-D\Delta c_{q,1} + QB(p, c_p) + QB(p, c_{q,0}) + QB(q_0, c_p) = Qh, \quad (9.35)$$

$$c'_{q,j-2} - D\Delta c_{q,j} + QB(p, c_p) +$$

$$+ QB(p, c_{q,j-1}) + QB(q_{j-1}, c_p) + QB(q_{j-2}, c_{q,j-2}) = Qh,$$

for $j \geq 2$. By summing the relations proved for $h_{k,m}$ we obtain

$$|c_{q,j}| \leq \varsigma_j L_j^{\frac{3}{2}} \delta_j^{\frac{1}{2}}, \quad \|c_{q,j}\| \leq \varsigma'_j L_j^{\frac{3}{2}} \delta_j^{2 \frac{1}{2}}, \quad |c'_{q,j}| \leq \varsigma''_j L_j^{\frac{3}{2}} \delta_j^{\frac{3}{2}}, \quad j \geq 0.$$

Now, along the lines of [8], we define the induced trajectories for $c$

$$c_j(t) = c_p(t) + c_{q,j}(t)$$

and find the distance between these functions and $c(t)$. We define $\xi_j(t) = c_j(t) - c(t) = c_{q,j}(t) - c_q(t)$.

**Theorem 6.** There are some constants, $\zeta_j$, independent of $m$ but depending on $j$ and $\nu$, $|f|$, $\lambda_1$, $D$ and there is a moment $t_3$ such that for every $m$, $j$ and for $t \geq t_3$ the inequalities

$$|\xi_j| \leq \zeta_j L_1^{\frac{j+1}{2}} \delta_1^{\frac{j+3}{2}}, \quad (9.37)$$

$$\|\xi_j\| \leq \zeta'_j L_1^{\frac{j+1}{2}} \delta_1^{\frac{j+1}{2}}, \quad (9.38)$$

$$|\xi'_j| \leq \zeta''_j L_1^{\frac{j+1}{2}} \delta_1^{\frac{j+3}{2}}, \quad (9.39)$$
hold for $\forall j \geq 0$.

**Proof.** We start with the definition of $c_{q,0}$ and from the equation for $c_q$, written in the form

\[
\Delta c_{q,0} = -\frac{1}{D} \left[ Qh - Q (p \nabla c_p) \right], \\
\Delta c_q = -\frac{1}{D} \left[ Qh - Q (u \nabla c) - \frac{dc_q}{dt} \right].
\]

By substracting the two relations, we obtain

\[
\Delta (c_{q,0} - c_q) = \frac{1}{D} \left[ Q (p \nabla c_p) - Q (u \nabla c) - \frac{dc_q}{dt} \right] - \frac{1}{D} \left[ Q (p \nabla c_q) + Q (q \nabla c_p) + Q (q \nabla c_q) + \frac{dc_q}{dt} \right],
\]

whence, by using the same inequalities as above, we have

\[
|\xi_0| = |c_{q,0} - c_q| \leq \frac{1}{D\Lambda} \left| Q (p \nabla c_q) + Q (q \nabla c_p) + Q (q \nabla c_q) + \frac{dc_q}{dt} \right| \leq C L^\frac{1}{2} \delta^\frac{3}{2} = C L^\frac{1}{2} \delta^\frac{3}{2} = \kappa_0 L^\frac{1}{2} \delta^\frac{3}{2},
\]

i.e. the inequality (9.37) for $j = 0$. Similarly, we obtain (9.38) for $j = 0$.

As we can see from (9.40), $\xi_0$ is analytical in $t$ and can be extended to a function of a complex variable, analytical in a neigbourhood of the $t$ axis. Then, by using the Cauchy formula, it follows that $|\xi'_0|$ is of the same order as $|\xi_0|$.

Now, for $|\xi_1|$ we have

\[
\Delta (c_{q,1} - c_q) = \frac{1}{D} \left[ Q (p \nabla c_p) + Q (p \nabla c_{q,0}) + Q (q_0 \nabla c_p) - Q (u \nabla c) - \frac{dc_q}{dt} \right] - \frac{1}{D} \left[ Q (p \nabla (c_{q,0} - c_q)) + Q ((q_0 - q) \nabla c_p) - Q (q \nabla c_q) - \frac{dc_q}{dt} \right],
\]

and, by taking the norm and using the previous results, we get (9.37) and (9.38) for $j = 1$.

The inequality for $|\xi'_1|$ is obtained as that for $|\xi'_0|$.

From this point on, we proceed by induction, assuming that (9.37), (9.38) and (9.39) hold for every $j \leq k - 1$. 


We have
\[
\Delta (c_{q,k} - c_q) = \frac{1}{D} [Q (P \nabla (c_{q,k-1} - c_q)) + Q ((q_{k-1} - q) \nabla c_p) + \\
+ Q((q_{k-2} - q) \nabla c_{q,k-2}) + Q(q \nabla (c_{q,k-2} - c_q)) + c'_{q,k-2} - c'_q],
\]
hence
\[
|\xi_k| = |c_{q,k} - c_q| \leq \frac{1}{D} \delta \left[ CL^{\frac{1}{2}} \rho_1 L_c^{\frac{3}{2}} \delta^{\frac{k+1}{2}} + CL^{\frac{1}{2}} \eta_1 L_c^{\frac{3}{2}} \delta^{\frac{k+1}{2}}
\right. \\
\left. + CL^{\frac{1}{2}} \delta^{\frac{k+1}{2}} \delta^{\frac{k+1}{2}} \delta^{\frac{1}{2}} + CL^{\frac{1}{2}} \delta^{\frac{k+1}{2}} \delta^{\frac{1}{2}} + CL^{\frac{1}{2}} \delta^{\frac{k+1}{2}} \right]
\leq CL^{\frac{k+1}{2}} \delta^{\frac{k+1}{2}} ,
\]
\[
\|\xi_k\| \leq CL^{\frac{k+1}{2}} \delta^{\frac{k+1}{2}}.
\]
By using the Cauchy formula for the extension of \(\xi_k(t)\) to a complex variable (analytical in a neighbourhood of the time axis), we obtain for \(|\xi'_k|\) the same type of estimate as for \(|\xi_k|\). □

10. APPROXIMATE INERTIAL MANIFOLDS FOR OUR PROBLEM

Here we construct the a.i.m.s for the diffusion problem (2.1)-(2.5).

The first a.i.m. is the manifold \(\tilde{M}_0\) defined as the graph of the \(\Psi_0 : P\bar{t}_1 \times P\bar{t}_2 \to Q\bar{t}_1 \times Q\bar{t}_2\),
\[
\Psi_0 (X, X) = \left(- (\nu \Delta)^{-1} (Qf - QB(X)), - (D\Delta)^{-1} [Qh - QB(X, X)]\right).
\]
Denote the components of \(\Psi_0 (X, X)\) by \((\Psi_{u0}(X), \Psi_{c0}(X, X))\) (obviously \(\Psi_{u0}(X) = \Phi_0(X)\)).

The second a.i.m., denoted by \(\tilde{M}_1\), is the graph of \(\Psi_1 : P\bar{t}_1 \times P\bar{t}_2 \to Q\bar{t}_1 \times Q\bar{t}_2\), \(\Psi_1 (X, X) = (\Psi_{u1}(X), \Psi_{c1}(X, X))\), where \(\Psi_{u1}(X) = \Phi_1(X)\) and
\[
\Psi_{c1}(X, X) = -\frac{1}{D} \Delta^{-1} [Qh - QB(X, X) - QB(X, \Psi_{c0}(X, X)) - QB(\Psi_{u0}(X), X)].
\]

For \(j \geq 2\) we define the function \(\Psi_j : P\bar{t}_1 \times P\bar{t}_2 \to Q\bar{t}_1 \times Q\bar{t}_2\), \(\Psi_j (X, X) = (\Psi_{u,j}(X), \Psi_{c,j}(X, X))\), where \(\Psi_{u,j}(X) = \Phi_j(X)\) and
\[ \Psi_{c,j}(X,X) = - (D\Delta)^{-1} [Q h QB(X,X) - QB(X,\Psi_{c,j-1}(X,X)) - \\
QB(\Psi_{u,j-1}(X),X) - QB(\Psi_{u,j-2}(X),\Psi_{c,j-2}(X,X)) - \\
- D\Psi_{c,j-2}(X,X) (\Gamma_{u,j-2}(X),\Gamma_{c,j-2}(X,X))], \]

where \( D\Psi_{c,j-2}(X,X) \) is the differential of \( \Psi_{c,j-2}(X,X) \) and

\[ \Gamma_{c,j-2}(X,X) = D\Delta X - PB(X + \Psi_{u,j-2}(X),X + \Psi_{c,j-2}(X,X)) + Ph. \]

The graph of this function is the a.i.m. \( \tilde{M}_j \) of the diffusion problem.

**Theorem 7.** For \( t \) large enough, the distance between the solution of the diffusion equations and the a.i.m. \( \tilde{M}_j \) is bounded by

\[ \text{dist}_{\mathcal{H}_1^2} \left( (u(t),c(t)), \tilde{M}_j \right) \leq K_1 L^{1+j} \delta^{3+j}. \] (10.41)

**Proof.** For \( j = 0 \) and \( j = 1 \), by definition, \( \Psi_{u,j}(p) = q_j(p) \), \( \Psi_{c,j}(p,p) = q_{p,j}(p,p) \); the inequality (10.42) follows directly from the result (10.41) of [8] and from (2.1), \( j = 0 \), resp. \( j = 1 \) of our Theorem 6. For \( j \geq 2 \), we need a special proof for (10.42). We take as initial step in our inductive reasoning the case \( j = 2 \). We have

\[ \Psi_{c,2}(p,c_p) - c_q = \Psi_{c,2}(p,c_p) - c_{q,2} + c_{q,2} - c_q = \]

\[ = \Psi_{c,2}(p,c_p) - c_{q,2} + \xi_2, \]

such that the difference between \( \Psi_{c,2}(p,c_p) \) and \( c_{q,2} \) satisfies the identities

\[ D\Delta [\Psi_{c,2}(p,c_p) - c_{q,2}] = QB(p,\Psi_{c,1}(p,c_p) - c_{q,1}) + \]

\[ + QB(\Psi_{u,1}(p) - q_1,c_p) + \\
+ QB(\Psi_{u,0}(p),\Psi_{c,0}(p,c_p)) - QB(q_0,c_{q,0}) + \\
+ D\Psi_{c,0}(p,c_p) (\Gamma_{u,0}(p),\Gamma_{c,0}(p,c_p) - c_{q,0}) \]

\[ = D\Psi_{c,0}(p,c_p) (\Gamma_{u,0}(p),\Gamma_{c,0}(p,c_p) - c_{q,0}), \]

whence

\[ |\Psi_{c,2}(p,c_p) - c_{q,2}| \leq \left| (D\Delta)^{-2} [Q (\Gamma_{u,0}(p) \nabla c_p) + Q (p \nabla \Gamma_{c,0}(p,c_p)) \right| \\
- Q (p' \nabla c_p) - Q (p \nabla c'_p) \right| \]

\[ \leq \frac{D}{\lambda^2} \delta^2 |QB(\Gamma_{u,0}(p) - p',c_p) + QB(p,\Gamma_{c,0}(p,c_p) - c'_p)| \]

\[ \leq \frac{D}{\lambda^2} \delta^2 CL \frac{\lambda}{2} (||\Gamma_{u,0}(p) - p'|| ||c_p|| + ||p|| ||\Gamma_{c,0}(p,c_p) - c'_p||). \]
Now,
\[
\|\Gamma_{u,0}(p) - p'\| = \|PB(p + \Phi_0(p)) - PB(p + q)\| = \|PB(p, \Phi_0(p) - q) + PB(\Phi_0(p) - q, p) + \Phi(\Phi_0(p), \Phi_0(p)) - PB(q, q)\| \leq \Lambda^2 CL \frac{1}{2} (||p|| \Phi_0(p) - q) + ||\Phi_0(p) - q|| ||q|| + + ||PB(\Phi_0(p), \Phi_0(p)) - PB(q, q)|| \leq CL \frac{1}{2}\]
and
\[
\|\Gamma_{u,0}(p, c_p) - c'_p\| = \|PB(p) + \Psi_{u,0}(p), c_p + \Psi_{c,0}(p, c_p)\) - PB(u, c)\| \leq \|PB(\Psi_{u,0}(p) - p, c_p) + PB(p, \Psi_{c,0}(p, c_p) - c_q)\| + + ||PB(\Psi_{u,0}(p), \Psi_{c,0}(p, c_p)) - PB(c_p, c_p)|| \leq \Lambda^2 CL \frac{1}{2} (||\Psi_{u,0}(p) - p|| ||c_p|| + ||p|| ||\Psi_{c,0}(p, c_p) - c_q||) + + \Lambda^2 C \left(L^2 ||\Psi_{u,0}(p) - c_p|| ||c_p|| + + ||PB(\Psi_{u,0}(p), \Psi_{c,0}(p, c_p)) - PB(c_p, c_p)|| \leq CL \frac{1}{2},\right.
\]
whence
\[
|\Psi_{c,2}(p, c_p) - c_{q,2}(p, c_p)| \leq CL \frac{1}{2} \delta^2.
\]
Assume that for every \(k \leq j - 1\)
\[
|\Psi_{c,k}(p, c_p) - c_{q,k}(p, c_p)| \leq CL \frac{k+1}{2} \delta^{\frac{k+3}{2}}.
\]
For \(j > 2\), we have
\[
|D\Delta(\Psi_{c,j}(p, c_p) - c_{q,j}(p, c_p))| \leq CL \frac{1}{2} ||p|| ||\Psi_{c,j-1}(p, c_p) - c_{q,j-1}|| + CL \frac{1}{2} ||c_p|| ||\Psi_{u,j-1}(p) - q_{j-1}|| + + C ||\Psi_{u,j-2}(p) - q_{j-2}|| \right| ||\Delta(\Psi_{u,j-2}(p) - q_{j-2})\right| \frac{1}{2} ||c_{q,j-2}|| + + C ||\Psi_{u,j-2}(p)\left| ||\Delta\Psi_{u,j-2}(p)\right| \frac{1}{2} ||\Psi_{c,j-2}(p, c_p) - c_{q,j-2}|| + + ||D\Psi_{c,j-2}(p, c_p)(\Gamma_{u,j-2}(p), \Gamma_{c,j-2}(p, c_p)) - c'_{q,j-2}\right| (10.42)
\]
A careful analysis of the order of magnitude and the use of the induction hypothesis and of the inequalities (2.6)-(2.8), (2.11)-(2.13) leads us to the conclusion that the terms in the right-hand side of (10.42) have the same order of magnitude with respect to \(\delta\), as \(\delta \to 0\).
Hence, finally,

$$\left| \Psi_{c,j} (p,c_p) - c_q \right| = \left| \Psi_{c,j} (p,c_p) - c_{q,j} + c_{q,j} - c_q \right| \leq \left| \Psi_{c,j} (p,c_p) - c_{q,j} \right| + \left| \xi_j \right| \leq CL^{\frac{j+1}{2}} \delta^j \frac{1}{\sqrt{2}}.$$

Since

$$\text{dist} \left( (u(t), c(t)), \tilde{M}_j \right) \leq C \left( |\Phi_j (p) - q| + |\Psi_{c,j} (p,c_p) - c_q (p,c_p)| \right),$$

the conclusion of the theorem follows. □

Acknowledgements

Work supported by the Romanian Ministry of Education and Research within the frame of the grant CEEEX-05-D11-25/5.11.2005

References


The paper formalizes the OR refinement of finite state machines and develops an efficient method to integrate the test sets of the component machines in order to test the flattened machine. The approach supports component-software development since it constructs test sets for an assembly of systems from readily available test sets of the prefabricated parts.

**Keywords:** finite state machine, OR state refinement, test generation, verification, W-method

**2000MSC:** 68N32

## 1. INTRODUCTION

The test set generation problem for protocols modeled by *deterministic finite state machines* has attracted a great deal of interest and has generated a substantial literature. Given two deterministic finite state machines $S$ and $I$, the former representing the specification and the latter the implementation, a test set is a set of input sequences that, if $S$ and $I$ are not equivalent, contains at least one sequence that produces different results when applied to the two machines. The test set is generated from the specification $S$ and, in principle, little information is available about the implementation $I$. For instance, in the so called W-method [7] the only information about the implementation is an upper bound on the number of its states. Originally, the W-method was developed in the context of *completely specified* deterministic finite state machines and was recently revised for the case in which both the specification and the implementation may be *partially specified*. The W-method has also been adapted to finite state machines based specification languages, such as stream X-machines [11], [10] and Harel statecharts [3], [4].

On the other hand, in practice a finite state machine specification is usually constructed through a process of refinement; the equivalent flattened finite state machine of such a refinement can be constructed and the W-method can be applied to test the flattened machine. This approach might not always be practical, for large scale systems the flattened machine can be extremely complex and this will result in a test set of unmanageable size.
The paper formalizes the OR refinement of Finite State Machines (in section 4) and develops an efficient testing strategy for this kind of refinement.

2. PRELIMINARIES

Definition 2.1 A Deterministic Finite State Machine (DFSM for short) is a quintuple \( M = (Q, \mathcal{L}, \mathcal{O}, h, M_0) \), where \( Q \) is a finite set of states, \( \mathcal{L} \) is a finite set of input symbols, \( \mathcal{O} \) is a finite set of output symbols, \( h : Q \times \mathcal{L} \to Q \times \mathcal{O} \) is the behavior (partial) function, \( M_0 \) is the initial state.

A DFSM is often described by a state-transition diagram as in Example 2.1.

Let \( q, r \in Q, a \in \mathcal{L}, o \in \mathcal{O} \). We write \( q \xrightarrow{a/o} r \) to denote a transition from \( q \) to \( r \) with label \( a/o \) if \( h(q, a) = (r, o) \). As a response to an arbitrary input sequence \( x = a_1 \cdots a_n, n \geq 0, a_i \in \mathcal{L} \) for all \( i, 1 \leq i \leq n \), a state \( q \in Q \) of the DFSM \( M \) may produce an output sequence \( t = o_1 \cdots o_n, o_i \in \mathcal{O} \) for all \( i, 1 \leq i \leq n \), if there are the following \( n \) transitions \( q_i \xrightarrow{a_i/o_i} q_{i+1} \), \( 1 \leq i \leq n \) and \( q_1 = q \). It is said that the input \( a_1 \cdots a_n \) leads the machine from the state \( q \) to the state \( r \) or that the state \( r \) is reachable from \( q \). We denote this by \( q \Rightarrow a_1/o_1 \cdots a_n/o_n \Rightarrow q_{n+1} \) or simply by \( q \Rightarrow a_1/o_1 \cdots a_n/o_n \Rightarrow \) if the reached state \( q_{n+1} \) may be omitted. Moreover, if the produced output string \( o_1 \cdots o_n \) is not significant, it may be omitted, writing either \( q \Rightarrow a_1 \cdots a_n \Rightarrow q_{n+1} \) or \( q \Rightarrow a_1 \cdots a_n \Rightarrow . \)

For an input sequence \( a_1 \cdots a_n \), we say that two states \( q, r \in Q \) give identical response to \( a_1 \cdots a_n \) if: \( q \Rightarrow a_1/o_1 \cdots a_n/o_n \Rightarrow \) iff \( r \Rightarrow a_1/o_1 \cdots a_n/o_n \Rightarrow \) for all \( o_1, \ldots, o_n \in \mathcal{O} \). Note that the definition includes the case where no such \( o_1, \ldots, o_n \) exist (i.e. for partially specified DFSM).

Let \( A \subseteq \mathcal{L}^* \) be an arbitrary set of input sequences. Two states \( q, r \in Q \) are said to be \( A - \) equivalent if \( q \) and \( r \) give identical responses for each input sequence \( x \in A \). We denote this by \( q \equiv_A r \). Otherwise, they are said to be \( A - \) distinguishable (denoted by \( q \not\equiv_A r \)). If \( A = \mathcal{L}^* \) (i.e. \( q \) and \( r \) give identical responses for any input sequence) then \( q \) and \( r \) are said to be equivalent.

Let \( M \) and \( N \) be two DFSMs with the same sets of input (\( \mathcal{L} \)) and output (\( \mathcal{O} \)) symbols and \( A \subseteq \mathcal{L}^* \). Then \( M \) and \( N \) are said to be \( A \)-equivalent if their initial states \( M_0 \) and \( N_0 \) are \( A \)- equivalent. Otherwise, \( M \) and \( N \) are said to be \( A \)-distinguishable. If \( A = \mathcal{L}^* \) then \( M \) and \( N \) are said to be equivalent.

A finite set \( R \subseteq \mathcal{L}^* \) is said to be a state cover of a DFSM \( M \) if the empty sequence \( \epsilon \) is contained in \( R \), and, for every state \( q \in Q \) other than \( M_0 \), \( R \) contains an input sequence \( x \in R \) that may lead the machine from the initial state \( M_0 \) to \( q \) (i.e. \( M_0 \Rightarrow x \Rightarrow q \)). A finite set \( W \subseteq \mathcal{L}^* \) is said to be a characterization set of a DFSM \( M \) if any two distinct states \( q, r \in Q \) are \( W \)-distinguishable.

Example 2.1 Consider a simple tape recorder capable of playing and recording a tape. For simplicity, we consider the tape to be infinite. The DFSM
Finite state machine testing from an OR state refinement design

Fig. 1. DFSM model of a tape recorder.

The inputs to this DFSM are events: \( L_i = \{ \text{play}, \text{stop}, \text{rec}, \text{on}, \text{off} \} \). The outputs are the operations performed by the tape recorder: \( L_o = \{ \text{Play}, \text{Idle}, \text{Rec}, \text{Off} \} \). The initial state OFF is pointed at by a transition from a blob.

Then \( R = \{ \epsilon, \text{on}, \text{on} \cdot \text{rec}, \text{on} \cdot \text{play} \} \) is a state cover of this DFSM and \( W = \{ \text{on}, \text{rec}, \text{play} \} \) is a characterization set.

A DFSM \( M \) is said to be minimal if any other equivalent DFSM has at least the same number of states as \( M \); a DFSM \( A \) is minimal if and only if there exists a state cover and a characterization set of \( A \) [8].

Let \( M = (Q, L_i, L_o, h, M_0) \) and \( M' = (Q', L_i, L_o, h', M'_0) \) be two DFSMs. Then a bijective function \( f : Q \rightarrow Q' \) is called an isomorphism from \( M \) to \( M' \) if for every states \( q_1, q_2 \in Q \), input symbol \( a \in L_i \) and output symbol \( o \in L_o \), \( q_1 \rightarrow a/o \rightarrow q_2 \) if and only if \( f(q_1) \rightarrow a/o \rightarrow f(q_2) \). Any two equivalent minimal DFSMs are isomorphic [8], so for any DFSM \( M \), there exists an unique (up to an isomorphism) minimal DFSM \( M' \) equivalent to \( M \). \( M' \) is called the minimal DFSM of \( M \).

3. THE W-METHOD

Let \( S \) be a DFSM specification and \( I \) a DFSM model of the implementation. The W-method generates a test set (a finite set of sequences) that can determine any error in \( I \) with respect to \( S \), provided that the number of states of \( I \) is bounded by an integer \( m \) that may be larger than the number \( n \) of states of \( S \). The set is generated from the specification \( S \) and no information is available about \( I \) except the above mentioned upper bound.
The W-method was first developed in the context of completely specified DFSMs [7]. In [1], the method was extended to cope with (possibly) partially specified DFSMs.

**Definition 3.1** Let $S = (Q_S, L_S, h_S, S_0)$ and $I = (Q_I, L_I, h_I, I_0)$ be two DFSMs. A finite set $Y \subseteq L^*$ is called a test set for $S$ and $I$ if: $S$ and $I$ are $Y$-equivalent $\implies$ $S$ and $I$ are equivalent.

The W-method involves the selection of two sets of input sequences:

- $W \subseteq L^*$, a characterization set of $S$
- $R \subseteq L^*$, a state cover set of $S$

The method makes the following assumptions about the specification $S$ and the implementation $I$:

- $S$ is deterministic and minimal
- $I$ is deterministic and the number of states of $I$ is bounded by an integer $m$.

For the case where $S$ and $I$ are assumed to be completely specified the W-method provides a test set

$$Y_{cs} = R \cdot L_i[m - n + 1] \cdot W,$$

with $L_i[j] = \{\epsilon\} \cup L_i \cup \cdots \cup L_i^j$ for $j \geq 0$ [7].

However, $Y$ is not always a valid test set when partially specified DFSMs are considered, as shown by the following example. Consider the partially specified DFSM specification $S = (\{0, 1\}, \{a, b\}, \{a, b\}, h, 0)$ with $h(0, a) = (1, a)$ and $h$ undefined elsewhere. Then $R = \{\epsilon, a\}$ is a state cover of $S$ and $W = \{a\}$ is a characterization set of $S$. Let $I = (\{0, 1\}, \{a, b\}, \{a, b\}, h', 0)$ be an implementation of $S$ with $h'(0, a) = (1, a)$, $h'(0, b) = (1, b)$ and $h'$ undefined elsewhere. Since $S$ and $I$ have the same number of states ($n = m = 2$), the W-method gives $Y_{cs} = R \cdot L_i[1] \cdot W = \{a, a \cup a, b \cup a, a \cup a \cup a, a \cup b \cup a\}$. However, the two DFSMs are $Y_{cs}$-equivalent but they are not equivalent. Indeed, the single word which distinguishes them is $b$, which is not contained in $Y_{cs}$.

This situation occurs because when $S$ and $I$ are completely specified whenever $S$ and $I$ are $\{s\}$-equivalent for an input sequence $s$, they are also $\{t\}$-equivalent for any prefix $t$ of $s$. This is no longer true when at least one of $S$ and $I$ is partially specified. Therefore, one way of extending the W-method to the general case where the machines may be partially specified is to take all prefixes of the sequences in $Y_{cs}$. It has been proved, however, that a smaller subset of prefixes is sufficient [1], this is

$$Y = R \cdot L_i[m - n + 1] \cdot (W \cup \{\epsilon\}).$$
4. OR STATE REFINEMENT

One way of refining an existing DFSM specification is to replace one or more states with other DFSMs. The semantics of this transformation is that given by the expansion of OR states in state charts [3] and is illustrated by the following example.

Example 4.1 The original model in Example 2.1 can be refined by allowing the tape recorder to pause while playing or recording. For instance, the behavior of the RECORDING state is a two-state machine describing whether the tape recorder is actively recording or waiting for the user to press the continue command. The refinement of the two states, PLAYING and RECORDING (called in what follows OR states), is graphically illustrated in fig. 2. The flattened DFSM for the refined model is presented in fig. 3. Any transition arriving to the OR state in the original specification will now be directed to the initial state of the DFSM that replaces that state. Any transition leaving the OR state will now leave all the states of the replacement DFSM.

Definition 4.1 A DFSM $M' = (Q', L_i', L_o', h', M_0')$ is called an OR state refinement of a DFSM $M = (Q, L_i, L_o, h, M_0)$ if $Q \subseteq Q'$, $L_i \subseteq L_i'$, $L_o \subseteq L_o'$, $M_0 = M_0'$ and there is a partition $E = \{E_q\}_{q \in Q}$ of $Q'$ with $q \in E_q$ for all $q \in Q$ such that the following hold:

- $h(q, a) = h'(q', a)$ for all $q \in Q$, $a \in L_i$, $q' \in E_q$
- $\pi_1(h'(q', a)) \subseteq E_q$ for all $q \in Q$, $a \in L_i - L_i'$, $q' \in E_q$, where $\pi_1 : Q' \times L_o' \to Q'$ denotes the projection function

Fig. 2. Refinement of OR states.
Fig. 3. Flattened DFSM for OR state refinement.

\[ L_{i0} = L_i' - L_i \] is called the refining input alphabet.

For \( q \in Q \), the DFSM \( M_q = (E_q, L_{i0}, L_{d0}, h_q, q) \), where \( h_q \) is the restriction of \( h' \) to \( E_q \times L_{i0} \), is called the refining machine for \( q \).

For simplicity, in Definition 4.1, the refined state and the initial state of the corresponding refining machine are considered to be identical. For clarity, they are given different names in our examples, where only the states \( \text{PLAYING} \) and \( \text{RECORDING} \) are refined. So the partition \( E \) is defined by \( E_{\text{PLAYING}} = \{ \text{PLAY}, \text{PAUSE}_{\text{PLAY}} \} \), \( E_{\text{RECORDING}} = \{ \text{RECORD}, \text{PAUSE}_{\text{REC}} \} \) and \( E_q = \{ q \} \) for the other states. The refining input alphabet is \( L_{i0} = \{ \text{continue}, \text{pause} \} \).

5. OR STATE REFINEMENT TESTING

Consider the problem of generating test sets from a specification of the form of an OR state refinement of a DFSM \( S = (Q_S, L_i, L_o, h_S, S_0) \). The W method can be directly applied to the flattened specification \( S' = (Q'_S, L'_i, L'_o, h'_S, S'_0) \) and, furthermore, a state cover \( R_{S'} \) and a characterization set \( W_{S'} \), can be constructed from a state cover \( R_S \) and a characterization set \( W_S \) of the original DFSM \( S \) and state covers \( R_q, q \in Q_S \) and characterization sets \( W_q, q \in Q_S \) of the refining machines using the following formulae

\[ R_{S'} = R_S \otimes \mathcal{R}, \]

where \( \mathcal{R} = \{ R_q \mid q \in Q_S \} \) is a set that contains a state cover \( R_{M_q} \) for each state \( q \in Q_s \) and for \( A \subseteq L_i^* \) and \( B = \{ B_q \mid q \in Q_S \} \subseteq 2^{L_i^*} \), \( A \otimes B = \{ a \cdot b \mid a \in A, b \in B_q, S_0 \Rightarrow a \Rightarrow q \} \). That is \( R_{S'} \) is obtained by concatenating each
sequence in $R_S$ that forces $S$ into $q$ from the initial state $S_0$ with all sequences in $R_q$

$$W_{S'} = W \cup \bigcup_{q \in Q_S} W_q.$$  

The proofs are straightforward.

On the other hand, since the specification has been developed through a refinement process, one would also expect the implementation to be build incrementally: initially a developer may only have a skeleton of the system then step by step fill it with details. For instance the DFSM in fig. 1 can be built first. Afterwards, when the functionality contained in the RECORDING (or PLAYING) refining machine is added it can be considered not to have an effect outside this state, (i.e. any transition leaving an OR state in $I$ will leave all the states of the refining machine in $I'$ and have the same destination as the original transition), so implementation can also be modelled by an OR state refinement $I' = (Q'_I, L_i', L_0', h'_I, I'_0)$ of some DFSM $I = (Q_I, L, L_0, h_I, I_0)$.

We can now take advantage of this situation and show how the test sets for $S$ and the refining machines can be reused in the construction of the new test set.

Let $S'$, the specification, be an OR state refinement of $S$ and $I'$, the implementation, be an OR state refinement of $I$, as above. Our aim is to construct a test set $Y'$ for $S'$ and $I$'. Let $n$ be the number of states of $S$ and $m$ the upper bound for the number of states of $I$. For $q \in Q_S$ let $S_q$ be the refining machine for $q$ and for $p \in Q_I$ let $I_p$ be the refining machine for $p$. Let $n_q$ be the number of states of $S_q$, $m_p$ the upper bound for the number of states of $I_p$ and $k = \max_{p \in Q_I} m_p$. We assume that $S$ is minimal and $S_q$ is minimal for all $q \in Q_S$. Without loss of generality all states of $I$ are assumed to be reachable from the initial state $I_0$ (otherwise the non-reachable states may be removed). Similarly, for $p \in Q_I$, all states of $I_p$ are assumed to be reachable from the initial state $p$.

Then the test set $Y'$ we are after is

$$Y' = Y \cup U,$$

where $Y = R \cdot L[m - n + 1] \cdot (W \cup \{\epsilon\})$ is a test set for $S$ and $I$ and $U$ is constructed using the following sets:

- $Y_q = R_q \cdot L_0[k - n_q + 1] \cdot (W_q \cup \{\epsilon\})$ for all $q \in Q_I$, where $R_q$ is a state cover and $W_q$ a characterization set of $S_q$; then $Y_q$ is a test set for $S_q$ and $I_p$ for all $p \in Q_p$

- $T = R \cdot L[m - n]$

Then

$$U = T \otimes Y,$$
Lemma 5.1 For $Y$ as above, if $S$ and $I$ are equivalent, then $T$ is a state cover of $I$.

Proof. We prove by induction on $0 \leq j \leq m - n$ that $T_j = R \cdot Li[j]$ reaches at least $n + j$ states of $I$. For $j = 0$ the statement follows since $R$ is a state cover of $S$, which is the minimal DFSM of $I$. Assume the statement true for $j$. Then either $T_j$ reaches all the states of $I$ or $T_{j+1}$ will reach at least one more state than $T_j$. Hence $T_{j+1}$ reaches at least $n + j + 1$ states of $I$. \qed

Theorem 5.1 For $S, S', I, I'$ and $Y'$ as above, $Y'$ is a test set for $S'$ and $I'$. Proof. Assume that $S'$ and $I'$ are $Y'$-equivalent. Then $S'$ and $I'$ are $Y'$-equivalent, hence $S$ and $I$ are $Y'$-equivalent. Since $Y$ is a test set for $S$ and $I$, $S$ and $I$ are equivalent. Furthermore, from Lemma 5.1 it follows that $T$ is a state cover of $I$.

Let $q \in Q_S$ and $p \in Q_I$ be two equivalent states of $S$ and $I$, respectively. Then there exists $t \in T$ such that $S_0 \Rightarrow t \Rightarrow q$ and $I_0 \Rightarrow t \Rightarrow p$. Since $S'$ and $I'$ are $U'$-equivalent it follows that $S_q$ and $I_p$ are $Y'$-equivalent. Hence $S_q$ and $I_p$ are equivalent.

Therefore, we have proved that $S$ and $I$ are equivalent and for any pair of equivalent states $q \in Q_S$, $p \in Q_I$, $S_q$ and $I_p$ are equivalent. Then from Definition 4.1 it follows that $S'$ and $I'$ are equivalent. \qed

It can be shown that the implementation refinement approach considerably reduces the size of the test set generated.

6. CONCLUSIONS

The paper formalizes the OR refinement for finite state machines and it shows how if the implementation is also constructed through a process of refinement, then the test process can be distributed into smaller chunks (i.e., the original machine and the refining machines), thus diminishing the effort devoted to testing and the size of the final test set.

The partial OR (p-OR) state refinement of finite state machines will also be considered in a future paper. The generalization of these test generation methods for refined DFSMs to other finite state machine-based specification languages, such as Harel statecharts, will also be investigated.
References


SHARPENING OF CRITICAL SPECTRAL PARAMETER AT DYNAMIC BIFURCATION
BY PSEUDOPERTURBATION METHOD

Boris V. Loginov*, Olga V. Makeeva**, Elena V. Foliadova**

*Ulyanovsk State Technical University, **Ulyanovsk State Pedagogical University, Russia

Abstract

Taking into account our previous works [2]-[5] at the suggested by the M.K. Gavurin [1] pseudoperturbation method the problem of sharpening the approximately given critical value of the spectral parameters and relevant Jordan chains in Poincaré-Andronov-Hopf bifurcation is solved.

1. THE STATEMENT OF THE PROBLEM

Let $E_1$, $E_2$ be two real Banach spaces and consider the differential equation

$$\frac{dx}{dt} = Bx - R(x, \varepsilon). \quad (1)$$

Here $A, B : E_1 \to E_2$ are, for simplicity, linear (bounded) operators, $R(x, \varepsilon)$ is definite and continuous in a neighborhood of the point $(0,0) \in E_1 + R^l$ together with its Fréchet derivative $R_x(x, \varepsilon)$, and $R(0,0) = 0$, $R_x(0,0) = 0$.

We consider the sufficiently general case when the $A$-spectrum $\sigma_A(B)$ of the operator $B$ intersects the imaginary axis at the points $\pm i\alpha$ of multiplicity $n$ [6, 7, 11, 13]. Let the elements $u^{(1)}_{1k}, u^{(1)}_{2k} \in E_1$ (elements $v^{(1)}_{1k}, v^{(1)}_{2k} \in E_2^\ast$) be such that

$$B(\alpha) \Phi^{(1)}_k \equiv \begin{pmatrix} B & \alpha A \\ -\alpha A & B \end{pmatrix} \begin{pmatrix} u^{(1)}_{1k} \\ u^{(1)}_{2k} \end{pmatrix} = 0,$$

$$B^\ast(\alpha) \Psi^{(1)}_k \equiv \begin{pmatrix} B^\ast & -\alpha A^\ast \\ \alpha A^\ast & B^\ast \end{pmatrix} \begin{pmatrix} v^{(1)}_{1k} \\ v^{(1)}_{2k} \end{pmatrix} = 0 \quad , \quad k = 1, \ldots, n, \quad (2)$$

i.e. the zero-subspace $N(B(\alpha))$ and $N(B^\ast(\alpha))$ of the operators $B(\alpha)$ and $B^\ast(\alpha)$ have the forms

$$N(B(\alpha)) = \text{span} \left\{ \Phi^{(1)}_{1k} = \begin{pmatrix} u^{(1)}_{1k} \\ u^{(1)}_{2k} \end{pmatrix}, \Phi^{(1)}_{2k} = \begin{pmatrix} -u^{(1)}_{2k} \\ u^{(1)}_{1k} \end{pmatrix}, \quad k = 1, \ldots, n \right\},$$

$$N(B^\ast(\alpha)) = \text{span} \left\{ \Psi^{(1)}_{1k} = \begin{pmatrix} v^{(1)}_{2k} \\ -v^{(1)}_{1k} \end{pmatrix}, \Psi^{(1)}_{2k} = \begin{pmatrix} v^{(1)}_{1k} \\ v^{(1)}_{2k} \end{pmatrix}, \quad k = 1, \ldots, n \right\}.$$
Carrying out the complexification of the equation (1), consider it in the spaces \( \mathcal{E}_k = E_k + iE_k, k = 1, 2 \), and suppose that the operator \( R(x, \varepsilon) \) admits a sufficiently smooth extension on these spaces. Then the elements \( u_k^{(1)} = u_k^{(1)} + iu_2^{(1)}, \bar{u}_k^{(1)} \) and \( v_k^{(1)} = v_k^{(1)} + iv_2^{(1)}, \bar{v}_k^{(1)} \) are the eigenvalues of the following eigenvalue problem [11]

\[
Bu_k^{(1)} = i\alpha Au_k^{(1)}, \quad B\bar{u}_k^{(1)} = -i\alpha A\bar{u}_k^{(1)}, \quad B^*v_k^{(1)} = -i\alpha A^*v_k^{(1)}, \quad B^*\bar{v}_k^{(1)} = i\alpha A^*\bar{v}_k^{(1)}.
\]

The Poincaré substitution \( t = \frac{x}{\alpha+\mu^2}, x(t) = y(\tau) \) in the equation (1) allows us to extend the classical Lyapounov-Schmidt method for the investigation of the Poincaré-Andronov-Hopf bifurcation problem [11, 13], by introducing the spaces \( Y(Z) \) of \( 2\pi \)-periodic continuously differentiable functions of \( \tau \) with values in \( \mathcal{E}_1 \) (\( 2\pi \)-periodic continuous functions of \( \tau \) with values in \( \mathcal{E}_2 \)) and duality

\[
\langle\langle y, f \rangle\rangle = \frac{1}{2\pi} \int_{0}^{2\pi} \langle y(\tau), f(\tau) \rangle d\tau \quad y \in Y, f \in Y^* (y \in Z, f \in Z^*).
\]

Then the operators \( (\mathcal{B}y)(\tau) = -\alpha A\frac{dy}{d\tau} + By(\tau) \) and \( (\mathcal{B}^*z)(\tau) = \alpha A^* \frac{dz}{d\tau} + B^*z(\tau) \) have \( 2\pi \)-dimensional zero subspaces

\[
N(\mathcal{B}) = \text{span} \left\{ \varphi_k^{(1)} = \varphi_k^{(1)}(\tau) = u_k^{(1)} e^{i\tau}, \bar{\varphi}_k^{(1)} = \bar{u}_k^{(1)} e^{-i\tau}, k = 1, \ldots, n \right\},
\]

\[
N(\mathcal{B}^*) = \text{span} \left\{ \psi_k^{(1)} = \psi_k^{(1)}(\tau) = v_k^{(1)} e^{i\tau}, \bar{\psi}_k^{(1)} = \bar{v}_k^{(1)} e^{-i\tau}, k = 1, \ldots, n \right\}.
\]

Let now the zero-subspaces \( N(\mathcal{B}) \) and \( N(\mathcal{B}^*) \) have \( A \)-Jordan (\( A^* \)-Jordan) chains \( \varphi_k^{(j)} = u_k^{(j)} e^{i\tau}, \bar{\varphi}_k^{(j)} = \bar{u}_k^{(j)} e^{-i\tau}, j = 1, \ldots, p_k \) of the length \( p_j \). According to the lemma on the biorthogonality of the Jordan chains (for linear operator-functions of spectral parameter) [8, 12], they can be chosen to satisfy the equalities

\[
\langle\langle \varphi_k^{(j)}, \gamma_s^{(l)} \rangle\rangle = \delta_{ks} \delta_{jl}, j(l) = 1, \ldots, p_k(p_s), k, s = 1, \ldots, n, \gamma_s^{(l)} = A^* \psi_s^{(p_s+1-l)},
\]

\[
z_k^{(j)} = A^* \varphi_k^{(p_k+1-j)}.
\]

For the operator-function (2) the relations 6 read (see (3)-(5))

\[
\langle A u_{1k}^{(p_k+1-j)}, v_{1s}^{(l)} \rangle + \langle A u_{2k}^{(p_k+1-j)}, v_{2s}^{(l)} \rangle = \delta_{ks} \delta_{jl}, \quad \langle A u_{1k}^{(p_k+1-j)}, v_{1s}^{(l)} \rangle = \langle A u_{1k}^{(p_k+1-j)}, v_{2s}^{(l)} \rangle,
\]

\[
\langle A u_{2k}^{(p_k+1-j)}, v_{1s}^{(l)} \rangle + \langle A u_{2k}^{(p_k+1-j)}, v_{2s}^{(l)} \rangle = \delta_{ks} \delta_{jl}, \quad \langle A u_{2k}^{(p_k+1-j)}, v_{1s}^{(l)} \rangle = \langle A u_{2k}^{(p_k+1-j)}, v_{2s}^{(l)} \rangle.
\]
which mean the biorthogonality property for \( A \)-Jordan chains of the operator-function \( B(\alpha - \varepsilon) = B(\alpha) - \varepsilon A, A = \begin{pmatrix} 0 & A \\ -A & 0 \end{pmatrix}, A^* = \begin{pmatrix} 0 & -A \\ A & 0 \end{pmatrix} \), i.e.

\[
\langle A\Phi_{pk}^{(p_k + 1 - l)}, \Phi^{(l)}_{ps} \rangle = \delta_{ps} \delta_{ks} \delta_{jl}, \quad \langle \Phi^{(j)}_{pk}, A^* \Phi^{(p_k + 1 - l)}_{ps} \rangle = \delta_{ps} \delta_{ks} \delta_{jl} \quad (8)
\]

The aim of this article is the sharpening of sufficiently good approximations \( \alpha_0, v_0^{(s)}, v_0^{(s)} \), \( |\alpha - \alpha_0| \leq \varepsilon, \|v_0^{(s)} - v_0^{(s)}\| \leq \varepsilon, \|v_0^{(s)} - v_0^{(s)}\| \leq \varepsilon, s = 1, \ldots, p, j = 1, \ldots, n \) to the critical value of the bifurcation parameter \( \alpha \) and generalized Jordan chains

\[
(B - i\alpha A)u_j^{(k)} = A^*v_j^{(k-1)}, \quad (B + i\alpha A)\bar{u}_j^{(k-1)} = -A^*v_j^{(k-1)},
\]

\[
(B^* + i\alpha A^*)v_j^{(k)} = -A^*v_j^{(k-1)}, \quad (B^* - i\alpha A^*)\bar{v}_j^{(k-1)} = A^*v_j^{(k-1)},
\]

where \( (A u_j^{(k)}, v_{\sigma_0}^{(p_\sigma + 1 - l)}) = \delta_{j\sigma_0} \delta_{kl}, \) by means of the pseudoperturbation method \([2, 3, 5]\).

The above-mentioned discussion allows us to solve this problem both on the base of operator-function \( B(\alpha) \) with relations (8) and operator-function \( B \pm i\alpha A \) with relations (9). The cases of nonlinear dependence on \( \alpha \) can be investigated by means of the linearization [5] of the spectral parameter.

The obtained results are supported by RFBR grant 06-01-01692.

2. GENERAL SCHEME OF PSEUDOPERTURBATION OPERATOR CONSTRUCTION

A. Operator-function \( B - i\alpha A \).

Similarly to \([3, 5]\) the following assertion can be proved

**Lemma 1A.** Passing if necessary to linear combinations, the systems \( \{\gamma_{k0}^{(l)}\}_{k,l=1}^{n,p_k} \), \( \zeta_{k0}^{(l)} = A^*v_{k0}^{(p_k + 1 - l)} \), \( \gamma_{k0}^{(l)} = A u_{k0}^{(p_k + 1 - l)} \) satisfying the biorthogonality relations \( (u_j^{(k)}, v_{k0}^{(l)}, \sigma_{k0}^{(l)}, \tau_{k0}^{(l)}) = \delta_{ik} \delta_{jl} \), \( (z_j^{(k)}, v_{k0}^{(l)}, \sigma_{k0}^{(l)}, \tau_{k0}^{(l)}) = \delta_{ik} \delta_{jl} \), can be determined.

Computing the expressions \( \sigma_{k0}^{(l)} = (B - i\alpha_0 A)u_k^{(1)}, \sigma_{k0}^{(l)} = (B - i\alpha_0 A)u_k^{(1)} - A u_{k0}^{(j-1)} \), \( \tau_{k0}^{(l)} = (B^* + i\alpha_0 A^*)v_{k0}^{(l)} \), \( \tau_{k0}^{(l)} = (B^* + i\alpha_0 A^*)v_{k0}^{(l)} + A^*v_{k0}^{(j-1)} \), 1 \( \leq \) \( k \leq p_k \), define two pseudoperturbation operators

\[
D_0 x = \sum_{k=1}^{n} \sum_{j=1}^{p_k} \langle x, \gamma_{k0}^{(j)} \rangle \sigma_{k0}^{(j)} + \sum_{r=1}^{n} \sum_{s=1}^{p_r} \langle \sigma_{k0}^{(j)}, v_{r0}^{(s)} \rangle \gamma_{r0}^{(s)} + \sum_{k=1}^{n} \sum_{j=1}^{p_k} \langle x, \gamma_{k0}^{(j)} \rangle \sigma_{k0}^{(j)}
\]

\[
D_0 x = \sum_{k=1}^{n} \sum_{j=1}^{p_k} \langle x, \gamma_{k0}^{(j)} \rangle \sigma_{k0}^{(j)} + \sum_{k=1}^{n} \sum_{j=1}^{p_k} \langle x, \tau_{k0}^{(j)} \rangle \sigma_{k0}^{(j)} + \sum_{r=1}^{n} \sum_{s=1}^{p_r} \langle u_{r0}^{(s)} \gamma_{r0}^{(s)} \rangle \tau_{r0}^{(s)} \quad (10)
\]
Theorem 1. The pseudoperturbation operators (10) have the following properties

\[ D_0 u_{k0}^{(j)} = \sigma_{k0}^{(j)}, \quad D_0 v_{k0}^{(j)} = \tau_{k0}^{(j)}, \quad j = 1, \ldots, p_k, \quad k = 1, \ldots, n \] (11)

Corollary 1B. The pseudoperturbation operators have the required properties

\[ (B - i\alpha_0 A - D_0) u_{k0}^{(1)} = 0, \quad (B - i\alpha_0 A - D_0) u_{k0}^{(j)} = A u_{k0}^{(j-1)}, \]

\[ (B^* + i\alpha_0 A^* - D_0^*) v_{k0}^{(1)} = 0, \quad (B^* + i\alpha_0 A^* - D_0^*) v_{k0}^{(j)} = -A^* v_{k0}^{(j-1)}, \quad j > 1. \] (12)

The proofs are performed by a direct computation [3, 5]. Thus the given sufficiently good approximation to pure imaginary eigenvalue and generalized Jordan chains become exact for the perturbed operator.

B. Operator-function \( \mathcal{B}(\alpha) \).

Lemma 1B. Passing, if necessary, to linear combinations, the systems

\[ \{ \gamma_{\mu k0}^{(l)} \}_{k, l=1}^{n, p_k}, \quad \gamma_{\mu k0}^{(l)} = \begin{pmatrix} 0 & -A \\ A & 0 \end{pmatrix} \Psi_{\mu k0}^{(p_k+1-l)}, \]

\[ \{ z_{\mu 0}^{(j)} \}_{i, j=1}^{n, p_i}, \quad z_{\mu 0}^{(j)} = \begin{pmatrix} 0 & A \\ -A & 0 \end{pmatrix} \Phi_{\mu 0}^{(i+1-j)}, \quad \mu = 1, 2, \]

satisfy the biorthogonality relations (8).

Computing the expressions \( \sigma_{\mu k0}^{(1)} = \mathcal{B}(\alpha_0) \Phi_{\mu k0}^{(1)}, \sigma_{\mu k0}^{(j)} = \mathcal{B}(\alpha_0) \Phi_{\mu k0}^{(j)} - A \Phi_{\mu k0}^{(j-1)}, \)

\( \gamma_{\mu k0}^{(1)} = B^*(\alpha_0) \Psi_{\mu k0}^{(1)}, \quad \gamma_{\mu k0}^{(j)} = B^*(\alpha_0) \Psi_{\mu k0}^{(j)} - A^* \Psi_{\mu k0}^{(j-1)}, \quad 1 < l \leq p_k, \quad \mu = 1, 2 \)

introduce four pseudoperturbation operators \( D_{\mu 0}, \mu = 1, 2 \) of the form (10).

Theorem 1B and its Corollary 1B are proved by direct computation.

3. **Newton-Kantorovich Iterational Procedure for the Determination of Critical Value of Parameter \( \alpha \) and Generalized Jordan Chains**

Since up to the notation the description of the Newton-Kantorovich iterative process for the operator-function \( \mathcal{B}(\alpha) \), repeats the relevant part of the articles [3, 5], here we present it only for the operator-function \( B - i\alpha A \).

Using the first equality of Corollary 1A rewrite the equation \( (B - i\alpha_0 A - D_0) u_k = 0 \) in the form of the system

\[ (B - i\alpha_0 A - D_0) u - i(\alpha - \alpha_0) A u + D_0 u + \sum_{k=1}^{n} \langle u, \gamma_{k0}^{(1)} \rangle z_{k0}^{(1)} = \sum_{k=1}^{n} \xi_{k0}^{(1)} z_{k0}^{(1)}, \] (13)

\[ \xi_{k0}^{(1)} = \langle u, \gamma_{k0}^{(1)} \rangle. \] (14)

According to the E. Schmid lemma and general perturbation theory for a sufficient accuracy of the initial approximations there exists the bounded
operator $\Gamma_0 = \tilde{B}_0^{-1} = [(B - i\alpha_0 A - D_0) + \sum_{k=1}^{n} \alpha \gamma^{(1)}_{k0}]^{-1}$. Following the scheme of the articles [3, 5], set

$$l_{jk}(it - i\alpha_0) = \langle ((it - i\alpha_0)A - D_0)[I - \Gamma_0(it - i\alpha_0)A + \Gamma_0D_0]^{-1}u^{(1)}_j, \gamma^{(1)}_{k0} \rangle = \langle \{I - [I - \Gamma_0((it - i\alpha_0)A - D_0)]^{-1}\}u^{(1)}_j, \gamma^{(1)}_{k0} \rangle.$$

The function $f(t) = \text{det}[l_{jk}(it - i\alpha_0)]$ has $t = \alpha$ as $K = \sum_{s=1}^{n} p_s$ - multiple root, since $-\frac{1}{s!} \frac{d^s l_{jk}(it - i\alpha_0)}{d t^s} = l_{jk,s}(it - i\alpha_0) = \langle [I - (it - i\alpha_0)\Gamma_0 A + \Gamma_0 D_0]^{-1}(\Gamma_0 A[I - (it - i\alpha_0)\Gamma_0 A + \Gamma_0 D_0]^{-1})^s u^{(1)}_j, \gamma^{(1)}_{k0} \rangle = 0$

for $s < K$, i.e. $\frac{df(t)}{dt} = 0$ and $\frac{d^{K} f(t)}{dt^K} = det[\frac{d^s l_{jk}(it - i\alpha_0)}{d t^s}] = 0$ by virtue of the relation $-\frac{1}{p_j} \frac{d p_j l_{jk}(it - i\alpha_0)}{d \alpha} = \langle u^{(1)}_j, \gamma^{(1)}_{k0} \rangle + O(\|D_0\|) = 0$.

In order to determine of $\alpha$ it is possible to apply to the equation $f^{(K-1)}(t) = 0$ the modified or basic Newton-Kantorovich method, taking as the initial approximation $t = \alpha_0$. In fact [9], since $|f^{(K)}(\alpha_1) - f^{(K)}(\alpha_2)| \leq L|\alpha_1 - \alpha_2|$, for $\alpha_1, \alpha_2 \in S(\alpha_0, \rho)$ - the ball of small radius $\rho$ and centered at $\alpha_0$, $L$ is some constant, then the equation $f^{(K-1)}(t) = 0$ is reduced, at every step, to the solution of $K\nu$ linear equations. In fact, the calculation of $f(i\alpha_0 - i\alpha_0)$ is reduced to the calculation of $l_{jk}(\alpha_0 - \alpha_0)$, i.e. to the solving of the equation $x^{(\nu-1)}_j - \Gamma_0(i\alpha_0 - i\alpha_0)A + \Gamma_0 D_0 x^{(\nu-1)}_j = u^{(1)}_j$, which is equivalent to the equation

$$\langle B - i\alpha \nu-1 A \rangle x^{(\nu-1)}_j + \sum_{r=1}^{n} \langle x^{(\nu-1)}_j, \gamma^{(1)}_{r0} \rangle = z^{(1)}_{j0}. \quad (15)$$

In order to compute the sequential derivatives $l_{jk,s}(i\alpha_0 - i\alpha_0)$, $j,k = 1,\ldots,n$, $s = 1,\ldots,p_j - 1$ we must solve the equations

$$(B - i\alpha \nu-1 A)x^{(\nu-1)}_j + \sum_{r=1}^{n} \langle x^{(\nu-1)}_j, \gamma^{(1)}_{r0} \rangle = Ax^{(\nu-1)}_j,$$

$$(B - i\alpha \nu-1 A)x^{(\nu-1)}_j + \sum_{r=1}^{n} \langle x^{(\nu-1)}_j, \gamma^{(1)}_{r0} \rangle = Ax^{(\nu-1)}_j.$$
The basic Newton-Kantorovich method requires \( n(K+1) \) linear equations.

Remark 1. According to results in [10] these equations are stable with respect to computation error of the operator and right-hand sides.

After the computation of the exact value \( \alpha \), the elements of the generalized Jordan chains can be found directly from the equations

\[
(B - i\alpha A)x_{js} + \sum_{r=1}^{n} \langle x_{js}, \gamma_{r0}^{(1)} \rangle z_{r0}^{(1)} = Ax_{js-1} = \begin{cases} z_{j0}^{(1)}, s = 1 \\ Ax_{js-1}, s = 2, p_j \end{cases}
\]

Remark 2. The cases of nonlinear dependence on \( \alpha \), which arise in Poincaré-Andronov-Hopf bifurcation for differential equations of the order \( m > 1 \), can be investigated on the base of the linearization [4].

4. THE CASE OF DIFFERENTIAL EQUATION OF HIGHER ORDER

Consider the Poincaré-Andronov-Hopf bifurcation for the s-th order differential equation

\[
A_s \frac{d^s x}{dt^s} + A_{s-1} \frac{d^{s-1} x}{dt^{s-1}} + \ldots + A_1 \frac{dx}{dt} = Bx - R(x, \frac{dx}{dt}, \ldots, \frac{d^{s-1} x}{dt^{s-1}}, \varepsilon) \quad (16)
\]

with bounded linear operators \( A_k, B : E_1 \to E_2, k = 1, s \). The nonlinear operator \( R(x, x', \ldots, x^{(s-1)}, \varepsilon) \) is definite and continuous in a neighborhood of the point \((0, \ldots, 0, 0) \in \mathbb{R}^s \) together with its Frechét derivatives \( R_{x(k)} \) and \( R(0, \ldots, 0, 0) = 0, R_{x(k)}(0, \ldots, 0, 0) = 0 \). The linearized equation (17) can be reduced to the equation \( A \frac{dX}{dt} = BX \) in the space of vector-functions \( X = (x_1, x_2, \ldots, x - s) \in \mathbb{R}^s \)

\[
A = \begin{pmatrix} C & 0 & 0 & \ldots & 0 & 0 \\ 0 & C & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & C & 0 \\ 0 & 0 & 0 & \ldots & 0 & A_s \end{pmatrix}, \quad B = \begin{pmatrix} 0 & C & 0 & \ldots & 0 & 0 \\ 0 & 0 & C & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & C \\ B & -A_1 & -A_2 & \ldots & -A_{s-2} & -A_{s-1} \end{pmatrix}.
\]
Sharpening of critical spectral parameter by pseudoperturbation method

where \( C \in L(E_1, E_2) \) is an arbitrary invertible operator and \( x_k = \frac{d^{k-1}x_0}{dx^{k-1}} \), \( k = \frac{1}{s} \). Therefore the nonlinear operator \( R(x, x', \ldots, x^{s-1}, \varepsilon) \) should be considered as representable in the form \( R(X, \varepsilon) \). According to Section 1, it is natural to consider in \( sE_1 \to sE_2 \) the eigenvalue problem

\[
B(\alpha) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \equiv \begin{pmatrix} B & \alpha A \\ -\alpha A & B \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0, \quad u_j = (u_{j1}, u_{j2}, \ldots, u_{js}), \ j = 1, 2.
\]

As before, we consider the sufficiently general case when the \( A \)-spectrum \( \sigma_A(B) \) of the operator \( B \) intersects the imaginary axis at the points \( \pm i\alpha \) of the multiplicity \( n \), i.e. there exist elements \( u^{(1)}_{jk} = (u^{(1)}_{j1k}, u^{(1)}_{j2k}, \ldots, u^{(1)}_{jsk}) \in sE_1, j = 1, 2 \) (elements \( v^{(1)}_{jk} = (v^{(1)}_{j1k}, v^{(1)}_{j2k}, \ldots, v^{(1)}_{jsk}) \in sE_2 \) such that

\[
B(\alpha) \Phi^{(1)}_k = \begin{pmatrix} B & \alpha A \\ -\alpha A & B \end{pmatrix} \begin{pmatrix} u^{(1)}_{1k} \\ u^{(1)}_{2k} \end{pmatrix} = 0,
\]

\[
(B^*(\alpha) \Psi^{(1)}_k = \begin{pmatrix} B^* & \alpha A^* \\ -\alpha A^* & B^* \end{pmatrix} \begin{pmatrix} v^{(1)}_{1k} \\ v^{(1)}_{2k} \end{pmatrix} = 0)
\]

\( k = \frac{1}{n} \) and the zero-subspaces \( N(B(\alpha)) \) and \( N(B^*(\alpha)) \) of the operators \( B(\alpha) \) and \( B^*(\alpha) \) have the forms (3).

The construction of the pseudoperturbation operator by means of the corresponding Newton-Kantorovich iterative process for the operator-function \( B(\alpha) \), up to the notation, repets the relevant parts of the articles [3, 5] and, therefore, is omitted.

References


EXISTENCE RESULTS FOR A CLASS OF NONLINEAR VOLterra INTEGRODIFFERENTIAL PROBLEM
Rodica Luca–Tudorache
Department of Mathematics, "Gh.Asachi" Technical University
rluca@hostingcenter.ro

Abstract The existence of the strong and weak solutions of a Volterra integrodifferential system with some nonlinear boundary conditions and initial data is proved.

Keywords: Volterra integrodifferential equations, boundary condition, maximal monotone operator, compact operator, weak and strong solutions.

2000 MSC: 47H05, 45K05

1. INTRODUCTION

We shall investigate the nonlinear Volterra integrodifferential system

\begin{equation}
\begin{aligned}
\frac{\partial u}{\partial t} + a_2(x) \frac{\partial^2 v}{\partial x^2} + a_1(x) \frac{\partial v}{\partial x} + a_0(x)v + \alpha(x,u) + \int_0^t b(t-s)A(u(s,x)) \, ds &= f(t,x) \\
\frac{\partial v}{\partial t} - \frac{\partial^2}{\partial x^2} (a_2(x)u) + \frac{\partial}{\partial x} (a_1(x)u - a_0(x)u + \beta(x,v) + \int_0^t c(t-s)B(v(s,x)) \, ds &= g(t,x),
\end{aligned}
\end{equation}

with the boundary condition

\begin{equation}
\begin{aligned}
\left( \begin{array}{c}
\col( a_1(0)u(t,0) - \frac{\partial}{\partial x}(a_2u)(t,0) - a_1(1)u(t,1) + \frac{\partial}{\partial x}(a_2u)(t,1), \\
a_2(0)u(t,0), -a_2(1)u(t,1)
\end{array} \right), S(w'(t)) \right) & \in \\
\in -G \left( \begin{array}{c}
\col( v(t,0), v(t,1), \frac{\partial v}{\partial x}(t,0), \frac{\partial v}{\partial x}(t,1)
\end{array} \right), \quad t > 0
\end{aligned}
\end{equation}

and the initial data

\begin{equation}
\begin{aligned}
\{ u(0,x) = u_0(x), \quad v(0,x) = v_0(x), \quad 0 < x < 1 \\
w(0) = w_0.
\end{aligned}
\end{equation}
Here $S$ is a positive diagonal $m$-matrix and $G$ is an operator in the space $\mathbb{R}^{m+4}$ which satisfies some appropriate assumptions. This problem has applications in the theory of integrated circuits and in the elastic beam theory (see [2,5,7] for further references). For $A = B \equiv 0$ the above problem has been studied in [3] concerning the existence, uniqueness and regularity properties of the solutions, and in [4,2] for the asymptotic behavior and the existence of periodic solutions. In this paper we deduce some existence results for the weak and strong solutions for the problem $(S)+(BC)+(IC)$, by applying several theorems from [8]. For the basic concepts and results in the theory of nonlinear evolution equations of monotone type and of Volterra integrodifferential equations we refer the reader to [1,6,8].

Introduce the assumptions used in the sequel.

(H1) The functions $a_k \in W^{k,\infty}(0,1)$, $k = 0, 2$ and $a_2(x) \neq 0$, $\forall x \in [0,1]$.

(H2) The functions $x \to \alpha(x,p)$ and $x \to \beta(x,p)$ are in $L^2(0,1)$ for any fixed $p \in \mathbb{R}$. Moreover, the functions $p \to \alpha(x,p)$ and $p \to \beta(x,p)$ are continuous and nondecreasing from $\mathbb{R}$ into $\mathbb{R}$, for a.a. $x \in (0,1)$.

(H3) The functions $b, c : [0,a] \to \mathbb{R}$ are continuous $(a > 0)$.

(H4) The functions $A, B : \mathbb{R} \to \mathbb{R}$ are continuous and there exist the constants $d, \bar{d} > 0$ and $e, \bar{e} \in \mathbb{R}$ such that

$$|A(u)| \leq d|u| + e, \quad |B(u)| \leq \bar{d}|u| + \bar{e}, \quad \forall u \in \mathbb{R}.$$

(H5) $G : D(G) \subset \mathbb{R}^{m+4} \to \mathbb{R}^{m+4}$ is a maximal monotone mapping (possibly multivalued), $D(G) \neq \emptyset$. Moreover, $G = \left( \begin{array}{cc} G_{11} & G_{12} \\ G_{21} & G_{22} \end{array} \right)$ with

$$G_{11} : D(G_{11}) \subset \mathbb{R}^4 \to \mathbb{R}^4, \quad G_{12} : D(G_{12}) \subset \mathbb{R}^m \to \mathbb{R}^4,$$

$$G_{21} : D(G_{21}) \subset \mathbb{R}^4 \to \mathbb{R}^m, \quad G_{22} : D(G_{22}) \subset \mathbb{R}^m \to \mathbb{R}^m.$$

(H6) $S = \text{diag}(s_1, \ldots, s_m)$ with $s_j > 0$, $j = 1,m$.

The assumption (H5) is a technical one; it is automatically satisfied if $G$ is a matrix.

2. PRELIMINARY RESULTS

Let us write the problem $(S)+(BC)+(IC)$ as a Volterra integrodifferential problem in a certain Hilbert space. For this, let us consider the spaces $X = (L^2(0,1))^2, \mathbb{R}^m$ and $Y = X \times \mathbb{R}^m$ with the corresponding scalar products $\langle f, g \rangle_X = \int_0^1 f_1(x)g_1(x) \, dx + \int_0^1 f_2(x)g_2(x) \, dx$, $f = \text{col}(f_1, f_2)$, $g = \text{col}(g_1, g_2) \in X$,

$$\langle x, y \rangle_s = \sum_{i=1}^m s_i x_i y_i, \quad x, y \in \mathbb{R}^m,$$
If (H1), (H2), (H5) and (H6) hold, then for every

\[
\langle f, x \rangle, \langle g, y \rangle \rangle_Y = \langle f, g \rangle_X + \langle x, y \rangle, \quad \left( f, \frac{g}{x} \right) \in Y.
\]

Now we define the operator \( A : D(A) \subset Y \to Y \),

\[ D(A) = \left\{ y = \text{col}(u, v, w); \ u, v \in H^2(0,1), \ w \in \mathbb{R}^m, \ \text{col}(\delta_0v, w) \in D(G), \right\} \]

where \( \delta_0v = \text{col}(w_1, \ldots, w_m) \), \( \delta_0v = \text{col}(v(0), v(1), v'(0), v'(1)) \), \ \delta_1u = \text{col}(a_1(0)u(0) - (a_2u)'(0), -a_1(1)u(1) + (a_2u)'(1)), \ a_2(0)u(0), -a_2(1)(u(1))\),

\[ A \left( \begin{pmatrix} u \\ v \\ w \end{pmatrix} \right) = \begin{pmatrix} a_2v'' + a_1v' + a_0v \\ -\lambda_0 = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \in D(A). \]

We also define the operator \( B : D(B) \subset Y \to Y \), \( \text{B}(\text{col}(u, v, w)) = \text{col}(\alpha(\cdot, u), \beta(\cdot, v), \ 0), \ \text{col}(u, v, w) \in D(B), \ \text{with} \ D(B) = \left\{ y = \text{col}(u, v, w) \in Y, \ \mathbb{B}(y) \in Y \right\}. \)

If the assumptions (H1), (H2), (H5) and (H6) hold, then \( D(A) \neq \emptyset, \ D(A) = X \times \{G(1) \cap G(22]\) and \( D(A) \subset D(B) [3]. \)

**Lemma 2.1** If (H1), (H2), (H5) and (H6) hold, then the operator \( A + B \) is maximal monotone.

For the proof of Lemma 1 see [3].

**Lemma 2.2** If (H1), (H2), (H5) and (H6) hold, then for every \( \lambda > 0 \) the operator \((I + \lambda(A + B))^{-1}\) is compact in \( Y \).

**Proof.** Without loss of generality we suppose that \( G \) is single-valued. Let be \( \lambda > 0 \) and \( M = \{ \text{col}(f^i, g^i, h^i); i \in I \} \) a bounded set in \( Y \). We prove that the set \( (I + \lambda(A + B))^{-1}(M) \) is bounded in \((H^2(0,1))^2 \times \mathbb{R}^m \). For this purpose, we firstly remark that the set \{\text{col}(p^i, q^i, r^i); i \in I\} is bounded in \( Y \), where

\[
\begin{pmatrix} p^i \\ q^i \\ r^i \end{pmatrix} = (I + \lambda(A + B))^{-1} \begin{pmatrix} f^i \\ g^i \\ h^i \end{pmatrix} = \begin{pmatrix} p^i \\ q^i \\ r^i \end{pmatrix} + \lambda(A + B) \begin{pmatrix} p^i \\ q^i \\ r^i \end{pmatrix}, \quad i \in I,
\]

then the operator \((I + \lambda(A + B))^{-1}\) is nonexpansive in \( Y \).

The last relation is equivalent to the system

\[
\begin{aligned}
\frac{1}{\lambda} p^i &= a_2(q^i)'' + a_1(q^i)' + a_0q_i + \alpha(x, p^i(x)) = \frac{1}{\lambda} f^i, \\
\frac{1}{\lambda} q^i &= a_2(q^i)' + a_2 p^i - \alpha(x, q^i(x)) = \frac{1}{\lambda} g^i, \\
\frac{1}{\lambda} r^i &= S^{-1}G_{21}(\delta_0q^i) + S^{-1}G_{22}(r^i)] = \frac{1}{\lambda} h^i, \quad j = \overline{1, m} \\
\end{aligned}
\]

(1)

We use similar arguments as those used in [3, Theorem 1] and by (1) we deduce that the sets \{\text{col}(q^i)'' + a_1(q^i)' - a_0p^i; i \in I\} and \{a_2(q^i)' + a_1(q^i)'} +
$a_0q^i; \ i \in I$ are bounded in $L^1(0,1)$. Then, by [7, Lemma B] it follows that the sets
$$\{(p^i)^{(j)}; \ i \in I\} \text{ and } \{(q^i)^{(j)}; \ i \in I, \ j = 0, 1 \text{ are bounded in } C([0,1])$$
and the sets
$$\{(p^i)''; \ i \in I\} \text{ and } \{(q^i)''; \ i \in I\} \text{ are bounded in } L^1(0,1).$$

By (H2) we deduce that the sets $\{\alpha(\cdot, p^i); \ i \in I\}, \ {\beta(\cdot, q^i); \ i \in I\}$ are bounded in $L^2(0,1)$ and then (1)_{1,2} give us that the sets $\{(p^i)''; \ i \in I\}$ and $\{(q^i)''; \ i \in I\}$ are bounded in $L^2(0,1)$. So, the set $\{(p^i, q^i, r^i); \ i \in I\}$ is bounded in $(H^2(0,1))^2 \times \mathbb{R}^m$ and therefore the operator $(I + \lambda(A + B))^{-1}$ is compact in $Y$. Q.E.D.

We define now the operator $K : [0,a] \rightarrow L(Y),
$$K(t) = b(t) 0 0 0 c(t) 0 0 0 I, \ \forall t \in [0,a],$$
where $A, B : L^2(0,1) \rightarrow L^2(0,1)$ are defined by $A(u)(x) = A(u(x)), B(v)(x) = B(v(x))$, for a.a. $x \in (0,1)$.

Using the above operators, problem (S)+(BC)+(IC) can be equivalently expressed as a nonlinear Volterra integrodifferential problem in the space $Y$

$$\begin{align*}
(P) & \ \left\{ \begin{array}{l}
\frac{dy}{dt}(t) + (A + B)(y(t)) + \int_0^t K(t - s)C(y(s)) \, ds \ni F(t, \cdot), \
y(0) = y_0.
\end{array} \right.
\end{align*}$$

where $y(t) = \text{col}(u(t), v(t), w(t)), F(t, \cdot) = \text{col}(f(t, \cdot), g(t, \cdot), 0), y_0 = \text{col}(u_0, v_0, w_0)$.

**Definition 2.1** A function $y = \text{col}(u, v, w) : [0, b] \rightarrow Y (b > 0)$ is called a weak (strong) solution on $[0, b]$ to the problem (P) with $y_0 = \text{col}(u_0, v_0, w_0) \in \overline{D(A)}$ if $y(t) \in \overline{D(A)}, \ \forall t \in (0, b), \ \text{for a.a. } t \in (0, b), \ \text{the function } h : [0, b] \rightarrow Y \ \text{defined by}$

$$h(t) = -\int_0^t K(t - s)C(y(s)) \, ds + F(t, \cdot), \ \forall t \in [0, b]$$

belongs to $L^1(0, b; Y)$ and $y$ is a weak (respectively strong) solution in the known sense [1] to the problem

$$\begin{align*}
\begin{array}{l}
\frac{dy}{dt}(t) + (A + B)(y(t)) \ni h(t), \ 0 \leq t \leq b \\
y(0) = y_0.
\end{array}
\end{align*}$$

For the definition in a general case (for mild solutions in a Banach space) see [8].
We say that \( y = \text{col}(u, v, w) \) is a weak (strong) solution of the problem (S)+(BC)+(IC) if \( y \) is a weak (respectively strong) solution in the above sense to the problem (P).

3. EXISTENCE OF WEAK AND STRONG SOLUTIONS TO PROBLEM (S)+(BC)+(IC)

First we present an existence result for the weak solutions of our problem.

**Theorem 3.1** Suppose that the assumptions (H1)-(H6) hold. If \( \text{col}(u_0, v_0, w_0) \in D(A), \ f, g \in L_{\text{loc}}^1(\mathbb{R}_+; L^2(0, 1)) \), then there exists \( T > 0 \), \( T \in (0, a] \) such that the problem (S)+(BC)+(IC) has at least one weak solution \( y = \text{col}(u, v, w) \in C([0, T]; Y) \).

**Proof.** Under the assumption (H3) the operator \( K \) is continuous. Indeed, let \( t_0 \in [0, a] \) and \( t_j \to t_0 \), as \( j \to \infty \), in \( \mathbb{R} \). Then

\[
\|K(t_j) - K(t_0)\|_{L(Y)} = \sup_{\|y\|_Y \leq 1} \|K(t_j) - K(t_0)(y)\|_Y \\
\leq \sup_{\|y\|_Y \leq 1} \left[ \|b(t_j) - b(t_0)\|_{L^2(0, 1)} + \|c(t_j) - c(t_0)\|_{L^2(0, 1)} \right] = \\
= \sup_{\|y\|_Y \leq 1} \left[ |b(t_j) - b(t_0)| + |c(t_j) - c(t_0)| \right] \leq 0, \text{ as } t \to \infty
\]

(b, c are continuous in \( t_0 \)). This yields \( K(t_j) \to K(t_0) \), as \( j \to \infty \), in \( L(Y) \).

If the assumption (H4) holds, then the operator \( \mathcal{C} \) is well-defined, that is, for every \( y \in Y \) it follows that \( \mathcal{C}(y) \in Y \). Indeed, let \( y = \text{col}(u, v, w) \) be an arbitrary element from \( Y \), fixed for the moment. Then the functions \( \tilde{A}(u) \) and \( \tilde{B}(v) \) are measurable and

\[
\|\mathcal{C}(y)\|_Y = \sqrt{\|\tilde{A}(u)\|_{L^2(0, 1)}^2 + \|\tilde{B}(v)\|_{L^2(0, 1)}^2} = \sqrt{\int_0^1 |A(u(x))|^2 \, dx + \int_0^1 |B(v(x))|^2 \, dx} \leq \\
\leq \sqrt{2 \int_0^1 (d^2|u(x)|^2 + \varepsilon^2) \, dx + 2 \int_0^1 (d^2|v(x)|^2 + \varepsilon^2) \, dx} < +\infty,
\]

\((u, v) \in L^2(0, 1)\). Therefore \( \mathcal{C}(y) \in Y \). Moreover, the operator \( \mathcal{C} \) is continuous, that is if \( y_j \to y \), as \( j \to \infty \) in \( Y \) then \( \mathcal{C}(y_j) \to \mathcal{C}(y) \), as \( j \to \infty \).

We shall prove that if \( u_j \to u \) as \( j \to \infty \) in \( L^2(0, 1) \), then \( \tilde{A}(u_j) \to \tilde{A}(u) \) as \( j \to \infty \), in \( L^2(0, 1) \). The proof of this conclusion is based on Vitali’s theorem: first, we show that \( \tilde{A}(u_j) \to \tilde{A}(u) \), as \( j \to \infty \), in measure and then, using (H4) we show that the sequence \( \{(\tilde{A}(u_j) - \tilde{A}(u))^2\} \) is uniformly integrable. Then, it follows that \( (\tilde{A}(u_j) - \tilde{A}(u))^2 \to 0 \), as \( j \to \infty \) in \( L^1(0, 1) \), hence \( \tilde{A}(u_j) - \tilde{A}(u) \to 0 \), as \( j \to \infty \), in \( L^2(0, 1) \).
By Lemma 1 and Lemma 2 the operator \( A + B \) is maximal monotone and for every \( \lambda > 0 \) the operator \( (I + \lambda (A + B))^{-1} \) is compact in the space \( Y \). Now, using [8, Theorem 5.1.1] we deduce that for \( \text{col}(u_0, v_0, w_0) \in \overline{D(A)} \), \( f, g \in L^1_{\mathrm{loc}}(\mathbb{R}^+; L^2(0, 1)) \), there exists \( T > 0 \), \( T \in (0, a] \) such that the problem \( (P) \) (and also the problem \( (S)+(BC)+(IC) \)) has at least one weak solution \( y = \text{col}(u, v, w) \in C([0, T]; Y) \). Q.E.D.

By Lemma 1, Lemma 2 and [8, Corollary 5.1.1] we obtain

**Corollary 3.1** Suppose that the assumptions \((H1), (H2), (H4)-(H6)\) hold. If \( b, c \in C^1([0, a]), y_0 = \text{col}(u_0, v_0, w_0) \in D(A), f, g \in W^{1,1}_{\text{loc}}(\mathbb{R}^+; L^2(0, 1)) \), then there exists \( T > 0 \), \( T \in (0, a] \) such that the problem \( (S)+(BC)+(IC) \) has at least one strong solution \( y = \text{col}(u, v, w) \in W^{1,\infty}(0, T; Y) \), that is \( y(t, \cdot) \in D(A), \forall t \in [0, T), y \) satisfies the equations of the system \( (S) \), the boundary condition \( (BC) \) and the initial data \( (IC) \) for a.a. \( x \in (0, 1) \) and for all \( t \in [0, T) \), where \( \partial u/\partial t, \partial v/\partial t \) and \( \partial w/\partial t \) are replaced by \( \partial^+ u/\partial t, \partial^+ v/\partial t \), respectively \( \partial^+ w/\partial t \). Moreover \( u, v \in L^1(0, T; H^2(0, 1)) \).

Now we introduce the assumptions

\((H3)’\) The functions \( b, c : [0, \infty) \to \mathbb{R} \) are continuous.

\((H7)\) There exists at least one solution \( y = \text{col}(u, v, w) \in D(A) \) to the equation

\[
A(y) + B(y) \ni 0.
\]

**Theorem 3.2** Assume that the assumptions \((H1)-(H2), (H3)’, (H4)-(H7)\) hold. If \( \text{col}(u_0, v_0, w_0) \in \overline{D(A)}, f, g \in L^1_{\text{loc}}(\mathbb{R}^+; L^2(0, 1)) \), then the problem \( (S)+(BC)+(IC) \) has at least one weak solution \( y = \text{col}(u, v, w) \) defined on the positive half-axis.

**Proof.** By \((H7)\) we have that \( F = (A + B)^{-1}(0) \neq \emptyset \) and using \((H4)\) we obtain

\[
\|C(y)\|_Y \leq \sqrt{d_0 \left( \|u\|_{L^2(0, 1)}^2 + \|v\|_{L^2(0, 1)}^2 \right)} + e_0 \leq \tilde{d}_0 \|y\|_Y + \tilde{e}_0, \forall y \in Y,
\]

where \( d_0 = 2 \max\{d^2, \tilde{d}^2\}, e_0 = 2(e^2 + \tilde{e}^2) \), \( \tilde{d}_0 = \sqrt{d_0}, \tilde{e}_0 = \sqrt{e_0} \).

In what follows, we shall prove that every weak solution of the problem \( (P) \), \( \text{col}(u, v, w) : [0, t] \to \overline{D(A)} \) \( (t > 0) \) is bounded on \([0, l)\). For, let \( \gamma = \text{col}(p, q, r) \in F \). We multiply the equation \( (P)_1 \) by \( y(t) - \gamma \) in the space \( Y \)

\[
\frac{1}{2} \frac{d}{dt} \|y(t) - \gamma\|_Y^2 < \int_0^t K(t-s)C(y(s)) \, ds, y(t) - \gamma < \gamma \leq \gamma \leq \gamma \Rightarrow \frac{1}{2} \frac{d}{dt} \|y(t) - \gamma\|_Y^2 \leq \left( \int_0^t \|K(t-s)\|_{L(Y)} \|C(y(s))\|_Y \, ds + \|F(t, \cdot)\|_Y \right) \|y(t) - \gamma\|_Y \]

\[
\Rightarrow \frac{1}{2} \frac{d}{dt} \|y(t) - \gamma\|_Y^2 \leq \int_0^t ([|b(t-s)| + |c(t-s)|]) \left( \tilde{d}_0 \|y(s)\|_Y + \tilde{e}_0 \right) \, ds +
\]
+\| F(t, \cdot) \|_{Y} \| y(t) - \gamma \|_{Y}, \ 0 < t < l \\
\Rightarrow \frac{d}{dt} \| y(t) - \gamma \|_{Y} \leq \int_{0}^{t} \left( |b(t-s)| + |c(t-s)| \right) (\tilde{d}_{0} \| y(s) - \gamma \|_{Y} + \tilde{d}_{0} \| \gamma \|_{Y} + \tilde{e}_{0}) \ ds + \\
+ \left\| \left( \begin{array}{c} f(t, \cdot) \\ g(t, \cdot) \end{array} \right) \right\|_{X}, \ 0 < t < l.
\]

We integrate the above inequality over \([0, t]\)
\[
\| y(t) - \gamma \|_{Y} \leq \| y_{0} - \gamma \|_{Y} + \int_{0}^{t} \int_{0}^{s} \left( |b(s-\theta)| + |c(s-\theta)| \right) (\tilde{d}_{0} \| y(\theta) - \gamma \|_{Y} + \tilde{d}_{0} \| \gamma \|_{Y}) \ ds + \\
+ \tilde{e}_{0} \ d\theta \ ds + \int_{0}^{t} \left\| \left( \begin{array}{c} f(s, \cdot) \\ g(s, \cdot) \end{array} \right) \right\|_{X} \ ds, \ 0 \leq t < l.
\]

From the obtained inequality, using the assumptions of the theorem we deduce that there exist the constants \(C_{1}, C_{2} > 0\) such that
\[
\| y(t) - \gamma \|_{Y} \leq C_{1} + C_{2} \int_{0}^{t} \| y(s) - \gamma \|_{Y} \ ds,
\]

\((C_{1}, C_{2} \text{ are dependent on } b, c, [0, l], \gamma \text{ and } l).\)

Now, by Bellman’s inequality, it follows that \(y\) is bounded on \([0, l]\). Using Lemma 1, Lemma 2 and [8, Theorem 5.1.3] we deduce that the problem (S)+(BC)+(IC) has at least one weak solution defined on the positive half-axis (that is a weak solution from Theorem 1 can be extended on \([0, \infty])\). Q.E.D.

**Remark.** If, in addition to the assumptions (H1)-(H2), (H5)-(H6), we suppose that

a) there exist \(m_{1}, m_{2} > 0\) such that
\[
(\alpha(x, p_{1}) - \alpha(x, p_{2}))(p_{1} - p_{2}) \geq m_{1}(p_{1} - p_{2})^{2}, \\
(\beta(x, p_{1}) - \beta(x, p_{2}))(p_{1} - p_{2}) \geq m_{2}(p_{1} - p_{2})^{2},
\]

for a.a. \(x \in (0, 1)\) and for all \(p_{1}, p_{2} \in \mathbb{R}_{+}\),

b) there exists \(M > 0\) such that for all \(x, y \in D(G), x = \text{col}(x^{a}, x^{b}), y = \text{col}(y^{a}, y^{b}) \in \mathbb{R}^{4} \times \mathbb{R}^{m}\) and \(w_{1} \in G(x), w_{2} \in G(y)\) we have
\[
< w_{1} - w_{2}, x - y >_{\mathbb{R}^{m \times 4}} \geq M \| x^{b} - y^{b} \|_{\mathbb{R}^{m}},
\]

then the equation (2) has a unique solution \(y = \text{col}(u, v, w) \in D(A)\) and (H7) is fulfilled. Indeed, under the above assumptions the operator \(A + B\) is strongly monotone, so it is coercive. Therefore \(R(A + B) = Y\) and we obtain the above conclusion.

**References**


SURVEY ON COMPUTER REALIZATION OF BRANCHING EQUATION CONSTRUCTION ON ALLOWED GROUP SYMMETRY AND SUBGROUP INARIANT SOLUTIONS

Oleg V. Makeev
Ulyanovsk State Technical University, Russia

Abstract
In order to study bifurcation problems, the nonlinear equations of stationary and dynamical branching are written in operator form in Banach spaces. The Lyapounov-Schmidt asymptotical method reduces their solving to the construction and investigation of finite-dimensional nonlinear algebraic systems. Multiple degeneracy of linearized operator frequently stipulated by the group of symmetries of the nonlinear problem entails technical difficulties at the construction of the branching equation (BEq). Herein the computer programs are suggested for the construction of the general form of the Lyapounov-Schmidt branching equation both for stationary and Andronov-Hopf bifurcation on the allowed group of symmetries with its subsequent investigation. Special attention is paid to planar and spatial crystallographic groups. For a discrete group of symmetries the program for the determination of subgroup structure and dual to it structure of branching systems is presented with applications to nonlinearly perturbed Helmholtz equation.

1. INTRODUCTION

Many applied problems of critical phenomena can be investigated as bifurcational symmetry breaking problems. At multivariate branching the nonlinear problem often has one or several families of solutions depending on free parameters. The existence of such solutions as a rule is related to the presence of the original problem symmetry. In bifurcational symmetry breaking problems the original nonlinear equation is invariant with respect to the Euclidean space $\mathbb{R}^s$ ($s \geq 1$) motion group and its solution invariant with respect to this group is the rest state or uniform motion. At the loss of stability there arise solutions with cellular structure, i.e. periodical solutions with crystallographic group of symmetries (semi-direct product $G = G_1 \rtimes \tilde{G}^1$ of $s$-parametrical continuous shift group $G_1(\alpha_1, \ldots, \alpha_s)$ and the group $\tilde{G}^1$ of the elementary cell of periodicity), which are mutually transformed by the action of the group $\tilde{G}^1$. These problems can be solved by the construction and investigation of the relevant branching equation (BEq).
The general scheme for the construction of sufficiently smooth (analytical) BEq of a general form on the allowed group of symmetries both for stationary and dynamic situations was developed in a series of B. V. Loginov’s papers [2, 9] on the base of group analysis methods for differential equations [8]. However, at the realization of this scheme to concrete applications serious difficulties have arisen. Usually, because of the high order of degeneracy, undergone by the linear operator at the bifurcation point, the investigator can not account for all possible factors and it is impossible to perform this scheme in manual. Whence the need for a computer realization of the indicated scheme. Such a realization is considered here. Computer programs for BEq construction and investigation are realized for BEq with simple cubic lattice symmetry. The applications to statistical crystal theory are known [2, 3, 9].

In various applications the problem of subgroup invariant solutions arises. Computer program is proposed by which all subgroups of discrete symmetries with known composition law can be found. Here it is applied to the concrete problem of the nonlinearly perturbed Helmholtz equation for square lattice symmetry.

The base consists of the group symmetry inheritance theorem for branching equation and the methods of group analysis of differential equations. Partially, the description of the proposed programs of the construction of general BEq is presented in [5]. In the following the terminology, designation of indicated works and results of program application are used without any additional explanation. For the sake of simplicity of presentation here only dynamic (Andronov-Hopf) bifurcation is considered.

Let $E_1$ and $E_2$ be two Banach spaces. The problem about periodical solutions of differential equation with a small parameter $\varepsilon$ reads

$$A \frac{dx}{dt} = Bx - R(x, \varepsilon), \quad R(0, \varepsilon) \equiv 0, \quad R_x(0, 0) = 0. \quad (1)$$

Here $A$ and $B$ are densely defined closed linear Fredholm operators $A: D_A \subset E_1 \to E_2$, $B: D_B \subset E_1 \to E_2$. Moreover $D_A \supset D_B$ and $A$ is subordinated to $B$ (i.e. $\|Ax\|_{E_2} \leq \|Bx\|_{E_2} + \|x\|_{E_1}$ on $D_B$) or $D_B \supset D_A$ and $B$ is subordinated to $A$ (i.e. $\|Bx\|_{E_2} \leq \|Ax\|_{E_2} + \|x\|_{E_1}$ on $D_A$). $R(x, \varepsilon)$ is nonlinear operator sufficiently smooth in a neighborhood of $(0, 0) \subset E_1 + R^1$.

The theorem about symmetry inheritance by relevant Lyapounov-Schmidt BEq $0 = f(\xi, \hat{\xi}, \mu, \varepsilon) = \{f_j(\xi, \hat{\xi}, \mu, \varepsilon)\}_{j=1}^n : \Xi^n \to \Xi^n$ at Andronov-Hopf bifurcation is expressed by the equality

$$f(Ag_\xi, \hat{A}g_\xi, \mu, \varepsilon) = Bg_f(\xi, \hat{\xi}, \mu, \varepsilon). \quad (2)$$

Here $Ag$ and $Bg$ are $n$-dimensional representations of the group $G$, which is allowed by the original nonlinear equation, in the zero-subspace $N$ and defect subspace $N^*$ of the operator linearized at the bifurcation point.
The equality (2) means that the manifold \( F : f \equiv f(\xi, \bar{\xi}) = 0 \) is invariant to the transformation group \( \tilde{\xi} = A_g \xi, \tilde{f} = B_g f \). Consequently, for the construction of the general BEq on the allowed group of symmetries the Lie-Ovsyannikov invariants and invariant manifolds techniques [8] may be applied.

If the left-hand side of the branching system is assumed to be continuous on \( \xi \)-variables, then for the construction of the general BEq it is sufficient to have only a complete system of functional independent invariants. In the analytic case, at the BEq expansion on homogeneous forms not all invariants can be expressed via powers of basic ones. But the use of additional invariants leads to the repetition of BEq summands. Therefore, when using additional invariants, the expansion of the BEq by invariant monomials should be factorized subject to constraints between the used invariants. This factorization with respect to the expression inside the brackets will be designated by the symbol \( \cdots \). Thus, according to the group-analysis scheme of the BEq, for the construction of its general form one must:
a) find the complete system of functionally independent invariants, allowed by BEq continuous group, b) introduce the additional invariants, c) establish the constraints between the used invariants and d) make the factorization of the BEq expansion by these constraints, i.e. construct the symbol \( \cdots \) out. All tasks of this logic-combinatorial problem are realized by means of computer programs for examples of construction of BEqs, which admit the simple cubic lattice symmetry.

The obtained results are entered into RFBR- and INTAS-06 applications.

2. **LYAPOUNOV-SCHMIDT BEQ CONSTRUCTION WITH SIMPLE CUBIC LATTICE SYMMETRY**

Here the construction of BEq for symmetry breaking problems of Andronov-Hopf bifurcation is presented when spatial cell of periodicity is an octahedron. Let \( \dim N = 12 \) for the choice of its basis of the form \( \{ \varphi_j = e^{i[1]_j q_1}, \varphi_j, \}_{1}^{6} \) with inverse lattice vectors \( \mathbf{l}_{2k - 1} = 2 \pi \mathbf{e}_k, \mathbf{l}_{2k} = -\mathbf{l}_{2k - 1}, k = 1, 2, 3 \). The BEq is invariant with respect to the reflection-rotation octahedron group, which is generated by

\[
C_4^{(1)} \cong (1)(2)(3546), \quad C_4^{(2)} \cong (1625)(3)(4), \\
C_4^{(3)} \cong (1324)(5)(6), \quad J \cong (12)(34)(56).
\]

The BEq continuous symmetry group has the form

\[
A_{g(a)} = \text{diag}\{e^{i(2\pi \alpha_1 + \alpha_0)}, e^{-i(2\pi \alpha_1 + \alpha_0)}, e^{i(-2\pi \alpha_1 + \alpha_0)}, e^{-i(-2\pi \alpha_1 + \alpha_0)}, e^{i(2\pi \alpha_2 + \alpha_0)}, e^{-i(2\pi \alpha_2 + \alpha_0)}, e^{i(-2\pi \alpha_2 + \alpha_0)}, e^{-i(-2\pi \alpha_2 + \alpha_0)}, e^{i(2\pi \alpha_3 + \alpha_0)}, e^{-i(2\pi \alpha_3 + \alpha_0)}, e^{i(-2\pi \alpha_3 + \alpha_0)}, e^{-i(-2\pi \alpha_3 + \alpha_0)}\}.
\]
By differentiating on $\alpha_0, \ldots, \alpha_3$ and reducing on common multiplier we receive the Lie algebra basis that corresponds to rotations in the coordinate planes $(\xi_{2k-1}, \xi_{2k}), \ k = 1, 2, 3$

\[
X_0 = (\xi_1, -\xi_1, \xi_2, -\xi_2, \xi_3 - \xi_3, \xi_4 - \xi_4, \xi_5 - \xi_5, \xi_6 - \xi_6),
\]
\[
X_1 = (\xi_1, -\xi_1, -\xi_2, \xi_2, 0, 0, 0, 0, 0, 0, 0, 0),
\]
\[
X_2 = (0, 0, 0, 0, \xi_3 - \xi_3, -\xi_3 - \xi_4, \xi_4, 0, 0, 0, 0, 0),
\]
\[
X_3 = (0, 0, 0, 0, 0, 0, 0, \xi_5, -\xi_5, -\xi_6, \xi_6).
\]

Then the basic invariants are defined by the next system of first order partial differential equations

\[
X^i_\nu \frac{\partial I}{\partial \xi_i} + F^j_\nu \frac{\partial I}{\partial f_j} = 0, \ \nu = 0, 1, 2, 3. \tag{3}
\]

Hereinafter in every monomial expression the symbol $\xi$ will be omitted, i.e. for example, the notation $\xi_1 \xi_2 \xi_3 \xi_4 = 1 \ 2 \bar{3} \bar{4}$ is used. The computer program selects six invariants of second order $\xi_k \xi_k, \ k = 1, \ldots, 6$ and next six invariants of fourth order

\[
1 \ 2 \bar{3} \bar{4}; \ 1 \ 2 \bar{5} \bar{6}; \ 1 \bar{2} \ 3 \ 4; \\
1 \bar{2} \ 5 \ 6; \ 3 \bar{4} \ 5 \bar{6}; \ 3 \ 4 \ 5 \ 6.
\]
Three standard constraints between fourth order invariants

\[
\begin{align*}
&1\,2\,3\,4\times\bar{1}\,\bar{2}\,5\,6 = 1\,\bar{1}\times2\,\bar{2}\times3\,\bar{3}\,\bar{4}\,\bar{5}\,\bar{6}, \\
&1\,2\,3\,4\times3\,4\,\bar{5}\,\bar{6} = 3\,\bar{3}\times4\,\bar{4}\times1\,\bar{2}\,5\,\bar{6}, \\
&1\,\bar{2}\,5\,\bar{6}\times3\,4\,\bar{5}\,\bar{6} = 5\,\bar{5}\times6\,\bar{6}\times1\,2\,3\,4
\end{align*}
\]

are separating the system of three fourth order invariants that are used for the BEq general construction

\[
1\,2\,3\,\bar{4}, \quad 1\,\bar{2}\,5\,\bar{6}, \quad 3\,4\,\bar{5}\,\bar{6}.
\]

This system is subordinated to one nonstandard constraint

\[
1\,2\,3\,\bar{4}\times1\,\bar{2}\,5\,\bar{6}\times3\,4\,\bar{5}\,\bar{6} = 1\,\bar{1}\times2\,\bar{2}\times3\,\bar{3}\times4\,\bar{4}\times5\,\bar{5}\times6\,\bar{6}.
\]

Thus the BEq corresponding to adopted symmetry takes the form \(f_1(\xi, \bar{\xi}, \mu, \varepsilon)\)

\[
\equiv \sum_{p_{a}, q_{\beta}} a_{p_{a} q_{\beta}} (\xi_{1} \bar{\xi}_{1} \xi_{2} \bar{\xi}_{2})^{p_{a}} (\xi_{3} \bar{\xi}_{3} \xi_{4} \bar{\xi}_{4})^{q_{\beta}} = 0.
\]

The symbol \([\cdots]^{\text{out}}\) occurs in the following way

\[
f_{1}(\xi, \bar{\xi}, \mu, \varepsilon) \equiv \sum_{p_{a}, q_{\beta}j_{1}, j_{2}} (\xi_{1} \bar{\xi}_{1})^{p_{a}} (\xi_{3} \bar{\xi}_{3})^{q_{\beta}} [a_{p_{a} q_{\beta}j_{1}, j_{2}} \xi_{1} + a_{p_{a} q_{\beta}j_{2}, j_{1}} \bar{\xi}_{1} + \cdots]
\]

where \(p_{j}\) consecutively takes the values \(C_{4}^{(3)}, C_{4}^{(2)}\). The first equation of the written branching system is invariant with respect to the octahedron subgroup generated by substitutions \(e, C_{4}^{(1)}, C_{4}^{(1)} \circ C_{4}^{(2)}, J \circ C_{4}^{(1)} \circ C_{4}^{(3)} \circ C_{4}^{(2)}, J \circ C_{4}^{(1)} \circ u_{13}, J \circ u_{58} = J \circ C_{4}^{(1)} \circ C_{4}^{(3)}\) that preserve the number 1 that gives the symmetry relation between its coefficients.
3. BEQ INVESTIGATION

At the Andronov-Hopf bifurcation investigation (i.e. in dynamical branching) difficulties arise too. They are generated by the fact that the branching system includes, except for $n$ complex variables, the unknown small additional contribution $\mu$ to the oscillation frequency that defines the limit cycle onset. All these variables should be calculated from the branching system. In the V. I. Yudovich articles (for example see [7]) the asymptotic approach for the determination of periodical branching solution was suggested. It is based on subspaces invariant techniques relatively to the BEQ left-hand side. After the solutions determination in a certain subspace as a consequence of group symmetry we have the solutions as trajectories of this subspace. Hence there arise the need of computer program for the construction of the subspace system. For the description of this program we refer to [5].

In the example of BEQ (2.1) with simple cubic lattice symmetry in case of the octahedron cell the computer program gives 1795 invariant subspaces. However their complete system consists of the following 46 subspaces

$$\Xi_1 = (\xi_1, 0, 0, 0, 0, 0), \Xi_2 = (\xi_1, \xi_1, 0, 0, 0, 0), \Xi_3 = (\xi_1, 0, \xi_1, 0, 0, 0), \Xi_4 = (\xi_1, \xi_2, 0, 0, 0, 0), \Xi_5 = (\xi_1, 0, \xi_2, 0, 0, 0), \Xi_6 = (\xi_1, \xi_1, 0, 0, 0, 0), \Xi_7 = (\xi_1, 0, \xi_1, 0, 0, 0),$$

$$\Xi_8 = (\xi_1, 0, \xi_2, 0, 0, 0), \Xi_9 = (\xi_1, 0, \xi_1, \xi_1, 0, 0), \Xi_{10} = (\xi_1, \xi_2, 0, 0, 0, 0), \Xi_{11} = (\xi_1, 0, \xi_2, 0, 0, 0), \Xi_{12} = (\xi_1, 0, \xi_1, \xi_1, 0, 0), \Xi_{13} = (\xi_1, \xi_1, 0, 0, 0, 0), \Xi_{14} = (\xi_1, 0, \xi_1, \xi_1, 0, 0), \Xi_{15} = (\xi_1, \xi_1, \xi_2, 0, 0, 0), \Xi_{16} = (\xi_1, \xi_2, 0, 0, 0, 0),$$

$$\Xi_{17} = (\xi_1, 0, \xi_1, 0, 0, \xi_1), \Xi_{18} = (\xi_1, \xi_2, \xi_1, 0, 0, 0), \Xi_{19} = (\xi_1, \xi_2, 0, \xi_1, 0, 0), \Xi_{20} = (\xi_1, 0, \xi_1, \xi_2, 0, 0), \Xi_{21} = (\xi_1, \xi_1, \xi_2, 0, 0, 0), \Xi_{22} = (\xi_1, \xi_1, \xi_2, 0, 0, 0),$$

$$\Xi_{23} = (\xi_1, 0, \xi_2, 0, 0, 0), \Xi_{24} = (\xi_1, \xi_2, \xi_2, 0, 0, 0), \Xi_{25} = (\xi_1, 0, \xi_2, \xi_2, 0, 0, 0), \Xi_{26} = (\xi_1, 0, \xi_2, \xi_3, 0, 0), \Xi_{27} = (\xi_1, 0, \xi_2, 0, 0, \xi_3), \Xi_{28} = (\xi_1, 0, \xi_1, \xi_2, \xi_2), \Xi_{29} = (\xi_1, 0, \xi_1, \xi_1, \xi_1),$$

$$\Xi_{30} = (\xi_1, \xi_1, 0, \xi_1, \xi_1, 0), \Xi_{31} = (\xi_1, 0, \xi_1, \xi_2, \xi_3), \Xi_{32} = (\xi_1, \xi_1, \xi_2, \xi_2, 0, 0), \Xi_{33} = (\xi_1, \xi_1, \xi_2, \xi_3, 0, 0), \Xi_{34} = (\xi_1, 0, \xi_1, \xi_2, \xi_3, 0, 0),$$

$$\Xi_{35} = (\xi_1, 0, \xi_1, \xi_2, \xi_3, 0, 0), \Xi_{36} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_3), \Xi_{37} = (\xi_1, 0, \xi_2, \xi_3, 0, 0), \Xi_{38} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_3),$$

$$\Xi_{39} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_3), \Xi_{40} = (\xi_1, 0, \xi_1, \xi_2, \xi_2, 0, 0), \Xi_{41} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_1), \Xi_{42} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_1), \Xi_{43} = (\xi_1, \xi_1, \xi_2, 0, 0, \xi_1),$$

$$\Xi_{44} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_1), \Xi_{45} = (\xi_1, \xi_1, \xi_2, \xi_3, \xi_3, \xi_3), \Xi_{46} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_2, 0, 0),$$

$$\Xi_{47} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_2, 0, 0), \Xi_{48} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_2, 0, 0), \Xi_{49} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_2, 0, 0), \Xi_{50} = (\xi_1, \xi_2, \xi_1, \xi_1, \xi_2, 0, 0).$$

The main part of the BEQ (2.1) has the form

$$i \mu \xi_1 + a \xi_1 \varepsilon + (b + c) \xi_1 |\xi_1|^2 + 2d_1 |\xi_2|^2 + e \xi_1 \xi_2^2 = 0,$$

As an example of invariant subspaces techniques consider this branching system in the subspace $\Xi_{21} = (\xi_1, \xi_2, \xi_2, 0, 0, 0)$
After the determination of $\xi_1 \neq 0$, $\xi_2 \neq 0$, and let us search for the solutions in the form

$$\xi_1 = r_1 e^{1/2} \exp(i\theta_1), \quad \xi_2 = r_2 e^{1/2} \exp(i\theta_2), \quad \mu = \nu e.$$ Dividing the first equation by $\varepsilon^{3/2} r_1 \exp(i\theta_1)$, the second one by $\varepsilon^{3/2} r_2 \exp(i\theta_2)$, we get the system

$$\begin{align*}
(b + c)r_1^2 + (2d + e \exp(i\alpha))r_2^2 &= -i\nu - a, \\
(2d + e \exp(-i\alpha))r_1^2 + (b + c)r_2^2 &= -i\nu - a, \\
\alpha &= 2(\theta_2 - \theta_1),
\end{align*}$$

whence

$$\begin{align*}
r_1^2 &= \frac{\Delta_1}{\Delta} = \frac{\Delta_1 \bar{\Delta}}{|\Delta|^2}, \\
r_2^2 &= \frac{\Delta_2}{\Delta} = \frac{\Delta_2 \bar{\Delta}}{|\Delta|^2},
\end{align*}$$

where

$$\begin{align*}
\Delta &= (b + c)^2 - (2d - b - c + e \exp(i\alpha))(2d - b - c + e \exp(-i\alpha)), \\
\Delta_1 &= (i\nu + a)(2d - b - c + e \exp(i\alpha)), \\
\Delta_2 &= (i\nu + a)(2d - b - c + e \exp(-i\alpha)).
\end{align*}$$

Since $r_1^2$ and $r_2^2$ are real, it follows the next system of two equations for the determination of $\nu$ and $\alpha$: $\text{Im}(\Delta_1 \bar{\Delta}) = \text{Im}(\Delta_2 \bar{\Delta}) = 0$, $\text{Im}(\Delta_1 \bar{\Delta}) = \text{Im}\{((b + c)^2 - 4d^2 - e^2 - 4de \cos \alpha) \cdot (i\nu + a)(2d - b - c + e \cos \alpha + i \sin \alpha)\} = 0$, $\text{Im}(\Delta_2 \bar{\Delta}) = \text{Im}\{((b + c)^2 - 4d^2 - e^2 - 4de \cos \alpha) \cdot (i\nu + a)(2d - b - c + e \cos \alpha - i \sin \alpha)\} = 0$.

This system reads, equivalently, in the form

$$\begin{align*}
p\nu + (u + sv) \cos \alpha + (v + r\nu) \sin \alpha + q &= 0, \\
p\nu + (u + sv) \cos \alpha - (v + r\nu) \sin \alpha + q &= 0,
\end{align*}$$

where

$$\begin{align*}
p &= \text{Re}(2d - b - c), \\
q &= \text{Im}(a(2d - b - c)), \\
u &= \text{Re}(a e), \\
v &= \text{Re}(a e), \\
r &= -\text{Im}(e).
\end{align*}$$

After the determination of $\nu$ from every equation of the system, it follows the system for the determination of $\alpha$, i.e.

$$\nu = -\frac{q + u \cos \alpha + v \sin \alpha}{p + s \cos \alpha + r \sin \alpha} = -\frac{q + u \cos \alpha - v \sin \alpha}{p + s \cos \alpha - r \sin \alpha},$$

Hereinafter by using the scheme [1] in Maple 6.0 one gets

$$\alpha = 0, \quad \pi, \quad \pm \text{Arg} \left( \frac{\sqrt{2}qrvp - v^2p^2 + u^2r^2 - q^2r^2 + v^2s^2 - 2uruvs}{ur - vs} - i\frac{qu - vp}{ur - vs} \right).$$
Finally, $|\xi_1| = \frac{\Delta_1}{4}$, $|\xi_2| = \frac{\Delta_2}{4}$.

The construction of the asymptotic periodical branching solution in other subspaces is made as indicated above and it is not given here in view of its awkwardness.

4. Solutions Invariant to Subgroups of Discrete Symmetries

The general theory for finding the subgroup invariant solutions of ordinary differential equations was suggested in [10]. Then this theory was developed for bifurcation problems and it was realized in concrete problems for the construction of solutions invariant to normal divisors of discrete symmetry [2]. Here this theory is realized in the form of computer program for the construction of the subgroup structure for the discrete symmetry of nonlinear equation with subsequent formation of the relevant branching systems of solutions invariant to arbitrary subgroups. As a simple illustration of this abstract theory we consider here the equations

\[
\Delta u + \lambda^2 \sinh u = 0, \quad (6)
\]
\[
\Delta u + \lambda^2 \sin u = 0 \quad (7)
\]

and search for periodic solutions with square lattice of periodicity.

Applications of these equations to low temperature plasma theory are given in [6].

At the choice of the base $N(B) = \{\varphi_j\}_1^4 = \{\frac{1}{2\pi} e^{i\lambda_0(x+y)}, \frac{1}{2\pi} e^{-i\lambda_0(x+y)}, \frac{1}{2\pi} e^{i\lambda_0(y-x)}, \frac{1}{2\pi} e^{-i\lambda_0(y-x)}\}$ the four dimensional BEq inheriting the square symmetry group $D_4$ has the form [9]

\[
f_1(\xi, \varepsilon) \equiv a_0(\varepsilon)\xi_1 + \sum_{|p|\geq 2} a_{p_1p_2}(\varepsilon)\xi_1(\xi_1\xi_2)^{p_1}(\xi_3\xi_4)^{p_2} = 0,
\]
\[
f_2(\xi, \varepsilon) \equiv r^2 \circ f_1(\xi, \varepsilon) = f_1(r^2 \circ \xi, \varepsilon) = 0,
\]
\[
f_3(\xi, \varepsilon) \equiv r \circ f_1(\xi, \varepsilon) = f_1(r \circ \xi, \varepsilon) = 0,
\]
\[
f_4(\xi, \varepsilon) \equiv r^3 \circ f_1(\xi, \varepsilon) = f_1(r^3 \circ \xi, \varepsilon) = 0; \quad r = (1324), s = (13)(24). \quad (8)
\]

The leading terms of the branching system (4.0) are the following

\[
\xi_1 \varepsilon + A\xi_1^2 \xi_2 + B\xi_1 \xi_3 \xi_4 + \ldots = 0,
\]
\[
\xi_2 \varepsilon + A\xi_2^2 \xi_1 + B\xi_2 \xi_3 \xi_4 + \ldots = 0,
\]
\[
\xi_3 \varepsilon + A\xi_3^2 \xi_4 + B\xi_3 \xi_1 \xi_3 + \ldots = 0,
\]
\[
\xi_4 \varepsilon + A\xi_4^2 \xi_3 + B\xi_4 \xi_2 \xi_4 + \ldots = 0. \quad (9)
\]

where $A = \pm \lambda_0^2/4$, $B = \pm \lambda_0^2/2$, $\lambda_0^2 = \pi^2/a^2$, $a$ is the lattice width (the upper sign is related to equation (6) and the lower sign to (7))
The origin is the discrete square-group \( D_4 = \{ e, r, s, r^2, s, sr, sr^2, sr^3 \} \), \( (r^ks = sr^{4-k}, \ k = 1, 2, 3) \) generated by the permutations of basic element indices and the structure \( L(D_4) \) of all its subgroups. If \( H_0 = D_4 \supset H_1 \supset H_2 \supset \ldots \supset H_m = \{ H_k \}^m \) is some chain of subgroups of the length \( m \) then there exists the basis \( R_m \) in \( N \), in which the representation \( A_g \) for every subgroup \( H_i \) is splitting into irreducible ones. The set of all BEqs for \( H \)-invariant solutions forms the dual to inclusion structure \( L' \) to \( L(D_4) \): BEq of solutions which are invariant with respect to the more narrow subgroup contains the BEq of solutions which are invariant with respect to a more wide subgroup. For two chains \( A = \{ H_k \}^n \) and \( g^{-1}Ag = \{ g^{-1}H_k g \}^n \) of similar subgroups the connection between the \( H_k \)-invariant element subspaces and BEqs of \( H_k \)-invariant solutions respectively is realized by the element \( g \).

The basis application of this abstract theoretical result application is the computer program for the subgroup structure of discrete symmetry construction. By means of this program, for every subgroups chain from the complete branching system the subsystem determining solutions invariant with respect to every of its representatives is selected.

Consider its algorithm. The program for the subgroups construction of finite discrete symmetry uses the Kelly table. The proposed program begins the subgroups structure construction with the bottom upwards from trivial subgroup \( \{ e \} \). Furthermore the ability for subgroups verification on normal divisor property is realized. Let describe the algorithm steps.

1. Let some subgroup \( H_k = \{ e, g_{r_1}, \ldots, g_{r_k} \} \) be found. Add to this subset some element \( g \) that does not enter the set \( I_k \) (in the first step \( I_k = H_k \)).
2. Form the set that consists of all kinds of products of elements in \( H_{k+1} = H_k \cup g \). As the result of the formation of such products one gets some set \( H_{k+1} = \{ e, g_{r_1}, \ldots, g_{r_k}, g, g_{r_{k+1}}, \ldots, g_{r_{k+s}} \} \). The order of this set is the initial group order or its divisor.
3. In case of equality the set \( H_{k+1} \) coincides with the investigated group. Let us come back to the point 1 and add to the set \( I_k \) the element \( g \).
4. Otherwise we have find the subgroup. Add to the set \( I_k \) elements from the subgroup \( H_{k+s} \). For the found subgroup \( H_{k+s} \) repeat all actions from the point 1.

Thus the suggested algorithm contains the recursion, leading to the determination of the subgroup structure.
In the structure $L(D_4)$ the suggested program selects the following subgroup chains

$A_1$ : $D_4 = A_{1,0} \supset A_{1,1} = \{e, rs, sr, r^2\} \supset A_{1,2} = \{e, rs\}$,

$s^{-1} A_1 s = sA_1 s$ : $D_4 \supset sA_{1,1} s = \{e, rs, sr, r^2\} \supset sA_{1,2} s = \{e, sr\}$,  

conjugate with $A_1$ meaning the reflection $s$

$A_2$ : $D_4 = A_{2,0} \supset A_{2,1} = A_{1,1} = \{e, rs, sr, r^2\} \supset A_{2,2} = \{e, r^2\}$;

$A_3$ : $D_4 = A_{3,0} \supset A_{3,1} = \{e, s, r^2, r^2 s\} \supset A_{3,2} = \{e, r^2 s\}$,

$r^{-1} A_3 r = r^3 A_3 r$ : $D_4 \supset r^3 A_{3,1} r = \{e, s, r, r^2 s\} \supset r^3 A_{1,2} r = \{e, s\}$, conjugate with $A_3$

meaning the counterclockwise rotation $r$

$A_4$ : $D_4 = A_{4,0} \supset A_{4,1} = A_{3,1} = \{e, s, r, r^2 s\} \supset A_{4,2} = \{e, r^2\}$;

$A_5$ : $D_4 = A_{5,0} \supset A_{5,1} = \{e, r, r^2, r^3\} \supset A_{5,2} = A_{2,2} = \{e, r^2\}$.

The subgroups $\{e, rs, sr, r^2\}$, $\{e, s, r^2, r^2 s\}$, $\{e, r, r^2, r^3\}$, $\{e, r^2\}$ are normal divisors (on the figure they are shown by semiboldface lines).

Fig. 2. Structure of $L(D_4)$.

Let $H$ be a subgroup of $D_4$. The subspace of $N(B)$ consisting of elements which are invariant to $H$ is determined by means of the projection operator \cite{4}

$$P(H) = \frac{1}{|H|} \sum_{g \in H} T(g), \quad (10)$$

where $T(g)$ is the representation of $D_4$ by the matrix group generated by

$$T(s) = \begin{pmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix} \quad \text{and} \quad T(r) = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix}.$$

For every chain of subgroups the projection operators and hence the subgroup-invariant elements in $N(B)$ are determined. On the base of the formulas for
the basis of \( N(B) \) changing and relevant transformations of BEq [2] we are going to construct the branching systems for the solution of the equations which are invariant with respect to subgroups.

Let us consider the chain \( A_1 \). The projection operator \( P(A_{1,1}) \) transforms the \( N(B) \) basis into the subspace of \( A_{1,1} \)-invariant elements \( \text{span}(\varphi_1^\times = \frac{\varphi_1 + \varphi_2}{\sqrt{2}}, \varphi_2^\times = \frac{\varphi_1 - \varphi_2}{\sqrt{2}}) \). Adding to the base of this subspace the elements \( \varphi_3^\times = \frac{-i(\varphi_1 + i\varphi_2)}{\sqrt{2}}, \varphi_4^\times = \frac{-i(\varphi_1 + i\varphi_2)}{\sqrt{2}} \) the BEq in the new base

\[
\begin{align*}
  a_0\eta_1 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_1 (\eta_1^2 + \eta_3^2) p_1 (\eta_2^2 + \eta_4^2) p_2 &= 0, \\
  a_0\eta_2 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_2 (\eta_2^2 + \eta_3^2) p_1 (\eta_1^2 + \eta_4^2) p_2 &= 0, \\
  a_0\eta_3 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_3 (\eta_3^2 + \eta_4^2) p_1 (\eta_1^2 + \eta_2^2) p_2 &= 0, \\
  a_0\eta_4 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_4 (\eta_4^2 + \eta_3^2) p_1 (\eta_1^2 + \eta_3^2) p_2 &= 0 
\end{align*}
\]

is obtained. Then the BEq for \( A_{1,1} \)-invariant solutions follows from (4.0) by setting \( \eta_3 = 0, \eta_4 = 0 \) in accordance with the form of the projector \( P(A_{1,1}) \)

\[
\begin{align*}
  a_0\eta_1 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_1^2 p_1 + \eta_2^2 p_2 &= 0, \\
  a_0\eta_3 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1 + p_2} \eta_3^2 p_1 + \eta_2^2 p_2 &= 0 \tag{12}
\end{align*}
\]

The leading part of (4.0) is

\[
\begin{align*}
  \eta_1 \varepsilon + \frac{A}{2} \eta_1^3 + \frac{B}{2} \eta_1 \eta_2^2 &= 0, \\
  \eta_2 \varepsilon + \frac{A}{2} \eta_2^3 + \frac{B}{2} \eta_2 \eta_3^2 &= 0 \tag{13}
\end{align*}
\]

with the solutions

1) \( \eta_2 = 0, \eta_1 = \pm \sqrt{-\frac{2c}{A}}, \text{sign}\ \varepsilon = -\text{sign}\ A; u = \eta_1 \varphi_1^\times = \pm \frac{1}{\alpha} \sqrt{-\frac{\varepsilon}{\pi}} \cos \lambda_0 (x + y) + O(|\varepsilon|), \)

2) \( \eta_1 = 0, \eta_2 = \pm \sqrt{-\frac{2c}{A}}, \text{sign}\ \varepsilon = -\text{sign}\ A; u = \eta_2 \varphi_2^\times = \pm \frac{1}{\alpha} \sqrt{-\frac{\varepsilon}{\pi}} \cos \lambda_0 (y - x) + O(|\varepsilon|), \)
3) $\eta_1^2 = \eta_2^2 = -\frac{2\varepsilon}{A+B}$, $\text{sign } \varepsilon = -\text{sign } (A+B)$; $u = \pm \frac{2}{\pi} \sqrt{-\frac{\varepsilon}{A+B}} \cos \lambda_0 x \cos \lambda_0 y + O(|\varepsilon|)$.

For the chain $sA_1 s$ of conjugate subgroups the transformation $s$ transforms the branching equations for solutions invariant with respect to subgroups of the first chain into the BEqs for solutions which are invariant relatively to subgroups of the second chain.

The subspace of $N(B)$ elements invariant to the subgroup $A_{1,2}$ span$\{\varphi_1^x = \varphi_1 + \varphi_2, \varphi_3^x = \varphi_3, \varphi_4^x = \varphi_4\}$ is transformed by $s$ in the subspace span$\{\varphi_1, \varphi_2, \varphi_3^x = \varphi_1 + \varphi_2, \varphi_4^x = \varphi_4\}$ of $N(B)$ elements invariant with respect to the subgroup $sA_{1,2} s$. The BEq for solutions invariant to $A_{1,2}$

$$a_0 \eta_1 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1} \eta_1^{p_1+1} \eta_3^{p_2} \eta_4^{p_2} = 0$$

$$a_0 \eta_3 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_2} \eta_1^{p_1} \eta_3^{p_1+1} \eta_4^{p_1+1} = 0$$

is transformed by $s$ in the BEq for solutions invariant to $sA_{1,2} s$

$$a_0 \eta_1 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_3} \eta_1^{p_1+1} \eta_2^{p_1} \eta_3^{2p_2} = 0,$$

$$a_0 \eta_2 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_1} \eta_1^{p_1} \eta_2^{p_1+1} \eta_3^{2p_2} = 0,$$

$$a_0 \eta_3 + \sum_{|p| \geq 1} \frac{a_{p_1 p_2}}{2p_2} \eta_1^{p_2} \eta_2^{p_2} \eta_3^{2p_1+1} = 0.$$

(14)
Moreover the branching systems form the structure dual to the structure of subgroups of $D_4$: BEq of solutions which are invariant to more narrow subgroup contains the BEq of solutions which are invariant to a more wide subgroup. In fact, the BEq of $A_{1,1}$-invariant solutions ($sA_{1,1}s$-invariant solutions) setting in (4.0) $\eta_3 = \eta_4 = \frac{\eta_3}{\sqrt{2}}$ (correspondingly in (4.0) $\eta_1 = \eta_2 = \frac{\eta_1}{\sqrt{2}}$ or $\eta_1 = \eta_2 = \frac{\eta_2}{\sqrt{2}}$).

The relevant leading part of (4.0) has the form
\[
\eta_1 \left( \varepsilon + \frac{A}{2} \eta_1 + B \eta_3 \eta_4 \right) = 0, \\
\eta_3 \left( \varepsilon + A \eta_3 \eta_4 + \frac{B}{2} \eta_3^2 \right) = 0, \\
\eta_4 \left( \varepsilon + A \eta_3 \eta_4 + \frac{B}{2} \eta_4^2 \right) = 0
\]
(16).

Consequently, the $A_{11}$-invariant solutions of the equations (6), (7) are represented by the formula $\eta_1 \varphi_1^1 + \eta_3 \varphi_3^1 + \eta_4 \varphi_4^1$, where the vector $\left( \eta_1^*, \eta_3^*, \eta_4^* \right)$ passes the solution set of the previous system. The $sA_{11}s$-invariant solutions are,
\[
s(\eta_1^* \varphi_1^1 + \eta_3^* \varphi_3^1 + \eta_4^* \varphi_4^1) = \eta_1^* \varphi_1^1 + \eta_3^* \varphi_3^1 + \eta_4^* \varphi_4^1 = \sqrt{2} \eta_1^* \varphi_1^1 + \eta_3^* \varphi_3^1 + \eta_4^* \varphi_4^1
\]
respectively. Among the set of the solutions of the system (4.0) should be mentioned
1) $\eta_3 = \eta_4 = 0$, $\eta_1 = \pm \sqrt{-\frac{2\varepsilon}{A}}$, $\text{sign } \varepsilon = -\text{sign } A$

2) $\eta_3 = \eta_4 = \pm \sqrt{-\frac{\varepsilon (A + B - 1)}{A (A + B)}}$, $\eta_1 = \pm \sqrt{-\frac{\varepsilon}{A + B}}$, $\text{sign } \varepsilon = -\text{sign } (A + B)$,

3) $\eta_1 = \pm \sqrt{-\frac{\varepsilon}{A + B}}$, $\eta_3 \eta_4 = -\frac{\varepsilon}{A + B}$, $\text{sign } \varepsilon = -\text{sign } (A + B)$.

References


ON A SPARSE MATRIX MEMORIZING METHOD

Vlad Monescu, Silviu Dumitrescu, Bogdana Pop
"Transilvania" University of Braşov, Dept. of Computer Sci.

Abstract  A method to memorize a sparse matrix is discussed. In order to access the elements, the matrix’s binary converted blocks are taken into consideration. The method is compared with some classic ones and computational results are presented.

Keywords: sparse matrix, binary conversion, block matrix.

1. INTRODUCTION

The idea to take into account the large number of zeros of a matrix and their location was initiated in the second half of nineteenth century by electrical engineers. A $n \times m$ matrix is a sparse matrix if the number of nonzero entries is much smaller than $n \times m$. There are two problems: how to retain in minimum memory space a sparse matrix and how to access its elements. Sparse matrices arise in optimization problems, solutions to partial differential equations, structural and circuit analysis and computational fluid dynamics. Sparse matrices can be huge; dimensions on the order of 100,000 are not uncommon. Only by exploiting sparsity can we hope to be able to manipulate such a matrix on a computer.

2. THE ALGORITHM

We propose a method to memorizing a sparse matrix. To this end, let $A$ be a sparse matrix, $A \in M_{n,m}(\mathbb{R})$. For the following construction choose the numbers $p, q \in \mathbb{N}^*$. The rows of matrix are divided in groups of $p$ rows and the columns are divided in groups of $q$ columns. If $n \mod p \neq 0$ or $m \mod q \neq 0$ then supplementary rows, respectively columns are added containing null values. The matrix we obtain by adding rows or columns is equivalent with the initial matrix. Each of $(\lceil \frac{n-1}{p} \rceil + 1) \times (\lceil \frac{m-1}{q} \rceil + 1)$ blocks of elements is $p \times q$ binary converted. All positions containing nonzero values are considered [1]. The block which has in the upper-left corner the element $a_{i,j}$ is

$$
\begin{array}{cccc}
  a_{r,s} & a_{r,s+1} & \cdots & a_{r,s+q-1} \\
  a_{r+1,s} & a_{r+1,s+1} & \cdots & a_{r+1,s+q-1} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{r+p-1,s} & a_{r+p-1,s+1} & \cdots & a_{r+p-1,s+q-1}
\end{array}
$$
We transform this block in a sequence of bits \( b_{pq}b_{pq-1}...b_{1}b_{0} \), where

\[
b_{pq-i(q-1)-j} = \begin{cases} 
1, & a_{r+i,s+j} \neq 0 \\
0, & a_{r+i,s+j} = 0 
\end{cases}
\]

Thus we obtain a new matrix \( T \) which has nonnegative integer elements as a result of conversion in \( p \times q \) bits in a matricial disposal. If we denote by \( N \) the number of nonzero elements in \( A \), then \( N + (\lfloor \frac{n-1}{p} \rfloor + 1) \times (\lfloor \frac{m-1}{q} \rfloor + 1) \) memory locations are needed to memorize the sparse matrix \( A \) using this method.

Example.

For the matrix \( A = \begin{pmatrix} 11 & 0 & 0 & 14 & 5 & 0 & 0 & 8 \\
0 & 0 & 0 & 5 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 7 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
0 & 0 & 0 & 0 & 10 & 0 & 0 & 0 \\
17 & 0 & 0 & 0 & 0 & 0 & 58 & 0 \end{pmatrix} \), if \( p = 4 \) and \( q = 4 \)

we have the following configuration

\[
A' = \begin{pmatrix}
1 & 0 & 0 & 1 & \cdot & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & \cdot & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & \cdot & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & \cdot & 0 & 1 & 0 & 0 \\
. & . & . & . & . & . & . & . & . \\
0 & 0 & 0 & 0 & \cdot & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & \cdot & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \cdot & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \cdot & 0 & 0 & 0 & 0 
\end{pmatrix}.
\]

It has been added a row with 0. Each group of 16 bits represents a nonnegative integer number in binary conversion. Thus we obtain

\[
T = \begin{pmatrix} 37145 & 37124 \\
128 & 6176 \end{pmatrix}.
\]

The vector of nonzero elements (ordered by their appearances in their blocks) is

\[
w^T = \begin{pmatrix} 11 & 14 & 5 & 6 & 4 & 7 & 5 & 8 & 1 & 1 & 17 & 4 & 10 & 58 \end{pmatrix}.
\]

Once we have \( T \), and \( w \) it is necessary to access the elements of matrix \( A \). Let it be \( v, b, k \in \mathbb{N} \). We denote

- \( N_b(v) \) - number of nonzero bits of \( v \) in the \( b \)-bits binary representation,
- \( P_{b,k}(v) \) - position of the \( k \)-th nonzero bit of \( v \) in the \( b \)-bits binary representation,
On a sparse matrix memorizing method

- \( B_{b,k}(v) \) - value of the \( k \)-th bit of \( v \) in the \( b \)-bits binary representation.

If we have to access the element \( a_{i,j} \) and

\[
B_{b,pq-(i \mod p)q-j \mod q}(t_{(i-1)/p}+1,t_{(j-1)/q}+1) = 0
\]

then \( a_{i,j} = 0 \), else the position \( s \) of the element \( a_{i,j} \) in \( w \) is given by

\[
s = \sum_{k=1}^{\lfloor \frac{i-1}{p} \rfloor} \sum_{l=1}^{\lfloor \frac{m-1}{q} \rfloor} N_{pq}(t_{k,l}) + \sum_{l=1}^{\lfloor \frac{i-1}{p} \rfloor} N_{pq}(t_{(i-1)/p+1,l}) - N(i \mod p)q+j \mod q(t_{(i-1)/p+1,(j-1)/q+1})
\]

This method is very easy to be implemented in a programming language using bitwise operations. Also, the memory space required to retain the matrix using this algorithm is

\[
N + (\lfloor \frac{n-1}{p} \rfloor + 1) \times (\lfloor \frac{m-1}{q} \rfloor + 1).
\]

This storage method is obviously better than classical methods if the numbers \( p \) and \( q \) are big enough.

Denoting

- \( r \), the number which represents the memorising index of the given matrix \( A \), \( (r = \frac{N}{nm}, N \neq nm) \) and
- \( r' \), the memorising index of the matrix \( A \) through this method,

then the inequality

\[
pq > \frac{nm}{(r-1)N}
\]

is a condition to obtain a better memorising index.

3. Conclusions

A new method to memorizing a sparse matrix was developed in this paper. Blocks of matrices were binary converted and were taken into consideration when the elements were accessed. The solving method was compared with other algorithms and a condition of efficiency was formulated. Computational results were presented in order to authenticate the proposed algorithm.

References

REGENERATIVE PROCESS WITH A RARE EVENT

Bogdan-Gh. Munteanu
"Transilvania" Univ. of Brașov, Dept. of Mathematical Analysis and Probability
munteanu77@yahoo.com

Abstract The distribution of time to failure in a system with reparable components, under the assumption that the repair time of failed components has an arbitrary cumulative distribution function is investigated.

Keywords: regenerative process, law limit, asymptotic theorems

2000 MSC: 62E17, 60F15, 60G50, 60K05, 60K10.

1. INTRODUCTION

For a repair time with an arbitrary cumulative distribution function, the techniques of birth and death processes is needed. We will assume that the system with renewable components enters from time to time into the brand new state and that the time instants of these entrances generate a regenerative process. Let the entrances take place at epochs $T_0 = 0$, $T_i$, $i = 1, 2, ...$. Then the random intervals $\xi_i = T_i - T_{i-1}$, $i = 1, 2, ...$ are independent identically distributed random variables. Then the random intervals $T_k = \sum_{i=1}^{k} \xi_i$, $k \geq 1$, are called the regeneration points, and the sequence $\xi_i$, $i \geq 1$ is called a regenerative process (fig.1). We assume that $\xi_i$ are positive random variables.

![Fig. 1. The regenerative process $\xi_i$; • designates regeneration points.](image-url)
2. DISTRIBUTION OF A REGENERATIVE PROCESS DURATION

Consider a simple model based on a regenerative process leading to the exponential distribution. Assume that some event \( A \) can happen with probability \( p \) during a regeneration period. The probability \( p \) remains constant and is not affected by the previous history of the process. Consider a random variable \( Y = \xi_1 + \xi_2 + \ldots + \xi_N \), where \( N \) has a geometric distribution

\[
P(N = k) = (1 - p)^{k-1}p, \quad k \geq 1.
\]

Therefore, \( Y \) represents the duration of a regenerative process which is stopped at the end of that period on which the event \( A \) had appeared for the first time. We stress that random variable \( N \) is independent of random variables \( \xi_j \), \( j \geq 1 \). The next proposition finds cumulative distribution function of \( Y \).

**Proposition 2.1** [2]. Let \( E[\xi_i] = \mu \) and \( Y = \xi_1 + \xi_2 + \ldots + \xi_N \). Then \( Y \sim \text{Exp}(p/\mu) \).

**Proof.** We apply the Laplace transform technique, i.e.

\[
\mathcal{L}(Y) = E[e^{-zY}] = \sum_{k=1}^{\infty} P(N = k) \cdot E[e^{-z(\xi_1+\xi_2+\ldots+\xi_N)} | N = k] = \sum_{k=1}^{\infty} (1 - p)^{k-1}p \cdot [\varphi(z)]^k = \frac{p \varphi(z)}{1 - (1 - p)\varphi(z)},
\]

where \( \varphi(z) = E[e^{-z\xi_i}] \). As \( E[\xi_i] = \mu \), then \( \xi_i \sim \text{Exp}(1/\mu) \), so that

\[
E[e^{-z\xi_i}] = \int_0^\infty e^{-zt}d(1 - e^{-t/\mu}) = -\frac{e^{-(1/z+1/\mu)t}}{z+1/\mu} \bigg|_0^\infty = \frac{1}{1+2\mu}.\]

Substituting it into (2.0), we obtain that \( \mathcal{L}(Y) = \left(1 + \frac{2\mu}{p}\right)^{-1} \) which means that \( Y \sim \text{Exp}(p/\mu) \), or \( P(Y \leq t) = 1 - e^{-pt/\mu} \). Therefore, the sum of a geometrically distributed number of exponential sumands has an exact exponential distribution. We prove that if the summands \( \xi_i \) have an arbitrary distribution then under some appropriate condition, including \( p \to 0 \), the sum \( \xi_1 + \xi_2 + \ldots + \xi_n \) converges in distribution to an exponential random variable.

We need the following lemma:

**Lemma 2.1** Let \( Y \sim \mathcal{G}(p) \) and \( Z = pY \). Then \( Z \xrightarrow{d} \text{Exp}(1) \) as \( p \to 0 \).

**Proof.** Denote by \( \lfloor t/p \rfloor \) the integer part of \( t/p \).

\[
\lim_{p \to 0} P(Z > t) = \lim_{p \to 0} P(pY > t) = P(Y \geq \lfloor t/p \rfloor) = \lim_{p \to 0} (1 - p)^{\lfloor t/p \rfloor} = \lim_{p \to 0} \left(1 - p\right)^{-1/p} - p[t/p] = e^{-\lim_{p \to 0} p[t/p]} = e^{-t}
\]
Regenerative process with a rare event

Theorem 2.1 Let:

i. \( X_i, \ i = 1, 2, \ldots \) be a random sequence satisfying the strong law of large numbers: 
\[
P \left( \lim_{n \to \infty} \bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \mu \right) = 1, \quad 0 < \mu < \infty;
\]

ii. the random variable \( N \sim \mathcal{G}(p) \), \( N \) is independent on \( X_i \), \( Y = X_1 + X_2 + \ldots + X_N \).

Then \( \frac{pY}{\mu} \xrightarrow{d} \text{Exp}(1) \) as \( p \to 0 \).

Proof. We have \( \frac{pY}{\mu} = \frac{p}{\mu} N \cdot \frac{\sum_{i=1}^{N} X_i}{N} = pN \cdot \frac{\bar{X}_N}{\mu} \). By Lemma 2.1, we have \( pN \xrightarrow{d} \text{Exp}(1) \). Further, \( P \left( \frac{\bar{X}_N}{\mu} \to 1 \right) = 1 \) as \( N \to \infty \) (the strong law of large numbers). Thus, by Slutsky’s theorem in [6], \( \frac{pY}{\mu} = pN \cdot \frac{\bar{X}_N}{\mu} \xrightarrow{d} \text{Exp}(1) \).

Remark that Theorem 2.1 does not require the existence of a finite second moment for random variables \( X_i \).

3. EXAMPLE-TWO UNIT SYSTEM WITH STANDBY AND UNRELIABLE REPLACEMENT OF FAILED UNIT

A system has two identical units; one is operating and the second is in cold standby. The average lifetime of the operating unit is \( E[\tau] = \mu \) and \( E[\tau^2] = \mu_2 \). When the operating unit fails, its place is taken immediately by the standby unit and the failed unit goes to repair. The repair is, in fact, an instantaneous replacement of the failed unit by a new one taken from the storage. There is a small probability \( p \) that the unit is not available at the storage and then the repair will be delayed for a long time. In that case, the system fails at the instant of the failure of the operating unit. Assume that system operation began with one unit being sent for replacement. It is easy to see that the system lifetime is:

\[
\tau^* = \tau_1 + \tau_2 + \ldots + \tau_N
\]

where \( N \sim \mathcal{G}(p) \), that is \( P(N = k) = (1 - p)^{k-1}p \), \( k \geq 1 \). If \( p \) is small, then one can assume that, by Theorem 2.1, the exponential approximation is valid for the system lifetime \( P \left( \frac{\tau^*}{\mu} \leq t \right) \approx 1 - e^{-t} \), or \( P \left( \tau^* > t \right) \approx e^{-\frac{t}{\mu}} \).

Example. Let \( p = 0.01 \), \( \mu = 10 \) hr, \( \mu_2 = 175 \) hr, \( t = 500 \) hr. Then \( P(\tau^* > 500) = P \left( \frac{\tau^*}{\mu} > \frac{500}{\mu} \right) \approx e^{-0.5} \approx 0.606 \). In order to see how accurate the exponential approximation is, let us make use of the following result
established by Brown in [1]

\[
\sup_{t \geq 0} \left| P\left( \sum_{i=1}^{N} \tau_i > t \right) - e^{-tp/\mu} \right| \leq \frac{\mu_2 p}{\mu^2}.
\]

(2)

For our data we obtain that \( \mu_2 p/\mu^2 = 0.0175 \), which means that the approximation (2) is quite accurate.

On the other side, in [3], Theorem 2.1, show that

\[
\sup_{x \geq 0} \left| P\left( \frac{\tau^*}{E(\tau^*)} > x \right) - e^{-x} \right| < \frac{1 - \sqrt{1 - 4a_2}}{1 + \sqrt{1 - 4a_2}}
\]

(3)

The right-hand side to be approximated with \( a_2 \), if \( a_2 \to 0 \). The Remark 2.5. in [3] clarifies the probabilistic meaning of the quantity \( a_2 \) expressed by the relation \( a_2 = 1 - \frac{E(\tau^*)}{E(\tau^*)^{\frac{1}{2}}} \), or \( a_2 = 1 - \frac{\mu^2}{2\mu^2} \). In the our example, \( a_2 = 0.125 \).

Thus, the approximation (3) is very accurate.

**Remark 3.1** The relation (3) establishes the deviation from standard exponential distribution of the distribution of normalized system lifetime, as soon as the approximation (2) provides that the cumulative distribution function of system lifetime \( \tau^* \) is rather close to \( 1 - e^{\frac{\mu x}{\mu}} \), that is the exponential distribution with the parameter \( \frac{\mu}{\mu} \).

**References**


STABILITY OF LIMIT CYCLES IN A CALCIUM OSCILLATIONS DYNAMICS MODEL

Ileana Rodica Nicola
Department of Mathematics I, University Politehnica of Bucharest
nicola_rodica@yahoo.com

Abstract The stability of the limit cycles of a nonlinear delay system of ordinary differential equations is studied. The system is a model proposed by Borghans in 1997 [2] describing calcium oscillations in nonexcitable cells. The unique positive equilibrium point of the associated dynamical system can loose its stability on the account of a Hopf bifurcation. We then investigate the stability of the bifurcating limit cycles by using the center manifold theorem.

Keywords: dynamical system, time delay, bifurcation, limit cycle.

2000 MSC: 34K18, 37G10, 37M20, 37N25, 92C35.

1. INTRODUCTION

This paper is devoted to the analysis of the nonlinear delay system of ordinary differential equations (SODE)

\[
\begin{align*}
\frac{dZ}{dt} &= -kZ(t) + V_0 + \beta V_1 + k_f(Y(t)) - V_{M_2} \frac{(Z(t))^2}{k_2^2 + (Z(t))^2} + V_{M_3} \frac{(Z(t))^m}{k_2^2(t)^m + (Z(t))^m k_f^2 + (Y(t))^2 k_A^4 + (A(t-\tau_0))^4} \\
\frac{dY}{dt} &= -k_f Y(t) + V_{M_2} \frac{(Z(t))^2}{k_2^2 + (Z(t))^2} - V_{M_3} \frac{(Z(t))^m}{k_2^2(t)^m + (Z(t))^m k_f^2 + (Y(t))^2 k_A^4 + (A(t-\tau_0))^4} \\
\frac{dA}{dt} &= \beta V_{M_4} - V_{M_5} \frac{(A(t-\tau_0))^p}{k_2^p + (A(t-\tau_0))^p k_d^n + (Z(t))^n} - \varepsilon A(t-\tau_0).
\end{align*}
\]

(1)

With the autonomous SODE (1.1) we associate the initial condition

\[Z(0) = Z_0, Y(0) = Y_0, A(\theta) = \varphi(\theta), \theta \in [-\tau, 0], \tau \geq 0,\]

where \(\varphi : [-\tau, 0] \rightarrow \mathbb{R}\) is a differentiable function which describes the behavior of the flow in the \(O\) direction.
These equations arise from a model of calcium oscillations dynamics. The non-delayed model was previously proposed and investigated by Houart et al. in [6]; it is based on the mechanism of Ca\(^{2+}\)-induced Ca\(^{2+}\)-release, that takes into account the Ca\(^{2+}\)-stimulated degradation of inositol 1,4,5-trisphosphate (InsP\(_3\)) by a 3-kinase. In the vast majority of cells, an external stimulus initiates the synthesis of InsP\(_3\), starting an intracellular chain reaction, which culminates with the release of Ca\(^{2+}\) from an internal store of the cell, in the cytosol. Two mechanisms are responsible for calcium oscillation: the autocatalitic nature of Ca\(^{2+}\) release in the cytosol and the increased InsP\(_3\) degradation, due to the Ca\(^{2+}\)-stimulated of the InsP\(_3\) 3-kinase.

The variables involved in the process are \(Z\), \(Y\) and \(A\), representing the concentration of free Ca\(^{2+}\) in the cytosol, in a certain internal pool of the cell, respectively the InsP\(_3\) concentration.

The biological interpretation of the parameters is as follows: \(V_0\) refers to a constant input of Ca\(^{2+}\) from the extracellular medium and \(V_1\) is the maximum rate of stimulus-induced influx of Ca\(^{2+}\) from the extracellular medium. Parameter \(\beta\) reflects the degree of stimulation of the cell by an agonist. The rates \(V_2\) and \(V_3\) refer, respectively, to pumping of cytosolic Ca\(^{2+}\) into the internal stores and to the release of Ca\(^{2+}\) from these stores into the cytosol in a process activated by citosolic calcium (CICR); \(V_{M2}\) and \(V_{M3}\) denote the maximum values of these rates. Parameters \(k_2\), \(k_Y\), \(k_Z\) and \(k_A\) are treshold constants for pumping, release, and activation of release by Ca\(^{2+}\) and by InsP\(_3\); \(k_f\) is a rate constant measuring the passive, linear leak of \(Y\) into \(Z\); \(k\) relates to the assumed linear transport of citosolic calcium into the extracellular medium; \(V_{M4}\) is the maximum rate of stimulus-induced synthesis of InsP\(_3\). \(V_5\) is the rate of phosphorylation of InsP\(_3\) by the 3-kinase; \(V_{M5}\) is the maximum value of this rate, while \(k_5\) is a half-saturation constant. \(m\), \(n\) and \(p\) are Hill coefficients related to the cooperative processes and \(\varepsilon\) is the phosphorilation rate of InsP\(_3\) by the 5-phosphatase.

The study of the Hopf bifurcation was previously performed in [13] for three sets of values of the parameters, namely the set for which the oscillations exhibit chaotic behaviour, bursting behaviour, respectively birhitmicity. We proved that in the case when the perturbation \(A(t - \tau)\) is obtained via the Dirac distribution, the calcium oscillations model may undergo a Hopf bifurcation for some critical value of the time delay \(\tau = \tau_0\). We further develop only the “bursting” case, that is, we consider the following values of the parameters

\[
\beta = 0.46, \ n = 2, \ m = 4, \ p = 1, \ K_2 = 0.1\mu M, \ k_5 = 1\mu M, \\
k_A = 0.1\mu M, \ k_d = 0.6\mu M, \ k_Y = 0.2\mu M, \\
k_z = 0.3\mu M, \ k = 0.1667s^{-1}, \ k_f = 0.0167s^{-1}, \ v = 0.0167s^{-1}, \\
V_0 = 0.0333\mu M\ s^{-1}, \ V_1 = 0.0333\mu M\ s^{-1}, \ V_{M2} = 0.1\mu M\ s^{-1}, \\
V_{M3} = 0.3333\mu M\ s^{-1}, \ V_{M4} = 0.0417\mu M\ s^{-1}, \ V_{M5} = 0.5\mu M\ s^{-1}.
\]
We mention that, in this case, using the Maple 8 software package, we find the equilibrium point \((z^*, y^*, A^*) = (0.2916496701; 0.2344675015; 0.1989819160)\), the eigenvalues \(\lambda = \pm i\omega = \pm 0.08289718923\) and the bifurcation parameter \(\tau_0 = 21.25439515\). The existence of a Hopf bifurcation leads to the existence of a limit cycle when the bifurcation occurs; herein our aim is to show the stability of this limit cycle following the approach in [7].

2. STABILITY OF LIMIT CYCLES

Denote by \(\Lambda = \{\pm i\Omega_0\}\) the simple eigenvalues of the characteristic equation associated with the delayed SODE (for details, see [13]). Denote \(x(t) = (z(t), y(t), a(t))\). By the translation \(Z = z + Z^*, Y = y + Y^*, A = a + A^*\) the autonomous delay differential system (1) changes to

\[
\dot{x}(t) = X_i(z(t), y(t), a(t - \tau)),
\]

\(z(0) = Z_0 - Z^*, y_0 = Y_0 - Y^*, \varphi(\theta) = \varphi(0) - A^*, \tau \geq 0\), with the equilibrium point \((0, 0, 0)\). In order to characterize the orbits of the linearized delay-differential system, we can use the standard form

\[
\dot{x}_t(\theta) = Lx_t = \begin{cases}
\frac{d}{dt}x_t(\theta), & \theta \in [-\tau, 0) \\
\int_{-\tau}^{0} [d\eta(\theta)]x_t(\theta), & \theta = 0
\end{cases}
\]

where

\[
x_t \in \mathcal{B} = C^\prime([-\tau, 0], \mathbb{R}^3), \quad x_t(\theta) = x(t + \theta), \quad \theta \in [-\tau, 0).
\]

The mapping \(\eta : [-\tau, 0] \to \mathbb{R}^3 \times \mathbb{R}^3\) is a matrix whose elements are functions with bounded variation and the Lebesgue-Stieltjes integral can be written as

\[
\int_{-\tau}^{0} [d\eta(\theta)]x_t(\theta) = N_1x(0) + N_2x(-\tau).
\]

The spectrum of the linear operator \(L\) coincides with the set of eigenvalues of the characteristic equation associated to the equilibrium point of the SODE.

Consider the linear operator \(L : \mathcal{B} \to \mathcal{B}, L\varphi = N_1\varphi(0) + N_2\varphi(-\tau), \varphi \in \mathcal{B}\), where

\[
N_1 = \begin{pmatrix}
n_1^1 & n_1^4 & 0 \\
n_1^2 & -n_1^4 & 0 \\
n_1^3 & 0 & 0
\end{pmatrix}, \quad N_2 = \begin{pmatrix}
0 & 0 & n_2^1 \\
0 & 0 & -n_2^1 \\
0 & 0 & n_2^2
\end{pmatrix}
\]

(3)

and

\[
n_1^1 = 0.38860527, \quad n_2^1 = -0.55530527, \quad n_2^3 = -0.08796881,
\]
The following statements are true:
the bilinear form
the generalized eigenvectors of the infinitesimal generators
Proposition 2.1
with respect to the bilinear form ("scalar product")
determines the adjoint operator
with respect to the bilinear form ("scalar product")

Proposition 2.1 The following statements are true:
(i) the generalized eigenvectors of the infinitesimal generators \( A \) and \( A^* \), corresponding to the proper eigen \( \Lambda \), are
\[
\phi(\theta) = (f_1, f_2, f_3)^T e^{i\Omega_0 \theta}, \quad \bar{\phi}(\theta) = (\bar{f}_1, \bar{f}_2, \bar{f}_3)^T e^{-i\Omega_0 \theta}, \quad \theta \in [0, \tau_0]
\]
\[
f_1 = 0.151551 - i 0.172126, \quad f_2 = 0.497684 - i 0.132631,
\]
\[
f_3 = -0.763963 - i 1.018165
\]
respectively,
\[
\phi^*(s) = (f_1^*, f_2^*, f_3^*) e^{i\Omega_0 s}, \quad \bar{\phi}^*(s) = (\bar{f}_1^*, \bar{f}_2^*, \bar{f}_3^*) e^{-i\Omega_0 s}, \quad s \in [0, \tau_0],
\]
\[
f_1^* = -0.536353 + i 0.373245, \quad f_2^* = -0.413810 + i 0.479090,
\]
\[
f_3^* = 0.594553 - i 0.870017;
\]
(ii) the bilinear form \((\cdot, \cdot) : \mathbb{M}_A(A^*) \times \mathbb{B} \rightarrow \mathbb{C}\), with \( \Lambda = \{ \pm i \Omega_0 \} \) is defined by
\[
(\phi^*(s), \phi(\theta)) = \bar{f}_1^* \phi_1(0) + \bar{f}_2^* \phi_2(0) + \bar{f}_3^* \phi_3(0) - \alpha_0 F(\varphi_3)(-\tau_0),
\]
\[
(\bar{\phi}^*(s), \phi(\theta)) = f_1^* \phi_1(0) + f_2^* \phi_2(0) + f_3^* \phi_3(0) - \bar{\alpha}_0 F(\varphi_3)(-\tau_0),
\]
\[
s \in [0, \tau_0], \varphi \in \mathbb{B}, \theta \in [-\tau_0, 0], \alpha_0 = 0.072126 - i 0.049286,
\]
where \( \mathbb{M}_A(A^*) \) is the generalized vector space of \( A^* \) and
\[
F(\varphi_3)(-\tau_0) = \int_{0}^{-\tau_0} e^{-i\Omega_0 \xi} \phi(\xi) d\xi;
\]
(iii) the generalized eigenvectors of $\phi^*(s), \bar{\phi}^*(s)$ and $\phi(\theta), \bar{\phi}(\theta)$ satisfy the relations

$$\begin{cases}
(\phi^*(s), \phi(\theta)) = r_1, \\
(\bar{\phi}^*(s), \bar{\phi}(\theta)) = r_2, \\
(\phi^*(s), \bar{\phi}(\theta)) = \bar{r}_2, \\
(\phi^*(s), \bar{\phi}(\theta)) = \bar{r}_1, 
\end{cases}$$

where $r_1 = -2.058522 - i 2.065137, \quad r_2 = -1.441408 - i 1.972671$;

(iv) the vectors

$$\begin{align*}
\psi(s) &= f_{11} \phi^*(s) + \bar{f}_{12} \bar{\phi}^*(s), \\
\bar{\psi}(s) &= f_{12} \phi^*(s) + \bar{f}_{11} \bar{\phi}^*(s), 
\end{align*}$$

are generalized eigenvectors of the operator $A^*$. They satisfy the relations

$$\begin{align*}
(\psi(s), \phi(\theta)) &= 1, \\
(\bar{\psi}(s), \bar{\phi}(\theta)) &= 0, \\
(\psi(s), \bar{\phi}(\theta)) &= 0, \\
(\bar{\psi}(s), \bar{\phi}(\theta)) &= 1,
\end{align*}$$

where $\theta \in [-\tau, 0], \ s \in [0, \tau_0]$. (12)

The bilinear form is associated to the linearized delay differential system. The numbers $f_{ij}$ are the entries of the matrix $F$, the inverse of the matrix $E = (e_{ij})_{i,j=1,2}$, where

$$e_{11} = (\phi^*(s), \phi(\theta)), \ e_{12} = (\bar{\phi}^*(s), \phi(\theta)), \ e_{21} = \bar{e}_{12}, \ e_{22} = \bar{e}_{11}. $$

**Proof.** (i) The generalized eigenvectors of $A$, corresponding to $\Lambda$ are

$$\phi(\theta) = \phi(0)e^{i\Omega_0 \theta}, \quad \bar{\phi}(\theta) = \bar{\phi}(0)e^{-i\Omega_0 \theta},$$

where $\phi(0)$ is a nonzero solution of the following linear system

$$(i\Omega_0 I - N_1 - e^{-i\Omega_0 \tau_0} N_2) \phi(0) = 0.$$

Replacing $N_1$ and $N_2$ by their expressions (2) and using the software package Maple 8, we find

$$\phi(0) = (f_1, f_2, f_3)^T, $$

where $f_1 = -0.151551 - i 0.172126, \ f_2 = 0.497684 - i 0.132631 \text{ and } f_3 = -0.763963 - i 1.018165$.

From (14) and (12) it follows (5). For the operator $A^*$, the generalized eigenvectors corresponding to $\Lambda$ are

$$\phi^*(s) = \phi^*(0)e^{i\Omega_0 s}, \quad \bar{\phi}^*(s) = \bar{\phi}^*(0)e^{-i\Omega_0 s}, \quad s \in [0, \tau_0],$$

$$(16)$$
where $\phi^*(0)$ is a nonzero solution of the following linear system
\[
\phi^*(0)(i\Omega_0 I - N_1 - e^{i\Omega_0 \tau_0} N_2) = 0. \tag{17}
\]
After computations we get
\[
\phi^*(0) = (f_1^*, f_2^*, f_3^*), \tag{18}
\]
where $f_1^* = -0.536353 + i 0.373245$, $f_2^* = -0.413810 + i 0.479090$ and $f_3^* = 0.594553 - i 0.870017$.

(ii) The definition of the bilinear form implies
\[
(\phi^*(s), \varphi(\theta)) = \tilde{f}_1^* \varphi_1(0) + \tilde{f}_2^* \varphi_2(0) + \tilde{f}_3^* \varphi_3(0) - e^{i\Omega_0 \tau_0} [(\tilde{f}_1^* - \tilde{f}_2^*) n_1^2 + \tilde{f}_3^* n_2^2] F(\varphi_3)(-\tau_0) = \tilde{f}_1^* \varphi_1(0) + \tilde{f}_2^* \varphi_2(0) + \tilde{f}_3^* \varphi_3(0) - \alpha_0 F(\varphi_3)(-\tau_0),
\]
\[
= 0 = 0.072121 - i 0.049286,
\]
\[
(\phi^*(s), \varphi(\theta)) = f_1^* \varphi_1(0) + f_2^* \varphi_2(0) + f_3^* \varphi_3(0) - e^{-i\Omega_0 \tau_0} [(f_1^* - f_2^*) n_1^2 + f_3^* n_2^2] F(\varphi_3)(-\tau_0) = f_1^* \varphi_1(0) + f_2^* \varphi_2(0) + f_3^* \varphi_3(0) - \tilde{\alpha}_0 F(\varphi_3)(-\tau_0)
\]
In conclusion, we get the relation (7).

(iii) From (7), for $\varphi(\theta) = \phi(\theta)$, we have
\[
(\phi^*(s), \phi(\theta)) = \tilde{f}_1^* \phi_1(0) + \tilde{f}_2^* \phi_2(0) + \tilde{f}_3^* \phi_3(0) - \alpha_0 \int_0^{-\tau_0} e^{-i\Omega_0 \xi} f_3 e^{i\Omega_0 \xi} d\xi = \tilde{f}_1^* \phi_1 + \tilde{f}_2^* \phi_2 + \tilde{f}_3^* \phi_3 + \alpha_0 \tau_0 \phi_3 = r_1, \quad r_1 = -2.058522 - i 2.065137.
\]
\[
(\phi^*(s), \phi(\theta)) = f_1^* \phi_1 + f_2^* \phi_2 + f_3^* \phi_3 + \alpha_0 \tau_0 \phi_3 = r_2, \quad r_2 = -1.441408 - i 1.972671.
\]
\[
(\tilde{\phi}^*(s), \phi(\theta)) = \tilde{f}_1^* \phi_1 + \tilde{f}_2^* \phi_2 + \tilde{f}_3^* \phi_3 + \tilde{\alpha}_0 \tau_0 \phi_3 = (\tilde{\phi}^*(s), \phi(\theta))
\]
\[
(\tilde{\phi}^*(s), \phi(\theta)) = f_1^* \phi_1 + f_2^* \phi_2 + f_3^* \phi_3 + \tilde{\alpha}_0 \tau_0 \phi_3 = (\tilde{\phi}^*(s), \phi(\theta))
\]
These lead to the relation (9).

(iv) Because the function $\cdot, \cdot$ is bilinear and $EF = I$, we have the relations
\[
(\psi^*(s), \phi(\theta)) = f_{11}(\phi^*(s), \phi(\theta)) + f_{12}(\tilde{\phi}^*(s), \tilde{\phi}(\theta)) = f_{11} e_{11} + f_{21} e_{12} = 1
\]
\[
(\psi^*(s), \tilde{\phi}(\theta)) = f_{11}(\phi^*(s), \tilde{\phi}(\theta)) + f_{12}(\tilde{\phi}^*(s), \tilde{\phi}(\theta)) = f_{11} e_{21} + f_{21} e_{22} = 0
\]
\[
(\psi^*(s), \phi(\theta)) = (\psi^*(s), \phi(\theta)) = 0
\]
\[
(\psi^*(s), \tilde{\phi}(\theta)) = (\psi^*(s), \tilde{\phi}(\theta)) = 1.
\]
\[
\square
\]
Let $N = \mathcal{M}_A(\mathcal{A})$ be the vector space generated by the generalized eigenvectors $\Phi, \bar{\Phi}$.

**Definition 2.1** A submanifold $W^r_{loc}(0) = W^r(0, V)$ in the Banach space $\mathcal{B}$, tangent at $0 \in \mathcal{B}$ to the vector space $N$ and invariant with respect to the semigroup of operators $T(t)$ of the delayed system (1), where $V$ is an open set with $0 \in V \subset \mathcal{B}$, is called stable-unstable local centre manifold of the system (1).

In the sequel, let us denote the nonlinear part of the system (1) by $F(x(t), x(t - \tau)) = X(x(t), x(t - \tau)) - N_1x(t) - N_2x(t - \tau)$, where $x = X(x), x = (z, y, a)^T$ is the vector field given by the relation (1). Using the Taylor formula of order three at the point $(0,0,0)$, the autonomous delay differential system (1) can be written as

\begin{align*}
\dot{x}(t) &= N_1x(t) + N_2x(t - \tau_0) + P(x(t), x(t - \tau_0)) + O_1(|x(t)|^4), \\
&\quad x(t) = (z(t), y(t), a(t)),
\end{align*}

(19)

where $P$ is the vector whose components are Taylor polynomials of degrees two or three of the vector $F$ and has the form $P(x(t), x(t - \tau)) = (P_1(z(t), y(t), a(t - \tau)), P_2(z(t), y(t), a(t - \tau)), P_3(z(t), y(t), a(t - \tau)))^T$, where

\begin{align*}
P_1(z(t), y(t), a(t - \tau)) &= -0.537512z(t)^2 + 2.226838z(t)y(t) + 0.747183a(t - \tau)z(t) - 0.862223y(t)^2 + 0.3708534a(t - \tau)y(t) - 1.171530a(t - \tau)^2 - 9.088956z(t)^3 - 2.956420z(t)^2y(t) - 0.991983a(t - \tau)z(t)^2 - 6.246203z(t)y(t)^2 + 2.684263a(t - \tau)z(t)y(t) - 8.486912a(t - \tau)^2z(t) + 1.020635y(t)^3 - 1.039337a(t - \tau)y(t)^2 - 4.208734a(t - \tau)^2y(t) + 10.050654a(t - \tau)^3 \\
P_2(z(t), y(t), a(t - \tau)) &= -N_1(z(t), y(t), a(t - \tau)), \\
P_3(z(t), a(t - \tau)) &= -0.035519z(t) - 0.368724a(t - \tau)z(t) + 0.055411a(t - \tau)^2 = 0.244208z(t)^3 - 0.148882a(t - \tau)z(t) + 0.307531a(t - \tau)^2z(t) - 0.046240a(t - \tau)^3
\end{align*}

Introduce the function

\begin{align*}
w(\theta, z_c, \bar{z}_c) &= w_20(\theta)\frac{|z_c|^2}{2} + w_11(\theta)z_c\bar{z}_c + w_02(\theta)\frac{|\bar{z}_c|^2}{2} + O_1(|z_c|^3),
\end{align*}

$w_{02} = \bar{w}_{20}(\theta), w_{11}(\theta) \in \mathbb{R}, \theta \in [-\tau, 0]$.

Replacing, in $P(x(t), x(t - \tau))$, $x(t)$ by $z_c\phi(0) + \bar{z}_c\bar{\phi}(0) + w(0, z_c, \bar{z}_c)$ and $x(t - \tau)$
The following statements are true:

by \( z_c \phi(-\tau) + \bar{z}_c \bar{\phi}(-\tau) + w(-\tau, z_c, \bar{z}_c) \), we find

\[
F(z_c, \bar{z}_c) = F_{20} \frac{z_c^2}{2} + F_{11} z_c \bar{z}_c + F_{02} \frac{z_c^2}{2} + F_{21} \frac{z_c^2 \bar{z}_c^2}{2},
\]

where

\[
F_{20} = \begin{pmatrix} F_{20}^1 \\ -F_{20}^2 \\ F_{20}^3 \end{pmatrix}, \quad F_{11} = \begin{pmatrix} F_{11}^1 \\ -F_{11}^2 \\ F_{11}^3 \end{pmatrix}, \quad F_{02} = \begin{pmatrix} F_{02}^1 \\ -F_{02}^2 \\ F_{02}^3 \end{pmatrix}, \quad F_{21} = \begin{pmatrix} F_{21}^1 \\ -F_{21}^2 \\ F_{21}^3 \end{pmatrix}.
\]

By extracting from expansion of \( F \) the coefficients of the terms \( \frac{z_c^2}{2}, \frac{z_c \bar{z}_c}{2}, \frac{z_c^2}{2} \) respectively \( \frac{z_c^2 \bar{z}_c^2}{2} \), we find

\[
F_{20}^1 = -1.137957 + i 3.464395, \quad F_{20}^3 = -0.010181 + i 0.041143,
\]

\[
F_{11}^1 = -3.981737 - i 1.600830, \quad F_{11}^3 = 0.035900 - i 0.014433,
\]

\[
F_{02}^1 = 1.576785 - i 3.287968, \quad F_{02}^3 = 0.021130 - i 0.036741,
\]

\[
F_{21}^1 = -108.851161 + i 11.064876 - (1.444304 + i 1.706172)w_{11}^1 -
\]

\[
\quad - (1.606180 - i 2.701638)w_{11}^2 + (5.042470 + i 2.815720)w_{11}^3 - \\
\quad - (0.988250 - i 0.522131)w_{20}^1 - (0.241235 + i 1.552891)w_{20}^2 + \\
\quad + (1.814090 + i 2.246725)w_{20}^3 \]

\[
F_{21}^3 = -0.068091 + i 0.020966 + (0.872300 - i 0.426980)w_{11}^1 - \\
\quad - (0.108014 - i 0.037829)w_{11}^3 + (0.325032 - i 0.360775)w_{20}^1 + \\
\quad + (0.043053 + i 0.037695)w_{20}^3.
\]

**Proposition 2.2** The following statements are true:

(i) the central manifold \( W_{\text{loc}}^c(0) \) of the system (1) has the elements \( \bar{\varphi} \in \mathcal{B} \), where

\[
\bar{\varphi}(\theta) = z_c \phi(\theta) + \bar{z}_c \bar{\phi}(\theta) + w_{20}(\theta) \frac{z_c^2}{2} + w_{11}(\theta) z_c \bar{z}_c + w_{02}(\theta) \frac{z_c^2}{2} + \ldots, \theta \in [-\tau, 0]
\]

where \( z_c = x_1 + iy_1, \ (x_1, y_1) \in V_1 \subset \mathbb{R}^2, V_1 \) is a neighborhood of zero

and

\[
w_{20}(\theta) = -\frac{g_{20}}{i\Omega} \phi(0)e^{i\theta} - \frac{g_{02}}{3i\Omega} \phi(0)e^{-i\theta} + E_1 e^{2i\theta}
\]

\[
w_{11}(\theta) = \frac{g_{11}}{i\Omega} \phi(0)e^{i\theta} - \frac{g_{11}}{i\Omega} \bar{\phi}(0)e^{-i\theta} + E_2, \quad \theta \in [-\tau_0, 0]
\]

\[
w_{02}(\theta) = w_{20}(\theta),
\]
if the initial condition of (1) is

\[ \begin{align*}
g_{20} &= 0.302874 + i 0.912785 \\
g_{11} &= -1.079822 + i 0.272609 \\
g_{02} &= -0.171742 - i 0.947895 \\
g_{21} &= -40.208575 + i 35.475448 + (0.032165 - i 1.7317770)w_{11}^1 + \\
&+ 0.064066 + i 0.764149)w_{11}^2 + (0.514234 - i 1.094785)w_{11}^3 + \\
&+ (0.722945 + i 0.294077)w_{20}^1 - (0.268046 + i 0.274150)w_{20}^2 + \\
&+ (0.579122 + i 0.194816)w_{20}^3.
\end{align*} \]

where

\[ \begin{align*}
w_{20} &= (w_{20}^1, w_{20}^2, w_{20}^3)^T, \\
w_{11} &= (w_{11}^1, w_{11}^2, w_{11}^3)^T, \\
E_1 &= \begin{pmatrix}
-22.833534 - i 15.188009 \\
38.104506 - i 7.770249 \\
2.614298 - i 11.178749
\end{pmatrix}, \\
E_2 &= \begin{pmatrix}
12.423192 + i 4.994660 \\
-0.431636 - i 0.173536
\end{pmatrix}.
\end{align*} \] 

(ii) if the initial condition of (1) is

\[ \begin{align*}
\tilde{\varphi}(\theta) &= (\psi, \varphi_1)\phi(\theta) + \overline{(\psi, \varphi_1)}\phi(\theta) + w_{20}(\theta) \frac{(\psi, \varphi_1)^2}{2} + \\
&+ w_{11}(\theta)(\psi, \varphi_1)(\psi, \varphi_1) + w_{02} \frac{(\psi, \varphi_1)^2}{2},
\end{align*} \]

then the solution of autonomous delay differential system (1) in a neighborhood of the equilibrium point \((0,0,0)\) is (25):

\[ \begin{align*}
z(\tilde{\varphi})(t) &= 2x_1 Re(f_1) - 2y_1 Im(f_1) + r_{10}^k(x_1^2 - y_1^2) - i_{20}^k x_1 y_1 + r_{11}^k(x_1^2 + y_1^2) \\
y(\tilde{\varphi})(t) &= 2x_1 Re(f_2) - 2y_1 Im(f_2) + r_{20}^k(x_1^2 - y_1^2) - i_{20}^k x_1 y_1 + r_{11}^k(x_1^2 + y_1^2) \\
a(\tilde{\varphi})(t) &= 2x_1 Re(f_3) - 2y_1 Im(f_3) + r_{30}^k(x_1^2 - y_1^2) - i_{20}^k x_1 y_1 + r_{11}^k(x_1^2 + y_1^2)
\end{align*} \]

where

\[ \begin{align*}
r_{20}^k &= Re(w_{20}^k(0)), \\
i_{20}^k &= Im(w_{20}^k(0)), \\
r_{11}^k &= w_{11}^k(0), \forall k = 1,3
\end{align*} \]

and \((x_1(t), y_1(t))\) is the solution of the differential equation

\[ \begin{align*}
z_{c}(t) &= i\Omega_0 z_{c}(t) + \frac{g_{20}}{2} z_{c}(t)^2 + g_{11} z_{c}(t) z_{c}(t) + \frac{g_{02}}{2} z(t)^2 + \frac{g_{21}}{2} z(t)^2 z(t) \\
z_{c}(0) &= x_1(0) + iy_1(0),
\end{align*} \] 

(27)
where
\[ x_1(0) = \text{Re} (\psi, \varphi_1), \quad y_1(0) = \text{Im} (\psi, \varphi_1) \]
\[ \varphi_1(\theta) = (Z_0 - Z^*, Y - Y^*, \varphi(\theta) - A^*)^T. \]

**Proof.** (i) We use the following relations
\[ E_1 = -(N_1 + e^{-2i\Omega_0} N_2 - 2i\Omega_0 I)^{-1} F_{20}, \quad E_2 = -(N_1 + N_2)^{-1} F_{11}. \]
\[ g_{20} = \tilde{\psi}(0) F_{20}, \quad g_{11} = \tilde{\psi}(0) F_{11}, \quad g_{02} = \tilde{\psi}(0) F_{02}, \quad g_{21} = \tilde{\psi}(0) F_{21}, \]
where
\[ \tilde{\psi}(0) = f_1 \phi^*(0) + f_2 \tilde{\phi}^*(0) \]
is one of the eigenvectors of $A^*$ (see Proposition 2.1 (iv))
(ii) The solution of autonomous delay differential system (1) in a neighborhood of the equilibrium point $(0,0,0)$ is
\[
\begin{cases}
    z(\tilde{\varphi})(t) = z_c(t) f_1 + \bar{z}_c(t) f_1 + \frac{1}{2} z_c(t)^2 w_{20}^1(0) + z_c(t) \bar{z}_c(t) w_{11}^1(0) + \frac{1}{2} \bar{z}_c(t)^2 w_{02}^1(0) \\
y(\tilde{\varphi})(t) = z_c(t) f_2 + \bar{z}_c(t) f_2 + \frac{1}{2} z_c(t) w_{20}^2(0) + z_c(t) \bar{z}_c(t) w_{11}^2(0) + \frac{1}{2} \bar{z}_c(t) w_{02}^2(0) \\
a(\tilde{\varphi})(t) = z_c(t) f_3 + \bar{z}_c(t) f_3 + \frac{1}{2} z_c(t) w_{20}^3(0) + z_c(t) \bar{z}_c(t) w_{11}^3(0) + \frac{1}{2} \bar{z}_c(t) w_{02}^3(0),
\end{cases}
\]
where $f_i, i = 1, 3$ are given by (5). Replacing $z_c$ by $x_1 + iy_1$, we find (24).

**Remark 2.1.** Let $R_{20} = \text{Re} (g_{20}), \quad I_{20} = \text{Im} (g_{20}), \quad R_{11} = \text{Re} (g_{11}), \quad I_{11} = \text{Im} (g_{21}), \quad R_{21} = \text{Re} (g_{21}), \quad I_{21} = \text{Im} (g_{21})$. The autonomous delay differential system (1) can be rewritten in the form

\[
\begin{cases}
    \dot{x}_1(t) = -\Omega_0 y_1(t) + \frac{1}{2} (R_{20} + 2R_{11} + R_{02}) x_1(t) - \frac{1}{2} (R_{20} - 2R_{11} + R_{02}) y_1(t) + (I_0 - I_{20}) x_1(t) y_1(t) + R_{21} x_1(t) x_1(t)^2 + y_1(t)^2 - I_{21} y_1(t) (x_1(t)^2 + y_1(t)^2), \\
    \dot{y}_1(t) = -\Omega_0 x_1(t) + \frac{1}{2} (I_{20} + 2I_{11} + I_{02}) y_1(t) - \frac{1}{2} (I_{20} - 2I_{11} + I_{02}) x_1(t)^2 + (R_{20} - R_{02}) x_1(t) y_1(t) + R_{21} y_1(t) (x_1(t)^2 + y_1(t)^2) - I_{21} x_1(t) (x_1(t)^2 + y_1(t)^2), \\
    x_1(0) = \text{Re} (\psi, \varphi_1), \quad y_1(0) = \text{Im} (\psi, \varphi_1),
\end{cases}
\]
where $\varphi_1$ is given by (27).
Remark 2.2. The term $g_{21}$ depends explicitly on $w_{11}^1(0)$, $w_{11}^2(0)$, $w_{11}^3(-\tau)$, $w_{20}^1(0)$, $w_{20}^2(0)$ and $w_{20}^3(-\tau)$: using this and the relations (20) we find $g_{21} = -55.129357 + i 4.070772$.

Remark 2.3. The limit cycle can be characterized by the following numbers
\[
\mu_2 = -\frac{\text{Re}(C_1)}{\text{Re}(M)}, \quad T_2 = -\frac{\text{Im}(C_1) + \mu_2 \text{Im}(M)}{\Omega_0}, \quad \beta_2 = 2\text{Re}(C_1),
\]
where
\[
C_1 = \frac{i}{2\Omega_0} \left( g_{20}g_{11} - 2|g_{11}|^2 - \frac{1}{3}|g_{02}|^2 \right) + \frac{g_{21}}{2},
\]
\[
M = \left( \frac{d \quad }{d\tau} \right)_{\tau = \tau_0, \omega = \Omega_0}.
\]

One knows from [5] that the following properties hold: if $\mu_2 > 0$ ($< 0$) then the Hopf bifurcation is supercritical (subcritical) and the bifurcating periodic solution exists for $\tau > \tau_0$ ($\tau < \tau_0$); the solutions are orbitally stable (unstable) if $\beta_2 > 0$ ($< 0$), and the bifurcating periodic solution increases (decreases) if $T_2 > 0$ ($< 0$).

Consider the values of the parameters as in the case of ”bursting”. Using Maple 8 programming techniques, we find $\mu_2 = 10587.6533$, $\beta_2 = -44.2353$ and $T_2 = -180.9126$. Hence, the unique Hopf bifurcation of the equation (1.1) seems to be supercritical and the solutions orbitally stable, with decreasing period.

Remark 2.4. The cycle of the dynamics of calcium oscillations with time-delay around the equilibrium point $(Z^*, Y^*, A^*)$ is given by
\[
Z(\tilde{\varphi})(t) = z(\tilde{\varphi})(t) + Z^*, \quad Y(\tilde{\varphi})(t) = y(\tilde{\varphi})(t) + Y^*, \quad A(\tilde{\varphi})(t) = a(\tilde{\varphi})(t) + A^*,
\]
where $z(\tilde{\varphi})(t), y(\tilde{\varphi})(t), a(\tilde{\varphi})(t)$ are given by Proposition 2.2.

3. ACKNOWLEDGMENT

The present work was partially supported by the Grant CNCSIS MEN A1478/2005.

References


Mainly, this paper is a survey of an example of a few classical concepts and results of fractal geometry. In Section 1, it is emphasized the fact that the box counting dimension, more easy to calculate than any other dimension, has a key role in understanding the relationship between Hausdorff and packing measure and dimension. The section ends reminding two important results: the equality of packing dimension and upper box counting dimension on compact "self-similar" sets [2] and the density theorem [4]. In Section 2, an example of a plane set $E$ with $0 < 1 = \mathcal{H}^1(E) < 5/4 \leq c^1(E) \leq 2 < p^1(E) = 4 \leq \infty$, and therefore $\dim_H(E) = \dim_P(E) = 1$, is shown to be theoretically important. In Section 2.2, a plane set with zero linear Hausdorff measure ($\mathcal{H}^1$) and infinite packing linear measure ($p^1$) is presented following [7]. More precisely, we show that $\dim_H(E) < 1 < \dim_P(E)$. Then, by generalizing this example we construct a set with specific different Hausdorff and packing dimension: $0 = \dim_H(E) < 1 < \dim_P(E) = \log 7 / \log 3$.

1. **A FEW CLASSICAL RESULTS IN FRACTAL GEOMETRY IN $\mathbb{R}^N$**

Let $A \subset \mathbb{R}^n$ be a set. A $s$-dimensional Hausdorff measure $\mathcal{H}^s$ of $E$ ($s \geq 0$) is

$$\mathcal{H}^s(E) := \sup_{\delta > 0} \mathcal{H}^s_\delta(E),$$

where $\mathcal{H}^s_\delta(E) := \inf \left\{ \sum_{i=1}^{\infty} |E_i|^s, (E_i)_i \delta - \text{covering of } E \right\}$.

(1)

The diameter $d(E)$ of $E$ is denoted by $|E|$; a $\delta$-covering of a set $E$ is a family $(E_i)_{i \geq 1} \subset \mathcal{P}(\Omega)$ with $E \subset \bigcup_{i \geq 1} E_i$, $|E_i| \leq \delta$).

For every $E \subset \mathbb{R}^n$ there is an unique value associated with $E$:

$$\dim_H(E) := \inf \left\{ s \geq 0 : |\mathcal{H}^s(E) = 0 \right\} = \sup \left\{ s \geq 0 : |\mathcal{H}^s(E) = \infty \right\},$$

called the *Hausdorff dimension of the set* $E$. It is monotone, countably stable, invariant to Lipschitz transforms and behaves "well" on smooth manifolds, i.e we get the usual topological dimension.

The packing premeasure is defined as follows: for every $E \in \mathcal{P}_b(\mathbb{R}^n)$ (family of bounded sets in $\mathbb{R}^n$), $P^s(E) := \inf_{\delta > 0} P^s_\delta(E)$, where, for all $\delta > 0$, $P^s_\delta(E) = \inf$
dimension is monotone and countably stable. The centered covering measure \(C\), namely \(C(E) := \inf \left\{ \sum_{i=1}^{\infty} P^s(E_i) \mid (E_i)_{i \geq 1} \, \text{balls centered in } E, \, E \subset \bigcup_{i \geq 1} E_i \right\}\), is a measure on \(\mathbb{R}^n\). According to [4], \(C^s(E) = \sup \left\{ s \, p^s(E) = \infty \right\} = \inf \left\{ s \, p^s(E) = 0 \right\}\). The unique number determined as above is called packing dimension of the set \(E\). The packing dimension is monotone and countably stable.

There is a set function that brings together Hausdorff and packing measure, namely centered covering measure.

Define the mapping \(C^s : \mathcal{P}(\mathbb{R}^n) \to [0, \infty], \, C^s(E) := \sup_{\rho > 0} C^s_{\rho}(E)\), where

\[
C^s_{\delta}(E) = \inf \left\{ \sum_{i=1}^{\infty} |B_i|^s \mid (B_i)_{i \geq 1} \, \text{balls centered in } E, \, E \subset \bigcup_{i \geq 1} B_i \right\}.
\]

According to [4], \(C^s\) is not monotone, so, in order to become monotone one define \(c^s(E) = \sup \left\{ C^s(E) \mid F \subset E \right\}\), which is a measure on \(\mathbb{R}^n\), indeed. It is called the centered covering measure.

The relationships between the three measures are [4]

\[
\left( \frac{1}{2^s} \right) c^s(E) \leq \mathcal{H}^s(E) \leq C^s(E) \leq c^s(E) \leq p^s(E) \leq P^s(E), \, E \subset \mathbb{R}^n,
\]

whence \(0 \leq \dim_H(E) \leq \dim_P(E) \leq n\).

The main result used in this paper is

**Theorem 1.1 (Raymond-Tricot "density theorem")** [4] For every positive, finite, Borelian measure \(\mu\) and \(x \in \mathbb{R}^n\) we have

\[
p^s(E) < \infty \implies p^s(E) \inf_{x \in E} d^s_{\mu}(x) \leq \mu(E) \leq p^s(E) \sup_{x \in E} d^s_{\mu}(x),
\]

\[
c^s(E) < \infty \implies c^s(E) \inf_{x \in E} d^s_{\mu}(x) \leq \mu(E) \leq c^s(E) \sup_{x \in E} d^s_{\mu}(x),
\]

where

\[
d^s_{\mu}(x) := \lim_{r \to 0} \frac{\mu(B_r(x))}{(2r)^s}, \quad d^s_{\mu}(x) := \limsup_{r \to 0} \frac{\mu(B_r(x))}{(2r)^s}.
\]
Corollary 1.1 If $\mathcal{H}^s(E) < \infty$ and $\mu(F) = \mathcal{H}^s(E \cap F)$, then $d^s_\mu(x)$ and $\mathcal{H}^s(E \cap F)$ become the "Besicovitch" densities, denoted by $D^s(x, E)$, $D^s(x, F)$, and we have

\[ p^s(E) < \infty \implies p^s(E) \inf_{x \in E} D^s(x, E) \leq \mathcal{H}^s(E) \leq p^s(E) \sup_{x \in E} D^s(x, E), \]

\[ c^s(E) < \infty \implies c^s(E) \inf_{x \in E} D^s(x, E) \leq \mathcal{H}^s(E) \leq c^s(E) \sup_{x \in E} D^s(x, E). \]

The oldest notion of "fractal" dimension is the so called box-counting dimension. For $E \subset \mathbb{R}^n$ bounded and $\delta > 0$, let $N_\delta(E)$ the smallest number of sets of diameter at most $\delta$ which can cover $E$. The lower and upper box-counting dimensions of $E$ are defined as follows:

\[ \dim_B(E) := \liminf_{\delta \to 0} \frac{\log N_\delta(E)}{-\log \delta}, \]

\[ \overline{\dim}_B(E) := \limsup_{\delta \to 0} \frac{\log N_\delta(E)}{-\log \delta}. \]

If these are equal we refer to the common value as the box-counting dimension of $E$ and write $\dim_B(E) := \lim_{\delta \to 0} \frac{\log N_\delta(E)}{-\log \delta}$.

We get the same previous numbers if we denote by $N_\delta(E)$ the smallest number of closed balls of radius $\delta$ that cover $E$, or the smallest number of cubes of side $\delta$ that cover $E$, or the number of $\delta$-mesh cubes that intersect $E$, or the largest number of disjoint balls of radius $\delta$ with centers at the points of $E$.

The relations between Hausdorff, packing and box-counting dimensions are

\[ 0 \leq \dim_H(E) \leq \dim_P(E) \leq \dim_B(E) \leq n, \forall E \subset \mathbb{R}^n, \]

\[ 0 \leq \dim_H(E) \leq \dim_B(E) \leq \overline{\dim}_B(E) \leq n, \forall E \subset \mathbb{R}^n. \]

The second important result used in this paper is

**Theorem 1.2** (Equality between upper box-counting dimension and packing dimension of self similar sets) Let $E \subset \mathbb{R}^n$ be compact with the property that for every open set $V$ that intersects $E$, $\dim_B(V \cap E) = \dim_B(E)$. Then $\dim_P(E) = \overline{\dim}_B(E)$.

2. ESTIMATION OF HAUSDORFF AND PACKING LINEAR MEASURE AND DIMENSIONS FOR SOME PLANE SETS OF POINTS

2.1. AN EXAMPLE OF PLANE SET WITH

\[ 0 < \mathcal{H}^1(E) < P^1(E) < \infty \]

Let $E_0$ be a closed equilateral triangle of side 1. Following [7], we construct the set $E_1$ by replacing $E_0$ by 3 equal triangles of side $1/3$, as shown in fig.1.
Supposing that $E_n$ is constructed (it consists of $3^n$ equal triangles of side $1/3^n$), we obtain $E_{n+1}$ by replacing each triangle of $E_n$ by 3 equal triangles as in fig. 1. Finally we define $E := \cap_{n \geq 1} E_n$. Let $\mathcal{C}_n$ be the cover of $E$ by the triangles of $E_n$. Because $\mathcal{C}_n$ is a $1/3^n$-cover of $E$, we have $\mathcal{H}^1(E) \leq \sum_{T \in \mathcal{C}_n} |T| = 3^n \cdot (1/3^n) = 1$. But the projection of $E$ of one side of $E_0$ is a segment of length 1, so linear Hausdorff measure is equal to 1 and we know the behaviour of the Hausdorff measure with respect to contractions: $\mathcal{H}^1(\text{pr}(E)) = 1 \leq \mathcal{H}^1(E)$. So $\mathcal{H}^1(E) = 1$. On the other hand, in [7] it is proved that $\mathcal{D}^1(x, E) = \frac{1}{4} \mathcal{H}^1$-a.e. on $E$. From Corollary 1.1 we get $p^1(E) = 4$. Using $C^1(E) \geq 5/4$ (the complete algebraic computations can be found in [4]) and again by Corollary 1.1 we get $C^1(E) \leq c^1(E) \leq 25 \mathcal{H}^1(E)$, so

$$0 < 1 = \mathcal{H}^1(E) \leq \frac{5}{4} \leq c^1(E) < p^1(E) = 4 < \infty.$$ 

This example has two major consequences. First, there exists a plane set with finite distinct values for linear Hausdorff measure, linear packing measure and linear centered covering measure. Second, the set function $C^1$ is not monotone: if we consider $F$ to be the set of all centers of the triangles of each $E_n$, for all $n$, then $E \cup F$ can be covered by $3^n$ centered balls of diameter $(2/\sqrt{3})3^{-n}$, so $C^1(E \cup F) \leq 2/\sqrt{3} < 5/4 \leq C^1(E)$.

### 2.2. EXAMPLES OF SETS WITH $\dim_H(E) < 1 < \dim_P(E)$

**I.** Consider a sequence of positive integers $a_n > 0$, $n \geq 1$ and define $c_n = \sum_{k=1}^{n} a_k$, $b_n = \sum_{k=1}^{n} a_{2k}$. Let the initial set $E_0$ be a closed disc of diameter 1. The operation of replacing a closed disc of diameter $d$ with another concentric disc of diameter $d/3$ is called the ”operation” $O1$. The operation of replacing a closed disc of diameter $d$ with 7 discs of diameter $d/3$ ”packed” in the initial disc (as in fig.2) is called the ”operation” $O2$. At the first step, we apply to the disc $E_0$ operation $O1$ $a_1$ times and get $E_1$, the concentric disc of diameter $(1/3)^{a_1} = (1/3)^{c_1}$. Then, to $E_1$ we apply the operation $O2$ $a_2$ times to get $E_2$, formed with $7^{a_2} = 7^{b_2}$ discs of diameter $(1/3)^{a_1+a_2} = (1/3)^{c_2}$ ”packed”
into $E_1$. In general, if $E_{2n}$ is formed with $7^{b_n}$ discs of diameter $(1/3)^{c_2n}$, we get $E_{2n+1}$ applying $O_1$ $a_{2n+1}$ times to each disc from $E_{2n}$ and obtain $7^{b_n}$ discs of diameter $(1/3)^{c_2n+1}$, and then $E_{2n+2}$ applying $O_2$ $a_{2n+2}$ times to each disc of $E_{2n+1}$ obtaining $7^{b_n+1}$ discs of diameter $(1/3)^{c_2n+2}$. Finally, we define $E := \bigcap_{n \geq 0} E_n$.

For each $n \geq 0$, $E$ is covered by $E_{2n+1}$ and "packed" by $E_{2n}$, so $\mathcal{H}^1(E) \leq \lim inf_{n \to \infty} 7^{b_n} \cdot 3^{-c_2n+1}$, $P^1(E) \geq \lim sup_{n \to \infty} 7^{b_n} \cdot 3^{-c_2n}$. Let $a_n = N^{n-1}$, $n \geq 1$, so

$$b_n = \frac{N^{2n+2} - 1}{N^2 - 1}, \quad c_n = \frac{N^{n+1} - 1}{N - 1}, \quad 7^{b_n} \cdot 3^{-c_2n+1} = \left( \frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} \right)^{N^{2n+2}}.$$ 

$$3^{-c_2n} = \frac{3^{\frac{1}{N-1}}}{7^{\frac{1}{N^2-1}}} \cdot \left( \frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} \right)^{N^{2n+2}}.$$ 

But $\frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} < 1$ and $\frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N(N-1)}}} > 1$, for all $N \geq 2$, so

$$\mathcal{H}^1(E) \leq \lim inf_{n \to \infty} 7^{b_n} \cdot 3^{-c_2n+1} = 0,$$

$$P^1(E) \geq \lim sup_{n \to \infty} 7^{b_n} \cdot 3^{-c_2n} = \infty.$$ 

In this case let us prove $\dim_H(E) < 1 < \dim_P(E)$. This is equivalent to $\exists t < 1$, such that $\mathcal{H}^t(E) = 0$, $\exists s > 1$, such that $P^s(E) = \infty$, or, in other words, $\exists t < 1$, such that $\left( \frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} \right)^{N^{2n+2}} \to 0$ and $\exists s > 1$, such that

$$\frac{3^{\frac{1}{N^2-1}}}{7^{\frac{1}{N^2-1}}} \cdot \left( \frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} \right)^{N^{2n+2}} \to \infty.$$ 

But this is true, because $\exists t < 1$, such that $\frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N-1}}} < 1$, $\exists s > 1$, such that $\frac{7^{\frac{1}{N^2-1}}}{3^{\frac{1}{N(N-1)}}} > 1$, for example $t \in \left( \frac{1}{3}, \frac{10}{9} \right)$ and $s \in \left( 1, \frac{11}{10} \right)$. 

**Fig. 2.** The "operation" $O_2$
II. Put $a_n = n!$ in the previous example to get $0 = \dim_H(E) < 1 < \dim_P(E) = \ln 7 / \ln 3$.

In order to prove that $\dim_H(E) = 0$, we need to show that $\forall t < 1$, $H^t(E) = 0$, that is $t < 1$. With some effort, this fact can be proved directly by elementary computations. We give an alternative proof:

To prove that $\dim_H(E) = 0$, we need to show that $\forall t < 1$, $H^t(E) = 0$, that is

$$\forall t < 1, 7^{2!+4!+\ldots+(2n)!} 3^t (1!+2!+\ldots+(2n+1)!)\rightarrow 0.$$ With some effort, this fact can be proved directly by elementary computations. We give an alternative proof:

remark that $E_{2n+1}$ is a cover for $E$ by $7^{2n}$ balls of diameter $\delta_n := (1/3)^{2n+1}$, so $N_{\delta_n}(E) \leq 7^{2n}$ ($N_\delta(E)$ is the smallest number of sets of diameter at most $\delta$ that can cover $E$), so

$$\dim_B(E) \leq \liminf_{n\to\infty} \frac{\log N_{\delta_n}(E)}{-\log \delta_n} \leq \liminf_{n\to\infty} \frac{\log 7^{2n}}{\log 3^{2n+1}} = \frac{\log 7}{\log 3} \lim_{n\to\infty} \frac{b_n}{c_{2n+1}} = 0,$$

using definition of "lim inf" and Stolz-Cesaro criterion. Then $0 \leq \dim_H(E) \leq \dim_B(E) \leq 0$, so $\dim_H(E) = 0$.

On the other hand, since $E$ is self-similar, by Theorem 1.2, $\dim_P(E) = \dim_B(E)$. But, again, $E_{2n}$ is a packing for $E$ by $7^{bn}$ balls of diameter $\delta_n := (1/3)^{2n}$, so $N_{\delta_n}(E) \geq 7^{bn}$ (here $N_\delta(E)$ is considered to be the largest number of disjoint balls of diameter $\delta$ with centers in $E$) and

$$\dim_B(E) \geq \limsup_{n\to\infty} \frac{\log N_{\delta_n}(E)}{-\log \delta_n} \geq \limsup_{n\to\infty} \frac{\log 7^{bn}}{\log 3^{c_{2n}}} = \frac{\log 7}{\log 3} \lim_{n\to\infty} \frac{b_n}{c_{2n}} = \frac{\log 7}{\log 3},$$

using definition of "lim sup" and Stolz-Cesaro criterion. Then $\dim_B(E) = \dim_B(E) \geq \log 7 / \log 3$. But $\dim_B(E) = \log 7 / \log 3$ (supposing that $\dim_B(E) > \log 7 / \log 3$, it is easy to see that we contradict the definition of upper box-counting dimension). So, finally

$$0 = \dim_H(E) < \dim_P(E) = \ln 7 / \ln 3.$$

References

ERROR HANDLING IN SOFTWARE SYSTEMS: MODELLING AND TESTING WITH FINITE STATE MACHINES
Radu Oprișa

Abstract Using the W method a test set of input sequences for the implementation of a software system starting from the Finite State Machine corresponding to specifications (SFSM) is obtained. To avoid halting, FSM corresponding to implementation (IFSM) must be completely specified. For the size reasons, usually specifications do not contain error-handling description for each unexpected input symbol. Due to this reason, for each state the designer must add in SFSM a corresponding error state and transitions to it for all unexpected input symbols. With this mechanism we are sure that IFSM is accepting all unexpected input sequences. But this will not prevent the problem of system recovery from error. In order to be able to perform a recovery from error we must allow the system to rich the previous state from each error state by adding for each error state a transition to the previous state. With this modeling approach we have the advantage of detecting all possible errors and the certitude of full error recovery in software systems at runtime.

Keywords: Finite State Machine, W method

2000MSC: 68N30

1. INTRODUCTION

Because the market necessities are growing, the software systems became more and more complex. Due to this, the effort to verify their correctness is also severely increased. The cheapest way to check if the implementation of a software system with respect to its specifications is an automatic testing is to be performed. For this purpose several methods had been developed, but all of them have specific limitations. An interesting approach is to model both system specifications and implementation with Finite State Machine and to determine a set of input sequences which can be used to prove or to invalidate their equivalence.

2. FUNDAMENTALS

Before generating the test set we need to introduce some basic notions about Finite State Machines. The following notation will be used. If $A$ is a finite alphabet, denote by $A^*$ the set of all finite sequences with members in $A$. Let $\epsilon$ denote the empty sequence. For $a, b \in A$ the concatenation
of the sequence \(a\) with \(b\) is denoted by \(a \bullet b\). For \(U, V \subseteq A^*\) we denote \(U \bullet V = \{a \bullet b \mid a \in U, b \in V\}\).

**Definition 2.1** A Deterministic Finite State Machine (DFSM) is an association of five elements \(M = (Q, Li, Lo, h, M_0)\) where:

- \(Q\) is a finite set of states;
- \(Li\) is a finite set of input symbols;
- \(Lo\) is a finite set of output symbols;
- \(h : Q \times Li \rightarrow Q \times Lo\) is the behaviour (partial) function;
- \(M_0\) is the initial state.

**Definition 2.2** A DFSM \(M\) is said to be completely specified if \(h\) is a total function.

**Definition 2.3** Let \(M = (Q, Li, Lo, h, M_0)\) be a DFSM, \(A \subseteq Li^*\) a set of input sequences and \(q, r\) two states in \(Q\). Then \(q\) and \(r\) are called \(A\)-equivalent if both produce the same responses for each input sequence \(a \in A\). Otherwise, they are said to be \(A\)-distinguishable. If \(q\) and \(r\) give identical responses for any input sequence \(a \in Li^*\) then \(q\) and \(r\) are said to be equivalent.

**Definition 2.4** Let \(M = (Q_M, Li, Lo, h_M, M_0)\) and \(N = (Q_N, Li, Lo, h_N, N_0)\) be two DFSMs with the same sets of input and output symbols and \(A \subseteq Li^*\). Then \(M\) and \(N\) are said to be \(A\)-equivalent if their initial states \(M_0\) and \(N_0\) are \(A\)-equivalent. Otherwise, \(M\) and \(N\) are said to be \(A\)-distinguishable. If \(A = Li^*\) then \(M\) and \(N\) are said to be equivalent.

**Definition 2.5** A DFSM \(M\) is said to be minimal if any other equivalent DFSM has at least the same number of states as \(M\).

**Definition 2.6** A finite set \(S \subseteq Li^*\) is called a state cover of a minimal DFSM \(M = (Q, Li, Lo, h, M_0)\) if \(S\) contains the empty sequence and for every state \(q \in Q\), other than \(M_0\), \(S\) contains an input sequence \(a\) that may lead the machine from the initial state \(M_0\) to \(q\).

In other words, a state cover is a set of input sequences that enables us to access any state in the machine from the initial state.

**Definition 2.7** Let \(M = (Q, Li, Lo, h, M_0)\) be a minimal finite state machine. \(T \subseteq Li^*\) is called a transition cover of \(M\) if \(\forall q \in Q, \exists a \in T\) so that \(a\) leads the machine from the initial state \(M_0\) to \(q\) and \(\forall x \in Li, a \bullet \{x\} \in T\).

In a DFSM \(M = (Q, Li, Lo, h, M_0)\), for any state \(q \in Q\) there are sequences in \(T\) that take \(M\) from \(M_0\) in \(q\) and then attempt to exercise transitions with
all input symbols from $L_i$ whatever they exists or not. Remark that if $S$ is a state cover of $M$, then $T = S \cup [S \cdot L_i]$ is a transition cover of $M$.

**Definition 2.8** A finite set $W \subseteq L_i^*$ is called a characterisation set of a DFSM $M = (Q, L_i, L_o, h, M_0)$ if any two distinct states $q, r \in Q$ are $W-$ distinguishable.

**Definition 2.9** Let $S = (Q_S, L_i, L_o, h_S, S_0)$ and $I = (Q_I, L_i, L_o, h_I, I_0)$ be two DFSMs having the same sets of input and output symbols. A finite set $Y \subseteq L_i^*$ is called a test set for $S$ and $I$ if: $S$ and $I$ are $Y-$ equivalent $\Rightarrow S$ and $I$ are equivalent.

The following well-known theorem is determining a test set which can be used to find out if two DFSM are equivalent or not.

**Theorem 2.1** Let $M = (Q_M, L_i, L_o, h_M, M_0)$ and $N = (Q_N, L_i, L_o, h_N, N_0)$ be two minimal DFSM, $T$ and $W$ a transition cover and a characterisation set for one of these machines respectively. Then they are isomorphic (behave identically) if $M_0$ and $N_0$ are $T \cdot Z-$ equivalent.

In other words, if both DFSMs produce the same outputs when the sequences from $T \cdot Z$ are applied, then they are equivalent machines.

### 3. THE W METHOD

Let $S = (Q_S, L_i, L_o, h_S, S_0)$ and $I = (Q_I, L_i, L_o, h_I, I_0)$ be two DFSMs having the same sets of input and output symbols, where $I$ has at least the same number of states as $S$. Let $d$ be the difference between the number of states of $I$ and $S$. If $S$ and $I$ are deterministic and completely specified and $S$ is in addition minimal, then the W method determinates a test set which can be used to check if $S$ and $I$ are equivalent.

The test set is determined as follows $Y = S_c \cdot L_i[d+1] \cdot W$, where $S_c \subseteq L_i^*$ is a state cover of $S$, $W \subseteq L_i^*$ is a characterisation set of $S$, $L_i[k] = \{\epsilon\} \cup L_i \cup \ldots \cup L_i^k$ for $k \geq 0$.

### 4. ERROR STATES

Consider the DFSM corresponding to specifications of a software system (SFSM) from fig. 1. As we can see, this SFSM is not completely specified. This is a simple example, but especially for large systems, in practice frequently we have combinations of states and input symbols with no description of the behaviour in the specifications. If the implementation is performed with respect to this SFSM, then for all such combinations of states and input symbols, where the behaviour function is not defined, the program will halt. This
is an undesirable behaviour for any software system especially when a large volume of user input data is required.

**Example 4.1** Consider the specifications of a software system which should allow user to fill input data in a form with several fields. At the end, the user should press "Submit" button in order to process data. At runtime the user fills the form, but by mistake press the "ALT+F4" key combination instead of "Submit" button. Because nothing is written in the specification about the behaviour when "ALT+F4" key combination occurs, the implementation has an uncontrolled behaviour.

The most easy way for the designer to solve the problem of the uncontrolled behaviour in SFSM is to add a new state, an error state $E$, and also transitions to it for all possible states and input symbols which do not have corresponding transitions in the initial system. Fig. 2 presents the extended SFSM from fig. 1 with an error state and transitions to it. Now we have a controlled behaviour when an input symbol occurs for which we have no description in the specifications. But this is not enough for the user because the system will reach the $E$ state and no possible action is defined from this state. Actually, even if the implementation will not halt, it is still impossible to recover input data from the system.

In order to solve this problem we need a more complex approach. Instead of defining only one error state, we will add an error state $E_k$ for each state $S_k$ of the initial SFSM and transitions $T_{ke}$ for all unexpected input symbols from the SFSM state to the corresponding error state. In order to be able to perform also a system recovery from the error states we need to add supplementary
transitions $T_{kr}$ from error states to the previous state. The set of input symbols will be extended with a new one $ok$ which will be used for all $T_{kr}$ transitions with interpretation that user noticed the error state and allow system recovery. It is possible to avoid the use of a new input symbol and to reuse an existing one. Two output symbols will be added: $Err$ which is used for all transitions which ends at $E_k$ states, and $Cont$ which is used for $T_{kr}$ transitions. $Err$ is the system response when an error state is reached. To have only one response in case of any error in the system is not user friendly. Due to this fact, distinct output symbols can be added for each transition which ends at an error state. However, for test purposes one output symbol is enough.

The system from fig. 1 is extended as in fig. 3. To keep the figure simple only the error state for $S3$ is represented. This solves the problem of system recovery from error but the obtained SFSM still have a weak point. An error state is the initial state only for one transition which accepts only the input symbol $ok$ (or the reused one). If another input symbol occurs when the system is in an error state, we are again in situation to obtain an uncontrolled behaviour. This can be solved by adding transitions for each error state and each unexpected input symbol to the error state itself. Such an extension is presented in fig. 4. To keep the figure simple, only the error state for $S3$ is represented.

5. CONCLUSIONS

Using this modelling technique and W testing method, the designer will be sure that the implementation will perform all requested operations described in the specifications. Moreover, all other actions initiated by the user
which are not described in the specifications will lead to an error state. Another big advantage is that the system can be recovered from any error state and this will eliminate the possibility of losing data.

References


USING STOCHASTIC PETRI NETS IN PERFORMANCE ANALYZE OF DISTRIBUTED SYSTEMS

Norocel Petrache

Abstract

This paper shows how Stochastic Petri Nets (SPNs) (Petri Nets with timed transitions and random, negative exponentially distributed firing delays) can be used to investigate performances of distributed systems. Because SPN systems are isomorphic to continuous time Markov chains (CTMCs), their analysis can be performed by construction of the state transition rate matrix. A model of a simple shared memory system and the software tool GreatSPN are described.

1. PETRI NETS AND TEMPORAL CONCEPTS

The concept of time was intentionally avoided in the original work by C. A. Petri, because of the effect that timing may have on the behaviour of PNs: the association of timing constraints with the activities represented in PN models or systems may prevent certain transitions from firing, thus invalidating the important assumption that all possible behaviours of a real system are represented by the structure of the PN.

Temporal concepts were introduced in Petri nets models in the mid 1970s. This initiated a hot debate about the suitability of such an extension, and about the most appropriate ways to make it. Both Petri nets with timed places and timed transitions were proposed; in the latter case there was either a time delay (before the atomic occurrence of a transition) or a time interval (between the consumption of input tokens and the generation of output tokens). Whether deterministic or stochastic timing should be used was also an issue of the discussion.

The firing of a transition in a PN model corresponds to the event that changes the state of the real system. This change of state can be due to the completion of some activity. Transitions can be used to model activities, so that transition enabling periods correspond to activity executions and transition firings correspond to activity completions. Hence, time can be naturally associated with transitions. Transitions with associated temporal specifications are called timed transitions; they are graphically represented by boxes or thick bars.

A timed transition T can be associated with a local clock or timer. When T is enabled the associated timer is set to an initial value. The timer is then
Norocel Petrache
decremented at constant speed and the transition fires when the timer reaches
the value zero. The timer can thus be used to model the duration of an activity
whose completion induces the state change that is represented by the change
of marking produced by the firing of T.

It is important to note that the activity is assumed to be in progress while
the transition is enabled. This means that in the evolution of complex nets, an
interruption of the activity may take place if the transition loses its enabling
condition before it can actually fire. The activity may be resumed later on,
during the evolution of the system in the case of a new enabling of the associ-
ated transition. This may happen several times until the timer goes down to
zero and the transition finally fires.

When introducing time into PN models and systems, it would be extremely
useful not to modify the basic behaviour of the underlying untimed model.
By so doing, it is possible to study the timed PNs exploiting the properties
of the basic model as well as the available theoretical results. The addition
of temporal specifications therefore must not modify the unique and original
way of expressing synchronization and parallelism that is peculiar to PNs. In
untimed PN systems, the choice of which transition to fire in a free-choice
conflict is completely nondeterministic. In the case of timed PN systems, the
conflict resolution depends on the delays associated with transitions and is
obtained through the so-called race policy: when several timed transitions are
enabled in a given marking M, the transition with the shortest associated delay
fires first.

An important issue that arises at every transition firing when timed tran-
sitions are used in a model is how to manage the timers of all the transitions
that do not fire. From the modelling point of view, the different policies that
can be adopted link the past history of the systems to its future evolution
considering various ways of retaining memory of the time already spent on ac-
tivities. The question concerns the memory policy of transitions, and defines
how to set the transition timers when a state change occurs, possibly modify-
ing the enabling of transitions. Two basic mechanisms can be considered for
a timed transition at each state change:
- continue: the timer associated with the transition holds the present value
  and will continue later on the count-down;
- restart: the timer associated with the transition is restarted, i.e., its present
  value is discarded and a new value will be generated when needed.

The main reason for the introduction of temporal specifications into PN
models is the interest for the computation of performance indexes. The intro-
duction of temporal specifications in PN models must not reduce the modelling
capabilities with respect to the untimed case (it must be done so as not to
modify the basic behaviour and specifically the non-determinism of the choices
occurring in the net execution). If the temporal specification is given in a de-
Using stochastic Petri nets in performance analyze of distributed systems

terministic way, the behaviour of the model is deterministically specified, and
the choices due to conflicting transitions may be always solved in the same
way; this would make many of the execution sequences in the untimed PN
system impossible. Instead, if the temporal specifications are defined in a sto-
chastic manner, by associating independent continuous random variables with
transitions to specify their firing delays, the non-determinism is preserved (in
the sense that all the execution sequences possible in the untimed PN system
are also possible in the timed version of the PN system).

2. STOCHASTIC PETRI NETS

A simple sufficient condition to guarantee that the qualitative behaviour
of PN models with timed transitions is identical to the qualitative behaviour
of the underlying untimed PN model (that is, to guarantee that the possible
transition sequences are the same in the two models) is that the delays associ-
ated with transitions be random variables whose distributions have an infinite
support. If this is the case, a probabilistic metrics transforms the nondeter-
minism of the untimed PN model into the probabilism of the timed PN model,
so that the theory of discrete-state stochastic processes in continuous time can
be applied for the evaluation of the performance of the real system described
with the timed PN model.

Discrete-state stochastic processes in continuous time may be very diffi-
cult to characterize from a probabilistic viewpoint and virtually impossible
to analyse. Their characterization and analysis become reasonably simple
in some special cases that are often used just because of their mathematical
tractability. The simplicity in the characterization and analysis of discrete-
state stochastic processes in continuous time can be obtained by eliminating
the amount of memory in the dynamics of the process. This is the reason
for the adoption of the negative exponential probability density function (pdf)
for the specification of random delays. The negative exponential is the only
continuous pdf that enjoys the memoryless property, i.e., the only continuous
pdf for which the residual delay after an arbitrary interval has the same pdf
as the original delay.

Stochastic Petri Nets (SPNs) were proposed, simultaneously, by S. Natkin
[Nat80] and M. K. Molloy [Mol82]. The new net class paved the way to the
utilization of Petri nets for performance evaluation. In a SPN, a firing delay
is associated with each transition. Firing delays are instances of random vari-
able that have a negative exponential probability distribution. The rates are
sufficient to characterize the pdf of the transition delays (the only parameter
of the negative exponential pdf is its rate, obtained as the inverse of the mean).
The selection of the transition instance to fire among the set of enabled ones
follows a race policy (the transition that has drawn the least delay is the one
that fires). Each timed transition can be used to model the execution of an activity in a distributed environment; all activities execute in parallel (unless otherwise specified by the PN structure) until they complete. At completion, activities induce a local change of the system state that is specified with the interconnection of the transition to input and output places.

No special mechanism is necessary for the resolution of timed conflicts because the probability that two timers expire at the same instant is zero. The negative exponential pdf is a continuous function defined in the interval \([0, 1)\), that integrates to one. The lack of discontinuities in the function makes the probability of any specific value \(x\) being sampled equal to zero (obviously, the probability that a value is sampled between two distinct values \(x_1 \geq 0\) and \(x_2 > x_1\) is positive). Thus, the probability of two timers expiring at the same time is null (given the value sampled by the first timer, the probability that the second one samples the same value is zero). The memoryless property of the negative exponential pdf, at any time instant, makes the residual time until a timer expires statistically equivalent to the originally sampled timer reading. Thus, whether a new timer value is set or not at every change of marking makes no difference from the point of view of the probabilistic metrics of the SPN. A formal definition of SPN is the following.

**Definition 2.1** SPN = \((P, T, A, M, \lambda)\), where \(P\) is the set of places, \(T\) is the set of transitions, \(P \cap T = \emptyset\), \(P \cup T \neq \emptyset\), \(A\) is the set of input and output arcs, \(A \subseteq (P \times T) \cup (T \times P)\), \(M\) is the initial marking and \(\lambda\) is the set of transition rates.

A marking of a stochastic Petri net is a distribution of tokens in its places; a marking may be viewed as a mapping from the set of places to the natural numbers. A stochastic Petri net is said to be \(k\)-bounded if there exists a finite nonnegative integer \(k\) such that for every marking \(M\) in the reachability set, and for every place \(P_i\) \(M(P_i) < k\). With each marking we can associate a state of the system and in the following the terms state and marking are used with essentially the same meaning.

Molloy [Mol81] has proved that there is an isomorphism between \(k\)-bounded SPNs and finite Markov processes. Two stochastic systems are isomorphic if - there are one-to-one mappings between the state space of the two systems, and between the set of state transitions of the two systems, - the probability of a transition from one state to another in one system equals the probability of a transition between the corresponding states of the other system.

The SPNs are isomorphic to continuous time Markov chains due to the memoryless property of the negative exponential distribution of firing times. The SPN markings correspond to states of the corresponding Markov chain. Since the size of the state space of a SPN is equal to the size of Markov process space, the complexity of solving an SPN model is the same as in the case of
the model based upon the Markov process; the methodology used to find the steady-state solution for a Markov chain can be used for an SPN system in which each marking corresponds to a Markov state. Although the SPNs do not provide more modelling power than Markov processes, they can be used as a convenient description of the system being modelled.

3. THE STOCHASTIC PROCESS ASSOCIATED WITH A SPN

SPN systems are isomorphic to continuous time Markov chains (CTMCs); k-bounded SPN systems are isomorphic to finite CTMCs. The CTMC associated with a given SPN system is obtained by applying the following rules:

1. the CTMC state space  $S = \{s_i\}$ corresponds to the reachability set $RS(M_0)$ of the PN associated with the SPN ($s_i \leftrightarrow M_i$)

2. the transition rate from state $s_i$ (corresponding to marking $M_i$) to state $s_j$ ($M_j$) is obtained as the sum of the firing rates of the transitions that are enabled in $M_i$ and whose firings generate marking $M_j$.

It is possible to develop algorithms for the automatic construction of the infinitesimal generator (also called the state transition rate matrix) of the isomorphic CTMC, starting with the SPN description. Denoting this matrix by $Q$, with $w_k$ the firing rate of $T_k$, and with $E_j(M_i) = \{T_h \in E(M_i) : M_i[T_h > M_j]\}$ the set of transitions whose firings bring the net from marking $M_i$ to marking $M_j$, the components of the infinitesimal generator are

$$ q_{ij} = \begin{cases} 
-\sum_{k:T_k\in E(M_i)} w_k, & i = j \\
\sum_{k:T_k\in E_j(M_i)} w_k, & i \neq j 
\end{cases} $$

In SPN analysis, as in Markov analysis, ergodic (irreducible) systems are of special interest. A k-bounded Markov system is said to be ergodic if it generates an ergodic CTMC; in [FN85] is showed that a SPN system is ergodic if $M_0$, the initial marking, is a home state (a marking $M \in RS(M_0)$ is a home state if and only if $\forall N \in RS(M_0), M \in RS(N)$). For ergodic SPN systems, the steady-state probability of the system being in any state always exists and is independent of the initial state.

Let the row vector $\eta$ represent the steady-state probability distribution on markings of the SPN; to compute $\eta$ we solve the linear system expressed in matrix form as

$$ \begin{cases} 
\eta Q = 0, \\
\eta 1^T = 1,
\end{cases} $$
where $\mathbf{0}$ is row vector with all its components equal to zero and $1^T$ is a column vector with all its components equal to one, used to enforce the normalization condition. We will use $\eta_i$ instead of $\eta(M_i)$ to denote the steady state probability of marking $M_i$. The sojourn time is the time spent by the system in a given marking. Because a CTMC can be associated with the SPN system, the sojourn time in marking $M_i$ is exponentially distributed with rate $q_i$. The pdf of the sojourn time in a marking corresponds to the pdf of the minimum among the firing times of the transitions enabled in the same marking. The probability that a given transition $T_k \in E(M_i)$ fires first in marking $M_i$ has the expression: $P\{T_k/M_i\} = \frac{w_k}{q_i}$ and the average sojourn time in marking $M_i$ is given by $E[SJ_i] = \frac{1}{q_i}$.

To allow the efficiency of the SPN system to be evaluated, performance indices can be computed by using reward functions with different interpretations: the probability of a particular condition, the mean number of firings per unit of time of a given transition, the expected value of the number of tokens in a given place etc. If $r(M)$ is a reward function, the average reward can be computed with the following weighted sum

$$R = \sum_{i:M_i \in RS(M_0)} r(M_i)\eta_i.$$ 

In order to illustrate this analysis step, the following example [MBCDF94] is presented.

![Fig. 1. A shared memory system](image)

In fig. 1, two processors try to access a common shared memory; in the initial marking, the two processors are both in a locally active state ($p_{act1} + p_{act2} + p_{idle}$). Processor 1 works locally for an exponentially distributed random amount of time with average $\frac{1}{\lambda_1}$, and then requests an access to the common memory ($T_{req1}$ fires). If common memory is available, (place $p_{idle}$ is marked),
the acquisition of the memory starts and takes an average of $\frac{1}{\alpha_1}$ units of time to complete. After transition $T_{\text{str}}$ fires, processor 1 uses the common memory (which is not available for processor 2) for $\frac{1}{\mu_1}$ units of time (on average) and when transition $T_{\text{end}}$ fires, the net is returning to its initial state. A similar processing cycle is possible for processor 2.

The evolution of the net describes the interleavings between the activities of the processors. A conflict exists when transitions $T_{\text{str}}$ and $T_{\text{str}}$ are both enabled (both processors want to simultaneously access the common memory). Transition $T_{\text{str}}$ fires with probability $P\{T_{\text{str}}\} = \frac{\alpha_1}{\alpha_1 + \alpha_2}$, whereas transition $T_{\text{str}}$ fires with probability $P\{T_{\text{str}}\} = \frac{\alpha_2}{\alpha_1 + \alpha_2}$. The conflict is resolved when the first transition fires and the speed at which the PN model exits from this marking is the sum of the individual speeds of the two transitions. The reachability set is $\{M_0 = p_{\text{act}} + p_{\text{idle}} + p_{\text{act}}, M_1 = p_{\text{req}} + p_{\text{idle}} + p_{\text{act}}, M_2 = p_{\text{act}} + p_{\text{act}}, M_3 = p_{\text{req}} + p_{\text{req}}, M_4 = p_{\text{req}} + p_{\text{idle}}, M_5 = p_{\text{req}} + p_{\text{req}}, M_6 = p_{\text{act}} + p_{\text{act}}, M_7 = p_{\text{req}} + p_{\text{act}}\}$ and the infinitesimal generator of the Markov chain is $Q =$

$$
\begin{bmatrix}
-\lambda_1 - \lambda_2 & \lambda_1 & 0 & 0 & 0 & \lambda_2 & 0 & 0 \\
0 & -\alpha_1 - \lambda_2 & \alpha_1 & 0 & \lambda_2 & 0 & 0 & 0 \\
\mu_1 & 0 & -\lambda_2 - \mu_1 & \lambda_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\mu_1 & \mu_1 & 0 & 0 & 0 \\
0 & 0 & 0 & \alpha_1 & -\alpha_1 - \alpha_2 & 0 & 0 & \alpha_2 \\
0 & 0 & 0 & 0 & \lambda_1 & -\alpha_2 - \lambda_1 & \alpha_2 & 0 \\
0 & \mu_2 & 0 & 0 & 0 & 0 & -\lambda_1 - \mu_2 & \lambda_1 \\
0 & \mu_2 & 0 & 0 & 0 & 0 & 0 & -\mu_2
\end{bmatrix}
$$

The system of linear equations whose solution yields the steady-state distribution over the CTMC states is

$$
\begin{cases}
(\eta_0, \eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7)Q = 0, \\
\sum_{i=0}^{7} \eta_i = 1.
\end{cases}
$$

Assuming $\lambda_1 = 1, \lambda_2 = 2, \alpha_1 = \alpha_2 = 100, \mu_1 = 10$ and $\mu_2 = 5$, we obtain $\eta_0 = 0.61471, \eta_1 = 0.00842, \eta_2 = 0.07014, \eta_3 = 0.01556, \eta_4 = 0.00015, \eta_5 = 0.01371, \eta_6 = 0.22854, \eta_7 = 0.04876$.

Since one token is present in place $p_{\text{idle}}$ in markings $M_0, M_1, M_4$ and $M_5$, the average number of tokens in $p_{\text{idle}}$ is $E[M(p_{\text{idle}})] = \eta_0 + \eta_1 + \eta_4 + \eta_5 = 0.637$ and the utilization of the shared memory is $U[\text{shared memory}] = 1.0 - E[M(p_{\text{idle}})] = 0.363$.

Since $T_{\text{req}}$ is enabled only in $M_0, M_5$ and $M_6$, the rate of access to the shared memory from the first processor (the throughout of $T_{\text{req}}$) is $f_{\text{req}} = (\eta_0 + \eta_5 + \eta_6)\lambda_1 = 0.8570$. 

Using stochastic Petri nets in performance analyze of distributed systems 187
These results show that Stochastic Petri Nets can be used as a tool for computing performance indices that allow the efficiency of modelled systems to be evaluated.

Developed at Universita di Torino, GRaphical Editor and Analyzer for Timed and Stochastic Petri Nets (GreatSPN) [Chi91] is a software package for the modeling, validation, and performance evaluation of distributed systems using Stochastic Petri Nets. It runs on Linux (gcc, Motif 1.2 and X11R6 are required) and it offers graphical model editing, definition of timing and stochastic specifications, parameters, and performance measures, Markovian solvers for steady-state, graphical representation of performance results and graphical interactive simulation of stochastic models. The package is available for free for universities and non-profit organizations (http://www.di.unito.it/greatspn/index.html).

References


SOLVING METHOD FOR MULTIPLE OBJECTIVE FUZZY CONSTRAINTS MIXED VARIABLES OPTIMIZATION

Bogdana Pop, Silviu Dumitrescu, Vlad Monescu

"Transilvania" University of Brașov, Dept. of Computer Science

Abstract In the classical problems of mathematical programming the coefficients of the problems are assumed to be exactly known. This assumption is seldom satisfied by the great majority of real-life problems. Taking into account the multiple criteria optimization in the environment of fuzzy constraints and mixed variables, a solving method is developed here.

Keywords: fuzzy constraints, mixed variables, multiple criteria.

1. INTRODUCTION

In the real life the certainty in and the solidity of data accuracy are illusory. The possibility to obtain an optimal solution is also under the influences of some data missing. Uncertainty has to be taken into consideration when real systems are analyzed. The compromise to accept the uncertainty into the mathematical models must be done working with complex systems. In order to build and to solve real life problem it is natural to interpret the information in fuzzy manner. The philosophy of fuzzy sets theory is related to the model of human thinking and making decisions. Concepts of fuzzy sets theory "crowded" into a lot of research fields since 1980, when fuzzy logic has had a great success in the control systems theory. Real advantages of a fuzzy approach to solve optimization problems could be highlighted when a comparison is made with the stochastic methods to deal with imprecision ([4, 9]).

"The best compromise" (or Pareto-optimality, efficiency, non-domination) is the central concept in multiple criteria optimization only because an optimal solution for one objective function is not necessary an optimal solution for others. In the classical theory weak-efficiency, strong-efficiency, proper-efficiency are gradual definitions for the same essence. Starting with classical definitions, the following concepts appear in the fuzzy multiple objective optimization: M-Pareto optimality (related to the membership functions of "equal" goals), alpha-Pareto optimality (related to the multiple objective programming problems associated with each alpha-level set), M-alpha-Pareto optimality (related
to the multiple objective programming problems associated with each alpha-level set after that “fuzzy equal” goals were defined) [1, 2].

In [6], Perkgoz et al. focused on multiobjective integer programming problems with random variable coefficients in objective functions and/or constraints. In [11], Sakawa and Kato presented an interactive fuzzy satisfying method for nonlinear integer programming problems through a genetic algorithm. In [10] α-Pareto optimal solutions were determined for the multiobjective integer nonlinear programming problems having fuzzy parameters in the constraints together with the corresponding stability set of the first kind. The paper [5] presented a method useful in solving a special class of large-scale multiobjective integer problems based upon a combination of the decomposition algorithm coupled with the weighting method together with the branch-and-bound method.

Taking into account the multiple criteria optimization in the environment of fuzzy constraints and mixed variables, a solving method is developed here. In what follows a multiple objective fuzzy constraints mixed variables optimization problem (MOFCMVOP) will be taken into account.

2. THE MOFCMVOP MODEL

Consider the multiple objective integer programming problem with fuzzy constraints and mixed variables

\[ \text{"min"} \left\{ z(x) = (z_1(x), z_2(x), ..., z_p(x)) \mid x \in X \right\}, \]

where

(i) \( X = \{ x \in \mathbb{R}^n \mid Ax \leq b, \ x \in Z^k \times R^r \} \) is the feasible set of the problem,

(ii) \( A \) is an \( m \times n \) constraint matrix, \( x \) is an \( n \)-dimensional vector of mixed \((k−integer, r−real)\) decision variable and \( b \in \mathbb{R}^m \),

(iii) \( p \geq 2 \),

(iv) \( z_i(x), \ i = \overline{1,p} \) are the objective functions which could be linear, linear fractional or convex functions (in order to make the new method workable).

The term "min" used in Problem (1) is for finding all efficient solutions in a minimization sense in terms of the Pareto optimality. A possible way to handle constraints imprecision is to consider the following parametric problem

\[ \text{"min"} \left\{ z(x) = (z_1(x), z_2(x), ..., z_p(x)) \mid x \in X(\theta) \right\} \]

where \( X(\theta) = \{ x \in \mathbb{R}^n \mid Ax \leq b + \theta b', \ x \in Z^k \times R^r \} \) is the feasible set of the problem, \( \theta \in [0,1] \) and \( b' \) is a given perturbation vector [13].
3. THE SOLVING METHOD

In [7] an algorithm to solving multiobjective linear fractional programming problem (MOLFPP) is presented. In [8] the algorithm is used to solve a fuzzy multiple objective integer programming problem. We will use the above algorithm (under the name MultiObjAlg(θ)) to solve Problem (2) meaning the deterministic problem with the feasible set \( X(θ) \). In order to improve the interactivity of the method different values for \( θ \) will be considered.

Wang and Horng [14] proposed an approach to perform complete parametric analysis in integer programming by considering all possible candidates of \( θ \). They defined the principal candidates of \( θ \) as being that \( θ \) which makes \( b_i + θ b_i' \) an integer. We also work with these \( θ \)'s. In [13], Wang and Liao proposed a heuristic algorithm to analyze the same fuzzy problem. We will use their method to solving multiple objective problem with integer variables and fuzzy constraints.

For a fixed value \( θ \) we start defining Problem (3) as Problem (2) without the integrity restriction of variables.

\[
\text{"min" } \{ z(x) = (z_1(x), z_2(x), ..., z_p(x)) \mid Ax \leq b + θb' \}.
\]

We obtain an efficient solution for Problem (3) using MultiObjAlg(1). It is a solution for Problem (2) if and only if its components are integer numbers. In this case it is also solution for Problem (1) but with minimal degree in fuzzy environment. Consequently, our next goals are to transform the solution into a mixed (\( k- \)integer, \( r- \)real) one and also to improve its fuzzy degree. These goals could be attend modifying values \( x_i \) in \([x_i]\) or \([x_i]+1\) for \( i = 1, k \) such that the deviation of the perturbation vector decreases using a modified back\((p, r)\) procedure described in [8].

Taking into account the above remarks the solving algorithm could be described as follows.

- **Step 1.** Define the thresholds \( 0 = θ_1 < θ_2 < ... < θ_q = 1 \) using the principal candidates of parameters \( θ \). Put \( q = 1 \).

- **Step 2.** For \( j = q \) down to 1
  - Compute values \( x^j = (x_1, x_2, ..., x_n) \) using MultiObjAlg(\( θ_j \)).
  - if \( x^j \in Z^k \times R^r \) then favorable STOP with \( x^j \) the \( θ_j \)-fuzzy degree acceptable solution of the problem.
  - otherwise call back\((1, j, y^j)\) and obtain \( x'^j = (x'_1, x'_2, ..., x'_n) \). Identify \( j_{\text{min}} \) such that \( Ax' < b + θ_{j_{\text{min}}} b' \). Then \( x_{j_{\text{min}}} \) is the \( θ_{j_{\text{min}}} \)-fuzzy degree acceptable solution of the problem.

- **Step 3.** Describe the problem solution as \( (y^j)_{j=1}^{q} \).
References

ON THE APPLICABILITY OF THE LINDSTEDT-POINCARÉ METHOD IN THE TWO-DIMENSIONAL CASE

Nicolae - Doru Stânescu
Department of Applied Mechanics, University of Piteşti

Abstract In this note we give sufficient conditions for a system of two second order ordinary differential equations (ode’s) be transformed into the canonical form, i.e. to which the Lindstedt-Poincaré method applies. We also exemplify the theory by an application.

Keywords: Lindstedt-Poincaré method, canonical form

2000 MSC: 74B20, 70K42

1. ONE-DIMENSIONAL CASE

The Lindstedt-Poincaré method is applied to the canonical form of ode

\[ \ddot{u} + f(u) = 0, \]  

where \( f \) is a nonlinear function [1]. In addition, \( f \) is assumed to be analytical at the origin, therefore there exists the following expansion \( f = \sum_{i=1}^{\infty} a_i u^i \), with \( a_i = \frac{1}{i!} f^{(i)}(0) \). If the critical point is not \( u_0 = 0 \), the translation \( x = u - u_0 \) transforms \( f \) and \( a_i \) into \( f = \sum_{i=1}^{\infty} a_i x^i \) and \( a_i = \frac{1}{i!} f^{(i)}(u_0) \), respectively. Correspondingly, in the classical Lindstedt-Poincaré method, the solution of the equation \( \ddot{x} + \sum_{i=1}^{\infty} a_i x^i = 0 \) is written as a formal asymptotic expansion, \( x(t, \varepsilon) = \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \varepsilon^3 x_3(t) + \ldots \). With the notation \( \omega_0 = \sqrt{a_1} = \sqrt{f'(u_0)} \) and performing a change of the time variable \( \tau = \omega t \), where \( \omega = \omega_0 + \varepsilon \omega_1 + \varepsilon^2 \omega_2 + \ldots \), \( \varepsilon \) being a small dimensionless parameter, the equation becomes \( (\omega_0 + \varepsilon \omega_1 + \varepsilon^2 \omega_2 + \ldots) \frac{d^2}{d\tau^2} (\varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots) + \sum_{i=1}^{\infty} (\varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots) = 0 \).
2. TWO-DIMENSIONAL CASE

Consider now the two-dimensional case, i.e. of two nonlinear second order ode’s in two unknown functions

\[ \ddot{x} + f_1(x, y) = 0, \quad \ddot{y} + f_2(x, y) = 0, \]  
where the functions \( f_1 \) and \( f_2 \) are analytical and the following expansions hold

\[ f_1(x, y) = \left. \frac{\partial f_1}{\partial x} \right|_{(0,0)} x + \left. \frac{\partial f_1}{\partial y} \right|_{(0,0)} y + \frac{1}{2} \left( \left. \frac{\partial^2 f_1}{\partial x^2} \right|_{(0,0)} x^2 + 2 \left. \frac{\partial^2 f_1}{\partial x \partial y} \right|_{(0,0)} xy + \left. \frac{\partial^2 f_1}{\partial y^2} \right|_{(0,0)} y^2 \right) + \ldots, \]

\[ f_2(x, y) = \left. \frac{\partial f_2}{\partial x} \right|_{(0,0)} x + \left. \frac{\partial f_2}{\partial y} \right|_{(0,0)} y + \frac{1}{2} \left( \left. \frac{\partial^2 f_2}{\partial x^2} \right|_{(0,0)} x^2 + 2 \left. \frac{\partial^2 f_2}{\partial x \partial y} \right|_{(0,0)} xy + \left. \frac{\partial^2 f_2}{\partial y^2} \right|_{(0,0)} y^2 \right) + \ldots, \]

where the critical point is assumed to be \((x, y) = (0, 0)\). Then (1) becomes

\[
\begin{align*}
\ddot{x} + \left. \frac{\partial f_1}{\partial x} \right|_{(0,0)} x + \left. \frac{\partial f_1}{\partial y} \right|_{(0,0)} y + \bar{f}_1(x, y) &= 0, \\
\ddot{x} + \left. \frac{\partial f_2}{\partial x} \right|_{(0,0)} x + \left. \frac{\partial f_2}{\partial y} \right|_{(0,0)} y + \bar{f}_2(x, y) &= 0,
\end{align*}
\]

with the obvious notation. Let us find a linear transformation

\[ x = \alpha \xi_1 + \beta \xi_2, \quad y = \gamma \xi_1 + \delta \xi_2, \]  

i.e. deduce the coefficients \( \alpha, \beta, \gamma \) and \( \delta \), such that the equation in \( \xi_1 \) (respectively \( \xi_2 \)) has no term in \( \xi_2 \) (respectively \( \xi_1 \)). We have

\[ \begin{align*}
\alpha \ddot{\xi}_1 + \beta \ddot{\xi}_2 + \left. \frac{\partial f_1}{\partial x} \right|_{(0,0)} (\alpha \xi_1 + \beta \xi_2) + \left. \frac{\partial f_1}{\partial y} \right|_{(0,0)} (\gamma \xi_1 + \delta \xi_2) + \bar{f}_1(\xi_1, \xi_2) &= 0, \\
\gamma \ddot{\xi}_1 + \delta \ddot{\xi}_2 + \left. \frac{\partial f_2}{\partial x} \right|_{(0,0)} (\alpha \xi_1 + \beta \xi_2) + \left. \frac{\partial f_2}{\partial y} \right|_{(0,0)} (\alpha \xi_1 + \beta \xi_2) + \bar{f}_2(\xi_1, \xi_2) &= 0,
\end{align*} \]

where \( \bar{f}_1(\xi_1, \xi_2) = \bar{f}_1(\alpha \xi_1 + \beta \xi_2, \gamma \xi_1 + \delta \xi_2), \bar{f}_2(\xi_1, \xi_2) = \bar{f}_2(\alpha \xi_1 + \beta \xi_2, \gamma \xi_1 + \delta \xi_2). \)

Multiplying (4-1) by \( -\delta \), (4-2) by \( \beta \) and adding the obtained equations, we find

\[
\begin{align*}
(-\alpha \delta + \beta \gamma) \ddot{\xi}_1 + \xi_1 \left( -\alpha \delta \left. \frac{\partial f_1}{\partial x} \right|_{(0,0)} - \gamma \delta \left. \frac{\partial f_1}{\partial y} \right|_{(0,0)} + \alpha \beta \left. \frac{\partial f_2}{\partial x} \right|_{(0,0)} + \gamma \beta \left. \frac{\partial f_2}{\partial y} \right|_{(0,0)} \right) + \\
+ \xi_2 \left( -\beta \delta \left. \frac{\partial f_1}{\partial x} \right|_{(0,0)} - \delta^2 \left. \frac{\partial f_1}{\partial y} \right|_{(0,0)} + \beta^2 \left. \frac{\partial f_2}{\partial x} \right|_{(0,0)} + \beta \delta \left. \frac{\partial f_2}{\partial y} \right|_{(0,0)} \right) - \\
- \delta \bar{f}_1(\xi_1, \xi_2) + \beta \bar{f}_2(\xi_1, \xi_2) &= 0.
\end{align*}
\]
Multiplying (4-1) by \(-\gamma\), (4-2) by \(\alpha\) and adding the results, we have

\[
(-\beta\gamma + \alpha\delta) \ddot{\xi}_2 + \xi_1 \left(-\alpha\gamma \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} - \gamma^2 \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \alpha^2 \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} + \alpha\gamma \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} \right) +
+ \xi_2 \left(-\beta\gamma \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} - \alpha\gamma \frac{\partial f_2}{\partial y} \bigg|_{(0,0)} + \alpha\beta \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} + \alpha\delta \frac{\partial f_2}{\partial y} \bigg|_{(0,0)} \right) -
- \gamma f_1^2 (\xi_1, \xi_2) + \alpha f_2^2 (\xi_1, \xi_2) = 0.
\]

(7)

Imposing to the term in \(\xi_2\) in (5) and term in \(\xi_1\) in (6) to vanish it follows

\[
-\beta\delta \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} - \delta^2 \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \beta^2 \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} + \beta\gamma \frac{\partial f_2}{\partial y} \bigg|_{(0,0)},
-\alpha\gamma \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} - \gamma^2 \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \alpha^2 \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} + \alpha\gamma \frac{\partial f_2}{\partial y} \bigg|_{(0,0)}.
\]

(8)

Note that the two equations (7) coincide if \(\alpha \to \beta\) and \(\gamma \to \delta\).

We are lead to the general problem: which are the necessary and sufficient conditions for the equation \(A u v + B u^2 + C v^2 = 0\) have real solutions for \(u\) and \(v\)? Looking at this equation as a second degree equation in the unknown \(u\) if \(B \neq 0\), its discriminant is

\[
\Delta = (A^2 - 4BC) \ v^2.
\]

(9)

For \(B = 0\), one real solution always exists, so further on it is understood that \(B \neq 0\). Therefore, the equation has real solution if

\[
A^2 - 4BC \geq 0.
\]

(10)

With the notation

\[
A = -\frac{\partial f_1}{\partial x} \bigg|_{(0,0)} + \frac{\partial f_2}{\partial y} \bigg|_{(0,0)}, \quad B = -\frac{\partial f_1}{\partial y} \bigg|_{(0,0)}, \quad C = \frac{\partial f_2}{\partial x} \bigg|_{(0,0)},
\]

(11)

the condition (9) for the equation (7) reads

\[
\left( \frac{\partial f_2}{\partial y} \bigg|_{(0,0)} \right)^2 + \left( \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} \right)^2 - 2 \frac{\partial f_2}{\partial y} \bigg|_{(0,0)} \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} + 4 \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} \geq 0.
\]

(12)

Assuming that the condition (11) holds, the system (2) can be written as

\[
A_1 \ddot{\xi}_1 + B_1 \xi_1 + \ddot{\xi}_1 (\xi_1, \xi_2) = 0, \quad A_2 \ddot{\xi}_2 + B_2 \xi_2 + \ddot{\xi}_2 (\xi_1, \xi_2) = 0,
\]

(13)

with \(A_1 = -\alpha\delta + \beta\gamma, \ A_2 = -\beta\gamma + \alpha\delta, \ B_1 = -\alpha\delta \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} - \gamma\delta \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \alpha\beta \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} + \gamma\beta \frac{\partial f_1}{\partial y} \bigg|_{(0,0)}, \ B_2 = -\beta\gamma \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} - \alpha\delta \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \alpha\beta \frac{\partial f_1}{\partial x} \bigg|_{(0,0)} + \gamma\beta \frac{\partial f_1}{\partial y} \bigg|_{(0,0)} + \alpha\delta \frac{\partial f_2}{\partial x} \bigg|_{(0,0)} + \gamma\delta \frac{\partial f_2}{\partial y} \bigg|_{(0,0)},\)
\[ \alpha \delta \frac{\partial f_2}{\partial y} \Big|_{(0,0)}, \quad \bar{f}_1(\xi_1, \xi_2) = -\delta \bar{f}_1^1(\xi_1, \xi_2) + \beta \bar{f}_2(\xi_1, \xi_2) = 0, \quad \bar{f}_2(\xi_1, \xi_2) = -\gamma \bar{f}_1^1(\xi_1, \xi_2) + \alpha \bar{f}_2(\xi_1, \xi_2) = 0. \]

Since the condition \( D \equiv \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} \neq 0, \) always holds because (3) is a transformation, the system (12) becomes

\[ \ddot{\xi}_1 + a_{11} \xi_1 + g_1(\xi_1, \xi_2) = 0, \quad \ddot{\xi}_2 + a_{22} \xi_2 + g_2(\xi_1, \xi_2) = 0, \quad (14) \]

where \( a_{11} = \frac{B_1}{A_1}, \quad a_{22} = \frac{B_2}{A_2}, \quad g_1 = \frac{\bar{f}_1(\xi_1, \xi_2)}{A_1}, \quad g_2 = \frac{\bar{f}_2(\xi_1, \xi_2)}{A_2}, \) and it is referred to as the canonical form. Consequently, a sufficient condition for (1) to be transformed into the canonical form is (11) and \( B \neq 0 \) and \( C \neq 0. \)

Similarly, several conditions can be written in the case \( B = 0 \) (e.g. \( A \neq 0 \)) or \( C = 0. \)

3. APPLICATION

Let us transform the system

\[ \ddot{x} + 3x + 2y + x^3 = 0, \quad \ddot{y} + 2x + 3y + y^3 = 0. \quad (15) \]

into the canonical form.

The functions \( f_1 \) and \( f_2 \) are \( f_2 = 3x + 2y + x^3 \) and \( f_1 = 2x + 3y + y^3. \) The unique critical point is \((x, y) = (0, 0)\). We have \( \frac{\partial f_1}{\partial x} = 3 + 3x^2, \quad \frac{\partial f_1}{\partial y} = 2, \quad \frac{\partial f_2}{\partial x} = 2, \quad \frac{\partial f_2}{\partial y} = 3 + 3y^2, \) such that (10) imply \( A = 0, \quad B = -2, \quad C = 2, \) and the discriminant in the relation (8) is \( \Delta = 16v^2. \) We obtain \( u = v \) or \( u = -v. \)

Choosing \( \alpha = 1, \beta = 1, \gamma = -1, \delta = 1, \) (3) becomes \( x = \xi_1 + \xi_2, \quad y = -\xi_1 + \xi_2, \) and the canonical form is

\[ \ddot{\xi}_1 + \xi_1 + 3\xi_1^3 + 3\xi_1^2\xi_2 = 0; \quad \ddot{\xi}_2 + 5\xi_2 + 3\xi_1^2\xi_2 + \xi_2^3 = 0. \quad (16) \]

Throughout the article, it is assumed that the conditions ensuring the existence of periodic solutions are fulfilled.

References

Abstract  An algorithm is proposed to solve a multicriterial "bottleneck" transportation model.

Keywords: nonlinear programming, transportation model.

The transportation problem dealing with the total cost minimizing criterion, considered as a classical one, is well-known and sufficiently analyzed in the respective sources.

The transportation model of a "bottleneck" type is a specific problem within the transportation classical issue, the objective function of which is a nonlinear one. Special cases of these types of problems are investigated in many papers, e.g. [1], [2], [5], [6], where concrete algorithms used to solve them are carried out. The transportation model of the "bottleneck" type with two criteria, where the first one is providing the total transportation cost minimization and the second one, that is nonlinear, is strangling in time, is studied in article [7], where the authors propose the concrete algorithm to solve it. The special algorithm for solving transportation model of the "bottleneck" type with 3 criteria is presented in paper [8], where it is tested on a concrete example.

In our daily life the multiobjective fractional programming models are of great interest. We are often concerned about the optimization of the ratios like the summary cost of the total transportation expenditures to the maximal necessary time to satisfy the demands, the total benefits or production values into time unit, the total depreciation into time unit and many other important similar criteria, which may appear in order to evaluate the economical activities and make the correct managerial decisions. These problems led to the "bottleneck" transportation model with multiple fractional criteria, where the "bottleneck" criteria appear as a "minmax" time constraining. The common characteristic of these objective ratios is the identical denominators. Concrete algorithms for solving special models of transportation type with one criterion, where the objective function is a fractional one, are proposed in [3],[4].

The multicriterial transportation model of "bottleneck" type with two fractional criteria is defined as follows
\[
\min z_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} / \max_{i,j} \{ t_{ij} | x_{ij} > 0 \} \tag{1}
\]

\[
\min z_2 = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} x_{ij} / \max_{i,j} \{ t_{ij} | x_{ij} > 0 \} \tag{2}
\]

\[
\min z_3 = \max_{i,j} \{ t_{ij} | x_{ij} > 0 \} \tag{3}
\]

in the conditions

\[
\sum_{j=1}^{n} x_{ij} = a_i, \forall i = 1, m \tag{4}
\]

\[
\sum_{i=1}^{m} x_{ij} = b_j, \forall j = 1, n \tag{5}
\]

\[
\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \tag{6}
\]

\[
x_{ij} \geq 0, i = 1, m, j = 1, n \tag{7}
\]

where \( c_{ij} \) - cost of transporting a unit from the source \( i \) to destination \( j \), \( d_{ij} \) - deterioration of a unit while transporting from the source \( i \) to destination \( j \), \( a_i \) - availability at source \( i \), \( b_j \) - requirement at destination \( j \), \( x_{ij} \) - amount transported from source \( i \) to destination \( j \), \( t_{ij} \) - time of transporting a unit from source \( i \) to destination \( j \).

A non traditional algorithm of building numerous efficient solutions of the models is carried out here. It is useless to look for an optimal solution to settle the multicriterial mathematical models. Indeed, as it often occurs, there are no solutions at all.

That is why, one should better determine the multitude of non-dominant solutions, which are known as efficient solutions or optimal in the sense of Pareto.

In order to solve the multicriteria model the notion of an efficient solution has been introduced.

**Definition 0.1** The feasible solution for the multicriterial model is considered to be efficient iff there exists no other feasible solution, for which we obtain a better value at least for one criterion while the values of the rest criteria remain unmodified.

In order to solve the problem (1)- (7) by finding the set of the efficient basic solutions, we reduce it to the following model
The generalized algorithm for solving the fractional multiobjective transportation problem

\[
\min z_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij}, \quad \min z_2 = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}x_{ij}, \quad \min z_3 = \max_{i,j} \{t_{ij}|x_{ij} > 0\}
\]

in the conditions (4)-(7).

In [1] an algorithm to solve the model (8)-(10) in conditions (4)-(7) is proposed. The algorithm determines the multitude of extreme non-dominant solutions within the admissible space of solutions. The algorithm is theoretically and scientifically tested and proved in a concrete case.

The algorithm of solving the model (1)-(7) develops a procedure of building a multitude of all efficient, basic solutions. This set coincides with the set of the efficient basic solutions for the model (8)-(10) in conditions (4)-(7). That is why, in order to find the set of its basic efficient solutions, we reduce the multicriterial fractional transportation model of "bottleneck" type (1)-(7) to the problem (8)-(10) with the constraints (4)-(7).

**Theorem 0.1** The set of the efficient basic solutions of the model (1)-(7) and the model (8)-(10) coincide.

**Proof.** Let \( X^1 \) be an efficient basic solution for the model (1)-(7), and \( T^1 = \max_{i,j} \{t_{ij}/x_{ij}^1 > 0\} \). Taking into account the definition of the efficient solution, we state that for each available solution \( X^2 \) of this model and corresponding \( T^2 = \max_{i,j} \{t_{ij}/x_{ij}^2 > 0\} \), the following inequalities

\[
Z_1(X^1) < Z_1(X^2) \text{ and } Z_2(X^1) \leq Z_2(X^2), \text{ or } Z_1(X^1) \leq Z_1(X^2) \text{ and } Z_2(X^1) < Z_2(X^2),
\]

where \( T^2 \leq T^1, \ T^1 \geq 0, \ T^2 \geq 0, \) hold.

Suppose that the solution \( X^1 \) is not efficient for the model (8)-(10) in the conditions (4)-(7). Similarly to the previous reasoning, using the definition of the efficient solution, it follows that there exists the available solution \( X^2 \) of this model and corresponding \( T^2 \), for which the following inequalities

\[
\frac{Z_1(X^2)}{T^2} < \frac{Z_1(X^1)}{T^1} \text{ and } \frac{Z_2(X^2)}{T^2} \leq \frac{Z_2(X^1)}{T^1}, \text{ or } \frac{Z_1(X^2)}{T^2} \leq \frac{Z_1(X^1)}{T^1} \text{ and } \frac{Z_2(X^2)}{T^2} < \frac{Z_2(X^1)}{T^1}
\]

hold, where \( T^2 \leq T^1, \ T^1 \geq 0, \ T^2 \geq 0. \)

Multiplying inequalities (12) by \( T^1 \) and supposing \( k = \frac{T^1}{T^2} \), we obtain that the following inequalities

\[
kZ_1(X^2) < Z_1(X^1) \text{ and } kZ_2(X^2) \leq Z_2(X^1), \text{ or } kZ_1(X^2) \leq Z_1(X^1) \text{ and } kZ_2(X^2) < Z_2(X^1)
\]
hold, where \( T^2 \leq T^1, T^1 \geq 0, T^2 \geq 0 \).

Obviously \( k \geq 1 \), therefore from (13) we conclude that for the solution \( X^2 \) the following inequalities

\[
Z_1(X^2) < Z_1(X^1) \quad \text{or} \quad Z_1(X^2) \leq Z_1(X^1) \quad \text{and} \quad Z_2(X^2) \leq Z_2(X^1),
\]

(14)

hold, where \( T^2 \leq T^1, T^1 \geq 0, T^2 \geq 0 \), that contradicts (11).

Similarly, it can be proved that each efficient solution of the model (8)-(10) is also an efficient solution for the model (1)-(7).

The theorem is proved.

Generalizing this idea for the model with multiple number of fractional criteria with the “bottleneck” constraining criterion, we conclude that, in order to find the set of its efficient basic solutions, it may be reduced to the model

\[
\min z_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^1 x_{ij}, \quad \min z_2 = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^2 x_{ij}, \quad \ldots,
\]

\[
\min z_r = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^r x_{ij}, \quad \min z_{r+1} = \max_{i,j} \{t_{ij} \mid x_{ij} > 0\},
\]

(15)
in the conditions (4)-(7). The values \( C_{ij}^k \), \( k = 1, \ldots, r, i = 1, \ldots, m, j = 1, \ldots, n \) correspond to the concrete interpretation of the corresponding criteria.

If among the set of criteria from the model (15), there are some criteria of ”max” type, it is not difficult to reduce this case to the initial one. Obviously, the model (8)-(10) is a particular case of the model (15). Therefore the algorithm to solve the model (15) in the conditions (4)-(7) can be used to solve the model (8)-(10).

The truthfulness of the above theorem for the model (15) can be proved similarly.

The algorithm of finding the set of the efficient basic solutions for the model (15) is an interactive one. Initially, in order to find the first efficient basic solution of model (15), we consider at least \((m + n - 1)\) cells from the tables \( C^k \), \( k = 1, 2, \ldots, r \). The indexes’ order is maintained the same as in the table \( T \), where the cells are numbered according to the respective time values well arranged in the increasing order. Each iteration supposes a deep levels’ exploration and a completion of the multitude of efficient basic solutions for a new unblocked stochastic time-variable.

In the case when the same solutions have been found at upper level of other branch or when all possibilities of improvement have been spent at this level, the exploration procedure of each time instant chain is finite in depth and ends on every branch.
In the case when the solution of a certain configuration detains the form recorded in another link, which has been investigated earlier, its depth exploration has no justification, that is why it is eventually stopped.

We propose the logic scheme to construct the algorithm for solving the multicriteria transportation models of "bottleneck" type with a finite number of criteria, where
\[
\Delta_{ij} = (u_i + v_j) - c_{ij}, \quad n_i < p \quad (p \text{ is defined by the dimension of the problem}, \quad n_i \text{ is an index of ordering the cells by data from the table } T).
\]

**ALGORITHM**

1. Table \(T\) with the increasing order of time values which uses the \(k\) index is being well arranged. The index order is maintained for the respective cells from the tables \(C^k, \quad k = 1, \ldots, r\).

2. The adoption of an initial, basic solution in the first \(p = (m + n - 1)\) cells from the table \(C\) is performed. The other cells are considered to be blocked.

3. All configurations of basic solutions can be recorded at the level \(l = 0\), using only the non-blocked cells and providing the dozing in all those cells with \((i, j)|x_{ij} > 0\), for which the relation \(\Delta_{ij} \geq 0\) is to be true at least for one criterion.

   Each configuration of the solution is iteratively investigated, in this way obtaining the following records at the next level: \(l = l + 1\).

   If a certain level of a basic solution, which was previously obtained, is found, the latter will be not further studied. Since the problem covers a finite dimension, the multitude, of all basic solutions for the unblocked cells will be obtained by exploring a finite number of levels in depth.

4. If \(p < m \ast n\), the following \(p = p + 1\) cell is unblocked, and for this purpose the exploration of the basic solutions is revived, then we start with the level 0. The 4th step will be repeated until we get \(p = m \ast n\).

   The basic efficient solution set is selected out of the multitude of the basic solutions.

**Theorem 0.2** The set of all efficient basic solutions for the multiple criteria transportation problem of "bottleneck" type is found by applying the above algorithm.

Proof. Let \(L\) be a list of efficient basic solutions of model (15) being found by applying the above algorithm. We suppose that the efficient basic solution \(S_1\), that was not found using the above algorithm, exists and \(S_1 \notin L\). Let \(S_1\) correspond to \(T_1\). We will fix it on the branch that corresponds to the \(T_1\) beginning with the level 0, when the corresponding cells from table \(T\) are cleared. An wide exploration of the fixed branch leads to the registration of all basic solutions of branch \(T_1\). Thus, all basic solutions corresponding to time \(T_1\) are contained in this set. We will separate from the set \(L_{T_1}\) the efficient basic solutions corresponding to time \(T_1\). Obviously \(L_{T_1} \subset L\). As a result,
if $S_1 \in L_{T_1}$, then $S_1$ is a basic efficient solution found by applying the above algorithm or if $S_1 \notin L_{T_1}$, then $S_1$ is not a basic solution and moreover it is not a basic efficient solution. Therefore, either $S_1$ is not a basic solution or it is contained in list $L$. The theorem is proved.

**Example.** Consider the following 3-criteria problem.

<table>
<thead>
<tr>
<th>Time</th>
<th>Supply</th>
<th>Demand</th>
<th>Cost 1,2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>95</td>
<td>73</td>
<td>52</td>
</tr>
<tr>
<td>68</td>
<td>66</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>37</td>
<td>63</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>14</td>
<td>16</td>
</tr>
</tbody>
</table>

Using the above proposed algorithm we have found the following 11 efficient basic solutions:

$S_1=(176,207,68)$; $S_2=(164,276,68)$; $S_3=(178,203,68)$;
$S_4=(172,213,68)$; $S_5=(158,283,68)$; $S_6=(208,167,73)$;
$S_7=(202,173,73)$; $S_8=(156,200,95)$; $S_9=(176,175,95)$;
$S_{10}=(143,265,95)$; $S_{11}=(186,171,95)$.

The authors of the article [1], using their own algorithm for this example, have obtained 9 efficient extreme solutions.

**References**


LISC – A MATLAB PACKAGE FOR LINEAR DIFFERENTIAL PROBLEMS

Damian Trif
Babes-Bolyai University of Cluj-Napoca
dtrif@math.ubbcluj.ro

Abstract  A MATLAB package, based on the Lanczos tau method, useful in the Lyapunov-Schmidt (LS) method for nonlinear second order differential boundary value problems of the type $L u = N u$ is presented. Future applications to evolution problems or bifurcation studies are also taken into account.

1. INTRODUCTION

The Lyapunov-Schmidt (LS) method, elaborated in the years 1906-1908 and reformulated in a modern mathematical language by L. Cesari [3] after 1963 applies to some nonlinear equations of the type $L u = N u$, in the presence of boundary conditions, considered on the domain of the linear operator $L$.

This method could be easily extended to the case of a nonlinear evolution equation on a Hilbert space $H$ (usually a $L^2$ space) of the form $\frac{d u}{d t} = F(u) \equiv L u + N u$, where the domain of $F$ is dense in $H$. We can also use the Lyapunov-Schmidt method to discuss the bifurcation problem $x - \lambda A x + G(x, \lambda) = 0$, $x \in X$, where $X$ is a Banach space, $A : X \rightarrow X$ is a linear compact operator and $G : X \times \mathbb{R} \rightarrow X$ is a continuous mapping such that $G(x, \lambda) = o(\|x\|)$ for all $\lambda \in \mathbb{R}$.

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, by Cesari, the alternative method) is also very useful to the effective approximation of these solutions.

We will present LISC, a MATLAB package based on the Lanczos-tau method, for Sturm-Liouville problems with applications to the Lyapunov-Schmidt method for nonlinear equations. 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 occurring in some examples, proves that the LS method behaves better than other known methods such as bvp4c or sbvp.
We assume that the linear part $L$ of the equation $Lu = Nu$ is a Sturm-Liouville operator

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

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

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

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

A theoretical constructive variant of the LS method can be found in [7], [8].

Summarizing, the approximating algorithm is:

a) we are looking for an approximate solution of the equation $Lu = Nu$ of the form

$$u = \sum_{k=1}^{m} c_k \varphi_k + \sum_{k=m+1}^{N} c_k \varphi_k,$$

where $0 \leq m \leq N$;

b) by fixing $v = \sum_{k=1}^{m} c_k \varphi_k$, we generate the associate function $y(v)$ performing the iterations

$$y^0 = v, \quad y^{s+1} = v + H_m Ny^s = v + \sum_{k=m+1}^{N} C_k^s \varphi_k, \quad s = 0, 1, ..., S;$$

c) with $y = y^{S+1}$ as an approximation of the associated function, we can write the system $Lu^* = P_m Ny$ of the determining equations, with the unknowns $c_1, ..., c_m$. This system of the form $F(c_1, ..., c_m) = 0$ is then numerically solved, by a suitable method, for instance by the Newton's method. Every evaluation of the function $F$ means the reiteration of the b) step. Finally, thus determined $u^*$ generates, also by the b) iterations, an approximation of the solution of the equation $Lu = Nu$.

We remark that in the case of the 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 $c_k, k = 1, ..., N$ are determined from the equations $(Lu^* - Nu^*, \varphi_k) = 0, k = 1, ..., N$ i.e.

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

These equations are got from the determining equations for $m = N$. If $m = 0$ the system of the determining equations disappears. The function associated
with a certain \( u^* \) satisfies the equation \( y = L^{-1}Ny \). Therefore, in this case, the algorithm is reduced 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.

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

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

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

There exists a Matlab package MATSLISE of V. Ledoux (2004), based on the works of L. Ixaru which uses the so called CP methods to calculate the eigenfunctions of Sturm-Liouville or Schrödinger operators but this package works slowly. A more interesting package is MWRtools of R. A. Adomaitis (1998-2001) [1] which uses spectral methods to calculate the eigenfunctions of the Sturm-Liouville operator in order to solve some linear boundary value problems.

In the next section we propose a Chebyshev-tau method to solve the Sturm-Liouville problem in order to get a good basis \( \varphi_i \), and the corresponding Matlab package.

3. CHEBYSHEV-TAU METHOD

3.1. THE PROBLEM

The module Chebyshev of our package solves problems of the type

\[
p_2(x)u'' + p_1(x)u' + p_0(x)u = g(x), \quad x \in (a, b) \tag{1}
\]

\[
\alpha_{11}u(x_{11}) + \alpha_{12}u'(x_{12}) = \beta_1,
\]

\[
\alpha_{21}u(x_{12}) + \alpha_{22}u'(x_{22}) = \beta_2 \tag{2}
\]

using the Chebyshev-tau method.

For the moment let us suppose that \( a = -1, b = 1 \). A powerful method to solve (1) is to express \( u \) as a Chebyshev series

\[
u(x) = c_0 T_0(x) + c_1 T_1(x) + ..., \]

where \( T_i(x) = \cos(i \cos^{-1}(x)) \) is the standard Chebyshev polynomial of order \( i \) [5], [4].

For the practical implementation, we define the vectors \( c \) and \( t \) by \( c^T = (c_0, c_1, c_2, ...) \), \( t^T = (\frac{T_0}{2}, T_1, T_2, ...) \) so that \( u(x) = c^T t = t^T c \). By using the properties of Chebyshev polynomials we have \( x \cdot T_i = \frac{T_{i-1}}{2} + \frac{T_{i+1}}{2} \), hence, \( x \cdot u(x) = c^T X^T t = (Xc)^T t \), where \( X_{0,1} = 1, X_{ii} = 0, X_{i,i-1} = X_{i,i+1} = 1/2 \).
Then, in general, $x^m u(x) = (X^m c)^T t$ and $f(x) u(x) = (f(X)c)^T t$ for analytical functions $f$. Moreover, $\frac{u(x)}{x^m} = (X^{-m} c)^T t$ 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, in general, the matrix version $\text{funm}(X)$ of the scalar function $f(x)$ or $X^{-m} = [\text{inv}(X)]^m$ are no longer sparse matrices.

Similarly, let $D$ be the differentiation matrix defining $\frac{d^m u}{dx^m} = (D^m c)^T t$. The matrix $D$ is an upper triangular and its entries are $D_{ii} = 0$, $D_{ij} = 0$ for $(j-i)$ even and $D_{ij} = 2$ otherwise.

Applying these formulae to equation (1), 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

$$g(x) = g_0 \frac{T_0(x)}{2} + g_1 T_1(x) + ...$$

The condition (2) can be formulated in a similar manner. We define $t_{ij} = t^T (x_{ij})$ so that it can be written in the form $h_i^T c = \beta_i$, $i = 1, 2$, where

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

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

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

of the form $Ac = b$, 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.

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

4. **EXAMPLES FOR LS METHOD**

The package LISC is under construction. Here we present some test problems for nonlinear equations, using 50 grid points and 21 eigenfunctions. The tutorials and the m-files of the package can be obtained by e-mail from the author (dtrif@math.ubbcluj.ro) or from Matlab Central - File Exchange.
For the moment, it is based on a Sturm-Liouville solver LiScEig for a bounded interval (for the $\varphi_i$ basis) and the files `as.m` and `lisc.m` which implement the LS method.

**Example 1.** Solve the boundary value problem

\[ x'' + x = 0.5t - 0.5x^3, \]
\[ x(0) = 0, \]
\[ x(1) + x'(1) = 0. \]

The eigenvalues of the linear part are $\lambda_i = 1 - \ell_i^2$, where $\sin \ell_i = -\ell_i \cos \ell_i$, $\varphi_i = c_i \sin (\ell_i t)$. The errors between our solution and the solution calculated by `bvp4c` of Matlab is less than $14 \times 10^{-6}$.

**Example 2.** Consider the two-point problem for the Emden equation (with a singularity)

\[ y'' + \frac{2}{x} y' + y^5 = 0, \]
\[ y'(0) = 0, \quad y(1) = \frac{\sqrt{3}}{2}. \]

The general Sturm-Liouville operator now has the form $\frac{1}{x^2} (x^2 u')' = \lambda u$. The closed-form solution reads $y(x) = \left(1 + \frac{x^2}{3}\right)^{-\frac{1}{2}}$ and the error is of order $1.e-5$.

**Example 3.** If comes from the combustion theory [6],

\[ -u'' + Mu' = DY_o Y_f e^{-\frac{\theta}{T_0 + u}}, \]
\[ Mu(0) - u'(0) = 0, \quad u(1) = 0, \]

where $Y_o = -u + Y_o e^{M(x-1)}$, $Y_f = -u + 1 - e^M (x-1)$, $T_0 = 0.118, \quad Y_o = 0.21, \quad \theta = 2.6$. The computed solution for $M = 6$ and $D = 5.5 \times 10^8$ (with a high gradient at the burning front) agrees well with those of [6].

5. **CONCLUSIONS**

The comparison between LISC and SBVP 1.0 of Auzinger [2] and `bvp4c` of Matlab (see Matlab help) shows a computing time for LISC about 1-2 times larger. 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 computation of the numerical solution. But we have good reasons to use LS method.
We can build a database with known eigenfunctions. In the problems with parameters, where we have (for example) bifurcations, or in evolution problems, we can use repeatedly the same eigenfunctions. The eigenfunctions carry physical information, so that our LS solution has a better structure for studies. LS method could be easily extended to 2D or 3D (evolution) problems, with non-invertible linear part. 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


OPTIMAL CONTROL AND STOCHASTIC UNIFORM OBSERVABILITY

Viorica Mariela Ungureanu
"Constantin Brancusi" Univ., Târgu -Jiu
vio@utgjiu.ro

Abstract: G. Da Prato and I. Ichikawa solved in [3] the quadratic control problem (1), (2), under stabilizability and detectability conditions. We replace the detectability condition with the uniform observability property and we obtain an optimal control and the minimal value of the cost functional (2). So, we generalize the results obtained by T. Morozan in [7] for finite dimensional case and we also conclude that our result is distinct from the one of G. Da Prato and I. Ichikawa. We use the above results to give a method for the numerical computation of the optimal cost. Under the same hypotheses we solve the tracking problem (1), (3).

Keywords: quadratic control, tracking, Riccati equation, detectability, stabilizability and observability

2000 MSC: 93E20, 49K30, 93B07.

1. NOTATION AND STATEMENT OF THE PROBLEM

Let $H, U, V$ be separable real Hilbert spaces. Denote by $L(H, V)$ the Banach space of all bounded linear operators from $H$ into $V$ (if $H = V$ we put $L(H, V) = L(H)$). Denote by $\mathcal{H}$ the subspace of $L(H)$ formed by all self-adjoint operators. Write $\langle \cdot, \cdot \rangle$ for the inner product and $\| \cdot \|$ for norms of elements and operators. Then $S \in L(H)$ is called nonnegative ($S \geq 0$) if $S$ is self-adjoint and $\langle Sx, x \rangle \geq 0$ for all $x \in H$. We set $L^+(H) = \{ S \in L(H), S \geq 0 \}$. For each interval $J \subset \mathbb{R}_+$, we denote by $C_s(J, L(H))$ the space of all strongly continuous mappings $G(t) : J \subset \mathbb{R}_+ \rightarrow L(H)$ and by $C_b(J, L(H))$ the subspace of $C_s(J, L(H))$, which consist of all mappings $G(t)$ such that $\sup_{t \in J} \| G(t) \| < \infty$. Let $E$ is a Banach space; denote by $C(J, E)$ the space of all continuous mappings $G(t) : J \subset \mathbb{R}_+ \rightarrow E$. We need the following assumption:

$P_1$: a) $A(t)$, $t \in [0, \infty)$ is a closed linear operator on $H$ with constant domain $D$ dense in $H$;

b) there exist $M > 0$, $\eta \in (\frac{1}{2} \pi, \pi)$ and $\delta \in (-\infty, 0)$ such that $S_{\delta, \eta} = \{ \lambda \in C; |\arg(\lambda - \delta)| < \eta \} \subset \rho(A(t))$, for all $t \geq 0$ and $\| R(\lambda, A(t)) \| \leq \frac{M}{|\lambda - \delta|}$ for all $\lambda \in S_{\delta, \eta}$;

c) there exist numbers $\alpha \in (0, 1)$ and $\tilde{N} > 0$ such that
It is known [4] that if $P_1$ holds, then the family $A(t), t \geq 0$ generates an evolution operator $U(t, s)$. For any $n \in \mathbb{N}$ we have $n \in \rho(A(t))$. The operators $A_n(t) = n^2 R(n, A(t)) - nI$ are called the Yosida approximations of $A(t)$. If we denote by $U_n(t, s)$ the evolution operator relative to $A_n(t)$, for each $x \in H$ one has

$$
\lim_{n \to \infty} U_n(t, s)x = U(t, s)x \text{ uniformly on any bounded subset of } \{(t, s); t \geq s \geq 0\}.
$$

Let $(\Omega, F, F_t, t \in [0, \infty), P)$ be a stochastic basis and assume that $P_1$ holds. Consider the following stochastic equation with control

$$
dy(t) = A(t)y(t) + B(t)u(t)dt + \sum_{i=1}^{m} G_i(t)y(t)dw_i(t), \quad y(s) = x \in H, \quad (1)
$$

denoted by $\{A, B; G_i\}$, where $u \in U_{ad} = \{u \in L^2(\mathbb{R}_+ \times \Omega, U), u \text{ is } F_\tau-\text{adapted}\}, w_i$’s are independent real Wiener processes relative to $F_\tau$. Suppose that the hypotheses

$$
P_2 : B \in C_b([0, \infty), L(U, H)), B^* \in C_b([0, \infty), L(H, U)), C \in C_b([0, \infty), L(H, V)), C^*C, G_i \in C_b([0, \infty), L(H)), K(t) \in C_b([0, \infty), L^+(U)) \text{ and there exists } \delta_0 > 0 \text{ such that }$$

$$
K(t) \geq \delta_0 I \text{ for all } t \in [0, \infty)
$$

are fulfilled. Denote

$$
\tilde{Z} = \sup_{r \in [0, \infty)} \|Z(r)\| \text{ for } Z = B, C, G_i, K.
$$

Assume that $P_1, P_2$ hold. It is known that under uniform observability and stabilizability conditions the Riccati equation of stochastic control (3) has a unique, uniformly positive, bounded on $\mathbb{R}_+$ and stabilizing solution (see Theorem 3.2). We use this result to solve both a quadratic control problem and a tracking problem associated with (1).

**Quadratic control problem.** We look for an optimal control $u \in U_{ad}$, which minimize the following quadratic cost

$$
I_\lambda(u) = E \int_s^\infty \|C(t)y(t)\|^2 + <K(t)u(t), u(t)> dt, \quad (2)
$$

where $U_{ad} = \{u \in U_{ad} \text{ such as } E \|y(t)\|^2 \to 0 \text{ as } t \to \infty\}$.

**Tracking problem.** Given a signal $r \in C_b(\mathbb{R}_+, H)$ we want to minimize the cost

$$
J(s, u) = \lim_{t \to \infty} \frac{1}{t-s} E \int_s^t \|C(\sigma)(y(\sigma) - r(\sigma))\|^2 + (K(\sigma)u(\sigma), u(\sigma)) d\sigma \quad (3)
$$

in a suitable class of controls $u$ subject to the equation $\{A, B; G_i\}$. 
2. BOUNDED SOLUTIONS OF LINEAR STOCHASTIC EQUATIONS

Assume that $P_1, P_2$ hold. With (1) associated the equation

$$dy(t) = A(t)y(t)dt + \sum_{i=1}^{m} G_i(t)y(t)dw_i(t), \quad y(s) = x \in H,$$

(4)

denoted by $\{A; G_i\}$.

It is known [3] that $\{A; G_i\}$ has a unique mild solution in $C([s, T], L^2(\Omega; H))$ that is adapted to $F_t$; namely the solution of

$$y(t) = U(t, s)x + \sum_{i=1}^{m} \int_{s}^{t} U(t, r)G_i(r)y(r)dw_i(r).$$

(5)

Let $y(t, s; x)$ be the mild solution of $\{A; G_i\}$.

We have the following definition (see [5] for the autonomous case):

**Definition 2.1** We say that $\{A; G_i\}$ is uniformly exponentially stable if there exist constants $M \geq 1, \omega > 0$ such that

$$E\|y(t, s; x)\|^2 \leq Me^{-\omega(t-s)} \|x\|^2 \quad \text{for all} \quad t \geq s \geq 0 \quad \text{and} \quad x \in H.$$

**Definition 2.2** [3] We say that $\{A, B; G_i\}$ is stabilizable if there exists $F \in C_b([0, \infty), L(H, U))$ such that $\{A+BF; G_i\}$ is uniformly exponentially stable.

**Lemma 2.1** Assume $P_1$ and $F \in C_b([0, \infty), L(H, U))$. If $h \in C_b(\mathbb{R}_+, H)$, then the equation

$$g_n'(t) = -(A_n^* + F^*)g_n(t) - h(t), g(T) = x_0 \in H$$

(6)

where $A_n, n \in \mathbb{N}$ are the Yosida approximations of $A$ and the weak differentiability is considered, has a unique solution. The functions $(t, x) \to \langle g_n'(t), x \rangle, n \in \mathbb{N}$ are continuous on $[0, \infty) \times H$. Moreover if $V_n(t, s) \ (\text{resp. } V(t, s))$ is the evolution operator generated by $A_n + F$ (resp. $A + F$), then we have

$$\langle g_n(t), y \rangle \to_{n \to \infty} \langle V_{A_n+F}(T, t)x_0, y \rangle + \left( \int_{t}^{T} V_{A_n+F}^*(\sigma, t)h(\sigma)d\sigma, y \right),$$

(7)

uniformly with respect to $t \in [0, T]$.

**Proof.** Since $\frac{\partial V_{A_n+F}(t, s;x)}{\partial s} = -V_{A_n+F}(t, s)(A_n(s) + F(s))x$, it is easy to see that $g_n(t) = V_{A_n+F}^*(T, t)x + \int_{t}^{T} V_{A_n+F}^*(\sigma, t)h(\sigma)d\sigma$ is the unique solution
of the equation (6). Using Lemma 3 in [9] it follows that for each $y \in H$, $\lim_{n \to \infty} V_{A_n+F}(t,s)y = V_{A+F}(t,s)y$ uniformly with respect to $0 \leq s \leq t \leq T$ and (7) holds.

**Remark 2.1** Assume that $P_1$ holds and $\{A, B; G_i\}$ is stabilizable with the stabilizing sequence $F \in C_b([0,\infty), L(H,U))$ and let $h \in C_b(\mathbb{R}_+, H)$. Since $\{A+BF, G_i\}$ is uniformly exponentially stable it follows that there exist constants $M \geq 1, \omega > 0$ such as $\|V_{A+BF}(t,s)\| \leq Me^{-\omega(t-s)}$ for all $t \geq s \geq 0$.

Hence, the integral

$$g(s) = \int_s^T V_{A+BF}^*(\sigma, s)h(\sigma)d\sigma$$

(8)

is convergent in $H$ and $g(s)$ is bounded on $\mathbb{R}_+$. Moreover

$$\left(\int_t^T V_{A+BF}^*(\sigma, t)h(\sigma)d\sigma, x\right) \to_{T \to \infty} \left(\int_s^\infty V_{A+BF}^*(\sigma, s)h(\sigma)d\sigma, x\right).$$

If we consider the solution of (6) with the initial condition $g_n(T) = g(T)$ it is not difficult to see that $\langle g_n(s), y \rangle \to_{n \to \infty} \langle g(s), y \rangle$ for all $y \in H$.

### 3. THE RICCATI EQUATION OF STOCHASTIC CONTROL AND THE UNIFORM OBSERVABILITY

Consider the system $\{A; G_i; C\}$ formed by equation $\{A; G_i\}$ and the observation relation $z(t) = C(t)y(t, s; x)$.

**Definition 3.1** [7] The system $\{A; G_i; C\}$ is uniformly observable if there exist $\tau > 0$ and $\gamma > 0$ such that $E \int_s^{s+\tau} \|C(t)y(t, s; x)\|^2 dt \geq \gamma \|x\|^2$ for all $s \in \mathbb{R}_+$ and $x \in H$.

In the deterministic case it is known (see [6] for the autonomous case) that uniform observability implies detectability. In [9] we proved that this assertion is not true in the stochastic case.

Let us consider the Riccati equation

$$P'(s) + A^*(s)P(s) + P(s)A(s) + \sum_{i=1}^m G_i^*(s)P(s)G_i(s) + C^*(s)C(s) - P(s)B(s)(K(s))^{-1}B^*(s)P(s) = 0.$$  

(9)
If $A_n(t), n \in \mathbb{N}$ are the Yosida approximations of $A(t)$, then we introduce the approximate equation

$$P'_n(s) + A_n(s)P_n(s) + P_n(s)A_n(s) + \sum_{i=1}^{m} G_i^*(s)P_n(s)G_i(s) + C^*(s)C(s) - P_n(s)B(s)(K(s))^{-1}B^*(s)P_n(s) = 0. \quad (10)$$

**Lemma 3.1** [3] Let $0 < T < \infty$ and let $R \in L^+(H)$. Then there exists a unique mild (resp. classical) solution $P$ (resp. $P_n$) of (3) (resp. (3)) on $[0, T]$ such that $P(T) = R$ (resp. $P_n(T) = R$). They are given by

$$P(s)x = U^*(t, s)RU(t, s)x + \int_{s}^{t} U^*(r, s)\sum_{i=1}^{m} G_i^*(r)Q(r)G_i(r) + C^*(r)C(r) - P(r)B(r)(K(r))^{-1}B^*(r)P(r)]U(r, s)dxdr \quad (11)$$

$$P_n(s)x = U^*_n(T, s)RU_n(T, s)x + \int_{s}^{T} U^*_n(r, s)\sum_{i=1}^{m} G_i^*(r)P_n(r)G_i(r) + C^*(r)C(r) - P_n(r)B(r)(K(r))^{-1}B^*(r)P_n(r)]U_n(r, s)dxdr \quad (12)$$

and for each $x \in H$, $P_n(s)x \to P(s)x$ uniformly on any bounded subset of $[0, T]$. Moreover, if we denote these solutions by $P(s, T; R)$ and $P_n(s, T; R)$ respectively, then they are monotone in the sense that $P(s, T; R_1) \leq P(s, T; R_2)$ if $R_1 \leq R_2$.

We say [3] that $P$ is a mild solution on an interval $J$ of (3) if $P \in C_s(J, L^+(H))$ and if $P(s, T; P(s))$ satisfies (11) for all $s \leq t, s, t \in J$. Moreover, if $P$ is a mild solution on $\mathbb{R}_+$ of (3) and $\sup_{s \in \mathbb{R}_+} ||P(s)|| < \infty$, then $P$ is said to be a bounded solution.

**Remark 3.1** From the above lemma it is easy to deduce that for all $\alpha, \beta \in \mathbb{R}_+, \alpha \leq \beta$ we have

$$P(s, \alpha, 0) \leq P(s, \beta, 0) \quad (13)$$

for all $s \in [0, \alpha]$.

**Definition 3.2** A mild solution of (3) is called stabilizing for $\{A, B; G_i\}$ if $\{L = A - BS; G_i\}$ is uniformly exponentially stable, where $S(t) = K^{-1}(t)B^*(t)P(t)$.

We introduce the following hypothesis:
Assume that the hypotheses of the above theorem hold. a) From Lemma 4.1 and the stochastic version of the Theorem 3.1 from [3], it is easy to see that the unique uniformly positive solution of the equation (3) has a bounded solution on $[0, \infty)$. Hence the evolution operator $U_{t, s}$ associated with the operator $L$ satisfies the inequality $E \|z(t, s; x)\|^2 \leq \beta e^{-\alpha(t-s)} \|x\|^2$ for all $t \geq s \geq 0$ and $x \in H$, where $\alpha = -\ln(1 - \frac{\beta}{\gamma})$ and $\beta$ depends only on $M_0, \omega, \tau$ (which are introduced above) and $m$. Hence the evolution operator $U_{t, s}$ satisfies the inequality $\|U_{t, s}\|^2 \leq \beta e^{-\alpha(t-s)}$ too.
Assume that the hypotheses of the Theorem 11 holds. Theorem 13 gives an estimate for the convergence rate of the sequence \( P(., T, 0) \) to the unique solution of the Riccati equation (3).

**Theorem 3.3** Assume that \( P_3 \) holds, \( \{A, G_i; B\} \) is stabilizable and \( \{A, G_i; C\} \) is uniformly observable. If \( P(.,) \) is the unique nonnegative and stabilizing solution of the Riccati equation (8), then

\[
\|P(t) - P(t, T, 0)\| \leq \varepsilon
\]

for all \( t \leq T \leq T \) and \( T \geq T_{\varepsilon, T} = -\alpha \ln \frac{\varepsilon}{\hat{C} e^{\alpha T}} \).

**Proof.** If \( P(.,) \) is the solution of the Riccati equation (8) such as the system \( \{L = A - BS; G_i\} \) is uniformly exponentially stable, then \( [3] \) \( P(.,) - P(., T, 0) \) is a mild solution of the equation

\[
Z' + L^* Z + ZL + \sum_{i=1}^{m} G_i^* Z G_i + ZB(K)^{-1}B^* Z = 0
\]

with the final condition \( Z(T) = P(T) \). Denote \( \hat{K} = \beta \left( \frac{2M^2}{\delta} + \sum_{i=1}^{m} \tilde{G}_i^2 \right) e^{\alpha T} \) and \( \hat{C} = M \beta e^{\alpha T} \), where \( T > 0 \) is fixed, we use the integral equation (see (11)) satisfied by \( Z \), the statement a) of the above remark and the Gronwall’s inequality to get

\[
\|P(t) - P(t, T, 0)\| \leq \hat{C} e^{-\alpha T} + \int_{t}^{T} \hat{K} e^{-\alpha u} \|P(u) - P(u, T, 0)\| \, du
\]

and

\[
\|P(t) - P(t, T, 0)\| \leq \hat{C} e^{-\alpha T} e^{\hat{K} e^{-\alpha T}} \text{ for all } t \leq T \leq T. \tag{14}
\]

Now, obviously, for all \( t < T, \varepsilon > 0 \) and \( T \geq T_{\varepsilon, T} = -\alpha \ln \frac{\varepsilon}{\hat{C} e^{\alpha T}} \) we have \( \|P(t) - P(t, T, 0)\| \leq \varepsilon \).

## 4. OPTIMAL QUADRATIC CONTROL

A consequence of Theorem 3.2, Theorem 4.2 and the stochastic version of Theorem 3.1 from [3] is the following theorem:

**Theorem 4.1** Assume that the hypotheses of the Theorem 11 are fulfilled and consider the control problem (1), (2). The optimal control is given by the feedback law

\[
\tilde{u}(t) = -K(t)^{-1} B^* (t) P(t) y(t),
\]
where \( P \) is the unique bounded positive solution of (3) \( (y(t) \) is the corresponding solution of (1)) and the optimal cost is

\[
I_s(\bar{u}) = \langle P(s)x, x \rangle.
\]  

If all operators in (1) and (2) are time invariant, then \( P(\cdot) \) is constant.

The following conclusion provides a numerical method for the computation of the optimal cost.

**Conclusion 1** Assume that the hypotheses of the above theorem hold and let \( H = \mathbb{R}^n \). From Remark 3.2 we deduce an algorithm which allows us to obtain the optimal cost (15) in the finite dimensional case. Let \( \varepsilon > 0 \) and \( t \in [0, T] \) be fixed.

1. We can use the Runge Kutta method to obtain the solutions \( P(t, \alpha, 0) \), \( \alpha \in \mathbb{R}_+ \) of the Riccati equation (3). We compute the constant \( M \) from the statement a) of Remark 12 (as a limit of an monotonously increasing sequence of real numbers). Consequently we can obtain all the constants from relation (14).

3. Using the constant \( T_{\varepsilon, T} \), introduced in Section 3, we see that \( P(t, T_{\varepsilon, T}, 0) \) is a good approximation (with the error \( \varepsilon \)) of the solution \( P(t), t \in [0, T] \) of the Riccati equation (3). Hence \( \langle P(t, T_{\varepsilon, T}, 0)x, x \rangle \) is an approximation of the optimal cost (15).

5. **TRACKING PROBLEM**

Consider the set of admissible controls \( \mathcal{U}_{ad} = \{ u \text{ is an } U\text{-valued random variable, } F_s \text{-measurable such as } \lim_{t \to \infty} \frac{1}{t-s}E \int_s^t \| u(\sigma) \|^2 d\sigma < \infty \text{ and } \sup_{t \geq s} \| y(t) \|^2 < \infty, \text{ where } y(t) \text{ is the solution of (1)} \} \}.

**Theorem 5.1** Assume that the hypotheses of Theorem 3.2 hold. Let \( P \) be the unique and bounded on \( \mathbb{R}_+ \) solution of the Riccati equation (3) and \( g(t) \) be given by (8), where \( F(t) = -K^{-1}(t)B^*(t)P(t) \) and \( h(t) = C^*(t)C(t)r(t) \). Then the optimal control is

\[
u(\sigma) = K^{-1}(\sigma)B^*(\sigma)[g(\sigma) - P(\sigma)y(\sigma)]
\] and the optimal cost is

\[
J(s) = \inf_{u \in \mathcal{U}_{ad}} J(s, u) = \\
\lim_{t \to \infty} \frac{1}{t-s} \left[ \int_s^t \| C(\sigma)r(\sigma) \|^2 d\sigma - \int_s^t \| K^{-1/2}B^*(\sigma)g(\sigma) \|^2 d\sigma \right].
\]
Proof. Let $P_n(s) = P_n(s, t_1; P(t_1))$ and let $g_n(s)$ be the solution of (6) with the final condition $g_n(t_1) = g(t_1)$, where $F_n(t) = -B(t)S_n(t)$, $S_n(t) = K^{-1}(t)B^*(t)P_n(t)$ and $h(t) = C^*(t)C(t)r(t)$. Consider the function $F_n(t, x) = \langle P_n(t)x, x \rangle - 2\langle g_n(t), x \rangle$, which is continuous together its partial derivatives $F_t, F_x, F_{xx}$ on $[0, \infty) \times H$, according to Lemmas 3.1 and 2.1. Let $u \in U_{ad}$ and let $y_n(t)$ be its response, where the approximate system associated with (1) is considered [3]. Using the Ito’s formula for $F_n(t, x)$ and $y_n(t)$ we get

$$E\{P_n(t_1)y_n(t_1), y_n(t_1)\} - 2E\{g_n(t_1), y_n(t_1)\} - \langle P(s)x, x \rangle + 2\langle g_n(s), x \rangle =$$

$$= -\int_{t_1}^t \|C(\sigma)[y_n(\sigma) - r(\sigma)]\|^2 + \langle K(\sigma)u(\sigma), u(\sigma) \rangle \, d\sigma +$$

$$E\int_{t_1}^t \left\| C^{1/2}\{S_n(\sigma)y_n(\sigma) + u(\sigma) - K^{-1}B^*(\sigma)g_n(\sigma)\} \right\|^2 \, d\sigma$$

$$+ \int_{t_1}^t \|C(\sigma)r(\sigma)\|^2 \, d\sigma - \int_{t_1}^t \left\| K^{1/2}B^*(\sigma)g_n(\sigma) \right\|^2 \, d\sigma$$

$$+ 2E\int_{t_1}^t \langle (S_n(\sigma) - S(\sigma))y_n(\sigma), B^*g_n(\sigma) \rangle \, d\sigma.$$  

Passing to the limit as $n \to \infty$ we obtain

$$E\{P(t_1)y(t_1), y(t_1)\} - 2E\{g(t_1), y(t_1)\} - \langle P(s)x, x \rangle + 2\langle g(s), x \rangle =$$

$$= -\int_{s}^{t_1} \|C(\sigma)[y(\sigma) - r(\sigma)]\|^2 + \langle K(\sigma)u(\sigma), u(\sigma) \rangle \, d\sigma +$$

$$E\int_{s}^{t_1} \left\| C^{1/2}\{S(\sigma)y(\sigma) + u(\sigma) - K^{-1}B^*(\sigma)g(\sigma)\} \right\|^2 \, d\sigma$$

$$+ \int_{s}^{t_1} \|C(\sigma)r(\sigma)\|^2 \, d\sigma - \int_{s}^{t_1} \left\| K^{1/2}B^*(\sigma)g(\sigma) \right\|^2 \, d\sigma.$$  

Since $P(t)$ and $g(t)$ are bounded on $\mathbf{R}_+$, we multiply the last relation by $\frac{1}{t_1-s}$ and taking the limit as $t_1 \to \infty$ and, then, the infimum we get the conclusion.

$\square$

References


AN INVESTIGATION OF MATHEMATICAL MODELS OF A PIPELINE - PRESSURE SENSOR MECHANICAL SYSTEM

Petr A. Velmisov, Yuliya V. Pokladova
Ulyanovsk State Technical University, Russia
velmisov@ulstu.ru

Abstract
The problem of the dynamics of an elastic element of a sensor, that is the component of the pipeline - pressure sensor mechanical system is studied. Similar problems, but with other position of a sensor (fig.1), are considered in [1]-[3].

Keywords: elastic system, finite difference methods
2000 MSC: 774S20

Consider three system in fig.1.

Fig. 1.

For each of these heuristic models the governing equation was obtained by connecting the pressure law of the working medium on the input of the pipeline and the deflection function of the plate $AB$.

The equation for the first model (fig.1) has the form

$$L(\omega) = P_0(y, t) - \frac{1}{y_0} \int_0^{y_0} P(y, t)dy -$$

$$- \frac{2\rho}{y_0} \sum_{n=1}^{\infty} \cos(\lambda_n y) \frac{\tanh(\lambda_n x_0)}{\lambda_n} \int_a^b \dot{\omega}(y, t) \cos(\lambda_n y)dy -$$

219
\[
-\frac{2}{y_0} \sum_{n=1}^{\infty} \frac{\cos(\lambda_n y)}{\cosh(\lambda_n x_0)} \int_{0}^{y_0} P(y, t) \cos(\lambda_n y) dy - \frac{\rho x_0}{y_0} \int_{a}^{b} \ddot{\omega}(y, t) dy.
\]

The equation for the second model (fig. 1) reads

\[
L(\omega) = -P_0(x, t) + \frac{1}{y_0} \int_{0}^{y_0} P(y, t) dy - 2\rho \sum_{n=1}^{\infty} \frac{\cos(\lambda_n x)}{\lambda_n} \int_{a}^{b} \dot{\omega}(x, t) \cos(\lambda_n x) dx - 2\rho \sum_{n=1}^{\infty} \frac{\cosh(\nu_n x)}{\cosh(\nu_n x_0)} \int_{0}^{y_0} P(y, t) \cos(\nu_n y) dy.
\]

For the third model (fig. 1) the mentioned relationship is

\[
L(\omega) = P_0(y, t) - P(t) + \frac{\rho}{\pi} \int_{a}^{b} \ddot{\omega}(\tau, t) \ln \left| \sin \frac{\pi \tau}{y_0} - \sin \frac{\pi y}{y_0} \right| d\tau.
\]

The mathematical model represented in fig. 2 differs from the model in fig. 1, by another position of a sensor. Further, this problem is solved in the linear approximation, relevant for small deformations of an elastic element, small perturbations of velocity potential for the working medium in the domain \( G \) and small deflections of the plate \( AB \) as the composite element of the sensor.

The working medium is an ideal incompressible liquid or gas. Velocity field is supposed to be planar, the length of a pipeline - infinite. The governing equation follows by, connecting the deflection function \( \omega(x, t) \) of the plate \( AB \) (fig. 2) and the law \( P(t) \) at the pressure change of the working medium on the input of the pipeline \( (x = +\infty) \).

Let \( \varphi(x, y, t) \) be the velocity potential of the working medium \( (t - \text{time}) \); \( \omega(x, t) \) is the deflection (the deformation) of the elastic plate \( AB \). The lin-
earized equation and the boundary conditions satisfied by these functions read

\[ \varphi_{xx} + \varphi_{yy} = 0, \quad (x, y) : 0 < x < +\infty, 0 < y < y_0; \] (1)

\[ \varphi_y(x, 0, t) = 0, \quad x \in (0, +\infty); \] (2)

\[ \varphi_x(0, y, t) = 0, \quad y \in (0, y_0); \] (3)

\[ \varphi_y(x, y_0, t) = 0, \quad x \in (0, a) \cup (b, +\infty); \] (4)

\[ \varphi_y(x, y_0, t) = \dot{\varphi}(x, y_0, t), \quad x \in (a, b); \] (5)

\[ \lim_{x \to +\infty} (\varphi_x^2 + \varphi_y^2) = 0; \] (6)

\[ \lim_{x \to +\infty} (P_* - \rho \varphi_t(x_0, y_0, t)) = P(t); \] (7)

\[ L(w) \equiv M \ddot{w} + D \dddot{w}'' + N \dddot{w}' + \alpha \dddot{w}''' + \beta \dddot{w} + \gamma w = \]
\[ = P_0(x, t) - \rho \varphi_t(x, y_0, t), \quad x \in (a, b). \] (8)

Here \( P(t) \) is the expression of the pressure change of a working medium on the input of the pipeline; \( P_0(x, t) \) is the distributed external load; \( P_* \) is the pressure of a working medium in the pipeline at the rest state; \( \rho \) is the density of a working medium; \( M \) is the mass of a unit length (linear density); \( D \) is the deflection rigidity; \( N \) is the compressing (stretching) stress; \( \alpha, \beta \) are coefficients of internal and external damping; \( \gamma \) is the coefficient of rigidity of the base.

By the complex variable functions theory the solving of this problem can be reduced to the investigation of the equation for deformation of an elastic element \[ \text{[4]} \]

\[ L(\omega) = P(t) - P_0(x, t) + \frac{\rho}{\pi} \int_a^b \ddot{\omega}(\tau, t) \ln \left| \cosh \frac{\pi \tau}{y_0} - \cosh \frac{\pi x}{y_0} \right| d\tau. \] (9)

We shall apply the Galerkin method for approximations of the m-th order

\[ \omega(x, t) = \sum_{k=1}^m \omega_k(t) \sin \beta_k(x - a), \quad \beta_k = \frac{\pi k}{b - a}. \]

From equation (8) we have

\[ L_k(\omega) \equiv M \left( \sum_{k=1}^m \ddot{\omega}_k(t) \sin \beta_k(x - a) \right) + D \left( \sum_{k=1}^m \beta_k^4 \omega_k(t) \sin \beta_k(x - a) \right) - \]

\[ -N \left( \sum_{k=1}^m \beta_k^2 \ddot{\omega}_k(t) \sin \beta_k(x - a) \right) + \alpha \left( \sum_{k=1}^m \beta_k^4 \dot{\omega}_k(t) \sin \beta_k(x - a) \right) + \]
\[ + \beta \left( \sum_{k=1}^{m} \omega_k(t) \sin \beta_k(x-a) \right) + \gamma \left( \sum_{k=1}^{m} \omega_k(t) \sin \beta_k(x-a) \right) + P_0(x,t) - P(t) - \frac{\rho}{\pi} \sum_{k=1}^{m} \dot{\omega}_k(t) I_k(x), \]

where \( I_k(x) = \int_a^b \sin \beta_k(\tau-a) \ln | \cosh \frac{\pi \tau}{y_0} - \cosh \frac{\pi x}{y_0} | \, d\tau \). Moreover, we must have

\[ \int_a^b L_*(\omega) \sin \beta_i(x-a) \, dy = 0, \quad i = 1, \ldots, m. \]

Then \( \omega_k(t) \) satisfy the system of \( m \) ordinary differential equations (\( i = 1, \ldots, m \))

\[ \sum_{k=1}^{m} A_{ik} \ddot{\omega}_k(t) + B_i \dot{\omega}_i(t) + C_i \omega_i(t) + F_i(t) = 0, \quad (10) \]

where

\[ A_{ik} = \begin{cases} M \frac{b-a}{2} - \frac{b}{a} \int_a^b I_k(x) \sin \beta_k(x-a) \, dx, & i = k \\ -\frac{\beta}{\pi} \int_a^b I_k(x) \sin \beta_k(x-a) \, dx, & i \neq k. \end{cases} \]

\[ B_i = \frac{b-a}{2} (\alpha \beta_i^4 + \beta), \quad C_i = \frac{b-a}{2} (D \beta_i^4 - N \beta_i^2 + \gamma), \]

\[ F_i(t) = -\int_a^b P_0(x,t) \sin \beta_i(x-a) \, dx - P(t) \frac{1 - (-1)^i}{\beta_i}. \]

In order to find \( \omega_k(0), \dot{\omega}_k(0) \) the initial conditions \( w(x,0) = u(x), \dot{w}(x,0) = v(x) \) are used. Compose the discrepancies

\[ R_1(\omega_k(0), x) = \sum_{k=1}^{m} \omega_k(0) \sin \beta_k(x-a) - u(x), \]

\[ R_2(\dot{\omega}_k(0), x) = \sum_{k=1}^{m} \dot{\omega}_k(0) \sin \beta_k(x-a) - v(x). \]

The initial conditions \( \omega_k(0), \dot{\omega}_k(0) \) can be obtained from the orthogonality conditions (\( i = 1, \ldots, m \))

\[ \int_a^b R_1(\omega_k(0), x) \sin \beta_i(x-a) \, dx = 0, \int_a^b R_2(\dot{\omega}_k(0), x) \sin \beta_i(x-a) \, dx = 0. \]

The system (10) in the matrix form reads \( A \ddot{\omega} + B \dot{\omega} + C \omega + F(t) = 0, \)
or, equivalently, $\ddot{\omega} = -A^{-1}B\dot{\omega} - A^{-1}C\omega - A^{-1}F(t)$, and reduce it to the normal form

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -A^{-1}Cy - A^{-1}Dx - A^{-1}F(t), \end{cases}$$

where $x = w$, $y = \dot{w}$. Then applying some difference method (the Euler-Cauchy and Runge-Kutta methods) the solution of the system can be found.

Assume now that the liquid in the pipeline be water and the elastic element be an aluminium plate.

**Example 1.** Let $P_0(x,t) = 0$, $P(t) = 10^5(1 + \sin(t))$, $w(x,0) = 0$, $\dot{w}(x,0) = 0$. Then using the mathematical system MathCAD 2000 for the values of parameters $m = 2$, $a = 0.1$, $b = 0.25$, $y_0 = 0.5$, $M = 42.4$, $D = 806.7$, $N = 10^3$, $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$, $\rho = 10^3$, we obtain the graph of the function $\omega(x,t) = \omega_1(t)\sin\beta_1(x-a) + \omega_2(t)\sin\beta_2(x-a)$ at $x = \frac{a+b}{2}$ (fig.3).

![Fig. 3.](image1)

**Example 2.** For $P(t) = 10^5\cos(t)$ and the above particular values of parameters, by means of the MathCAD 2000, we obtain the graph of the function $\omega(x,t) = \omega_1(t)\sin\beta_1(x-a) + \omega_2(t)\sin\beta_2(x-a)$ at $x = \frac{a+b}{2}$ (fig.4).

![Fig. 4.](image2)
Example 3. For \( P(t) = 10^5 e^{-10t} \) and the above values of parameters, by using MathCAD 2000, we obtain the graph of the function \( \omega(x, t) = \omega_1(t) \sin \beta_1(x - a) + \omega_2(t) \sin \beta_2(x - a) \) at \( x = \frac{a+b}{2} \) (fig. 5).

Example 4. Let \( P_0(x, t) = 0, \ P(t) = 10^5 e^{-100t}, \ w(x, 0) = 0, \ \dot{w}(x, 0) = 0. \) Then, by MathCAD 2000, for the values of parameters \( m = 2, \ a = 0, \ b = 0.2, \ y_0 = 1, \ M = 42.4, \ D = 806.7, \ N = 10^5, \ \alpha = 0.5, \ \beta = 0.5, \ \gamma = 0.5, \ \rho = 10^3 \), we obtain the graph of the function \( \omega(x, t) = \omega_1(t) \sin \beta_1(x-a) + \omega_2(t) \sin \beta_2(x-a) \) at \( x = \frac{a+b}{2} \) (fig. 6).

References

EXTENDING OBJECT-ORIENTED DATABASES FOR FUZZY INFORMATION MODELING

Cristina-Maria Vladarean
S.C. WATERS Romania S.R.L.
cristina.vladarean@waters.com

Abstract In this paper, based on the possibility distribution and semantic measure of fuzzy data, an extended object-oriented database model to handle imperfect as well as complex object in the real-world is introduced. Some major notions in object-oriented databases such as objects, classes, object-classes, relationships, subclass/superclass and multiple inheritances are extended under fuzzy information environment. A generic model for fuzzy object-oriented databases and some operations are presented.

Keywords: fuzzy data; object-oriented databases; semantic measure; fuzzy object-oriented database model.

1. INTRODUCTION

A major goal for database research has been the corporation of additional semantics into the data model. In real-world applications, information is often vague or ambiguous. Therefore, different kinds of incomplete information have been extensively introduced into relational databases. However, classical relational database model in its extension of imprecision and uncertainty does not satisfy the need of modeling complex objects with imprecision and uncertainly. So many researches have been concentrated on the development of some database models to deal with complex objects and uncertain data together.

For practical needs, fuzzy object-oriented databases were developed. Some major notions in object-oriented databases such as objects, classes, object-classes relationships, subclass/superclass, and multiple inheritances are extended under fuzzy information environment. Such a generic model is presented in this paper.

2. BASIC KNOWLEDGE

Fuzzy set and possibility distribution. Fuzzy data, originally described by Zadeh, are defined as follows. Let $U$ be a universe of discourse. Then a fuzzy value on $U$ is characterized by a fuzzy set $F$ in $U$. A membership function $\mu_F : U \rightarrow [0, 1]$ is defined for the fuzzy set $F$, where $\mu_F(u)$, for
each \( u \in U \), denotes the degree of membership of \( u \) in the fuzzy set \( F \). Thus, the fuzzy set \( F \) is described as follows:

\[
F = \{ \mu(u_1)/u_1, \mu(u_2)/u_2, \ldots, \mu(u_n)/u_n \}.
\]

When the \( \mu_F(u) \) above is explained to be a measure of possibility that a variable \( X \) has the value \( u \) in this approach, where \( X \) takes values in \( U \), a fuzzy value is described by a possibility distribution \( \pi_X \). Let \( \pi_X \) and \( F \) be the possibility distribution representation and the fuzzy set representation for a fuzzy value respectively. It is apparent that \( \pi_X = F \) is true.

In addition, a fuzzy data is represented by similarity relations in domain elements, in which the fuzziness comes from the similarity relations between two values in a universe of discourse, not from the status of an object itself. Similarity relations are thus used to describe the degree similarity of two values from the same universe of discourse. A similarity relation \( \text{Sim} \) on the universe of discourse \( U \) is a mapping:

\[
U \times U \rightarrow [0, 1]
\]

such that

(a) for \( \forall x \in U \), \( \text{Sim}(x, x) = 1 \) (reflexivity),

(b) for \( \forall x, y \in U \), \( \text{Sim}(x, y) = \text{Sim}(y, x) \) (symmetry),

(c) for \( \forall x, y, z \in U \), \( \text{Sim}(x, z) \geq \max_y(\min(\text{Sim}(x, y), \text{Sim}(y, z))) \) (transitivity).

Semantic measure of fuzzy data. The semantics of fuzzy data represented by possibility distribution corresponds to the semantic space and the semantic relationship between two fuzzy data can be described by the relationship between their semantic spaces. The semantic inclusion degree is then employed to measure semantic inclusion and thus measure semantic equivalence of fuzzy data.

Definition. Let \( \pi_A \) and \( \pi_B \) be two fuzzy data, and let their semantic spaces be \( SS(\pi_A) \) and \( SS(\pi_B) \), respectively. Let \( \text{SID}(\pi_A, \pi_B) \) denote the degree that \( \pi_A \) semantically includes \( \pi_B \). Then

\[
\text{SID}(\pi_A, \pi_B) = \frac{SS(\pi_B) \cap SS(\pi_A)}{SS(\pi_B)}.
\]

For two fuzzy data \( \pi_A \) and \( \pi_B \), the meaning of \( \text{SID}(\pi_A, \pi_B) \) is the percentage of the semantic space of \( \pi_B \) which is wholly included in the semantic space of \( \pi_A \).

Definition. Let \( \pi_A \) and \( \pi_B \) be two fuzzy data and \( \text{SID}(\pi_A, \pi_B) \) be the degree that \( \pi_A \) semantically includes \( \pi_B \). Let \( \text{SE}(\pi_A, \pi_B) \) denote the degree that \( \pi_A \) and \( \pi_B \) are equivalent to each other. Then

\[
\text{SE}(\pi_A, \pi_B) = \min(\text{SID}(\pi_A, \pi_B), \text{SID}(\pi_B, \pi_A)).
\]

Definition. Let \( U = u_1, u_2, \ldots, u_n \) be the universe of discourse. Let \( \pi_A \) and \( \pi_B \) be two fuzzy data on \( U \) based on possibility distribution and \( \pi_A(u_i), u_i \in U \), denote the possibility that \( u_i \) is true. Let \( \text{Res} \) be a resamblance relation on the domain \( U \), let \( \alpha \), for \( 0 \leq \alpha \leq 1 \), be a threshold corresponding to \( \text{Res} \). Then \( \text{SID}(\pi_A, \pi_B) \) is defined by
The notion of the semantic equivalence degree of attribute values can be extended to the semantic equivalence degree of tuples. Let \( t_i = < a_{i1}, a_{i2}, \ldots, a_{in} > \) and \( t_j = < a_{j1}, a_{j2}, \ldots, a_{jn} > \) be two tuples in fuzzy relational instance \( r \) over the schema \( R(A_1, A_2, \ldots, A_n) \). The semantic equivalence degree of tuples \( t_i \) and \( t_j \) is denoted by 
\[
SE(t_i, t_j) = \min \left( SE(t_i[A_1], t_j[A_1]), SE(t_i[A_2], t_j[A_2]), \ldots, SE(t_i[A_n], t_j[A_n]) \right).
\]

3. FUZZY OBJECTS AND CLASSES

The objects model real-world entities or abstract concepts. The objects have the properties of being attributes of the object itself or relationships also known as associations between the object and one or more other objects. An object is fuzzy because of a lack of information. Formally, objects that have at least one attribute whose value is a fuzzy set are fuzzy objects.

The objects having the same properties are gathered into classes that are organized into hierarchies. Theoretically, a class can be considered from two different viewpoints: (a) an extensional class, where the class is defined by the list of its object instances, and (b) an intensional class, where the class is defined by a set of attributes and their admissible values.

In addition, a subclass defined from its superclass by means of inheritance mechanism in the OODB can be seen as the special case of (b) above. Therefore, a class is fuzzy because of the following several reasons. First, some objects are fuzzy, which have similar properties. A class defined by these objects may be fuzzy. These objects belong to the class with the membership degree of \([0, 1]\). Second, when a class is intentionally defined, the domain of an attribute may be fuzzy and a fuzzy class is formed. For example, a class Old equipment is a fuzzy one because the domain of its attribute Using period is a set of fuzzy values such as long, very long and about 20 years. Third, the subclass produced by a fuzzy class by means of specializations and the superclass produced by some classes (in which there is at least one class who is fuzzy) by means of generalizations are also fuzzy.

The main difference between fuzzy classes and crisp classes is that the boundaries of fuzzy classes are imprecise. The imprecision in the class boundaries is caused by the imprecision of the values in the attribute domain. In the fuzzy OODB, classes are fuzzy because their attribute domain are fuzzy. The issue that an object fuzzily belongs to a class occurs since a class or an object is fuzzy. Similarly, a class is a subclass of another class with membership degree of \([0, 1]\) because of the class fuzziness. In the OODB, the above mentioned relationships are certain. Therefore, the evaluations of fuzzy object-class relationships and fuzzy inheritance hierarchies are the core of in-
formation modeling in the fuzzy OODB. In the following discussion, let us assume that the fuzzy attribute values of fuzzy objects and the fuzzy values in fuzzy attribute domain are represented by possibility distribution.

**Fuzzy object-class relationships.** In the fuzzy OODB, the following four situations can be distinguished for object-class relationships.

(a) **Crisp class and crisp object:** this situation is the same as the OODB, where the object belongs or not to the class certainly (such as for example the objects Car and Computer are for a class Vehicle, respectively).

(b) **Crisp class and fuzzy object:** although the class is precisely defined and has the precise boundary, an object is fuzzy since its attribute value(s) may be fuzzy. In this situation, the object may be related to the class with the special degree in $[0, 1]$ (such as for example the object whose position attribute may be graduate, research assistant, pr research assistant professor is for the class Faculty).

(c) **Fuzzy class and crisp object:** being the same as the case in (b), the object may belong to the class with the membership degree $[0, 1]$ (such as for example a Ph.D. student is for the Young student class).

(d) **Fuzzy class and fuzzy object:** in this situation, the object also belongs to the class with the membership degree in $[0, 1]$.

The object-class relationships in (b)-(d) above are called fuzzy object-class relationships. In fact, the situation in (a) can be seen as the special case of fuzzy object-class relationships, where the membership degree of the object to the class is one. It is clear that estimating the membership of an object to the class is crucial for the fuzzy object-class relationship when class is instantiated.

In the OODB, determining if an object belongs to a class depends on the fact whether its attribute values are respectively included in the corresponding attribute domains of the class. Similarly, in order to calculate the membership degree of an object to the class in a fuzzy object-class relationship, it is necessary to evaluate the degrees that the attribute domains of the class include the attribute values of the object. However, it should be noted that in a fuzzy object-class relationship, only the inclusion degree of object values with respect to the class domains is not accurate for the evaluation of the membership degree of an object to the class. The attributes play different roles in the definition and identification of a class. Some may be dominant and some not. Therefore, a weight $w_i$ is assigned to each attribute of the class according to its importance to the designer. Then the membership degree of an object to the class in a fuzzy object-class relationship should be calculated using the inclusion degree of object values with respect to the class domains and the weight of attributes.

Let $C$ be a class with attributes $\{A_1, A_2, \ldots, A_n\}$, $o$ be an object on attribute set $\{A_1, A_2, \ldots, A_n\}$, and let $o(A_i)$ denote the attribute value of $o$ on $A_i (1 \leq i \leq n)$. In $C$, each attribute $A_i$ is connected with a domain de-
noted $dom(A_1)$. The inclusion degree of $o(A_i)$ with respect to $dom(A_i)$ is denoted $ID(dom(A_i), o(A_i))$. In the following, let us investigate the evaluation of $ID(dom(A_i), o(A_i))$. As we know, $dom(A_i)$ is a set of crisp values in the OODB and may be a set of fuzzy subsets in fuzzy databases. Therefore, in a uniform OODB for crisp and fuzzy information modeling, $dom(A_i)$ should be the union of these two components, $dom(A_i) = cdom(A_i) \cup fdom(A_i)$, where $cdom(A_i)$ and $fdom(A_i)$, respectively, denote the sets of crisp values and fuzzy subsets. On the other hand, $o(A_i)$ may be a crisp value or a fuzzy value. The following cases may be identified for evaluating $ID(dom(A_i), o(A_i))$:

Case 1: $o(A_i)$ is a fuzzy value. Let $fdom(A_i) = \{f_1, f_2, \ldots, f_m\}$, where $f_j(1 \leq j \leq m)$ is a fuzzy value, and $cdom(A_i) = \{c_1, c_2, \ldots, c_k\}$, where $c_l(1 \leq l \leq k)$ is a crisp value. Then

$$ID(dom(A_i), o(A_i)) = \max(ID(cdom(A_i), o(A_i)), ID(fdom(A_i), o(A_i))) = \max(SID(f_1, \frac{1}{c_1}, \frac{1}{c_2}, \ldots, \frac{1}{c_k}), o(A_i)), \max_j(SID(f_j), o(A_i))),$$

where $SID(x, y)$ is used to calculate the degree that fuzzy value $x$ includes fuzzy value $y$.

Case 2: $o(A_i)$ is a crisp value. Then,

$$ID(dom(A_i), o(A_i)) = 1 \text{ if } o(A_i) \in cdom(A_i), \text{ else } ID(dom(A_i), o(A_i)) = ID(fdom(A_i), \{\frac{1}{o(A_i)}\}).$$

Now, let us define the formula to calculate the membership degree of the object $o$ to the class $C$ as follows, where $w(A_i(C))$ denotes the weight of attribute $A_i$ to class $C$:

$$\mu_C(o) = \frac{\sum_{i=1}^{m} ID(dom(A_i), o(A_i)) \times w(A_i(C))}{\sum_{i=1}^{m} w(A_i(C))}$$

In the above-given determination that an object belongs to a class fuzzily, it is assumed that the object and the class have the same attributes, namely class $C$ is with attributes $\{A_1, A_2, \ldots, A_n\}$ and object $o$ is also on $\{A_1, A_2, \ldots, A_n\}$. Such an object-class relationship is called direct object-class relationship. As we know, there exist subclass/superclass relationships in the OODB, where the subclass inherits some attributes and methods of the superclass, and defines some new attributes and methods. Any object belonging to the subclass must belong to the superclass since a subclass is the specialization of the superclass. So we have one kind of special object-class relationship: the relationship between superclass and the objects of the subclass. Such an object-class relationship is called indirect object-class relationship. Since the object and the class in indirect object-class relationship have different attributes, in the following we show how to calculate the membership degree of an object to the class in an indirect object-class relationship.

Let $C$ be a class with attributes $\{A_1, A_2, \ldots, A_k, A_{k+1}, \ldots, A_m\}$ and $o$ be an object on attributes $\{A_1, A_2, \ldots, A_k, A_{k+1}', \ldots, A_m', A_{m+1}, \ldots, A_n\}$. Here attributes $A_{k+1}', \ldots, A_m'$ are overridden from $A_{k+1}, \ldots, A_m$ and attributes $A_{m+1}, \ldots, A_n$ are special. Then we have
are in a direct object-class relationship. Then class a direct object-class relationship. Then class can only belong to it uniquely (except for the case of subclass/superclass).

or may not belong to a given class definitely. If it belongs to a given class, it only arises in fuzzy object-oriented databases. In the OODB, an object may two classes with different membership degrees simultaneously. This situation class/superclass relationship, it is possible that an object may belong to these degrees. This situation occurs in fuzzy inheritance hierarchies, which will be simultaneously belongs to the subclass and superclass with different membership degrees. Let

\[ \mu_C(o) = \frac{\sum_{i=1}^{k} ID(\text{dom}(A_i), o(A_i)) \times w(A_i(C)) + \sum_{j=1}^{m} ID(\text{dom}(A_j), o(A'_j)) \times w(A_j(C))}{\sum_{i=1}^{k} w(A_i(C))} \]

Based on the direct object-class relationship and the indirect object-class relationship, now we focus on arbitrary object-class relationship. Let \( C \) be a class with attributes \( \{A_1, A_2, \ldots, A_k, A_{k+1}, \ldots, A_m, A_{m+1}, \ldots, A_n\} \) and \( o \) be an object on attributes \( \{A_1, A_2, \ldots, A_k, A'_{k+1}, \ldots, A'_m, B_{m+1}, \ldots, B_p\} \). Here the attributes \( A'_{k+1}, \ldots, A'_m \) are overridden from \( A_{k+1}, \ldots, A_m, A_{m+1}, \ldots, A_n \). Attributes \( A_{m+1}, \ldots, A_n \) and \( B_{m+1}, \ldots, B_p \) are special. Then we have

\[ \mu_C(o) = \frac{\sum_{i=1}^{k} ID(\text{dom}(A_i), o(A_i)) \times w(A_i(C)) + \sum_{j=1}^{m} ID(\text{dom}(A_j), o(A'_j)) \times w(A_j(C))}{\sum_{i=1}^{k} w(A_i(C))} \]

Since an object may belong to a class with membership degree in \([0,1]\) in fuzzy object-class relationship, it is possible that an object that is in a direct object-class relationship and an indirect object-class relationship simultaneously belongs to the subclass and superclass with different membership degrees. This situation occurs in fuzzy inheritance hierarchies, which will be investigated in the next section. Also, for two classes that do not have subclass/superclass relationship, it is possible that an object may belong to these two classes with different membership degrees simultaneously. This situation only arises in fuzzy object-oriented databases. In the OODB, an object may or may not belong to a given class definitely. If it belongs to a given class, it can only belong to it uniquely (except for the case of subclass/superclass).

The situation where an object belongs to different classes with different membership degrees simultaneously in fuzzy object-class relationships is called multiple membership of object in this paper. Let us focus on how to handle the multiple membership of object in fuzzy object-class relationships. Let \( C_1 \) and \( C_2 \) be (fuzzy) classes and \( \alpha \) be a given threshold. Assume there exists an object \( o \). If \( \mu_{C_1}(o) \geq \alpha \) and \( \mu_{C_2}(o) \geq \alpha \), the conflict of the multiple membership occurs, namely, \( o \) belongs to multiple classes simultaneously. At this moment, which one in \( C_1 \) and \( C_2 \) is the class of object \( o \) depends on the following cases.

**Case 1:** There exists a direct object-class relationship between object \( o \) and one class in \( C_1 \) and \( C_2 \). Then the class in the direct object-class relationship is the class of the object \( o \).

**Case 2:** There is no direct object-class relationship but only an indirect object-class relationship between object \( o \) and one class in \( C_1 \) and \( C_2 \), say \( C_1 \). And there exists such a subclass \( C'_1 \) of \( C_1 \) that object \( o \) and \( C'_1 \) are in a direct object-class relationship. Then class \( C'_1 \) is the class of object \( o \).

**Case 3:** There is neither direct object-class relationship nor indirect object-class relationship between object \( o \) and classes \( C_1 \) and \( C_2 \). Or there exists only an indirect object-class relationship between object \( o \) and one class in \( C_1 \) and \( C_2 \), say \( C_1 \), but there is no such subclass \( C'_1 \) of \( C_1 \) that object \( o \) and \( C'_1 \) are in a direct object-class relationship. Then class \( C_1 \) is considered as the
class of object $o$ if $\mu_{C1}(o) > \mu_{C2}(o)$, else class $C2$ is considered as the class of object $o$.

It can be seen that in Cases 1 and 2, the class in direct object-class relationship is always the class of object $o$ and the object and the class have the same attributes. In Case 3, however, object $o$ and the class that is considered as the class of object $o$, say $C1$, have different attributes. It should be pointed out that class $C1$ and the object $o$ are definitely defined, respectively, viewed from their structures. For the situation in Case 3, the attributes of $C1$ do not affect the attributes of $o$ and vice-versa. There should be a class $C$, and $C$ and $o$ are in direct object-class relationship. But class $C1$ is more similar to $C$ than $C2$. Class $C$ is the class of object $o$ once $C$ is available.

4. FUZZY INHERITANCE HIERARCHIES

In the OODB, a new class, called subclass, is produced from another class, called superclass by means of inheriting some attributes and methods of the superclass, overriding some attributes and methods of superclass, and defining some new attributes and methods. Since a subclass is the specialization of the superclass, any one object belonging to the subclass must belong to the superclass. This characteristic can be used to determine if two classes have subclass/superclass relationship.

In the fuzzy OODB, however, classes may be fuzzy. A class produced from a fuzzy class must be fuzzy. If the former is still called subclass and the later superclass, the subclass/superclass relationship is fuzzy. In other words, a class is a subclass of another class with membership degree of $[0, 1]$ at this moment. Correspondingly, the method used in the OODB for determination of subclass/superclass relationship is modified as

(a) for (any) fuzzy object, if the member degree that it belongs to the subclass is less than or equal to the member degree that it belongs to the superclass, and

(b) the member degree that it belongs to the subclass is greater than or equal to the given threshold.

The subclass is then a subclass of the superclass with the membership degree, which is the minimum in the membership degrees to which these objects belong to the subclass.

Let $C1$ and $C2$ be (fuzzy) classes and $\beta$ be a given threshold. We say $C2$ is a subclass of $C1$ if $(\forall o) (\beta \leq \mu_{C2}(o) \leq \mu_{C1}(o))$. The membership degree that $C2$ is a subclass of $C1$ should be $\min_{\mu_{C2}(o) \geq \beta}(\mu_{C2}(o))$.

It can be seen that by utilizing the inclusion degree of objects to the class, we can assess fuzzy subclass/superclass relationship in the fuzzy OODB. It is clear that such assessment is indirect. If there is no object available, this
method is not used. In fact, the idea used in evaluating the membership degree of an object to a class can be used to determine the relationship between fuzzy subclass and superclass. We can calculate the inclusion degree of a (fuzzy) subclass with respect to the (fuzzy) superclass according to the inclusion degree of the attribute domains of the subclass with respect to the attribute domains of the superclass as well as the weight of attributes. In the following, a method for evaluating the inclusion degree of fuzzy attribute domains is given.

Let $C_1$ and $C_2$ be (fuzzy) classes with attributes $\{A_1, A_2, \ldots, A_k, A_{k+1}, \ldots, A_m\}$ and $\{A_1, A_2, \ldots, A_k, A'_{k+1}, \ldots, A'_{m+1}, A_{m+1} \}$, respectively. It can be seen that in $C_2$, attributes $A_1, A_2, \ldots, A_k$ are directly inherited from $A_1, A_2, \ldots, A_k$ in $C_1$, attributes $A'_{k+1}, \ldots, A'_{m}$ are overridden from $A_{k+1}, \ldots, A_{m}$ in $C_1$, and attributes $A_{m+1}, \ldots, A_n$ are special. For each attribute in $C_1$ or $C_2$, say $A_i$, there is a domain, denoted $dom(A_i)$. As shown above, $dom(A_i)$ should be $dom(A_i) = ddom(A_i) \cup fdom(A_i)$, where $ddom(A_i)$ and $fdom(A_i)$ denote the sets of crisp values and fuzzy subsets, respectively. Let $A_i$ and $A_j$ be attributes of $C_1$ and $C_2$, respectively. The inclusion degree of $dom(A_j)$ with respect to $dom(A_i)$ is denoted by $ID(dom(A_i), dom(A_j))$.

Then we identify the following cases and investigate the evaluation of $ID(dom(A_i), dom(A_j))$:

(a) when $i \neq j$ and $1 \leq i, j \leq k$, $ID(dom(A_i), dom(A_j)) = 0$,
(b) when $i = j$ and $1 \leq i, j \leq k$, $ID(dom(A_i), dom(A_j)) = 1$, and
(c) when $i = j$ and $k + 1 \leq i, j \leq m$, then

$$ID(dom(A_i), dom(A_j)) = ID(dom(A_i), dom(A'_j)) = \max(ID(dom(A_i), ddom(A'_j)), ID(dom(A_i), fdom(A'_j))).$$

Now we respectively define $ID(dom(A_i), ddom(A'_j))$ and $ID(dom(A_i), fdom(A'_j))$. Let $ddom(A'_j) = \{f_1, f_2, \ldots, f_m\}$, where $f_j (1 \leq j \leq m)$ is a fuzzy value, and $cdom(A'_j) = \{c_1, c_2, \ldots, c_k\}$, where $c_l (1 \leq l \leq k)$ is a crisp value. We can consider $\{c_1, c_2, \ldots, c_k\}$ as a special fuzzy value $\{\frac{1}{c_1}, \frac{1}{c_2}, \ldots, \frac{1}{c_k}\}$. Then we have the following, which can be calculated by using the definition given above in this paper:

$$ID(dom(A_i), cdom(A'_j)) = ID(dom(A_i), \{\frac{1}{c_1}, \frac{1}{c_2}, \ldots, \frac{1}{c_k}\}),$$
$$ID(dom(A_i), fdom(A'_j)) = \max_j(ID(dom(A_i), f_j)).$$

Based on the inclusion degree of attribute domains of the subclass with respect to the attribute domains of its superclass as well as the weight of attributes, we can define the formula to calculate the degree to which a fuzzy class is a subclass of another fuzzy class. Let $C_1$ and $C_2$ be (fuzzy) classes with attributes $\{A_1, A_2, \ldots, A_k, A_{k+1}, \ldots, A_m\}$ and $\{A_1, A_2, \ldots, A_k, A'_{k+1}, \ldots, A'_{m}, A_{m+1}, \ldots, A_n\}$, respectively, and $w(A)$ denote the weight of attribute $A$. Then the degree that $C_2$ is a subclass of $C_1$, written $\mu(C_1, C_2)$, is defined as follows
\[ \mu(C_1, C_2) = \frac{\sum_{i=1}^{m} ID(\text{dom}(A_i(C_1)), \text{dom}(A_i(C_2))) \times w(A_i))}{\sum_{i=1}^{m} w(A_i)} \]

In subclass-superclass hierarchies, a critical issue is multiple inheritance of class. Ambiguity arises when more than one of the superclasses have common attributes and the subclass does not declare explicitly the class from which the attribute was inherited.

Let class \( C \) be a subclass of classes \( C_1 \) and \( C_2 \). Assume that the attribute \( A_i \) in \( C_1 \), denoted by \( A_i(C_1) \), is common to the attribute \( A_i \) in \( C_2 \), denoted by \( A_i(C_2) \). If \( \text{dom}(A_i(C_1)) \) and \( \text{dom}(A_i(C_2)) \) are identical, there does not exist a conflict in the multiple inheritance hierarchy and \( C \) inherits attribute \( A_i \) directly. If \( \text{dom}(A_i(C_1)) \) and \( \text{dom}(A_i(C_2)) \) are not identical, however, the conflict occurs. At this moment, which among \( A_i(C_1) \), \( A_i(C_2) \) is inherited by \( C \) depends on the following rule:

If \( ID(\text{dom}(A_i(C_1)), \text{dom}(A_i(C_2))) \times w(A_i(C_1)) > \)

\[ ID(\text{dom}(A_i(C_2)), \text{dom}(A_i(C_1))) \times w(A_i(C_2)), \]

then \( A_i(C_1) \) is inherited by \( C \), else \( A_i(C_2) \) is inherited by \( C \).

Note that in fuzzy multiple inheritance hierarchy, the subclass has different degrees with respect to different superclasses, not being the same as the situation in classical object-oriented database systems.

5. FUZZY OBJECT-ORIENTED DATABASE MODEL AND OPERATIONS

Based on the discussion above, we have known that the classes in the fuzzy OODB may be fuzzy. Accordingly, in the fuzzy OODB, an object belongs to a class with a membership degree of \([0, 1]\) and a class is the subclass of another class with degree of \([0, 1]\) too. In the OODB, the specification of a class includes the definition of ISA relationships, attributes and methods implementations. In order to specify a fuzzy class, some additional definitions are needed. First, the weights of attributes to the class must be given. In addition to these common attributes, a new attribute should be added into the class to indicate the membership degree to which an object belongs to the class. If the class is a fuzzy subclass, its superclasses and the degree that the class is the subclass of the superclasses should be illustrated in the specification of the class. Finally, in the definition of a fuzzy class, fuzzy attributes may be explicitly indicated.

Formally, the definition of a fuzzy class is shown as follows:

```
CLASS class name WITH DEGREE OF degree
INHERITES superclass_1 name WITH DEGREE OF degree_1

\vdots

INHERITES superclass_k name WITH DEGREE OF degree_k
```
ATTRIBUTES
  Attribute_1 name: [FUZZY] DOMAIN dom_1: TYPE OF type_1
  WITH DEGREE OF degree_1
  ...
  Attribute_m name: [FUZZY] DOMAIN dom_m: TYPE OF type_m
  WITH DEGREE OF degree_m
  Membership Attribute name: membership_degree

WEIGHT
  w(Attribute_1 name)=w_1
  ...
  w(Attribute_m name)=w_m

METHODS
  ...
END

For non-fuzzy attributes, the data types include simple types such as integer, real, Boolean, string, and complex types such as set type and object type. For fuzzy attributes, however, the data types are fuzzy-type-based on simple or complex types, which allow the representation of descriptive form of imprecise information. Only fuzzy attributes have fuzzy type and fuzzy attributes are explicitly indicated in a class definition. Therefore, in the definition above, we declare only the base type (e.g., integer) of fuzzy attributes and the fuzzy domain. A fuzzy domain is a set of possibility distributions or fuzzy linguistic terms (each fuzzy term is associated with a membership function).

The change in database model impacts the operations on the database model. In the following, let us describe some operations based on the proposed fuzzy class model above. First, we briefly discuss several combination operations of the fuzzy classes. Finally, we investigate the issue of user request-queries based on the fuzzy classes. Depending on the relationships between the attribute sets of the combining classes, three kinds of combination operations can be identified: fuzzy product (×), fuzzy join (∞) and fuzzy union (∪). Let C1 and C2 be fuzzy classes and let Attr(C1) and Attr(C2) be their attribute sets, respectively. Assume a new class C is created by combining C1 and C2. Then, C = C1 × C2, if Attr′(C1) ∩ Attr′(C2) = ∅;
C = C1 ∞ C2, if Attr′(C1) ∩ Attr′(C2) ≠ ∅ and Attr′(C1) = Attr′(C2),
or C = C1 ∪ C2, if Attr′(C1) = Attr′(C2).

Here, Attr′(C1) and Attr′(C2) are obtained from Attr(C1) and Attr(C2) through removing the membership degree attributes from Attr(C1) and Attr(C2), respectively. In the following, μC is used to represent the membership degree attribute of C. Assume we have an object o of C. Then μC(o) is used to represent the value of o on μC. For a common attribute in C, say Ai,
\( o(A_i) \) is used to represent the value of \( o \) on \( A_i \). If we have a set of such common attributes, say \( \{A_i, A_j, \ldots, A_m\} \), then \( o(\{A_i, A_j, \ldots, A_m\}) \) is used to represent the values of \( o \) on attributes in \( \{A_i, A_j, \ldots, A_m\} \). Furthermore, \( o(C) \) is used to represent all values of \( o \) on the common attributes in \( C \). In the following, the formal definitions of fuzzy product, fuzzy join, and fuzzy union operations are given.

**Fuzzy product:** The fuzzy product of \( C_1 \) and \( C_2 \) is a new class \( C \), which consists of the common attributes of \( C_1 \) and \( C_2 \) as well as a membership degree attribute. The objects of \( C \) are created by the composition of objects from \( C_1 \) and \( C_2 \):

\[
C = C_1 \times C_2 = \{ o(\forall a') (\forall a'') (a' \in C_1 \land a'' \in C_2)
\land o(Attr'(C_1)) = o'(C_1))
\land o(Attr''(C_2)) = o''(C_2))
\land \mu_C(o) = \omega(\mu_{C_1}(o'), \mu_{C_2}(o'')) \}
\]

Here, operation \( \omega \) is undefined. Generally, \( \omega(\mu_{C_1}(o'), \mu_{C_2}(o'')) \) may be:

- \( \min(\mu_{C_1}(o'), \mu_{C_2}(o''))\) or \( \mu_{C_1}(o') \times \mu_{C_2}(o'') \).

**Fuzzy join:** The fuzzy join of \( C_1 \) and \( C_2 \) is a new class \( C \), where \( Attr'(C_1) \cap Attr'(C_2) \neq \emptyset \) and \( Attr'(C_1) \neq Attr'(C_2) \). Class \( C \) is composed of \( Attr'(C_1) \cup (Attr'(C_2) - (Attr'(C_1) \cap Attr'(C_2))) \) as well as a membership degree attribute. The objects of \( C \) are created by the composition of objects from \( C_1 \) and \( C_2 \), which are semantically equivalent on \( Attr'(C_1) \cap Attr'(C_2) \) under the given thresholds. It should be noted that, however, \( Attr'(C_1) \cap Attr'(C_2) \neq \emptyset \) implies \( C_1 \) and \( C_2 \) have the same weights of attributes for the attributes in \( Attr'(C_1) \cap Attr'(C_2) \). This is an additional requirement to be met in the case of the fuzzy join operation. Let \( \alpha \) be the given threshold. Then,

\[
C = C_1 \bowtie C_2 = \{ o(\exists a') (\exists a'') (a' \in C_1 \land a'' \in C_2)
\land SE(o'(Attr'(C_1) \cap Attr'(C_2)), o''(Attr'(C_1) \cap Attr'(C_2))) \geq \alpha
\land o(Attr'(C_1)) = o'(C_1))
\land o(Attr'(C_2) - (Attr'(C_1) \cap Attr'(C_2)))
= o''(Attr'(C_2) - (Attr'(C_1) \cap Attr'(C_2)))
\land \mu_C(o) = \omega(\mu_{C_1}(o'), \mu_{C_2}(o'')) \}
\]

Here, operation \( \omega \) is also undefined. Generally, \( \omega(\mu_{C_1}(o'), \mu_{C_2}(o'')) \) may be:

- \( \min(\mu_{C_1}(o'), \mu_{C_2}(o'')) \) or \( \mu_{C_1}(o') \times \mu_{C_2}(o'') \).

**Fuzzy union:** The fuzzy union of \( C_1 \) and \( C_2 \) requires \( Attr'(C_1) = Attr'(C_2) \), which implies that all corresponding attributes in \( C_1 \) and \( C_2 \) have the same weights. Let a new class \( C \) be the fuzzy union of \( C_1 \) and \( C_2 \). Then the objects of \( C \) are composed of three kinds of objects. The first two kinds of objects are such objects that directly come from one component class (e.g., \( C_2 \)) under the given thresholds. The last kind of objects is such that the objects are the results of merging the redundant objects from two component classes under the given thresholds. Let \( \alpha \) be the given threshold.
\[ C = C_1 \cup C_2 = \{ o | (\forall o') (o' \in C_2 \land o \in C_1 \land SE(o(C_1), o'(C_2)) < \alpha) \lor (\forall o') (o' \in C_1 \land o \in C_2 \land SE(o(C_2), o'(C_1)) < \alpha) \lor ((\exists o') ((\exists o'') o' \in C_1 \land (\exists o'') (o' \in C_1 \land o'' \in C_2) \land SE(o(C_1) \land o''(C_2)) \geq \alpha \land o = merge(o', o'')) ) \} \]

Here, \textit{merge} is an operation for merging two redundant objects of the class to form a new object of the class. Let \( o' \) and \( o'' \) be two objects of class \( C \) and \( o = merge(o', o'') \). Then \( o(C) = o'(C) \) or \( o(C) = o''(C) \) and \( \mu_C(o) = \max(\mu_{C_1}(o'), \mu_{C_2}(o'')) \).

\textbf{Query processing} refers to such procedure that the objects satisfying a given condition are selected and then they are delivered to the user according to the required formats. These format requirements include which attributes appear in the result and if the result is grouped and ordered over the given attribute(s). So a query can be seen as comprising two components, namely a Boolean selection condition and some format requirements. As a simple illustration, some format requirements are ignored in the following discussion. An SQL (structured query language) like query syntax is represented as

\[
\text{SELECT} \langle \text{attribute list} \rangle \text{FROM} \langle \text{class names} \rangle \text{WHERE} \langle \text{query condition} \rangle,
\]

where \textit{attributelist} is the list of attributes separated by commas. At least one attribute name must be specified in this list. Attributes that take place in \textit{attributelist} are selected from the associated classes which are specified in the \textit{FROM} clause. \textit{classnames} contains the class names separated by commas, classes from which the attributes are selected with the \textit{SELECT} clause.

Classical databases suffer from a lack of flexibility to query. The given query condition and the contents of the database are all crisps. A query is flexible if the databases contain imprecise and uncertain information, and the query condition is imprecise and uncertain. For the fuzzy object-oriented databases, it has been shown above that objects belong to a given class with membership degree \([0, 1]\). In addition, an object satisfies the given query condition also with membership degree \([0, 1]\) because fuzzy information occurs in the query condition and/or the object. Therefore, the query processing based on the proposed fuzzy object-oriented database model refers to such procedure that the objects satisfying a given threshold and a given condition under given thresholds simultaneously are selected from the classes. It is clear that the queries for the fuzzy object-oriented databases are threshold-based ones, which are concerned with the number choices of threshold. Therefore, an SQL like query syntax based on the fuzzy object-oriented database model is represented as follows:

\[
\text{SELECT} \langle \text{attribute list} \rangle \text{FROM} \langle \text{Class}_1 \text{ WITH threshold}_1, \ldots, \text{Class}_m \text{ WITH threshold}_m \rangle \text{WHERE} \langle \text{query condition WITH threshold} \rangle.
\]
Here, \textit{(query condition)} is a fuzzy condition and all thresholds are crisp numbers in $[0, 1]$. Utilizing such SQL, one can get such objects that belong to the class under the given thresholds and also satisfy the query condition under the given thresholds at the same time. Note that the item \textit{WITH threshold} can be ommitted. The default of the threshold is exactly 1 for such a case.

6. CONCLUSION

Incorporation of imprecise and uncertain information in database model has been an important topic of database research because such an information extensively exists in data and knowledge intensive application such as expert system, decision making etc. Besides that, these systems are characterized by complex object structures. Classical relational database model and its extension of imprecision and uncertainty do not satisfy the need of handling complex objects with imprecision and uncertainty. Fuzzy object-oriented databases are hereby introduced.

In this paper, based on the possibility distribution and the semantic measure method of fuzzy data, a fuzzy object-oriented database model was presented to cope with imperfect as well as complex objects in the real-world at a logical level.

References


