S. Bogdanović, M. Ćirić and Ž. Perović

Algebra, Logic & Discrete Mathematics

UNIVERSITY OF NIŠ
FACULTY OF PHILOSOPHY
Department of Mathematics

FILOMAT

9:3 (1995)

International conference on

ALGEBRA, LOGIC & DISCRETE MATHEMATICS

Edited by: S. Bogdanović, M. Ćirić and Ž. Perović

УНИВЕРЗИТЕТ У НИШУ - ФИЛОЗОФСКИ ФАКУЛТЕТ -

NIŠ

FILOMAT

a continuation of
ZBORNIK RADOVA FILOZOFSKOG FAKULTETA U NIŠU
SERIJA MATEMATIKA

This publication was supported by the Faculty of Philosophy, University of Niš

Published by the Faculty of Philosophy, University of Niš

Printed by "Prosveta", Niš

UNIVERZITET U NIŠU
FILOZOFSKI FAKULTET

FILOMAT

9:3 (1995)

Internacionalna konferencija

ALGEBRA, LOGIC & DISCRETE MATHEMATICS

Urednici: S. Bogdanović, M. Ćirić and Ž. Perović

NIŠ, 1995

PREFACE

The International Conference on Algebra, Logic and Discrete Mathematics took place in Niš, April 14–16, 1995, and was organized by the Faculty of Economics Niš, Faculty of Philosophy Niš, and Mathematical Institute SANU Belgrade.

This book contains most of the papers presented during the Conference.

Editors

CONTENTS

Invited Lectures

Contributions

INVITED LECTURES

FILOMAT (Niš) 9:3 (1995), 385–426

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

THEORY OF GREATEST DECOMPOSITIONS OF SEMIGROUPS (A SURVEY)

Miroslav Ćirić and Stojan Bogdanović

Dedicated to Professor L. N. Shevrin on his 60th birthday

Contents

Supported by Grant 0401B of RFNS through Math. Inst. SANU.

Introduction

As known, one of the best methods used in studying of structure of semigroups, as well as other algebras, is the *decomposition method*. The main idea of this method is to decompose a semigroup into components, possibly of simpler structure, to study the components in details and to establish mutual relationships between the components within the entire semigroup. We differentiate two general kinds of decompositions: *external decompositions*, where we include decompositions into a direct product and related concepts, and *internal decompositions*, by which we mean decompositions by equivalence relations. In this paper our attention will be aimed only to internal decompositions, which will be here called simply *decompositions*.

By a *kind of decompositions* we will mean a mapping $\mathfrak{T} : S \mapsto \mathfrak{T}_S$ by which to any semigroup S we associate a subset $\mathfrak{T}_S$, possibly empty, of the partition lattice Part(S) of S. But it is often of interest to consider such kinds of decompositions which can be applied on any semigroup, i.e. such that $\mathfrak{T}$ is nonempty subset of Part(S), for any semigroup S. For example, many kinds of decompositions have the property that for any semigroup S, $\mathfrak{T}_S$ contains the zero of Part(S), i.e. the one-component partition $\{S\}$. For that reason we define a *type of decompositions*, or a *decomposition type*, as a mapping $\mathfrak{T} : S \mapsto \mathfrak{T}_S$ by which to any semigroup S we associate a subset $\mathfrak{T}_S$ of the partition lattice Part(S) of S, containing its zero. In other words, a decomposition type $\mathfrak{T}$ is a collection of sets $\mathfrak{T}_S$ indexed by the set of all semigroups, and it is defined if for any semigroup S we define what are the elements of $\mathfrak{T}_S$. Of course, any type $\mathfrak{T}$ of decompositions induces a mapping $\mathfrak{T}' : S \mapsto \mathfrak{T}'_S$ by which to any semigroup S we associate a subset $\mathfrak{T}'_S$ of the lattice $\mathcal{E}(S)$ of equivalence relations on S, containing the universal relation on S, called a *type of equivalences*, and vice versa. For a given type $\mathfrak{T}$ of decompositions and a semigroup S, the elements of $\mathfrak{T}_S$ will be called *$\mathfrak{T}$-decompositions* of S, and related equivalence relations will be called $\mathfrak{T}$-equivalences on S, and S will be called *$\mathfrak{T}$-indecomposable* if the one-component partition $\{S\}$ is the unique $\mathfrak{T}$-decomposition of S, i.e. if the universal relation is the unique $\mathfrak{T}$-equivalence on S.

Consider a decomposition type $\mathfrak{T}$ and a semigroup S. Since $\mathfrak{T}_S$ is a subset of the lattice Part(S), then $\mathfrak{T}_S$ is a poset with respect to usual ordering of partitions, from where several very important questions follow:

(1) Does $\mathcal{T}_S$ have a greatest element?
(2) Is $\mathcal{T}_S$ a complete lattice?
(3) Does $\mathcal{T}_S$ a complete sublattice of the partition lattice on S?

Such problems have been treated first by T. Tamura and N. Kimura [112], 1954, and [113], 1955. After that, they have been considered by many authors. The aim of this paper is to make a survey of main ideas, concepts and results concerning greatest decompositions of semigroups of various types. We will talk about the mostly important decomposition types and the results concerning these.

We know that one of the most important algebraic theorems is the famous *Birkhoff's representation theorem*, proved by G. Birkhoff in [3], 1944, which says that any algebra can be decomposed into a subdirect product of subdirectly irreducible algebras. Of course, in Theory of semigroups similar theorems are also very important. A decomposition type $\mathfrak{T}$ will be called *atomic* if there exists the greatest $\mathfrak{T}$-decomposition and their components are $\mathfrak{T}$-indecomposable. But only four atomic types of decompositions of semigroups are known: *semilattice decompositions*, whose atomicity has been proved by T. Tamura [110], 1956, *ordinal decompositions*, whose atomicity has been proved by E. S. Lyapin [62], 1960, $\cup$-*decompositions*, whose atomicity has been proved by L. N. Shevrin [96], 1965, and *orthogonal decompositions*, whose atomicity has been established by S. Bogdanović and M. Ćirić in [10], 1995. In this paper these decomposition types will take an outstanding place.

This paper is divided into five chapters.

In the first chapter we introduce notions and notations that will be used in the further text, we give a classification of decomposition types and define the types that will considered in this paper, and we also present several general results concerning decompositions by congruences.

Because of the great importance and enormous quantity of the results concerning semilattice decompositions of semigroups, these results will be separated from the ones concerning band decompositions and they will be presented in Chapter 2.

Chapter 3 is devoted to the remaining significant types of band decompositions. Namely, in this chapter we make a survey of the results on matrix and normal band decompositions of semigroups.

In Chapter 4 we consider decompositions of semigroups with zero: orthogonal decompositions, decompositions into a left, right and matrix sum of semigroups, and quasi-semilattice decompositions.

Finally, in Chapter 5 we talk about yet other types of decompositions: $\cup$-decompositions, ordinal decompositions, I-matrix decompositions and semilattice-matrix decompositions.

1. Preliminaries

This chapter is divided into three sections. In Section 1.1 we introduce notions and notations that will be used in the further text. In Section 1.2 we make a classification of decompositions and we single out the most important decomposition types, which will be treated later. Finally, in Section 1.3 we consider decompositions by congruence relations and we present several general results concerning these decompositions.

1.1. Basic notions and notations

Throughout this paper, $\mathbf{Z}^+$ will denote the set of all positive integers. Further, $S = S^0$ means that S is a semigroup with zero 0, and $S \neq S^0$ means that S is a semigroup without zero. If $S = S^0$, we will write 0 instead $\{0\}$, and if A is a subset of S, then $A^\bullet = A - 0$, $A^0 = A \cup 0$ and $A' = (S - A)^0$. If A is a subset of a semigroup S, then $\sqrt{A} = \{x \in S \mid (\exists n \in \mathbf{Z}^+)\, x^n \in A\}$.

For a binary relation ξ on a set A, ξ^∞ will denote the transitive closure of ξ, ξ^{-1} will denote the relation defined by $a\,\xi^{-1}\,b \Leftrightarrow b\,\xi\,a$, and for $a \in A$, $a\xi = \{x \in A \mid a\,\xi\,x\}$ and $\xi a = \{x \in A \mid x\,\xi\,a\}$. By a *quasi-order* we mean a reflexive and transitive binary relation. If ξ is a quasi-order on a set A, then the relation $\widetilde{\xi}$ defined by $\widetilde{\xi} = \xi \cap \xi^{-1}$ is an equivalence relation called the *natural equivalence* of ξ. A relation ξ on a semigroup $S = S^0$ is called *left 0-restricted* if $0\xi = 0$. A *right 0-restricted* relation on S is defined dually, and a relation ξ on $S = S^0$ will be called *0-restricted* if it is both left and right 0-restricted, i.e. if $0\xi = \xi 0 = 0$. We say that a relation ξ on a semigroup S satisfies the *common multiple property*, briefly the *cm-property*, if for all $a, b, c \in S$, $a\,\xi\,c$ and $b\,\xi\,c$ implies $ab\,\xi\,c$. Similarly, for a relation ξ on a semigroup $S = S^0$ we say that ξ satisfies the *0-common multiple property*, briefly the *0-cm-property*, if for all $a, b, c \in S$, $ab \neq 0$, $a\,\xi\,c$ and $b\,\xi\,c$ implies $ab\,\xi\,c$.

Let K be a subset of a lattice L (not necessary complete). If K contains the meet of any its nonempty subset having the meet in L, then K is called a *complete meet-subsemilattice* of L. A *complete join-subsemilattice* is defined dually. If K is both complete meet-subsemilattice and complete join-subsemilattice of L, then it is called a *complete sublattice* of L. If L is a lattice with unity, then any sublattice of L containing its unity is called a *1-sublattice* of L. Dually we define a *0-sublattice* of a lattice with zero, and we define a sublattice of a lattice L with zero and unity to be a *0,1-sublattice* if it is both 0-sublattice and 1-sublattice of L. If any element of L is the meet of some nonempty subset of K, then K is called *meet-dense* in L.

An element a of a lattice L with the zero 0 is an *atom* of L if $a > 0$ and there exists no $x \in L$ such that $a > x > 0$. A complete Boolean algebra B is *atomic* if every element of B is the join of some set of atoms of B. If L is a distributive lattice with zero and unity, then the set $\mathfrak{B}(L)$ of all elements of L having a complement in L is a Boolean algebra and it is called the *greatest Boolean subalgebra* of L.

For a nonempty set A, $\mathcal{P}(A)$ will denote the *lattice of subsets* of A. Let A be a nonempty set and let L be a sublattice of $\mathcal{P}(A)$ containing its unity and having the property that any nonempty intersection of elements of L is also in A. Then for any $a \in A$ there exists the smallest element of L containing a (it is the intersection of all elements of L containing a), which will be called the *principal element* of L generated by a.

A subset A of a semigroup S is called *completely semiprime* if for $x \in S$, $x^2 \in A$ implies $x \in A$, *completely prime* if for $x, y \in S$, $xy \in A$ implies the either $x \in A$ or $y \in A$, *left consistent* if for $x, y \in S$, $xy \in A$ implies $x \in A$, *right consistent* if for $x, y \in S$, $xy \in A$ implies $y \in A$, and it is *consistent* if it is both left and right consistent. Clearly, the empty set has any of these properties and the sets of completely semiprime, completely prime, left consistent, right consistent and consistent subsets are complete sublattices of $\mathcal{P}(S)$. A consistent subsemigroup of a semigroup S will be called a *filter* of S. The empty set will be also defined to be a filter. By $\mathcal{F}(S)$ we denote the *lattice of filters* of S, which is a complete meet-subsemilattice of $\mathcal{P}(S)$, and therefore a complete lattice, but it is not necessary a sublattice of $\mathcal{P}(S)$. It is well known that a subset A of a semigroup S is a filter of S if and only its complement is either empty or a completely prime ideal of S. The principal element $\mathcal{F}(S)$, called the *principal filter*, generated by $a \in S$ will be denoted by $N(a)$.

In studying of semigroups with zero we use some similar notions. Namely, a subset A of a semigroup $S = S^0$ is called *left 0-consistent* if $A^\bullet$ is left consistent, *right 0-consistent* if $A^\bullet$ is right consistent, and it is *0-consistent* if $A^\bullet$ is consistent. Similarly, an equivalence relation θ on $S = S^0$ will be called *left 0-consistent* if for $x, y \in S$, $xy \neq 0$ implies $xy\,\theta\,x$, *right 0-consistent* if for $x, y \in S$, $xy \neq 0$ implies $xy\,\theta\,y$, and *0-consistent* if it is both left and right 0-consistent.

Let S be a semigroup. By $\mathcal{I}d(S)$ we denote the *lattice of ideals* of S. This lattice is a sublattice of $\mathcal{P}(S)$, it is also a complete join-subsemilattice of $\mathcal{P}(S)$, but it is not necessary a complete meet-subsemilattice, since the empty set is not included in $\mathcal{I}d(S)$. The principal element of $\mathcal{I}d(S)$, called the *principal ideal*, generated by $a \in S$ will be denoted by $J(a)$.Further, $\mathcal{LI}d(S)$ will denote the *lattice of left ideals* of a semigroup S defined in the following way: if $S = S^0$, then $\mathcal{LI}d(S)$ consists of all left ideals of S, and if S has no zero, then $\mathcal{LI}d(S)$ consists of the empty set and all left ideals of S. The *lattice of right ideals* of S, in notation $\mathcal{RId}(S)$, is defined dually. Lattices $\mathcal{LId}(S)$ and $\mathcal{RId}(S)$ are complete sublattices of $\mathcal{P}(S)$. The principal element of $\mathcal{LId}(S)$, called the *principal left ideal*, generated by $a \in S$ will be denoted by $L(a)$. The *principal right ideal* generated by $a \in S$, defined dually, will be denoted by $R(a)$. By $\mathcal{Id}^{\mathbf{cs}}(S)$ we denote the *lattice of completely semprime ideals* of S, which is a complete 1-subsemilattice of $\mathcal{Id}(S)$. The principal element of $\mathcal{Id}^{\mathbf{cs}}(S)$, called the *principal radical*, generated by $a \in S$ will be denoted by $\Sigma(a)$. By $\mathcal{RId}^{\mathbf{lc}}(S)$ and $\mathcal{LId}^{\mathbf{rc}}(S)$ we denote the lattice of left consistent right ideals and the lattice of right consistent left ideals of S, which are complete sublattices of $\mathcal{RId}(S)$ and $\mathcal{LId}(S)$, respectively.

For a nonempty subset A of a semigroup S define the relations P_A, R_A and L_A by:

$$\begin{aligned}
a\,P_A\,b &\Leftrightarrow (\forall x, y \in S)(xay \in A \Leftrightarrow xby \in A),\\
a\,R_A\,b &\Leftrightarrow (\forall y \in S)(ay \in A \Leftrightarrow by \in A),\\
a\,L_A\,b &\Leftrightarrow (\forall x \in S)(xa \in A \Leftrightarrow xb \in A).
\end{aligned}$$

Then P_A is a congruence on S called the *principal congruence* on S defined by A, R_A is a right congruence called the *principal right congruence* on S defined by A, and L_A is a left congruence called the *principal left congruence* on S defined by A. If $\mathcal{A}$ is a nonempty family of subsets of S, then $P(\mathcal{A})$ will denote the congruence which is the intersection of all principal congruence on S defined by elements from $\mathcal{A}$.

Let A be a nonempty set and let $X \in \mathcal{P}(A)$. The relation Θ_X on A defined by

$$a\,\Theta_X\,b \quad \Leftrightarrow \quad a, b \in X \text{ or } a, b \in A - X \qquad (a, b \in A),$$

is an equivalence relation on A whose classes are precisely the nonempty sets among the sets X and $A - X$. Clearly, when $X = \varnothing$ or $X = A$, then Θ_X is the universal relation on A. Also, for any $X \in \mathcal{P}(A)$, $\Theta_X = \Theta_{A_X}$. Further, for a nonempty subset $\mathcal{A}$ of $\mathcal{P}(A)$, $\Theta(\mathcal{A})$ will denote the equivalence relation on A defined by:

$$\Theta(\mathcal{A}) = \bigcap_{X \in \mathcal{A}} \Theta_X .$$

If $\mathcal{A}$ is a complete meet-subsemilattice of $\mathcal{P}(A)$, and it contains the unity of $\mathcal{P}(A)$, then $\Theta(\mathcal{A})$ can be alternatively defined by:

$$a\,\Theta(\mathcal{A})\,b \Leftrightarrow \mathcal{A}(a) = \mathcal{A}(b) \qquad (a, b \in S),$$

where for $x \in A$, $\mathcal{A}(x)$ denotes the principal element of $\mathcal{A}$ generated by x.

For a semigroup S, $\mathcal{Q}(S)$ will denote the lattice of quasi-orders on S, $\mathcal{E}(S)$ will denote the lattice of equivalence relations on S and $\mathrm{Con}\,(S)$ will denote the lattice of congruence relations on S. It is well-known that $\mathrm{Con}\,(S)$ is a complete sublattice of $\mathcal{E}(S)$ and $\mathcal{E}(S)$ is a complete sublattice of $\mathcal{Q}(S)$. By $\mathcal{E}^{\bullet}(S)$ we denote the lattice of 0-restricted equivalence relation on a semigroup $S = S^0$, which is the principal ideal of $\mathcal{E}(S)$ generated by the equivalence relation χ determined by the partition $\{S^{\bullet}, 0\}$.

An ideal A of a semigroup S is a *prime ideal* if for $x, y \in S$, $xSy \subseteq A$ implies that either $x \in A$ or $y \in A$, or, equivalently, if for all ideals M and N of S, $MN \subseteq A$ implies that either $M \subseteq A$ or $N \subseteq A$. A completely 0-simple semigroup with the property that the structure group of its Rees-matrix representation is the one-element group, is called a *rectangular 0-band*. Equivalently, a rectangular 0-band can be defined as a semigroup $S = S^0$ in which 0 is a prime ideal and for all $a, b \in S$, either $aba = a$ or $aba = 0$.

For undefined notions and notations we refer to the following books: G. Birkhoff [2], S. Bogdanović [4], S. Bogdanović and M. Ćirić [7], S. Burris and H. P. Sankappanavar [17], A. H. Clifford and G. B. Preston [35], [36], G. Grätzer [45], J. M. Howie [48], E. S. Lyapin [62], M. Petrich [72], [73], L. N. Shevrin [98], L. N. Shevrin and A. Ya. Ovsyanikov [102], [103], O. Steinfeld [105] and G. Szász [109].

1.2. A classification of decompositions

In this section we classify decompositions of semigroups into few classes and we single out the most important types of decompositions.

Let us say again that by a decompositions of a semigroup S we mean a family $\mathcal{D} = \{S_\alpha\}_{\alpha \in Y}$ of subsets of S satisfying the condition

$$S = \bigcup_{\alpha \in Y} S_\alpha, \qquad \text{where } S_\alpha \cap S_\beta = \varnothing, \text{ for } \alpha, \beta \in Y, \alpha \neq \beta.$$

Various special kinds of decompositions we obtain in two general ways: imposing some requirements on the structure of the components S_α, and imposing some requirements on products of elements from different classes.

The first general type of decompositions that we single out are *decompositions S onto subsemigroups*, determined by the property that any S_α is a subsemigroup of S. Clearly, decompositions onto subsemigroups correspond to equivalence relations satisfying the *cm*-property, so the following theorem can be easily proved:

Theorem 1.1. *The poset of decompositions of a semigroup S onto subsemigroups is a complete lattice which is dually isomorphic to the lattice of equivalence relations on S satisfying the cm-property.*

If to a decomposition of a semigroup S onto subsemigroups we impose an additional condition

$$ab \in \langle a\rangle \cup \langle b\rangle ,$$

for all elements $a, b \in S$ belonging to the different components, then we obtain so called $\cup$-*decompositions*. Decompositions of this type will be considered in Section 5.1.

The second general class of decompositions that we single out form decompositions whose related equivalence relations are congruences. Decompositions of this type are called *decompositions by congruences*. When the decomposition $\mathcal{D}$ is a decomposition by a congruence relation, then the index set Y is a factor semigroup of S and many properties of S are determined by structure of the semigroup Y. Special types of decompositions by congruences we obtain imposing some requirements on the structure of the related factor semigroup. If a class $\mathfrak{C}$ of semigroups and a semigroup S are given, then a congruence relation θ on S is called a $\mathfrak{C}$-*congruence* on S if the related factor S/θ is in $\mathfrak{C}$, the related decomposition is given a $\mathfrak{C}$-*decomposition*, and the related factor semigroup is called a $\mathfrak{C}$-*homomorphic image* of S. When there exists the greatest $\mathfrak{C}$-decomposition of S, i.e. the smallest $\mathfrak{C}$-congruence on S, then we say that the factor semigroup corresponding to this congruence is the *greatest* $\mathfrak{C}$-*homomorphic image* of S. A semigroup S is called $\mathfrak{C}$-*indecomposable* if the universal relation is the unique $\mathfrak{C}$-congruence on S. Of course, when the class $\mathfrak{C}$ contains the trivial (one-element) semigroup, then the $\mathfrak{C}$-decompositions determine a decomposition type.

If the decomposition $\mathcal{D}$ is both a decomposition by a congruence relation and a decomposition onto subsemigroups, then it is called a *band decomposition* of S and the related congruence relation is called a *band congruence* on S. Equivalently, the type of band decompositions is defined as the type of $\mathfrak{C}$-decompositions, where $\mathfrak{C}$ equals the variety $[x^2 = x]$ of bands. Moreover, by some subvarieties of the variety of bands we define the following very important special types of band decompositions and band congruences:

- *semilattice decompositions* and *congruences*, determined by the variety $[x^2 = x, xy = yx]$ of *semilattices*;
- *matrix decompositions* and *congruences*, determined by the variety $[x^2 = x, xyx = x] = [x^2 = x, xyz = xz]$ of *rectangular bands*;
- *left* (*right*) *zero band decompositions* and *congruences*, determined by the variety $[x^2 = x, xy = x]$ ($[x^2 = x, xy = y]$) of *left* (*right*) *zero bands*;

- *normal band decompositions* and *congruences*, determined by the variety $[x^2 = x, xyzx = xzyx] = [x^2 = x, xyzu = xzyu]$ of *normal bands*;
- *left (right) normal band decompositions* and *congruences*, determined by the variety $[x^2 = x, xyz = xzy]$ $([x^2 = x, xyz = yxz])$ of *left (right) normal bands*.

Also, *chain decompositions* and *congruences* are determined by the class of *chains* (linearly ordered semilattices). The decomposition $\mathcal{D}$ is called an *ordinal decomposition* if it is a chain decomposition, i.e. Y is a chain, and for all $a, b \in S$,

$$a \in S_\alpha,\ b \in S_\beta,\ \alpha < \beta \quad \Rightarrow \quad ab = ba = a.$$

These decompositions will be considered in Section 5.2. In the last chapter of this paper we will also consider I-matrix decompositions and semilattice-matrix decompositions, which will be precisely defined in Sections 5.3 and 5.4, respectively.

Semigroups with zero have a specific structure and in studying of such semigroups it is often convenient to represent a semigroup $S = S^0$ in the form:

$$S = \bigcup_{\alpha \in Y} S_\alpha, \qquad \text{where } S_\alpha \cap S_\beta = 0, \text{ for } \alpha, \beta \in Y, \alpha \neq \beta.$$

In this case, the partition $\mathcal{D}$ of S, whose components are 0 and $S_\alpha^\bullet$, $\alpha \in Y$, is called a *0-decomposition* of S. If, moreover, any S_α is a subsemigroup of S, we say that $\mathcal{D}$ is a *0-decomposition* of S *onto subsemigroups* and that S is a *0-sum* of semigroups S_α, $\alpha \in Y$, and the semigroups S_α will be called the *summands* of this decomposition. Equivalence relations corresponding to such decompositions are exactly the ones which satisfy the 0-*cm*-property, so the following theorem follows:

Theorem 1.2. *The poset of 0-decompositions of a semigroup $S = S^0$ onto subsemigroups is a complete lattice which is dually isomorphic to the lattice of equivalence relations on S satisfying the 0-cm-property.*

Special decompositions of this type may be determined by some properties of the index set Y. Namely, it is often convenient to assume that Y is a partial groupoid whose all elements are idempotents, and to require that the multiplication on S is carried by Y, by the following condition:

$$\begin{cases} S_\alpha S_\beta \subseteq S_{\alpha\beta} & \text{if } \alpha\beta \text{ is defined in } Y \\ S_\alpha S_\beta = 0 & \text{otherwise} \end{cases}, \qquad \text{for all } \alpha, \beta \in Y.$$

For example, if Y is a semigroup, i.e. a band, we obtain so called *0-band decompositions*. If the product $\alpha\beta$ is undefined, whenever $\alpha \neq \beta$, then $S_\alpha S_\beta = 0$, whenever $\alpha \neq \beta$, and such decompositions are called *orthogonal decompositions*. If Y is a left (right) zero band, then the corresponding decomposition is called a *decomposition into a left* (*right*) *sum* of semigroups. If Y is a nonempty subset of $I \times \Lambda$, where I and Λ are nonempty sets, and the partial multiplication on Y is defined by: for $(i, \lambda), (j, \mu) \in Y$, the product $(i, \lambda)(j, \mu)$ equals (i, μ), if $(i, \mu) \in Y$, and it is undefined, otherwise, then the decomposition $\mathcal{D}$ carried by Y is called a *decomposition into a matrix sum* of semigroups S_α, $\alpha \in Y$. Finally, if Y is an arbitrary poset and for $\alpha, \beta \in Y$, the product $\alpha\beta$ is defined as the meet of α and β, if it exists, then the related decomposition is called a *quasi-semilattice decomposition* of S.

1.3. Decompositions by congruences

Given a nonempty class $\mathfrak{C}$ of semigroups. Let $\mathrm{Con}_{\mathfrak{C}}(S)$ denotes the set of all $\mathfrak{C}$-congruences on S. Of course, $\mathrm{Con}_{\mathfrak{C}}(S)$ is a subset of $\mathrm{Con}(S)$ and it can be treated as a poset with respect to the usual ordering of congruences. Properties of posets of $\mathfrak{C}$-congruences inside the lattice $\mathrm{Con}(S)$ have been first investigated by T. Tamura and N. Kimura in [123], 1955, where they proved the following theorem:

Theorem 1.3. (T. Tamura and N. Kimura [123]) *If $\mathfrak{C}$ is a variety of semigroups, then $\mathrm{Con}_{\mathfrak{C}}(S)$ is a complete lattice, for any semigroup S.*

For the variety of semilattices, the previous theorem has been proved also by T. Tamura and N. Kimura [122], 1954 (see Theorem 2.1).

The problem of existence of the greatest decomposition of a given type has been solved in a special case, for so-called μ-decompositions, by T. Tamura [110], 1956. The solution of this problem in the general case has been given by N. Kimura [54], 1958, by the next theorem. Note that by an *algebraic class of* of semigroups we mean a class of semigroups closed under isomorphisms.

Theorem 1.4. (N. Kimura [54]) *Let $\mathfrak{C}$ be a nonempty algebraic class of semigroups. Then $\mathfrak{C}$ is closed under subdirect products if and only if $\mathrm{Con}_{\mathfrak{C}}(S)$ has the smallest element, for any semigroup S for which $\mathrm{Con}_{\mathfrak{C}}(S) \neq \varnothing$.*

As N. Kimura [54] noted, this theorem has been also found by E. J. Tully. Note that if $\mathrm{Con}_{\mathfrak{C}}(S)$ has the smallest elements, then it is a complete meet-subsemilattice of $\mathrm{Con}(S)$.

The converse of Theorem 1.3 has been proved in a recent paper of M. Ćirić and S. Bogdanović [24]. Namely, they proved the following theorem:

Theorem 1.5. (M. Ćirić and S. Bogdanović [24]) *Let $\mathfrak{C}$ be a nonempty algebraic class of semigroups. Then $\mathfrak{C}$ is a variety if and only if* $\mathrm{Con}_{\mathfrak{C}}(S)$ *is a complete sublattice of* $\mathrm{Con}(S)$, *for any semigroup S.*

By the proof of the previous theorem, given in [24], the next theorem also follows:

Theorem 1.6. (T. Tamura and N. Kimura [123]) *If $\mathfrak{C}$ is a variety of semigroups, then* $\mathrm{Con}_{\mathfrak{C}}(S)$ *is a principal dual ideal of* $\mathrm{Con}(S)$, *for any semigroup S.*

Note that Theorems 1.4, 1.5 and 1.6 holds also for any algebra.

The following theorem, proved by M. Petrich in [72], 1973, has been very useful in his investigations of some greatest decompositions of semigroups.

Theorem 1.7. (M. Petrich [72]) *Let $\mathfrak{C}$ be a variety of semigroups, $\mathfrak{D}$ the class of subdirectly irreducible semigroups from $\mathfrak{C}$ and S any semigroup. Then a congruence θ on a semigroup S, different from the universal congruence, is a $\mathfrak{C}$-congruence if and only if it is the intersection of some family of $\mathfrak{D}$-congruences.*

If we define the trivial semigroup to be subdirectly irreducible, then Theorem 1.7 says that $\mathrm{Con}_{\mathfrak{D}}(S)$ is meet-dense in $\mathrm{Con}_{\mathfrak{C}}(S)$.

2. Semilattice decompositions

Semilattice decompositions of semigroups have been first defined and studied by A. H. Clifford [29], 1941. After that they have been investigated by many authors and they have been systematically studied in several monographs: by E. S. Lyapin [62], 1960, A. H. Clifford and G. B. Preston [35], 1961, M. Petrich [72], 1973, and [73], 1977, S. Bogdanović [4], 1985, S. Bogdanović and M. Ćirić [7], 1993, and other.

First general results concerning semilattice decompositions of semigroups have been the results of T. Tamura and N. Kimura from [122], 1954. There they proved a theorem, given below as Theorem 2.1, by which it follows the existence of the greatest semilattice decomposition on any semigroup. This theorem initiated intensive studying of the greatest semilattice decompositions of semigroups and Section 2.1 is devoted to the results from this area. We present various characterizations of the greatest semilattice decomposition of a semigroup, the smallest semilattice congruence on a semigroup and the greatest semilattice homomorphic image of a semigroup, given by M. Yamada [132], 1955, T. Tamura [110], 1956, [112], 1964, and [117], 1972, M.

Petrich [69], 1964, and [72], 1973, M. S. Putcha [79], 1973, and [80], 1975, and M. Ćirić and S. Bogdanović [21]. We also quote the famous theorem of T. Tamura [110], 1956, on atomicity of semilattice decompositions, which is probably the most important result of the theory of semilattice decompositions of semigroups, and we give several characterizations of semilattice indecomposable semigroups given by M. Petrich [69], 1964, and [72], 1973, and T. Tamura [117], 1972. For the related results concerning decompositions of groupoids we refer to G. Thierrin [127], 1956.

Section 2.2 is devoted to lattices of semilattice decompositions of a semigroup, i.e. to lattices of semilattice congruence on a semigroup. We present characterizations of these lattices of T. Tamura [120], 1975, M. Ćirić and S. Bogdanović [23], and S. Bogdanović and M. Ćirić [12].

2.1. The greatest semilattice decomposition

As we noted above, the first general result concerning semilattice decompositions of semigroups is the one of T. Tamura and N. Kimura [122], 1954, which is given by the following theorem:

Theorem 2.1. (T. Tamura and N. Kimura [122]) *The poset of semilattice decompositions of any semigroup is a complete lattice.*

By the previous theorem it follows that any semigroup has a greatest semilattice decomposition. The first characterization of the greatest semilattice decomposition has been given by M. Yamada [132], 1955, in terms of P-subsemigroups. A subsemigroup T of a semigroup S is called a *P-semigroup* of S if for all $a_1, \dots, a_n \in S$,

$$a_1 \cdots a_n \in T \quad \Rightarrow \quad C(a_1, \dots, a_n) \subseteq T,$$

where $C(a_1, \dots, a_n)$ denotes the subsemigroup of S consisting of all products of elements $a_1, \dots, a_n \in S$ with each a_i appearing at least once [132]. Recall that $P(\mathcal{A})$ denotes the intersection of all principal congruences defined by elements of a nonempty family $\mathcal{A}$ of subsets of a semigroup.

Theorem 2.2. (M. Yamada [132]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = P(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of P-subsemigroups of S.*

As a consequence of the previous theorem it can be deduced the following theorem:

Theorem 2.3. (M. Yamada [132]) *The smallest semilattice congruence on a semigroup S equals the congruence $P(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all P-subsemigroups of S.*

Another approach to the greatest decompositions of semigroups, through completely prime ideals and filters, has been developed by M. Petrich [69], 1964. He proved the following four theorems:

Theorem 2.4. (M. Petrich [69]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of completely prime ideals of S.*

Theorem 2.5. (M. Petrich [69]) *The smallest semilattice congruence on a semigroup S equals the congruence $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all completely prime ideals of S.*

Theorem 2.6. (M. Petrich [69]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of filters of S.*

Theorem 2.7. (M. Petrich [69]) *The smallest semilattice congruence on a semigroup S equals the congruence $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all filters of S.*

Another proofs of the previous two theorems have been given by the authors in [21].

The role of completely prime ideals and filters in semilattice decompositions of semigroups can be explained by Theorem 1.7. Namely, the two-element chain Y_2 is, up to an isomorphism, the unique subdirectly irreducible semilattice, and any homomorphism of a semigroup S onto Y_2 determines a partition of S whose one component is a completely prime ideal and other is a filter of S. This approach has been used by M. Petrich in [72], 1973.

M. Petrich [69], 1964, also gave a method to construct the principal filters of a semigroup:

Theorem 2.8. (M. Petrich [69]) *The principal filter of a semigroup S generated by an element $a \in S$ can be computed using the following formulas:*

$$N_1(a) = \langle a \rangle, \quad N_{n+1}(a) = \langle \{x \in S \mid N_n(a) \cap J(x) \neq \varnothing\} \rangle, \ n \in \mathbf{Z}^+$$
$$N(a) = \bigcup_{n \in \mathbf{Z}^+} N_n(a).$$

The third approach to the greatest decompositions of semigroups is the one of T. Tamura from [117], 1972. Using the *division relation* $\mid$ on a semigroup S defined by:

$$a \mid b \quad \Leftrightarrow \quad b \in S^1 a S^1,$$

T. Tamura defined the relation $\longrightarrow$ on S by:

$$a \longrightarrow b \quad \Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, a \mid b^n,$$

and he gave an efficient characterization of the smallest semilattice congruence on a semigroup:

Theorem 2.9. (T. Tamura [117]) *The smallest semilattice congruence on a semigroup S equals the natural equivalence of the relation $\longrightarrow^\infty$.*

Another proof of this theorem has been given by T. Tamura [118], 1973.

Three different characterizations of the smallest semilattice congruence on a semigroup have been also obtained by M. S. Putcha in [79], 1973, and [80], 1975.

Theorem 2.10. (M. S. Putcha [80]) *The smallest semilattice congruence on a semigroup S equals the equivalence on S generated by the relation $xy \equiv xyx \equiv yx$, for all $x, y \in S^1$.*

Another proof of this theorem has been given by T. Tamura [119], 1973.

Theorem 2.11. (M. S. Putcha [80]) *The smallest semilattice congruence on a semigroup S equals the relation $—^\infty$, where $— = \longrightarrow \cap \longrightarrow^{-1}$.*

Theorem 2.12. (M. S. Putcha [79]) *The smallest semilattice congruence on a semigroup S equals the relation θ on S defined by: $a\,\theta\,b$ if and only if for all $x, y \in S^1$ there exists a semilattice indecomposable subsemigroup T of S such that $xay, xby \in T$.*

An approach to semilattice decompositions of semigroups, different to the one of M. Petrich and T. Tamura, has been developed by M. Ćirić and S. Bogdanović in [21]. As we will see later, the results obtained there explained the connections between the above presented results of M. Petrich and T. Tamura. M. Ćirić and S. Bogdanović [21] started from the completely semiprime ideals and they first gave the following representations of the principal radicals of a semigroup:

Theorem 2.13. (M. Ćirić and S. Bogdanović [21]) *The principal radical of a semigroup S generated by an element $a \in S$ has the following representation:*

$$\Sigma(a) = \{x \in S \mid a \longrightarrow^{\infty} x\}.$$

Theorem 2.14. (M. Ćirić and S. Bogdanović [21]) *The principal radical of a semigroup S generated by an element $a \in S$ can be computed using the following formulas:*

$$\Sigma_1(a) = \sqrt{SaS},\ \Sigma_n(a) = \sqrt{S\Sigma_n(a)S},\ n \in \mathbf{Z}^+,\ \Sigma(a) = \bigcup_{n \in \mathbf{Z}^+} \Sigma_n(a).$$

Recall that $\mathcal{I}d^{\mathrm{cs}}(S)$ denotes the lattice of all completely semiprime ideals of a semigroup S. By means of Theorems 2.13 and 2.9, the authors in [21] obtained the following characterization of the smallest semilattice congruence on a semigroup:

Theorem 2.15. (M. Ćirić and S. Bogdanović [21]) *The smallest semilattice congruence on a semigroup S equals the equivalence $\Theta(\mathcal{I}d^{\mathrm{cs}}(S))$.*

A characterization of the greatest semilattice homomorphic image of a semigroup has been given by M. Ćirić and S. Bogdanović [21], through principal radicals of a semigroup:

Theorem 2.16. (M. Ćirić and S. Bogdanović [21]) *If a, b is any pair of elements of a semigroup S, then*

$$\Sigma(a) \cap \Sigma(b) = \Sigma(ab),$$

i.e. the set Σ_S of all principal radicals of S, partially ordered by inclusion, is a semilattice and it is the greatest semilattice homomorphic image of S.

As a consequence of the previous theorem, the authors in [21] proved the next theorem without use of the Zorn's lemma arguments:

The next theorem, which gives a connection between Theorems 2.15 and 2.5, has been proved by M. Petrich [72], 1973. Another proof of this theorem, without use of the Zorn's lemma arguments, has been given by the authors in [21], as a consequence of Theorem 2.16.

Theorem 2.17. (M. Petrich [72]) *Any completely semiprime ideal of a semigroup S is the intersection of some family of completely prime ideals of S.*

In other words, Theorem 2.17 says that the set of completely prime ideals of a semigroup S is meet-dense in $\mathcal{I}d^{\mathrm{cs}}(S)$.

Another consequence of Theorem 2.16 is the next theorem which gives a representation of the principal filters better than the one from Theorem 2.8.

Theorem 2.18. (M. Ćirić and S. Bogdanović [21]) *The principal filter of a semigroup S generated by an element a has the following representation:*

$$N(a) = \{x \in S \mid x \longrightarrow^{\infty} a\}.$$

The components of the greatest semilattice decomposition of a semigroup are characterized by the next theorem, which is clearly a consequence of Theorems 2.13, 2.18 and 2.9.

Theorem 2.19. (M. Petrich [72]) *The component of the greatest semilattice decomposition of a semigroup S containing an element a of S is precisely the subsemigroup $\Sigma(a) \cap N(a)$.*

The most significant theorem of the theory of semilattice decompositions of semigroup is probably the theorem of T. Tamura [110], 1956, on atomicity of semilattice decompositions of semigroups, given here as Theorem 2.20. Note that another proofs of this theorem have been given by T. Tamura in [112], 1964, by means of the concept of "contents", in [117], 1972, using the relation $\longrightarrow^{\infty}$, in [118], 1973, and [120], 1975, by M. Petrich [69], 1964, by means of completely prime ideals, and by M. S. Putcha [79], 1973, using the relation defined in Theorem 2.12 and the subsemigroups of the form $C(a_1, \dots, a_n)$.

Theorem 2.20. (T. Tamura [110]) *Any component of the greatest semilattice decomposition of a semigroup is a semilattice indecomposable semigroup.*

Semilattice indecomposable semigroups have been described by T. Tamura [117] and M. Petrich [69], [72], by the following

Theorem 2.21. *The following conditions on a semigroup S are equivalent:*

(i) *S is semilattice indecomposable;*
(ii) $(\forall a, b \in S)\ a \longrightarrow^{\infty} b$*;*
(iii) *S has no proper completely semiprime ideals;*
(iv) *S has no proper completely prime ideals.*

The equivalence of conditions (i) and (ii) has been established by T. Tamura [117], 1972, (i) ⇔ (iii) has been proved by M. Petrich [69], 1964, and (i) ⇔ (iv) by M. Petrich [72], 1973.

Note that in the class of semilattice indecomposable semigroup the mostly investigated were Archimedean semigroups, defined by: $a \longrightarrow b$, for all elements a and b. Semilattices of such semigroups have been studied by many authors. The most important results from this area have been obtained by M. S. Putcha [79], 1973, T. Tamura [116], 1972, M. Ćirić and S. Bogdanović

[19], 1993, and [21], S. Bogdanović and M. Ćirić [6], 1992, and [14], and L. N. Shevrin [99] and [100], 1994. For more informations about semilattices of Archimedean semigroups the reader is also referred to the survey paper of S. Bogdanović and M. Ćirić [8], 1993, or their book [7], 1993.

2.2. The lattice of semilattice decompositions

T. Tamura [120] got an idea of studying semilattice decompositions of a semigroup through quasi-orders on this semigroup having some suitable properties. We say that a quasi-order ξ on a semigroup S is *positive* if $a\,\xi\,ab$ and $b\,\xi\,ab$, for all $a, b \in S$. These quasi-orders have been introduced by B. M. Schein [88], 1965, and they were since studied from different points of view by T. Tamura, M. S. Putcha, S. Bogdanović, M. Ćirić and other. By a *half-congruence* T. Tamura in [120], 1975, called a compatible quasi-order on a semigroup, and by a *lower-potent* quasi-order he called a quasi-order ξ on a semigroup satisfying the condition: $a^2\,\xi\,a$, for all elements a. Using these notions, T. Tamura proved the following theorem:

Theorem 2.22. (T. Tamura [120]) *The lattice of semilattice congruences on a semigroup S is isomorphic to the lattice of positive lower-potent half-congruences on S.*

As the authors noted in [23], the notion "lower-potent half-congruence" in Theorem 2.22 can be replaced by "quasi-order satisfying the *cm*-property". Recall from Section 1.1 that a relation ξ on a semigroup S satisfies the *common multiple property*, briefly the *cm-property*, if for all $a, b, c \in S$, $a\,\xi\,c$ and $b\,\xi\,c$ implies $ab\,\xi\,c$. Using this notion, introduced by T. Tamura in [116], 1972, Theorem 2.22 can be written as follows:

Theorem 2.23. *The lattice of semilattice congruences on a semigroup S is isomorphic to the lattice of positive quasi-orders on S satisfying the cm-property.*

Using the Tamura's approach, the authors in [23] connected semilattice decompositions of a semigroup with some sublattices of the lattice $\mathcal{I}d^{\mathrm{cs}}(S)$ of completely simple ideals of a semigroup. Recall from Section 1.1 that a subset K of a lattice L is *meet-dense* in L if any element of L can be written as the meet of some family of elements of K. We will say that a sublattice L of $\mathcal{I}d^{\mathrm{cs}}(S)$ satisfies the *completely prime ideal property*, shortly the *cpi-property*, if the set of completely prime ideals from L is meet-dense in L, i.e. if any element of L can be written as the intersection of some family of completely prime ideals from L. As we seen before, this property was proved for $\mathcal{I}d^{\mathrm{cs}}(S)$ by Theorem 2.17. M. Ćirić and S. Bogdanović [23] showed that the *cpi*-property plays a crucial role in semilattice decompositions of semigroups:

Theorem 2.24. (M. Ćirić and S. Bogdanović [23]) *The lattice of semilattice decompositions of a semigroup S is isomorphic to the lattice of complete 1-sublattices of $\mathcal{I}d^{\mathrm{cs}}(S)$ satisfying the cpi-property.*

Another connection of semilattice decompositions of a semigroup, with some sublattices of the lattice of subsets of a semigroup, has been established by S. Bogdanović and M. Ćirić in [12]. There they proved the following theorem:

Theorem 2.25. (S. Bogdanović and M. Ćirić [12]) *The lattice of semilattice decompositions of a semigroup S is isomorphic to the lattice of complete 1-sublattices of $\mathcal{P}(S)$ whose principal elements are filters of S.*

For more informations about the role of quasi-orders in semilattice decompositions of semigroups we refer to another survey paper of S. Bogdanović and M. Ćirić [16].

3. Band decompositions

Although the existence of the greatest band decomposition has been established by T. Tamura and N. Kimura in [123], 1955, by the theorem which is given here as Theorem 1.3, there are no sufficiently efficient characterizations of the greatest band decomposition of a semigroup in the general case. But, there are very nice descriptions of greatest decompositions for some special types of band decompositions, as semilattice decompositions, treated in the previous chapter, matrix decompositions, where left zero band and right zero band decompositions are included, and normal band decompositions, where left normal band and right normal band decompositions are included. This chapter is devoted to the results concerning the greatest matrix decomposition of a semigroup, which will be presented in Section 3.1, and to the results concerning the greatest normal band decomposition of a semigroup, which will be presented in Section 3.2.

Matrix decompositions, as well as left zero band and right zero band decompositions, have appeared first in studying of completely simple semigroups. Namely, by the famous Rees-Sushkevich theorem on matrix representations of completely simple semigroups, any completely simple semigroup can be decomposed into a matrix of groups, and also into a left zero band of right groups and into a right zero band of left groups. First general results concerning these decompositions have been obtained by P. Dubreil

[41], 1951, who constructed the smallest left zero band congruence on a semigroup, and by G. Thierrin [128], 1956, who characterized the components of the greatest left zero band decomposition of a semigroup. The general theory of matrix decompositions of semigroups has been founded by M. Petrich in [70], 1996. These results will be a topic of Section 3.1.

Normal bands have been introduced by M. Yamada and N. Kimura [133], 1958, whereas left normal bands have been first defined and studied by V. V. Vagner [129], 1962, and B. M. Schein [86], 1963, and [87], 1965. The general results concerning left normal band, right normal band and normal band decompositions of a semigroup, presented in Section 3.2, have been obtained by M. Petrich in [71], 1966.

For additional informations about matrix and normal band decompositions the reader is referred to the book of M. Petrich [73], 1977.

3.1. Matrix decompositions

As we noted before, the first general result concerning left zero band decompositions of a semigroup is the one of P. Dubreil [41], 1951. Define the relations $\overset{l}{\approx}$ and $\overset{r}{\approx}$ on a semigroup S by:

$$a\overset{l}{\approx}b \ \Leftrightarrow\ L(a)\cap L(b)\neq\varnothing, \qquad a\overset{r}{\approx}b \ \Leftrightarrow\ R(a)\cap R(b)\neq\varnothing,\ \ (a,b\in S).$$

The relation $\overset{r}{\approx}$ has been introduced in above mentioned paper of P. Dubreil, where he proved the following theorem:

Theorem 3.1. (P. Dubreil [41]) *The smallest left zero band congruence on a semigroup S equals the relation $\overset{r}{\approx}{}^{\infty}$.*

The components of the greatest left zero band decomposition of a semigroup have been first described by G. Thierrin [128], 1956, by the following theorem:

Theorem 3.2. (G. Thierrin [128]) *The components of the greatest left zero band decomposition of a semigroup S are the minimal left consistent right ideals.*

Other characterizations of the greatest left zero band decomposition of a semigroup have been obtained by M. Petrich in [70], 1966. In this paper he proved the following two theorems:

Theorem 3.3. (M. Petrich [70]) *A relation θ on a semigroup S is a left zero band congruence on S if and only if $\theta=\Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of left consistent right ideals of S.*

Theorem 3.4. *The smallest left zero band congruence on a semigroup S equals the relation $\Theta(\mathcal{RI}d^{\mathbf{lc}}(S))$.*

The key theorem in theory of matrix decompositions of semigroups is the next theorem, proved by M. Petrich in [70], 1966, which gives a connection between left zero band, right zero band and matrix congruences on a semigroup:

Theorem 3.5. (M. Petrich [70]) *The intersection of a left zero band congruence and a right zero band congruence on a semigroup S is a matrix congruence on S.*

Conversely, any matrix congruence on S can be written uniquely as the intersection of a left zero band congruence and a right zero band congruence on S.

Combining Theorems 3.1 and 3.5, the following characterization of the smallest matrix congruence on a semigroup follows:

Theorem 3.6. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\overset{l}{\approx}{}^{\infty} \cap \overset{r}{\approx}{}^{\infty}$.*

Combining Theorem 3.3 and its dual, M. Petrich [70] obtained the following two theorems:

Theorem 3.7. (M. Petrich [70]) *A relation θ on a semigroup S is a matrix congruence on S if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty subset $\mathcal{A}$ of $\mathcal{X}$, where $\mathcal{X} = \mathcal{LI}d^{\mathbf{rc}}(S) \cup \mathcal{RI}d^{\mathbf{lc}}(S)$.*

Theorem 3.8. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\Theta(\mathcal{X})$, where $\mathcal{X} = \mathcal{LI}d^{\mathbf{rc}}(S) \cup \mathcal{RI}d^{\mathbf{lc}}(S)$.*

M. Petrich in [70] also gave an alternative approach to matrix decompositions of semigroups, through so-called quasi-consistent subsemigroups. Namely, by a *quasi-consistent* subset of a semigroup S he defined a completely semiprime subset A of S satisfying the condition: for all $x, y, z \in S$, $xyz \in A$ if and only if $xy \in A$. Quasi-consistent subsemigroups of a semigroup M. Petrich connected with left consistent right ideals and right consistent left ideals by the following theorem:

Theorem 3.9. (M. Petrich [70]) *The intersection of a left consistent right ideal and a right consistent left ideal of a semigroup S is a quasi-consistent subsemigroup of S.*

Conversely, any quasi-consistent subsemigroup of S can be written uniquely as the intersection of a left consistent right ideal and a right consistent left ideal.

Using the previous theorem, matrix congruences on a semigroup can be characterized through quasi-consistent subsemigroups of a semigroup as follows:

Theorem 3.10. (M. Petrich [70]) *A relation θ on a semigroup S is a matrix congruence on S if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of the set of quasi-consistent subsemigroups of S.*

Theorem 3.11. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all quasi-consistent subsemigroups of S.*

Using Theorem 3.5 and the fact that the join of any left zero band congruence and any right zero band congruence on a semigroup equals the universal congruence on this semigroup, the lattice of matrix congruences on a semigroup can be characterized in the following way:

Theorem 3.12. *The lattice of matrix congruences on a semigroup S is isomorphic to the direct product of the lattice of left zero band congruences and the lattice of right zero band congruences on S.*

A characterization of the lattice of right zero band decompositions of a semigroup can be obtained through left consistent right ideals of a semigroup, modifying the results of S. Bogdanović and M. Ćirić [13] to semigroups without zero. For related results concerning semigroups with zero we refer to Section 4.2.

Until the end of this section we will consider only semigroups without zero, because the definition of the lattice $\mathcal{RI}d(S)$ is different for semigroups with and without zero, and the set of right consistent left ideals of a semigroup with zero is one-element.

Theorem 3.13. *The poset $\mathcal{RI}d^{\mathbf{lc}}(S)$ of left consistent right ideals of a semigroup $S \neq S^0$ without zero is a complete atomic Boolean algebra and $\mathcal{RI}d^{\mathbf{lc}}(S) = \mathfrak{B}(\mathcal{RI}d(S))$.*

Theorem 3.14. *The lattice of lęft zero band decompositions of a semigroup $S \neq S^0$ is isomorphic to the lattice of complete Boolean subalgebras of $\mathcal{RI}d^{\mathbf{lc}}(S)$.*

The role of left zero band decompositions of a semigroup in direct decompositions of the lattice of right ideals of this semigroup is demonstrated by the following two theorems:

Theorem 3.15. *The lattice $\mathcal{RI}d(S)$ of right ideals of a semigroup $S \neq S^0$ is a direct product of lattices L_α, $\alpha \in Y$, if and only if S is a left zero band of semigroups S_α, $\alpha \in Y$, and $L_\alpha \cong \mathcal{RI}d(S_\alpha)$, for any $\alpha \in Y$.*

Theorem 3.16. *If S_α, $\alpha \in Y$, are components of the greatest left zero band decomposition of a semigroup $S \neq S^0$, then the lattice $\mathcal{RI}d(S)$ can be decomposed into a direct product of its intervals $[0, S_\alpha]$, $\alpha \in Y$, which are directly indecomposable.*

3.2. Normal band decompositions

In the introduction of Chapter 3 we said that the general theory of normal band decompositions of semigroups, including here left normal band and right normal band decompositions, has been founded by M. Petrich in [71], 1966. The methods used in this paper has been obtained by combination of the methods which M. Petrich used in [69], in studying of semilattice decompositions, and the ones used in [70], in studying of matrix decompositions.

M. Petrich in [71] defined a *left* (*right*) *normal complex* of a semigroup S as a nonempty subset A of S which is a left (right) consistent right (left) ideal of the smallest filter $N(A)$ of S containing A, and he defined a *normal complex* of S as a subset A of S which is a quasi-consistent subsemigroup of $N(A)$. He also introduced the following relations on a semigroup S: for a nonempty subset A of S, Φ_A is the equivalence relation on S whose classes are nonempty sets among the sets A, $N(A) - A$ and $S - N(A)$, and for a nonempty family $\mathcal{A}$ of subsets of S, $\Phi(\mathcal{A})$ is the equivalence relation on S defined by:

$$\Phi(\mathcal{A}) = \bigcap_{A \in \mathcal{A}} \Phi_A .$$

Theorem 3.17. (M. Petrich [71]) *A relation θ on a semigroup S is a left normal band congruence on S if and only if $\theta = \Phi(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of left normal complexes of S.*

Theorem 3.18. (M. Petrich [71]) *The smallest left normal band congruence on a semigroup S equals the relation $\theta = \Phi(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all left normal complexes of S.*

In order to study normal band congruences on a semigroup through left normal band congruences and right normal band congruences, M. Petrich proved the following theorem, similar to Theorem 3.5 concerning matrix congruences:

Theorem 3.19. (M. Petrich [71]) *The intersection of a left normal band congruence and a right normal band congruence on a semigroup S is a normal band congruence on S.*

Conversely, any normal band congruence on S can be written as the intersection of the smallest left normal band congruence and the smallest right normal band congruence on S containing it.

Using this theorem, M. Petrich [71] characterized normal band congruences and the smallest normal band congruence on a semigroup by the following two theorems:

Theorem 3.20. (M. Petrich [71]) *A relation θ on a semigroup S is a normal band congruence on S if and only if $\theta = \Phi(\mathcal{A})$, for some nonempty subset $\mathcal{A}$ of $\mathcal{X}$, where $\mathcal{X}$ denotes the set of all left normal complexes and all right normal complexes of S.*

Theorem 3.21. (M. Petrich [71]) *The smallest normal band congruence on a semigroup S equals the relation $\theta = \Phi(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all left normal complexes and all right normal complexes of S.*

In Theorem 3.20, $\mathcal{X}$ cannot be replaced by the set of all normal complexes, but this can be done in Theorem 3.21:

Theorem 3.22. (M. Petrich [71]) *The smallest normal band congruence on a semigroup S equals the relation $\theta = \Phi(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all normal complexes of S.*

4. Decompositions of semigroups with zero

The first known type of decompositions of semigroups with zero have been orthogonal decompositions, called also 0-direct unions, which have been first defined and studied by E. S. Lyapin in [60] and [61], 1950, and Š. Schwarz [90], 1951. After that, they have been studied by many authors, mainly as orthogonal sums of completely 0-simple semigroups. General study of orthogonal decompositions of semigroups with zero has done by S. Bogdanović and M. Ćirić in [10], 1995, and [13]. The results obtained there will be a topic of Section 4.1. Among these results we emphasize Theorem 4.8 on atomicity of orthogonal decompositions.

Decompositions of a semigroup with zero into a left, right and matrix sum of semigroups have been first defined and studied by S. Bogdanović and M. Ćirić in [13]. The results concerning these decompositions obtained in this paper will be presented in Section 4.2. We also give Theorem 4.21 which establish connections between the decompositions into a left, right and matrix sum, and orthogonal decompositions inside the lattice of 0-decompositions of a semigroup with zero. Note also that some decompositions of semigroups with zero, similar to decomposition into a matrix sum, have been considered by O. Steinfeld in [105].

Quasi-semilattice decompositions of a semigroup with zero, which are carried by partially ordered sets, appeared recently in the paper of M. Ćirić and S. Bogdanović [26]. These decompositions will be considered in Section 4.3.

Note finally that decompositions into a left, right and matrix sum of semigroups, and quasi-semilattice decompositions of semigroups with zero are generalizations (or analogues) of left zero band, right zero band, matrix and semilattice decompositions, respectively, as showed by Theorems 4.22 and 4.28. Orthogonal sums have no such analogue.

4.1. Orthogonal decompositions

In studying of orthogonal decompositions of semigroups with zero, S. Bogdanović and M. Ćirić [10], 1995, has started from the notion of 0-consistent ideal. They defined a *0-consistent* ideal of a semigroup $S = S^0$ as an ideal A having the property that $A^\bullet$ is a consistent subset of S. They denoted by $\mathcal{I}d^{0c}(S)$ the set of all 0-consistent ideals of a semigroup $S = S^0$ and they proved the following theorem:

Theorem 4.1. (S. Bogdanović and M. Ćirić [10]) *The poset $\mathcal{I}d^{0c}(S)$ of all 0-consistent ideals of a semigroup $S = S^0$ is a complete atomic Boolean algebra and $\mathcal{I}d^{0c}(S) = \mathfrak{B}(\mathcal{I}d(S))$.*

Furthermore, any complete atomic Boolean algebra is isomorphic to the Boolean algebra of 0-consistent ideals of some semigroup with zero.

Using this theorem, S. Bogdanović and M. Ćirić [10] obtained the following theorem concerning orthogonal decompositions:

Theorem 4.2. (S. Bogdanović and M. Ćirić [10]) *Any semigroup $S = S^0$ has a greatest orthogonal decomposition and its summands are all the atoms in $\mathcal{I}d^{0c}(S)$.*

Another approach to orthogonal decompositions, through certain equivalence relations, has done by S. Bogdanović and M. Ćirić in [13]. A 0-restricted and 0-consistent equivalence relation on a semigroup $S = S^0$ will be called an *orthogonal equivalence*. This name will be justified by the role of these equivalences in orthogonal decompositions, which will be demonstrated in Theorem 4.4. Namely, the authors proved in [13] the following two theorems:

Theorem 4.3. (S. Bogdanović and M. Ćirić [13]) *The poset of orthogonal equivalences on a semigroup $S = S^0$ is a complete sublattice of the lattice $\mathcal{E}(S)$.*

Theorem 4.4. (S. Bogdanović and M. Ćirić [13]) *The poset of orthogonal decompositions of a semigroup $S = S^0$ is a complete lattice and it is dually isomorphic to the lattice of orthogonal equivalences on S.*

Note that the sumands in an orthogonal decomposition of a semigroup $S = S^0$ are precisely the nonzero classes of the related orthogonal equivalence, with the zero adjoined, and vice versa.

By Theorems 4.3 and 4.4 we deduce the following:

Theorem 4.5. *The lattice of orthogonal decompositions of a semigroup $S = S^0$ is a complete sublattice of the partition lattice of S.*

The lattice of orthogonal decompositions has been also characterized by some Boolean subalgebras of $\mathcal{I}d^{0c}(S)$ as follows:

Theorem 4.6. (S. Bogdanović and M. Ćirić [13]) *The lattice of orthogonal decompositions of a semigroup $S = S^0$ is isomorphic to the lattice of complete Boolean subalgebras of $\mathcal{I}d^{0c}(S)$.*

Note that any complete Boolean subalgebra of $\mathcal{I}d^{0c}(S)$ is atomic and its atoms are precisely the summands in the related orthogonal decomposition of S, and vice versa.

To describe the smallest orthogonal equivalence on a semigroup with zero, S. Bogdanović and M. Ćirić in [10] defined the relation $\sim$ on a semigroup $S = S^0$ by

$$a \sim b \quad \Leftrightarrow \quad J(a) \cap J(b) \neq 0, \quad \text{for } a, b \in S^\bullet, \qquad\qquad 0 \sim 0,$$

and they proved the following:

Theorem 4.7. (S. Bogdanović and M. Ćirić [10]) *The smallest orthogonal equivalence on a semigroup $S = S^0$ equals the relation $\sim^\infty$.*

Note also that the lattice of orthogonal equivalences on S is the principal dual ideal of the lattice $\mathcal{E}^\bullet(S)$ of 0-restricted equivalence relations on S, generated by $\sim^\infty$. Since $\mathcal{E}^\bullet(S)$ is the principal ideal of $\mathcal{E}(S)$, generated by the equivalence relation χ on S determined by the partition $\{S^\bullet, 0\}$, then the lattice of orthogonal equivalences on S is precisely the (closed) interval $[\sim^\infty, \chi]$ of $\mathcal{E}(S)$.

The main theorem of the theory of orthogonal decompositions of semigroups with zero is the theorem on atomicity of orthogonal decompositions, proved by S. Bogdanović and M. Ćirić in [10], 1995. This is the following theorem:

Theorem 4.8. (S. Bogdanović and M. Ćirić [10]) *The summands of the greatest orthogonal decomposition of a semigroup $S = S^0$ are orthogonally indecomposable semigroups.*

S. Bogdanović and M. Ćirić in [13] also observed that orthogonal decompositions of a semigroup $S = S^0$ are closely connected with direct decompositions of the lattice of ideals of S. This connection is demonstrated by the following three theorems:

Theorem 4.9. *The lattice $\mathcal{I}d(S)$ of ideals of a semigroup $S = S^0$ is a direct product of lattices L_α, $\alpha \in Y$, if and only if S is an orthogonal sum of semigroups S_α, $\alpha \in Y$, and $L_\alpha \cong \mathcal{I}d(S_\alpha)$, for any $\alpha \in Y$.*

Theorem 4.10. (S. Bogdanović and M. Ćirić [13]) *The lattice $\mathcal{I}d(S)$ of ideals of a semigroup $S = S^0$ is directly indecomposable if and only if S is orthogonally indecomposable.*

Theorem 4.11. (S. Bogdanović and M. Ćirić [13]) *If S_α, $\alpha \in Y$, are summands of the greatest orthogonal decomposition of a semigroup $S = S^0$, then the lattice $\mathcal{I}d(S)$ can be decomposed into a direct product of lattices $\mathcal{I}d(S_\alpha)$, $\alpha \in Y$, which are directly indecomposable.*

4.2. Decompositions into a left, right and matrix sum

In studying of decompositions of semigroups with zero into a left sum of semigroups, S. Bogdanović and M. Ćirić in [13] used the methods similar to the ones used in studying of orthogonal decompositions. At first, they considered equivalence relations on a semigroup with zero which we call here left sum equivalences. Namely, a 0-restricted, left 0-consistent equivalence relation on a semigroup $S = S^0$ will be called an *left sum equivalence*. *Right sum equivalences* on S are defined dually. These names will be explained by the role of these equivalences in decompositions of S into a left sum and a right sum of semigroups, respectively, as demonstrated in Theorem 4.13. But, first we give the following theorem:

Theorem 4.12. (S. Bogdanović and M. Ćirić [13]) *The poset of left sum equivalences on a semigroup $S = S^0$ is a complete sublattice of the lattice $\mathcal{E}(S)$.*

Theorem 4.13. (S. Bogdanović and M. Ćirić [13]) *The poset of decompositions of a semigroup $S = S^0$ into a left sum of semigroups is a complete lattice and it is dually isomorphic to the lattice of left sum equivalences on S.*

As in orthogonal decompositions, the sumands in a decomposition of a semigroup $S = S^0$ into a left sum are the nonzero classes of the related left sum equivalence, with the zero adjoined, and vice versa.

By Theorems 4.12 and 4.13 we obtain the following:

Theorem 4.14. *The lattice of decompositions of a semigroup $S = S^0$ into a left sum of semigroups is a complete sublattice of the partition lattice of S.*

To characterize the smallest left sum equivalence on a semigroup, the authors used the relation $\overset{r}{\sim}$ defined by G. Lallement and M. Petrich [59], 1966, on a semigroup $S = S^0$ by:

$$a \overset{r}{\sim} b \quad \Leftrightarrow \quad R(a) \cap R(b) \neq 0, \quad \text{for } a, b \in S^{\bullet}, \qquad\qquad 0 \overset{r}{\sim} 0.$$

The relation $\overset{\ell}{\sim}$ on S is defined dually. Using the above relation, S. Bogdanović and M. Ćirić [13] characterized the smallest left sum equivalence as follows:

Theorem 4.15. (S. Bogdanović and M. Ćirić [13]) *The smallest left sum equivalence on a semigroup $S = S^0$ equals the relation $\overset{r}{\sim}^{\infty}$.*

As in orthogonal equivalences, the set of left sum equivalences on a semigroup $S = S^0$ equals the interval $[\overset{r}{\sim}^{\infty}, \chi]$ of the lattice $\mathcal{E}(S)$.

Instead of 0-consistent ideals, used in studying of orthogonal decompositions, in studying of decompositions of a semigroup with zero into a left sum of semigroups, S. Bogdanović and M. Ćirić used in [13] the notion of the left 0-consistent right ideal. Namely, they defined a right ideal A of a semigroup $S = S^0$ to be *left 0-consistent* if $A^{\bullet}$ is a left consistent subset of S. The set of all left 0-consistent ideals of a semigroup they denoted by $\mathcal{RI}d^{\mathbf{l0c}}(S)$ and they proved the following two theorems:

Theorem 4.16. (S. Bogdanović and M. Ćirić [13]) *The poset $\mathcal{RI}d^{\mathbf{l0c}}(S)$ of all left 0-consistent right ideals of a semigroup $S = S^0$ is a complete atomic Boolean algebra and $\mathcal{RI}d^{\mathbf{l0c}}(S) = \mathfrak{B}(\mathcal{RI}d(S))$.*

Theorem 4.17. (S. Bogdanović and M. Ćirić [13]) *The lattice of decompositions of a semigroup $S = S^0$ into a left sum of semigroup is isomorphic to the lattice of complete Boolean subalgebras of $\mathcal{RI}d^{\mathbf{l0c}}(S)$.*

As in orthogonal decompositions, the summands in a decomposition of a semigroup $S = S^0$ into a left sum of semigroups are precisely the atoms in the related complete Boollean subalgebra of $\mathcal{RI}d^{\mathbf{l0c}}(S)$, which is atomic, and vice versa.

As S. Bogdanović and M. Ćirić in [13] observed, the previous two theorems can be applied to direct decompositions of the lattice of right ideals of a semigroup with zero:

Theorem 4.18. *The lattice $\mathcal{RI}d(S)$ of right ideals of a semigroup $S = S^0$ is a direct product of lattices L_α, $\alpha \in Y$, if and only if S is a left sum of semigroups S_α, $\alpha \in Y$, and $L_\alpha \cong \mathcal{RI}d(S_\alpha)$, for any $\alpha \in Y$.*

Theorem 4.19. (S. Bogdanović and M. Ćirić [13]) *If S_α, $\alpha \in Y$, are the summands of the greatest decomposition of a semigroup $S = S^0$ into a left sum of semigroups, then the lattice $\mathcal{LI}d(S)$ can be decomposed into a direct product of its intervals $[0, S_\alpha]$, $\alpha \in Y$, which are directly indecomposable.*

Note that the interval $[0, S_\alpha]$ in Theorem 4.19 cannot be replaced by the lattice $\mathcal{RI}d(S_\alpha)$, in contrast to Theorem 4.11.

In order to characterize decompositions of a semigroup with zero into a matrix sum of semigroups, S. Bogdanović and M. Ćirić consider in [13] equivalence relations that are the intersection of a left sum equivalence and a right sum equivalence, which will be called here *matrix sum equivalences*, and they proved the following theorems:

Theorem 4.20. (S. Bogdanović and M. Ćirić [13]) *The poset of matrix sum equivalences on a semigroup $S = S^0$ is a complete lattice.*

Theorem 4.21. (S. Bogdanović and M. Ćirić [13]) *The poset of decompositions of a semigroup $S = S^0$ into a matrix sum of semigroups is a complete lattice and it is dually isomorphic to the lattice of matrix sum equivalences on S.*

Note that the sumands in a decomposition of a semigroup $S = S^0$ into a matrix sum are exactly the nonzero classes of the related matrix sum equivalence, with the zero adjoined, and vice versa.

Note also that the previous two theorems give a connection between the decompositions into a left sum, decompositions into a right sum and decompositions into a matrix sum. The authors in [13] established a similar connection between the decompositions into a left sum, decompositions into a right sum and orthogonal decompositions. This connection is given by the following theorem:

Theorem 4.22. (S. Bogdanović and M. Ćirić [13]) *The join in $\mathcal{E}(S)$ of any left sum equivalence and any right sum equivalence on a semigroup $S = S^0$ is an orthogonal equivalence on S.*

Especially, the join of $\overset{r}{\sim}{}^{\infty}$ and $\overset{\ell}{\sim}{}^{\infty}$ equals $\sim^{\infty}$.

The above quoted results can be summarized by the following theorem:

Theorem 4.23. *In the partition lattice of a semigroup $S = S^0$, the meet of any decomposition of S into a left sum and any decomposition of S into a right sum of semigroups is an orthogonal decomposition, and its join is a decomposition into a matrix sum of semigroups.*

Especially, the meet of the greatest decomposition of S into a left sum and the greatest decomposition of S into a right sum of semigroups is the greatest *orthogonal decomposition of S, and its join is the greatest decomposition of S into a matrix sum of semigroups.*

Note finally that decompositions of a semigroup with zero into a left sum, right sum and matrix sum of semigroups can be treated as generalizations of left zero band, right zero band and matrix decompositions, respectively. This follows by the following theorem:

Theorem 4.24. *The lattice of left zero band (right zero band, matrix) decompositions of a semigroup S is isomorphic to the lattice of decompositions into a left (right, matrix) sum of semigroups of a semigroup T arising from S by adjoining the zero.*

4.3. Quasi-semilattice decompositions

Studying of quasi-semilattice decompositions of semigroups with zero began in the paper of M. Ćirić and S. Bogdanović [20], 1994. In this paper, some notions which appears in studying of semilattice decompositions of semigroups the authors modified for semigroups with zero. Namely, the authors defined a *0-positive* quasi-order on a semigroup $S = S^0$ as a quasi-order ξ having the property that for $a, b \in S$, $ab \neq 0$ implies $a \,\xi\, ab$ and $b \,\xi\, ab$, they defined a quasi-order ξ on S to satisfy the *0-cm-property* if for all $a, b, c \in S$, $ab \neq 0$, $a \,\xi\, c$ and $b \,\xi\, c$ implies $ab \,\xi\, c$, and they proved the following theorem:

Theorem 4.25. (M. Ćirić and S. Bogdanović [20]) *The poset of left 0-restricted positive quasi-orders on a semigroup $S = S^0$ satisfying the 0-cm-property and the poset of 0-restricted 0-positive quasi-orders on S satisfying the 0-cm-property are complete lattices and they are isomorphic.*

Further, M. Ćirić and S. Bogdanović defined in [20] a *completely 0-semiprime* ideal of a semigroup $S = S^0$ as an ideal A of S having the property that $A^\bullet$ is a completely semiprime subset of S. Similarly, A is said to be *completely 0-prime* if $A^\bullet$ is a completely prime subset. The set of all completely 0-semiprime ideals of S, denoted by $\mathcal{I}d^{\mathrm{c0s}}(S)$ is clearly a complete lattice. A sublattice L of $\mathcal{I}d^{\mathrm{c0s}}(S)$ is defined to satisfy the *c-0-pi-property* if the set of completely 0-prime ideals from L is meet-dense in L, i.e. if any element of L can be written as the intersection of some family of completely 0-prime ideals from L. Using these notions, M. Ćirić and S. Bogdanović [20] proved the following theorem:

Theorem 4.26. (M. Ćirić and S. Bogdanović [20]) *For a semigroup $S = S^0$, the poset of complete 0,1-sublattices of the lattice $\mathcal{I}d^{\mathrm{c0s}}(S)$ satisfying the c-0-pi-property is a complete lattice and it is dually isomorphic to the lattice of 0-restricted 0-positive quasi-orders on S satisfying the 0-cm-property.*

The investigation of quasi-semilattice decompositions of semigroups with zero M. Ćirić and Bogdanović continued in [26], where they proved the following three theorems that characterize the lattice of quasi-semilattice decompositions of a semigroup with zero:

Theorem 4.27. (M. Ćirić and S. Bogdanović [20]) *The poset of quasi-semilattice decompositions of a semigroup $S = S^0$ is a complete lattice and it is dually isomorphic to the lattice of 0-restricted 0-positive quasi-orders on S satisfying the 0-cm-property.*

Theorem 4.28. (M. Ćirić and S. Bogdanović [26]) *The lattice of quasi-semilattice decompositions of a semigroup $S = S^0$ it is dually isomorphic to the lattice of left 0-restricted positive quasi-orders on S satisfying the 0-cm-property.*

Theorem 4.29. (M. Ćirić and S. Bogdanović [26]) *The lattice of quasi-semilattice decompositions of a semigroup $S = S^0$ is isomorphic to the lattice of complete 0,1-sublattices of $\mathcal{I}d^{\mathbf{c0s}}(S)$ satisfying the c-0-pi-property.*

We finish this chapter by the theorem which give a connection between quasi-semilattice decompositions of semigroups with zero and semilattice decompositions. Note that this connection is incorporated in the name of quasi-semilattice decompositions.

Theorem 4.30. (M. Ćirić and S. Bogdanović [26]) *The lattice of semilattice decompositions of a semigroup S is isomorphic to the lattice of quasi-semilattice decompositions of the semigroup T arising from S by adjoining the zero.*

5. Yet other decompositions

In this paper we talk about yet other types of decompositions having the greatest one.

The topic of Section 5.1 will be $\cup$-decompositions, introduced and first studied by L. N. Shevrin [93], 1961, as a powerful tool in studying of lattices of subsemigroups of a semigroup. We quote the theorem considering the properties of the poset of $\cup$-decompositions, the theorem on atomicity of these decompositions, as Theorem 5.3, and also three theorems on application of $\cup$-decompositions to direct decompositions of the lattice of subsemigroups of a semigroup. For informations on other applications of $\cup$-decompositions, and related $\cup$-band decompositions, in studying of the lattice of subsemigroups of a semigroup the reader is referred to the books of L. N. Shevrin and A. Ya. Ovsyanikov [102], 1990, and [103], 1991, their survey article [101], 1983, and the book of M. Petrich [73], 1977. Note that L. N. Shevrin used the names "strong decomposition" and "strong band decomposition" for these decompositions. But, because the notion "strong band of semigroups" has been also used for other concepts of the semigroup theory, here we use the names used also in the book of M. Petrich [73], 1977.

Ordinal decompositions, treated in Section 5.2, came out from studying of linearly ordered groups in the papers of F. Klein-Barmen [55] and [56], 1942, and [57], 1948, and A. M. Kaufman [51] and [52], 1949. They have been introduced by A. M. Kaufman [51], 1949, where they have been called successively-annihilating sums (bands) of semigroups. General study of these decompositions has done by E. S. Lyapin in his book [62], 1960, where he showed that the poset of ordinal decompositions of any semigroup is a complete sublattice of the partition lattice of this semigroup, and proved the theorem on atomicity of ordinal decompositions, given here as Theorem 5.8. Here we also present the results of M. Ćirić and S. Bogdanović that characterize lattices of ordinal decompositions of semigroups. For more applications on applications of ordinal decompositions see the books: E. S. Lyapin [62], 1960, M. Petrich [73], 1977, and L. N. Shevrin and A. Ya. Ovsyanikov [102], 1990, and [103], 1991.

I-matrix decompositions have arisen in the paper of G. Lallement and M. Petrich [59], 1966, as a generalization of matrix decompositions. The very nice results obtained in this paper will be presented in Section 5.3. For some applications of such decompositions see the papers of J. Fountain and M. Petrich [43], 1986, and [44], 1989.

The last section of this chapter is devoted to semilattice-matrix decompositions of semigroups. These decompositions have been first studied by A. H. Clifford [29], 1941, who proved that unions of groups (completely regular semigroups) are semilattices of completely simple semigroups, which are in fact semilattices of matrices of groups. After that, semilattice-matrix decompositions have been studied by many authors, for example by P. Chu, Y, Guo and X. Ren [28], 1989, L. N. Shevrin [100], 1994, S. Bogdanović and M. Ćirić [11], 1995, and [15], and other. By the well-known theorem of D. McLean [64], 1954, and A. H. Clifford [30], 1954, on the decomposition of a band into a semilattice of rectangular bands, semilattice-matrix decompositions can be treated as generalizations of band decompositions, and in many papers these decompositions have been used to make preparations for band decompositions. Here we present some general properties of these decompositions discovered by M. Ćirić and S. Bogdanović in [27].

5.1. $\bigcup$-decompositions

We said in the introduction of this paper that general study of $\cup$-decompositions has been done by L. N. Shevrin in [96], 1965. There he has obtained the result that can be formulated in the following way:

Theorem 5.1. (L. N. Shevrin [96]) *The poset of $\cup$-decompositions of a semigroup S is a principal ideal of the partition lattice of S.*

In the same paper L. N. Shevrin considered also $\cup$-band decompositions and some their special types. Namely, for any subvariety $\mathcal{V}$ of the variety of bands a *$\cup$-$\mathcal{V}$-band decomposition* of a semigroup S is defined as a decomposition which is both $\cup$-decomposition and $\mathcal{V}$-band decomposition. By Theorems 5.1 and 1.6 the following theorem follows:

Theorem 5.2. *For any subvariety $\mathcal{V}$ of the variety of bands, the poset of $\cup$-$\mathcal{V}$-band decompositions of a semigroup S is a principal ideal of the partition lattice of S.*

L. N. Shevrin [96] also proved the theorem on atomicity of $\cup$-decompositions, which is given below.

Theorem 5.3. (L. N. Shevrin [96]) *The components of the greatest $\cup$-decomposition of a semigroup S are $\cup$-indecomposable.*

Among the numerous applications of $\cup$-decompositions in studying of lattices of subsemigroups of a semigroup we emphasize the application to decompositions of these lattices into a direct product, which is demonstrated by the following three theorems:

Theorem 5.4. (L. N. Shevrin [94]) *The lattice $\mathrm{Sub}\,(S)$ of subsemigroups of a semigroup S is a direct product of lattices L_α, $\alpha \in Y$, if and only if S has a $\cup$-decomposition into subsemigroups S_α, $\alpha \in Y$, and $\mathrm{Sub}\,(S_\alpha) \cong L_\alpha$, for any $\alpha \in Y$.*

Theorem 5.5. (L. N. Shevrin [96]) *The lattice $\mathrm{Sub}\,(S)$ of subsemigroups of a semigroup S is directly indecomposable if and only if S is $\cup$-indecomposable.*

Theorem 5.6. (L. N. Shevrin [96]) *If S_α, $\alpha \in Y$, are the components of the greatest $\cup$-decomposition of a semigroup S, then the lattice $\mathrm{Sub}\,(S)$ of subsemigroups of S can be decomposed into a direct product of lattices $\mathrm{Sub}\,(S_\alpha)$, $\alpha \in Y$, which are directly indecomposable.*

We advise the reader to compare the previous three theorems with Theorems 4.9–4.11, concerning direct decompositions of the lattice of ideals of a semigroup with zero, Theorems 4.18 and 4.19, concerning direct decompositions of the lattice of right ideals of a semigroup with zero, and Theorems 3.15 and 3.16, concerning direct decompositions of the lattice of right ideals of a semigroup without zero.

5.2. Ordinal decompositions

General study of ordinal decompositions has been made by E. S. Lyapin in his book [62] from 1960. There he showed the following property of the poset of ordinal decompositions:

Theorem 5.7. (E. S. Lyapin [62]) *The poset of ordinal decompositions of a semigroup S is a complete sublattice of the partition lattice of S.*

E. S. Lyapin [62] also proved the very important theorem on atomicity of ordinal decompositions, whose another proof has been given by M. Ćirić and S. Bogdanović in [25].

Theorem 5.8. (E. S. Lyapin [62]) *The components of the greatest ordinal decomposition of a semigroup S are ordinally indecomposable.*

To characterize the lattice of ordinal decompositions, M. Ćirić and S. Bogdanović [25] have used the next theorem, obtained in their earlier paper [23], which gives a characterization of the poset of chain decompositions of a semigroup through completely prime ideals.

Theorem 5.9. (M. Ćirić and S. Bogdanović [23]) *The poset of chain decompositions of a semigroup S is isomorphic to the poset of complete 1-sublattices of $\mathcal{I}d^{\mathrm{cs}}(S)$ consisting of completely prime ideals of S.*

Note that another characterization of the poset of chain decompositions can be given by filters as follows:

Theorem 5.10. (S. Bogdanović and M. Ćirić [12]) *The poset of chain decompositions of a semigroup S is isomorphic to the poset of complete 0,1-sublattices of $\mathcal{P}(S)$ consisting of filters of S.*

M. Ćirić and S. Bogdanović [25] defined a *strongly prime* ideal of a semigroup S as an ideal P of S having the property that for all $x, y \in S$, $xy = p \in P$ implies that either $x = p$ or $y = p$ or $x, y \in P$, and they proved that the set of all strongly prime ideals of a semigroup S, denoted by $\mathcal{I}d^{\mathrm{sp}}(S)$, is a complete 1-sublattice of the lattice $\mathcal{I}d(S)$ of ideals of S. Moreover, they gave the following characterization of the lattice of ordinal decompositions of a semigroup:

Theorem 5.11. (M. Ćirić and S. Bogdanović [25]) *The lattice of ordinal decompositions of a semigroup S is isomorphic to the lattice of complete 1-sublattices of $\mathcal{I}d^{\mathbf{sp}}(S)$.*

5.3. I-matrix decompositions

If θ is a congruence on a semigroup S and S/θ is a rectangular 0-band, then θ is said to be an *I-matrix congruence*, where I is an ideal of S which is the θ-class that is the zero of S/θ. The corresponding decomposition is an *I-matrix decomposition* of S, and I is called a *matrix ideal* of S. G. Lallement and M. Petrich [59] defined a *quasi-completely prime* ideal of a semigroup S as an ideal I satisfying the condition that for all $a, b, c \in S$, $abc \in I$ implies that either $ab \in I$ or $bc \in I$, and they proved the following theorem:

Theorem 5.12. (G. Lallement and M. Petrich [59]) *An ideal I of a semigroup S is a matrix ideal if and only if it is prime and quasi-completely prime.*

To characterize I-matrix congruences, G. Lallement and M. Petrich [59] introduced the following notions: if A is a nonempty subset of a semigroup S, then an equivalence relation θ on S is called a *left A-equivalence* if the following conditions hold:

(1) A is a θ-class of S;

(2) θ is a left congruence;

(3) for all $x, y \in S$, $xy \notin A$ implies $xy\,\theta\,x$.

A *right A-equivalence* is defined dually. Necessary and sufficient conditions for existence of a left A-equivalence and a right A-equivalence on a semigroup have been determined by the following theorem:

Theorem 5.13. (G. Lallement and M. Petrich [59]) *Let A be a subset of a semigroup S. Then there exists a left A-equivalence and a right A-equivalence if and only if A is a quasi-completely prime ideal of S.*

The following theorem has been also proved in [59]:

Theorem 5.14. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the poset of left I-equivalences on S is a complete sublattice of the lattice of left congruences on S.*

G. Lallement and M. Petrich [59] characterized the smallest I-equivalence in three ways. At first, they defined a *left I-complex* of a semigroup S as a nonempty subset A of S having the following properties:

(1) $A \cap I = \varnothing$;

(2) A is a left consistent subset of S;

(3) $A \cup I$ is a right ideal of S.

A *right I-complex* has been defined dually. For an element $a \in S - I$, let $C(a)$ they denoted the smallest left I-complex of S containing a, i.e. the intersection of all left I-complexes of S containing a, called the *principal left I-complex* of S generated by a, and they proved the following

Theorem 5.15. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the relation θ on S defined by:*

$$a\,\theta\,b \quad \Leftrightarrow \quad a, b \in I \ \text{ or } \ C(a) = C(b) \qquad\qquad (a, b \in S),$$

equals the smallest left I-equivalence on S.

The second and third characterization of the smallest left I-equivalence on a semigroup have been given by the following two theorems:

Theorem 5.16. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the relation θ on S defined by*

$$a\,\theta\,b \quad \Leftrightarrow \quad (\forall x \in S)\,(axa \in I \quad \Leftrightarrow \quad bxb \in I) \qquad\qquad (a, b \in S),$$

equals the smallest I-matrix congruence on S.

Theorem 5.17. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the smallest left I-equivalence equals the principal left congruence L_I.*

Following the ideas used by M. Petrich in studying of matrix decompositions, G. Lallement and M. Petrich proved in [59] the next theorem, similar to Theorem 3.5.

Theorem 5.18. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the intersection of a left I-equivalence and a right I-equivalence is a I-matrix congruence on S.*

Conversely, any I-matrix congruence on S can be written uniquely as the intersection of a left I-equivalence and a right I-equivalence on S.

Using the previous theorem and Theorem 5.14, G. Lallement and M. Petrich proved also in [59] the following two theorems:

Theorem 5.19. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S. Then the poset of I-matrix congruences on S is a complete sublattice of* $\mathrm{Con}\,(S)$.

Theorem 5.20. (G. Lallement and M. Petrich [59]) *Let I be a matrix ideal of a semigroup S and let θ denote the smallest I-matrix congruence on S. Then $\theta = R_I \cap L_I = R_K = L_K$, where $K = \{x \in S \mid x^2 \notin I\}$.*

5.4. Semilattice-matrix decompositions

Let a semigroup S be a semilattice Y of semigroups S_α, $\alpha \in Y$, and for any $\alpha \in Y$, let S_α be a left zero band (right zero band, matrix) of semigroups S_i, $i \in I_\alpha$. M. Ćirić and S. Bogdanović [27] called the partition $\{S_i \mid i \in I\}$, where $I = \bigcup_{\alpha\in Y} I_\alpha$, a *semilattice-left* (*semilattice-right, semilattice-matrix*) *decomposition* of S, or briefly *s-l-* (*s-r-, s-m-*)*decomposition*. If θ denotes the equivalence relation determined by this partition and if ϱ denotes the semilattice congruence determined by the partition $\{S_\alpha \mid \alpha \in Y\}$, then θ is called a *semilattice-left* (*semilattice-right, semilattice-matrix*) *equivalence on S carried by ϱ*, or briefly *s-l-* (*s-r-, s-m-*)*equivalence*, and ϱ is called a *carrier* of θ. Clearly, an equivalence relation θ on a semigroup S contained in a semilattice-congruence ϱ on S is a s-l-(s-r-, s-m-)equivalence carried by ϱ if and only if for all $a, b \in S$, $a\,\varrho\,b$ implies $ab\,\theta\,a$ ($a\,\varrho\,b$ implies $ab\,\theta\,b$, $a\,\varrho\,b$ implies $aba\,\theta\,a$).

M. Ćirić and S. Bogdanović studied in [27] some general properties of s-l-, s-r- and s-m-equivalences and their carriers, and they proved the next four theorems. Note that Theorem 5.24 is similar to Theorems 3.5, 3.19 and 5.18.

Theorem 5.21. (M. Ćirić and S. Bogdanović [27]) *The set of s-l-(s-r-, s-m-) equivalences on a semigroup S carried by a semilattice congruence ϱ on S is a closed interval of $\mathcal{E}(S)$.*

Theorem 5.22. (M. Ćirić and S. Bogdanović [27]) *The set of carriers of a s-l-(s-r-, s-m-)equivalence θ on a semigroup S is a convex subset, with the smallest element, of the lattice of semilattice congruences on S.*

Theorem 5.23. (M. Ćirić and S. Bogdanović [27]) *The poset of all a s-l-(s-r-, s-m-)equivalences on a semigroup S is a complete lattice.*

Theorem 5.24. (M. Ćirić and S. Bogdanović [27]) *The intersection of a s-l-equivalence and a s-r-equivalence on a semigroup S is a s-m-equivalence.*

Conversely, any s-m-equivalence can be written, uniquely up to a carrier, as the intersection of a s-l-equivalence and a s-r-equivalence.

References

[1] O. Anderson, *Ein Bericht uber Structur abstracter Halbgruppen*, Thesis, Hamburg, 1952.

[2] G. Birkhoff, *Lattice theory*, Amer. Math. Soc, Coll. Publ. Vol. 25, (3rd. edition, 3rd. printing), Providence, 1979.

[3] G. Birkhoff, *Subdirect unions in universal algebra*, Bull. Amer. Math. Soc. **50** (1944), 764–768.

[4] S. Bogdanović, *Semigroups with a system of subsemigroups*, Inst. of Math., Novi Sad, 1985.

[5] S. Bogdanović and M. Ćirić, *U_{n+1}-semigroups*, Contributions MANU **11** (1990), no. 1–2, 9–23.

[6] S. Bogdanović and M. Ćirić, *Semigroups in which the radical of every ideal is a subsemigroup*, Zb. rad. Fil. fak. Niš, Ser. Mat. **6** (1992), 129–135.

[7] S. Bogdanović and M. Ćirić, *Semigroups*, Prosveta, Niš, 1993. (in Serbian)

[8] S. Bogdanović and M. Ćirić, *Semilattices of Archimedean semigroups and (completely) π-regular semigroups* I *(A survey)*, Filomat (Niš) **7** (1993), 1–40.

[9] S. Bogdanović and M. Ćirić, *A new approach to some greatest decompositions of semigroups (A survey)*, Southeast Asian Bull. Math. **18** (1994), no. 3, 27–42.

[10] S. Bogdanović and M. Ćirić, *Orthogonal sums of semigroups*, Israel J. Math. **90** (1995), 423–428.

[11] S. Bogdanović and M. Ćirić, *Semilattices of weakly left Archimedean semigroups*, ibid, this volume.

[12] S. Bogdanović and M. Ćirić, *Positive quasi-orders with the common multiple property on a semigroup*, in: Proc. Math. Conf. in Priština 1994, Lj. D. Kočinac ed, Priština, 1995, pp. 1–6.

[13] S. Bogdanović and M. Ćirić, *Decompositions of semigroups with zero*, Publ Inst. Math. (Beograd) (to appear).

[14] S. Bogdanović and M. Ćirić, *Semilattices of left completely Archimedean semigroups* (to appear).

[15] S. Bogdanović and M. Ćirić, *A note on left regular semigroups*, Publ. Math. Debrecen (to appear).

[16] S. Bogdanović and M. Ćirić, *Quasi-orders and semilattice decompositions of semigroups (A survey)* (to appear).

[17] S. Burris and H. P. Sankappanavar, *A Course in Universal Algebra*, Springer-Verlag, 1983.

[18] M. Ćirić and S. Bogdanović, *Rédei's bands of periodic π-groups*, Zbornik rad. Fil. fak. Niš, Ser. Mat. **3** (1989), 31–42.

[19] M. Ćirić and S. Bogdanović, *Decompositions of semigroups induced by identities*, Semigroup Forum **46** (1993), 329–346.

[20] M. Ćirić and S. Bogdanović, *The lattice of positive quasi-orders on a semigroup* II, Facta Univ. (Niš), Ser. Math. Inform. **9** (1994), 7–17.

[21] M. Ćirić and S. Bogdanović, *Semilattice decompositions of semigroups*, Semigroup Forum (to appear).

[22] M. Ćirić and S. Bogdanović, *Orthogonal sums of 0-σ-simple semigroups*, Acta. Math. Hung (to appear).

[23] M. Ćirić and S. Bogdanović, *The lattice of positive quasi-orders on a semigroup*, Israel J. Math. (to appear).

[24] M. Ćirić and S. Bogdanović, *Posets of $\mathfrak{C}$-congruences*, Algebra Universalis (to appear).

[25] M. Ćirić and S. Bogdanović, *Ordinal sums of semigroups*, Univ. Beograd Publ. Elektr. Fak, Ser. Mat. (to appear).

[26] M. Ćirić and S. Bogdanović, *Quasi-semilattice decompositions of semigroups with zero* (to appear).

[27] M. Ćirić and S. Bogdanović, *Semilattice-matrix decompositions of semigroups* (to appear).

[28] P. Chu, Y. Guo and X. Ren, *The semilattice (matrix)-matrix(semilattice) decomposition of the quasi-completely orthodox semigroups*, Chinese J. of Contemporary Math. **10** (1989), no. 4, 425–438.

[29] A. H. Clifford, *Semigroups admitting relative inverses*, Annals of Math. **42** (1941), no. 2, 1037–1049.

[30] A. H. Clifford, *Bands of semigroups*, Proc. Amer. Math. Soc. **5** (1954), no. 2, 499–504.

[31] A. H. Clifford, *Naturally totally ordered commutative semigroups*, Amer. J. Math. **76** (1954), no. 3, 631–646.

[32] A. H. Clifford, *Review of M. Yamada* [132], Math. Reviews **17** (1956), 584.

[33] A. H. Clifford, *Totally ordered commutative semigroups*, Bull. Amer. Math. Soc. **64** (1958), no. 6, 305–316.

[34] A. H. Clifford, *Completion of semi-continuous ordered commutative semigroups*, Duke Math. J. **26** (1959), no. 1, 41–60.

[35] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, Vol 1, Amer. Math. Soc., 1961.

[36] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, Vol 2, Amer. Math. Soc., 1967.

[37] R. Croisot, *Demi-groupes inversifs et demi-groupes réunions de demi-groupes simples*, Ann. Sci. Ecole Norm. Sup. **70** (1953), no. 3, 361–379.

[38] R. Croisot, *Équivalences principales bilatères définies dans un demi-groupe*, J. Math. Pures Appl. **36** (1957), no. 9, 373–417.

[39] J. Dieudonné, *Sur le socle d'un anneaux et les anneaux simple infinis*, Bull. Soc. Math. France **70** (1942), 46–75.

[40] P. Dubreil, *Contribution a la theorie des demi-groupes*, Mem. Acad. Sci. Inst. France (2) **63** (1941), no. 3, 1–52.

[41] P. Dubreil, *Contribution a la theorie des demi-groupes* II, Rend. Mat. e appl. **10** (1951), 183–200.

[42] M. Ego, *Structure des demi-groupes dont le treillis des sous-demi-groupes est modulaire ou semi-modulaire*, C. R. Acad. Sci. Paris **254** (1962), 1723–1725.

[43] J. Fountain and M. Petrich, *Completely 0-simple semigroups of quotients*, J. Algebra **101** (1986), 365–402.

[44] J. Fountain and M. Petrich, *Completely 0-simple semigroups of quotients* III, Math. Proc. Camb. Phil. Soc. **105** (1989), 263–275.

[45] G. Grätzer, *Universal Algebra*, Van Nostrand, Princeton, 1968.

[46] J. A. Green, *On the structure of semigroups*, Ann. of Math. **54** (1951), 163–172.

[47] T. Hall, *On the natural order of T-class and of idempotents in a regular semigroup*, *Glasgow Math. J.* **11** (1970), 167–168.

[48] J. M. Howie, *An introduction to semigroup theory*, Acad. Press, New York, 1976.

[49] K. Iséki, *Contribution to the theory of semigroups* IV, Proc. Japan Acad. **32** (1956), 430–435.

[50] P. R. Jones, *Semilattice decompositions and the congruence lattice of semigroups*, preprint (1981).

[51] A. M. Kaufman, *Successively-annihilating sums of associative systems*, Uch. Zap. Leningrad. Gos. Ped. Inst. **86** (1949), 149–165. (in Russian)

[52] A. M. Kaufman, *Structure of some holoids*, Uch. Zap. Leningrad. Gos. Ped. Inst. **86** (1949), 167–182. (in Russian)

[53] A. V. Kelarev, *Radicals and 0-bands of semigroups*, Stud. Sci. Math. Hung. **27** (1992), 125–132.

[54] N. Kimura, *On some existence theorems on multiplicative systems.* I. *Greatest quotients*, Proc. Japan Acad. **36** (1958), no. 6, 305–309.

[55] F. Klein-Barmen, *Über gewisse Halverbände und kommutative Semigruppen* I, Math. Z. **48** (1942), 275–288.

[56] F. Klein-Barmen, *Über gewisse Halverbände und kommutative Semigruppen* II, Math. Z. **48** (1942), 715–734.

[57] F. Klein-Barmen, *Ein Beitrag zur Theorie der linearen Holoide*, Math. Z. **51** (1948), 355–366.

[58] O. B. Kozhevnikov, *A generalization of the concept of complete regularity*, in: Associativnye dejstvija, Gos. Ped. Inst., Leningrad, 1983, pp. 50–56. (in Russian)

[59] G. Lallement and M. Petrich, *Décompositions I-matricielles d'un demi-groupe*, J. Math. Pures et Appl. **45** (1966), 67–117.

[60] E. S. Lyapin, *Normal complexes of associative systems*, Izv. Akad. Nauk SSSR **14** (1950), 179–192. (in Russian)

[61] E. S. Lyapin, *Semisimple commutative associative systems*, Izv. Akad. Nauk SSSR **14** (1950), 367–380. (in Russian)

[62] E. S. Lyapin, *Semigroups*, Fizmatgiz, Moscow, 1960, English transl. Amer. Math. Soc. 1968 (2nd edition). (in Russian)

[63] E. S. Lyapin, *Identities of successively-annihilating bands of semigroups*, Izv. Vysh. Uchebn. Zav. Mat. **1 (200)** (1979), 38–45. (in Russian)

[64] D. Mc Lean, *Idempotent semigroups*, Amer. Math. Monthly **6** (1954), 110–113.

[65] H. Mitsch, *Semigroups and their lattice of congruences*, Semigroup Forum **26** (1983), 1–64.

[66] W. D. Munn, *Decompositions of the lattice of congruences on a semigroup*, Proc. Edinburgh Math. Soc. **23** (1980), 193–198.

[67] Y. S. Park, J. P. Kim and M. G. Sohn, *Semiprime ideals in semigroups*, Math. Japonica **33** (1988), no. 2, 269–273.

[68] M. Petrich, *The maximal semilattice decomposition of a semigroup*, Bull. Amer. Math. Soc. **69** (1963), 342–344..

[69] M. Petrich, *The maximal semilattice decomposition of a semigroup*, Math. Zeitschr. **85** (1964), 68–82.

[70] M. Petrich, *The maximal matrix decomposition of a semigroup*, Portugaliae Math. **25** (1966), 15–33.

[71] M. Petrich, *Homomorphisms of a semigroup onto normal bands*, Acta. Sci. Math. Szeged **27** (1966), 185–196.

[72] M. Petrich, *Introduction to semigroups*, Merill, Ohio, 1973.

[73] M. Petrich, *Lectures in semigroups*, Akad. Verlag, Berlin, 1977.

[74] R. S. Pierce, *Homomorphisms of semigroups*, Ann. of Math. **59** (1954), no. 2, 287–291.

[75] R. Plemmons, *Semigroups with a maximal semigroup with zero homomorphic image*, Notices Amer. Math. Soc. **11** (1964), no. 7, 751.

[76] R. J. Plemmons and T. Tamura, *Semigroups with a maximal homomorphic image having zero*, Proc. Japan Acad. **41** (1965), 681–685.

[77] G. B. Preston, *Matrix representation of inverse semigroups*, J. Austral. Math. Soc. **9** (1969), 29–61.

[78] G. B. Preston, *Representations of inverse semigroups*, J. London Math. Soc. **29** (1954), 411–419.

[79] M. S. Putcha, *Semilattice decompositions of semigroups*, Semigroup Forum **6** (1973), 12–34.

[80] M. S. Putcha, *Minimal sequences in semigroups*, Trans. Amer. Math. Soc. **189** (1974), 93–106.

[81] M. S. Putcha, *Positive functions from $\mathcal{S}$-indecomposable semigroups into partially ordered sets*, Check. Math. J. **26 (101)** (1976), 161–170.

[82] M. S. Putcha, *On the maximal semilattice decomposition of the power semigroup of a semigroup*, Semigroup Forum **15** (1978), 263–267.

[83] M. S. Putcha and A. H. Schoenfeld, *Applications of algebraic and combinatoric techniques to a problem in geometry*, J. Pure Appl. Algebra **7** (1976), 235–237.

[84] D. Rees, *On semi-groups*, Proc. Cambridge Phil. Soc. **36** (1940), 387–400.

[85] B. M. Schein, *On subdirectly irreducible semigroups*, Doklady AN SSSR **144** (1962), 999–1002, Correction – **148** (1963), p. 996. (in Russian)

[86] B. M. Schein, *A contribution to the theory of restrictive semigroups*, Izv. vysh. Uchebn. Zaved, Mat. **2 (33)** (1963), 152–154. (in Russian)

[87] B. M. Schein, *Restrictive bisemigroups*, Izv. vysh. Uchebn. Zaved, Mat. **1 (44)** (1965), 168–179. (in Russian)

[88] B. M. Schein, *On certain classes of semigroups of binary relations*, Sibirsk. Mat. Zhurn. **6** (1965), 616–635. (in Russian)

[89] B. M. Schein, *Homomorphisms and subdirect decompositions of semigroups*, Pacific J. Math. **17** (1966), 529–547.

[90] Š. Schwarz, *Semigroups having a kernel*, Check. Math. J. **1 (76)** (1951), 259–301. (in Russian)

[91] Š. Schwarz, *On Gallois connections in the theory of characters of commutative semigroups*, Check. Math. J. **4** (1956), 296–313. (in Russian)

[92] Š. Schwarz, *Semigroups satisfying some weakened forms of the cancellation laws*, Mat.-fyz. čas. **6** (1956), 149–158. (in Slovakian)

[93] L. N. Shevrin, *Semigroups with some types of lattices of subsemigroups*, DAN SSSR **138** (1961), no. 4, 796–798. (in Russian)

[94] L. N. Shevrin, *Semigroups whose lattice of subsemigroups is decomposable into a direct product of lattices*, 5th vsesoyuzn koll. po obsh. algebre (Abstracts), Novosibirsk, 1963, p. 59. (in Russian)

[95] L. N. Shevrin, *Semigroups of finite width*, Theory of semigroups and its applications, V. V. Vagner (editor), Izd. Saratov. Univ., 1965, pp. 325–351. (in Russian)

[96] L. N. Shevrin, *Strong bands of semigroups*, Izv. Vysh. Uchebn. Zaved. Mat. **6 (49)** (1965), 156–165. (in Russian)

[97] L. N. Shevrin, *On the isomorphism of semigroups with relatively complemented subsemigroups*, Mat. Zap. Ural. Gos. Univ. **5** (1965), 92–100. (in Russian)

[98] L. N. Shevrin, *Semigroups*, in: general Algebra, L. A. Skornyakov (editor), Nauka, Moscow, 1991, pp. 11–191. (in Russian)

[99] L. N. Shevrin, *Theory of epigroups* I, Mat. Sbornik **185** (1994), no. 8, 129–160. (in Russian)

[100] L. N. Shevrin, *Theory of epigroups* II, Mat. Sbornik **185** (1994), no. 9, 153–176. (in Russian)

[101] L. N. Shevrin and A. Ya. Ovsyanikov, *Semigroups and their subsemigroup lattices*, Semigroup Forum **27** (1983), no. 1, 1–154.

[102] L. N. Shevrin and A. Ya. Ovsyanikov, *Semigroups and their subsemigroup lattices*, Vol. 1, Izd. Ural. Univ., Sverdlovsk, 1990. (in Russian)

[103] L. N. Shevrin and A. Ya. Ovsyanikov, *Semigroups and their subsemigroup lattices*, Vol. 2, Izd. Ural. Univ., Sverdlovsk, 1991. (in Russian)

[104] O. Steinfeld, *On semigroups which are unions of completely 0-simple semigroups*, Check. Math. J. **16** (1966), 63–69.

[105] O. Steinfeld, *Quasi-ideals in rings and semigroups*, Akad. Kiadó, Budapest, 1978.

[106] R. Šulka, *The maximal semilattice decomposition of a semigroup, radicals and nilpotency*, Mat. časopis **20** (1970), 172–180.

[107] A. K. Sushkevich, *Über die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit*, Math. Ann. **99** (1928), 30–50.

[108] A. K. Sushkevich, *Theory of generalized groups*, GNTI. Kharkov–Kiev, 1937. (in Russian)

[109] G. Szász, *Théorie des treillis*, Akad. Kiadó, Budapest, et Dunod, Paris, 1971.

[110] T. Tamura, *The theory of construction of finite semigroups* I, Osaka Math. J. **8** (1956), 243–261.

[111] T. Tamura, *Operations on binary relations and their applications*, Bull. Amer. Math. Soc. **70** (1964), 113–120.

[112] T. Tamura, *Another proof of a theorem concerning the greatest semilattice decomposition of a semigroup*, Proc. Japan Acad. **40** (1964), 777–780.

[113] T. Tamura, *The theory of operations on binary relations*, Trans. Amer. Math. Soc. **120** (1965), 343–358.

[114] T. Tamura, *Maximal or greatest homomorphic image of a given type*, Canad. J. Math. **20** (1968), 264–271.

[115] T. Tamura, *The study of closets and free contents related to semilattice decomposition of semigroups*, in: Semigroups, K. W. Folley (editor), Acad. Press, New York, 1969, pp. 221–259.

[116] T. Tamura, *On Putcha's theorem concerning semilattice of Archimedean semigroups*, Semigroup Forum **4** (1972), 83–86.

[117] T. Tamura, *Note on the greatest semilattice decomposition of semigroups*, Semigroup Forum **4** (1972), 255–261.

[118] T. Tamura, *Semilattice congruences viewed from quasi-orders*, Proc. Amer. Math. Soc. **41** (1973), 75–79.

[119] T. Tamura, *Remark on the smallest semilattice congruence*, Semigroup Forum **5** (1973), 277–282.

[120] T. Tamura, *Quasi-orders, generalized Archimedeaness, semilattice decompositions*, Math. Nachr. **68** (1975), 201–220.

[121] T. Tamura, *Semilattice indecomposable semigroups with a unique idempotent*, Semigroup Forum **24** (1982), 77–82.

[122] T. Tamura and N. Kimura, *On decompositions of a commutative semigroup*, Kodai Math. Sem. Rep. **4** (1954), 109–112.

[123] T. Tamura and N. Kimura, *Existence of greatest decomposition of a semigroup*, Kodai Math. Sem. Rep. **7** (1955), 83–84.

[124] T. Tamura and J. Shafer, *Another proof of two decomposition theorems of semigroups*, Proc. Japan Acad. **42** (1966), 685–687.

[125] G. Thierrin, *Quelques propriétiés des sous-groupoides consistants d'un demi-groupe abélien*, C. R. Acad. Sci. Paris **236** (1953), 1837–1839.

[126] G. Thierrin, *Sur quelques propriétiés de certainess classes de demi-groupes*, C. R. Acad. Sci. Paris **241** (1955), 1192–1194.

[127] G. Thierrin, *Sur quelques décompositions des groupoides*, C. R. Acad. Sci. Paris **242** (1956), 596–598.

[128] G. Thierrin, *Sur la théorie des demi-groupes*, Comment. Math. Helv. **30** (1956), 211–223..

[129] V. V. Vagner, *Restrictive semigroups*, Izv. vysh. Uchebn. Zaved, Mat. **6 (31)** (1962), 19–27. (in Russian)

[130] P. S. Venkatesan, *On a class of inverse semigroups*, Amer. J. Math. **84** (1962), 578–582.

[131] P. S. Venkatesan, *On decompositions of semigroups with zero*, Math. Zeitsch. **92** (1966), 164–174.

[132] M. Yamada, *On the greatest semilattice decomposition of a semigroup*, Kodai Math. Sem. Rep. **7** (1955), 59–62.

[133] M. Yamada and N. Kimura, *Note on idempotent semigroups* II, Proc. Japan Acad. **34** (1958), 110–112.

University of Niš, Faculty of Philosophy, Ćirila i Metodija 2, 18000 Niš, Yugoslavia

E-mail address: `mciric@archimed.filfak.ni.ac.yu`

University of Niš, Faculty of Economics, Trg JNA 11, 18000 Niš, Yugoslavia

E-mail address: `root@eknux.eknfak.ni.ac.yu`

FILOMAT (Niš) 9:3 (1995), 427–448
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

WORD PROBLEMS FOR VARIETIES OF ALGEBRAS (A SURVEY)

Siniša Crvenković

1. Introduction

In the algebraic sense, a *word* is a formal expression, or finite string of symbols, built up in a more or less transparent way from certain primitive symbols, called constants, and certain other symbols which represent algebraic operations. A *word problem* is the problem of deciding in a given context, whether or not two given words represent the same element of the algebra. For such a problem to have a definite sense, certain assumptions must be made. Typically, one is concerned with some specific variety of algebras, such as groups or associative rings or the like. Word problems range all the way from triviality to algorithmic unsolvability.

The origin of the field of word problems may be traced back to R. Dedekind who in 1900 described the free modular lattice on three generators. At the begining of the century Axel Thue had formulated the word problem for finitely presented semigroups—or, as one now says, Thue systems—and solved various special cases of the general problem.

But negative results, unsolvability results in algebra, were impossible before the notion of an *algorithmically unsolvable problem* was formulated. In 1935–1936 A. Church and, independently, A. M. Turing gave equivalent precise mathematical definitions of the intuitive notion of algorithm. "*Turing machines*" and "*Church's Thesis*", led to Church's negative solution of the decision problem for first–order arithmetic; and, subsequently, to independent negative solutions by Church and Turing to Hilbert's *Entscheidungsproblem* for pure predicate logic. It seems that all unsolvability results in mathematics are, in final analysis, a translation of such classical results into a new setting.

In 1947 E. Post and A. A. Markov, independently, showed the word problem for semigroups unsolvable, constructing the bridge from logic to algebra.

Supported by Grant 0401B of RFNS through Math. Inst. SANU

This result was the first unsolvability result outside the foundations of mathematics.

Perhaps the most celebrated result is the unsolvability of the word problem for groups obtained by P. S. Novikov in 1952.

This paper surveys, unifies, and extends a number of results on the word problems in the context of *universal algebra.* From our point of view, the theory of free spectra is also a part of algebra dealing with words. A model–theoretic argument is used to prove unsolvability of many word problems for varieties of universal algebras. Most of the presented results are small contributions of the authors to the great topic and appeard, or will be published, elswhere.

2. Definitions

In the sequel, $\mathcal{L}$ denotes a first order language which contains the symbol of identity $\approx$, and has no relation symbols. If t_1, t_2 are terms of the language $\mathcal{L}$, then $t_1 \approx t_2$ is called an *equation* or an *identity.* The set of all identities of $\mathcal{L}$ is denoted by $Eq(\mathcal{L})$. If θ is a set of formulas of $\mathcal{L}$, then by $Mod(\theta)$ we denote the class of all algebras $\mathcal{A}$ such that $\mathcal{A} \models \theta$.

If G is a set of new contant symbols ($\mathcal{L} \cap G = \emptyset$), then by $\mathcal{L}_G$ we denote $\mathcal{L} \cup G$. Usually, a symbol from G and its interpretation is denoted by the same letter. Let $\mathcal{A}$ be an algebra and $G \subseteq A$. Then by $\mathcal{A}_G$ we denote the algebra $(\mathcal{A}, x)_{x \in G}$. If R is a set of identities in $\mathcal{L}_G$ with no variables, then (G, R) is called a *presentation* in $\mathcal{L}_G$.

Definition 2.1. *Let θ be a set of identities of $\mathcal{L}$, $\mathcal{V} = Mod(\theta)$ and (G, R) a presentation in $\mathcal{L}_G$. For an algebra $\mathcal{A}$ in $\mathcal{L}$ we say that it is* presented by *(G, R) in $\mathcal{V}$ if the following hold:*

(i) *$\mathcal{A}$ is generated by G;*

(ii) *$\mathcal{A}_G \models \theta \cup R$;*

(iii) *For any identity e in $\mathcal{L}_G$, with no variables, we have $\theta \cup R \models e$ provided $\mathcal{A} \models e$.*

If an algebra $\mathcal{A}$ is presented by (G, R) in $\mathcal{V}$, then we put $\mathcal{A} = \mathcal{P}_{\mathcal{V}}(G, R)$. For an algebra $\mathcal{B}$ we say that it is *finitely presented* in $\mathcal{V}$ if there are finite sets G and R such that $\mathcal{B}$ is presented by (G, R) in $\mathcal{V}$. Note that the algebra presented by (G, R) in $\mathcal{V}$ is unique up to isomorphism.

Example 2.2. *Let (G, R) be a presentation in $\mathcal{L}_G$. Let θ be a set of identities of $\mathcal{L}$ and $\mathcal{V}$ the variety defined by the set $\theta \cup R$. Then the free algebra $\mathcal{F}_{\mathcal{V}}(\emptyset)$ of the variety $\mathcal{V}$ on the empty set of free generators is an algebra presented by (G, R) in $\mathcal{V} = Mod(\theta)$.*

Let θ be a set of identities of $\mathcal{L}$, $\mathcal{V} = Mod(\theta)$, and $\mathcal{A}$ the algebra finitely presented by (G, R) in $\mathcal{V}$. *The word problem for $\mathcal{A} = \mathcal{P}_{\mathcal{V}}(G, R)$ in $\mathcal{V}$* asks if there is an algorithm to determine, for any identity e in $\mathcal{L}_G$ with no variables, whther or not $\mathcal{A} \models e$. If such an algorithm exists, the word problem is *solvable (decidable)*; otherwise it is *unsolvable (undecidable)*.

The following two options occur in the literature for what is meant by the solvability of the word problem for a variety $\mathcal{V}$:

(1) there is an algorithm which, given a finite presentation $\mathcal{P}_{\mathcal{V}}(G, R)$ solves the word problem for $\mathcal{P}_V(G, R)$ in $\mathcal{V}$;

(2) for each finite presentation $\mathcal{P}_V(G, R)$, there is an algorithm which solves the word problem for $\mathcal{P}_V(G, R)$ in $\mathcal{V}$.

We say that $\mathcal{V}$ has *uniformly solvable* word problem if (1) holds.

Varieties with uniformly solvable word problem include commutative semigroups and abelian groups, any finitely based locally finite or residually finite variety, and the variety of all algebras of a given finite type (see [21]).

Most of the examples which appear in the literature, of varieties with unsolvable word problem, provide a finite presentation for which the word problem is unsolvable. These include semigroups, groups and modular lattices.

In this paper we will apply the method of embedding to obtain several unsolvabilities of word problem.

3. Varieties with solvable but uniformly unsolvable word problems

Probably the first one who recognized the difference between the uniform solvability and solvability of the word problem was A. I. Mal'cev [27]. According to Benjamin Wells, A. Tarski was also interested in the existence of varieties with solvable but not uniformly solvable word problem.

An algebra $\mathcal{A}$ is *locally finite* if every finitely generated subalgebra is finite. A *variety $\mathcal{V}$ is lically finite* if every member of $\mathcal{V}$ is locally finite.

Let us recall some facts from mathematical logic. For an arbitrary first–order theory $\mathcal{K}$, we correlate with each symbol α of $\mathcal{K}$ a positive integer $\Gamma(\alpha)$, called the *Gödel number* of α. Thus, Γ is a one–one function from the set of symbols of $\mathcal{K}$, expressions of $\mathcal{K}$, and finite sequences of expressions of $\mathcal{K}$, into the set of positive integers.

A set of Gödel numbers is *recursive* if its characteristic function is a recursive function. Denote by $\mathcal{T}(x_1, x_2, \dots, x_n)$ the set of all n–ary terms in the language of a variety $\mathcal{V}$. According to Church's Thesis, an algebra $\mathcal{A}$ finitely presented by (G, R) in $\mathcal{V}$ has a *solvable word problem* if the set

$$\{\Gamma(p \approx q) | \, p, q \in \mathcal{T}(x_1, x_2, \dots, x_n),$$

$$p^{\mathcal{A}}(g_1, g_2, \dots, g_n) = q^{\mathcal{A}}(g_1, g_2, \dots, g_n),\ n \in N,\ g_1, g_2, \dots, g_n \in G\}$$

is recursive.

Proposition 3.1. *Let $\mathcal{B}$ be a finite algebra of a finite type. If $(c_1, c_2, \dots, c_n) \in B^n$ is a fixed n-tuple, then the set*

$$S = \{\Gamma(p \approx q) |\, p, q \in \mathcal{T}(x_1, x_2, \dots, x_n), p^{\mathcal{B}}(c_1, c_2, \dots, c_n) = q^{\mathcal{B}}(c_1, c_2, \dots, c_n)\}$$

is recursive.

Proof. The proof is straightforward. □

Proposition 3.2. *If a finitely presented algebra $\mathcal{A}$ is finite, then the word problem for $\mathcal{A}$ is solvable.*

Proof. Follows from the previous prposition □

Let $e, e_1, e_2, \dots, e_n$ (where $n \in N$) be identities of $\mathcal{L}$. Then the formula $e_1 \wedge e_2 \wedge \dots \wedge e_n \rightarrow e$ is called a *quasi-identity*. The set of all quasi-identities of $\mathcal{L}$ is denoted by $Q(\mathcal{L})$. If $\mathcal{K}$ is a class of algebras in a language $\mathcal{L}$ then $Q(\mathcal{K}) = \{q \in Q(\mathcal{L}) |\, \mathcal{K} \models q\}$. *The problem of quasi-identities for a class $\mathcal{K}$* asks if the set $Q(\mathcal{K})$ is recursive (i.e. the set of Gödel numbers of the elements of $Q(\mathcal{K})$). If so, the problem of quasi-identities is *solvable*; otherwise it is *unsolvable*.

By the Church's Thesis, the problem of solvability (decidability) of the problem of quasi-identities for a class $\mathcal{K}$ is equivalent to the problem of the existence of an algorithm which, for every quasi-identity $q \in Q(\mathcal{L})$, decides whether or not $\mathcal{K} \models q$.

Remark 3.3. *Let θ be a set of formulas of $\mathcal{L}$ and $\mathcal{K} = Mod(\theta)$. Then we have*

$$Q(\mathcal{K}) = \{q \in Q(\mathcal{L}) |\, \theta \vdash q\}.$$

Therefore, the problem of quasi-identities for such a class $\mathcal{K}$ is solvable iff there exists an algorithm which, for any $q \in q(\mathcal{K})$, decides whether or not $\theta \vdash q$.

The following proposition is a part of the folklore.

Proposition 3.4. *Let θ be a set of identities in some language $\mathcal{L}$ and let $\mathcal{K} = Mod(\theta)$. Then $\mathcal{K}$ has uniformly solvable word problem iff the problem of quasi-identities for $\mathcal{K}$ is solvable.*

Similarly to the case of quasi-identities, if $\mathcal{K}$ is a class of algebras in a language $\mathcal{L}$, then $Eq(\mathcal{K}) = \{e \in Eq(\mathcal{L}) |\, \mathcal{K} \models e\}$. The set $Eq(\mathcal{K})$ is called the *equational theory* of the class $\mathcal{K}$. We say that equational theory of a class $\mathcal{K}$ is *decidable (solvable)* if the set $Eq(\mathcal{K})$ is recursive (i.e. the set of Gödel numbers of the elements of $Eq(\mathcal{K})$ is recursive).

Proposition 3.5. *Let $\mathcal{V}$ be a locally finite variety of a finite type. Then $\mathcal{V}$ has a solvable word problem.*

Proof. Follows from Proposition 3.2. □

As we all know, the set of all recursive functions is countable. For a class of algebras $\mathcal{K}$, let $H_{Eq(\mathcal{K})}$ denote the characteristic function of the set $\{\Gamma(p \approx q) | \, p \approx q \in Eq(\mathcal{K})\}$. If $\mathcal{V}_1$ and $\mathcal{V}_2$ are two different varieties, then $H_{Eq(\mathcal{V}_1)} \neq H_{Eq(\mathcal{V}_2)}$. Therefore we have

Proposition 3.6. *If a class of varieties of the same type has uncountably many elements, then there is a variety from that class having undecidable equational theory.*

Corollary 3.7. *Let $\mathcal{V}$ be a locally finite variety of a finite type. If $\mathcal{V}$ has uncountably many subvarieties, then $\mathcal{V}$ has solvable but not uniformly solvable word problem.*

Proof. From Proposition 3.6. it follows that $\mathcal{V}$ has undecidable equational theory. This, of course, implies that the problem of $Q(\mathcal{V})$ is unsolvable which is equivalent with the uniform unsolvability of the word problem for $\mathcal{V}$. □

The student of A. Tarski, Benjamin Wells, in his Ph. D. thesis at the University of Berkeley (1982), presented the first examples of varieties having solvable but uniformly unsolvable word problem. This result appeared first in [36], later as Theorem 11.17. in [37], and recently as Theorem 1.7. in [38]. The last result is almost identical to our result even though they were obtained independently. Our construction is primarily based on an example appearing in the paper of Mekler, Nelson and Shelah [30].

Theorem 3.8 ([11]). *In a language of the type* $(2, 0, 1, 1)$ *there exists a variety having solvable word problem and undecidable equational theory. This variety is axiomatized by the following identities*

$$\begin{array}{ccc}
x \cdot 0 \approx 0 & f(0) \approx 0 & h(h(x)) \approx h(x) \\
x^2 \approx 0 & f(f(x)) \approx f(x) & h(x)y \approx 0 \\
xy \approx yx & f(x \cdot y) \approx 0 & f(h(x)) \approx h(x) \\
x \cdot (y \cdot z) \approx (x \cdot y) \cdot z & h(0) \approx 0 &
\end{array}$$

$$h^k(f(x_1)f(x_2)\cdots f(x_{\varphi(k)})) \approx 0,$$

where $\varphi(k)$ is a primitive recursive function such that $X = \{\varphi(k) | \, k \in N\}$ is a recursively enumerable nonrecursive set.

In [12] we proved the following

Theorem 3.9 ([12]). *In a language of the type* $(2,0,1,1)$ *there exists an infinite (isomorphic to* $\langle\omega,\leq\rangle$*) chain of varieties with solvable word problems and undecidable equational theories.*

The varieties from Theorem 3.9. are constructed in the following way. Let the variety defined in Theorem 3.8. be denoted by $\mathcal{V}_1$. Denote by ε_n, $n \geq 2$ the identity in $\{\cdot, f, h, 0\}$ of the form:

$$\varepsilon_n : f(x_1 f(x_2 \ldots f(x_n)\ldots)) \approx 0.$$

The variety whose set of definitional identities is same as the one for $\mathcal{V}_1$, with the exception of $f(x \cdot y) \approx 0$ being replaced by (ε_n), will be denoted by $\mathcal{V}_n$. Obviously,

$$\mathcal{V}_1 \subseteq \mathcal{V}_2 \subseteq \ldots \subseteq \mathcal{V}_n \subseteq \ldots$$

It is easy to prove that all the inclusions are strict.

Definition 3.10. *Ternary discriminator on a set* A *is the function*

$$t_A(a,b,c) = \begin{cases} c, & \textit{for } a = b \\ a, & \textit{otherwise.} \end{cases}$$

Definition 3.11. *A variety* $\mathcal{V}$ *in a language* $\mathcal{L}$ *is said to be a* discriminator *variety if there exists a term in* $\mathcal{L}$ *inducing ternary discriminator on the universe of every subdirectly irreducible algebra in* $\mathcal{V}$.

Following the idea of Ross Willard we were able to prove

Theorem 3.12 ([12]). *There exists a recursively axiomatized discriminator variety in a finitary language with solvable word problem and undecidable equational theory.*

Discriminator varieties are only a part of wider class of so called *EDPC* varieties, arising in the algebraization of different logical systems.

Corollary 3.13. *There exists a recursively based EDPC variety in a finitary language having solvable word problem and undecidable equational theory.*

This result rules out the possibility of obtaining the converse of the following result, due to Blok and Pigozzi:

Theorem 3.14 ([2]). *Let* $\mathcal{V}$ *be an EDPC variety having decidable equational theory. Then* $\mathcal{V}$ *has solvable word problem.*

B. Wells proved that there is a variety of a finite type, with a base of not more than 350 000 axioms, having solvable but not uniformly solvable word problem. Mekler, Nelson and Shelah in [30] also presented a finitely based variety of a finite type having the same properties. However these examples seem to be too complicated and are not from any well known class of algebras. Also their varieties have decidable equational theories. The following problem is still open

Problem 3.15. *Is there a finitely based variety with solvable word problem having undecidable equational theory?*

4. Embedding

There are several undecidability proofs in the literature that use the result of Post and Markov on the existence of a finitely presented semigroup with unsolvable word problem. For example, in [26] we proved unsolvability of the word problem for the variety of relation algebras. We used the following result of Kogalovskii [25].

Proposition 4.1. *If $\mathcal{K}_1$ and $\mathcal{K}_2$ are classes of algebras such that*

(i) $\mathcal{K}_1 \subseteq \mathcal{K}_2$,

(ii) *every algebra from $\mathcal{K}_2$ is embeddable into an algebra from $\mathcal{K}_1$,*

then the theories of quasi-identitics of $\mathcal{K}_1$ and $\mathcal{K}_2$ are the same.

Proof. See [27] and [25]. □

Corollary 4.2. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be varieties of algebras such that $\mathcal{K}_1 \subseteq \mathcal{K}_2$ and every algebra from $\mathcal{K}_2$ is embeddable into an algebra from $\mathcal{K}_1$. Then, $\mathcal{K}_1$ and $\mathcal{K}_2$ have equivalent uniform word problems.*

Proof. Direct consequence of Proposition 4.1. □

So, if $\mathcal{K}_2$ is the class of all semigroups, and $\mathcal{K}$ is some class of algebras, such that some reduct $\mathcal{K}_1$ of $\mathcal{K}$ satisfies conditions of Corollary 4.2., then $\mathcal{K}$ has uniformly unsolvable word problem. But, this is not enough to obtain the result about the solvability of the word problem for $\mathcal{K}$. The following theorem gives something more than Corollary 4.2.

Theorem 4.3 ([5]). *Let $\mathcal{V}$ be a variety with an associative operation $*$ in its language. If every semigroup can be embedded into the $*$-reduct of some algebra from $\mathcal{V}$, then $\mathcal{V}$ has unsolvable word problem.*

The proof of Theorem 4.3. has been given in [5] and [26]. If we analyze this proof, we see that condition that $\mathcal{V}$ has to have semigroups in it is not necessary. The same goes for any variety with unsolvable word problem.

Theorem 4.3. enables us to obtain several undecidability results in a uniform way. For example, this theorem gives results on unsolvability of the word problem for some varieties which are obtained from the algebras of binary relations.

For an algebra $\mathcal{A} = (A, F)$ we say that it is an *algebra of binary relations* if $A = \mathcal{P}(S^2)$, for some set S, and F is a set of operations on binary relations.

Let $\mathcal{R}_F$ be a class of algebras of binary relations such that F contains the operation of relative multiplication of binary relations "$\circ$". Then the variety $HSP(\mathcal{R}_F)$ has unsolvable word problem. For example, we have unsolvability of the word problem for the following:

(a) variety generated by the class of all semigroups of binary relations ($F = \{\circ\}$);
(b) variety generated by the class of all involutive semigroups of binary relations ($F = \{\circ, ^{-1}\}$);
(c) (representable) relation algebras of Tarski ($F = \{\cup, \cap, ^{-}, \circ, ^{-1}, \Delta\}$);
(d) relation algebras of Jónsson ($F = \{\cap, \circ, ^{-1}, \Delta\}$);
(e) Kleene algebras ($F = \{\cup, \emptyset, \circ, ^{-1}, \Delta, ^{\mathrm{rtc}}\}$), ($F = \{\cup, \emptyset, \circ, ^{-1}, \Delta\}$) and ($F = \{\cup, \emptyset, \circ, \Delta, ^{\mathrm{rtc}}\}$);
(f) no special name (e.g. $F = \{\cup, \circ\}$, $F = \{\cap, \circ\}$).

Theorem 4.3. can easily be applied in the following cases, thus having unsolvable word problems

(g) rings ($F = \{+, \cdot, 0, 1\}$),
(h) involutive semigroups ($F = \{\cdot, ^{-1}\}$),
(i) semirings ($F = \{+, \cdot\}$),
(j) variety generated by Baer *–semigroups ($F = \{\cdot, ^*\}$),
(k) variety generated by the class of all simple semigroups,
(l) variety generated by the class of all bisimple semigroups,
(m) inverse semigroups ($F = \{\cdot, ^{-1}\}$),
(n) rings with involution ($F = \{+, \cdot, ^*\}$).

5. Partial algebras

Let A be a set and $B \subseteq A^n$. Then $f : B$ — A is called a *partial operation* on A of type n. A *partial algebra* $\mathcal{A}$ is a pair (A, F), where A is a nonempty set and F is a collection of partial operations on A. In our considerations F will always be a finite set.

Let $\mathcal{A}$ be a partial algebra. Denote by $\Delta(\mathcal{A})$ the positive diagram of $\mathcal{A}$:

$$\Delta(\mathcal{A}) = \{ f(a_1, a_2, \dots, a_n) = a \mid f \in F,\ a_1, a_2, \dots, a_n \in A,$$

$f(a_1, a_2, \dots, a_n)$ is defined and equals a in $\mathcal{A}$.}

Of course, if $\mathcal{A}$ is finite, then $\Delta(\mathcal{A})$ is finite.

Suppose that $\mathcal{A}$ and $\mathcal{B}$ are partial algebras. $\varphi : A \to B$ is called a *homomorphism* of $\mathcal{A}$ into $\mathcal{B}$ if, whenever $f(a_1, a_2, \dots, a_n)$ is defined, then so is $f(\varphi(a_1), \varphi(a_2), \dots, \varphi(a_n))$ and

$$\varphi(f(a_1, a_2, \dots, a_n)) = f(\varphi(a_1), \varphi(a_2), \dots, \varphi(a_n)).$$

A homorphism is an *isomorphism* if φ is a bijection.

Let $\mathcal{A} = (A, F)$ be a partial algebra and let $\emptyset \neq B \subseteq A$. Then

(i) $\mathcal{B}$ is a *subalgebra* of $\mathcal{A}$ if it is closed under all the operations in $\mathcal{A}$ i.e. if $b_1, b_2, \dots, b_n \in B$ and $f(b_1, b_2, \dots, b_n)$ is defined in $\mathcal{A}$, then $f(b_1, b_2, \dots, b_n) \in B$.

(ii) $\mathcal{B}$ is a *relative subalgebra* of $\mathcal{A}$ if for all $f \in F$ and all $b_1, b_2, \dots, b_n$, $b \in B$, we have:
$f(b_1, b_2, \dots, b_n)$ is defined and equals b iff $f(b_1, b_2, \dots, b_n)$ is defined in $\mathcal{A}$ and $f(b_1, b_2, \dots, b_n) = b$ in $\mathcal{A}$.

It is not dificult to give an example of a partial algebra $\mathcal{A}$ and a set $B \subseteq A$, such that B is the carrier of some relative subalebra of $\mathcal{A}$ but not the carrier of any subalgebra in $\mathcal{A}$.

Let $\mathcal{K}$ be a class of algebras, A nonempty set, and F a set of partial operations on A. Then $\mathcal{A} = (A, F)$ is a *partial $\mathcal{K}$–algebra* if (A, F) is a relative subalgebra of an algebra $\mathcal{B}$ in $\mathcal{K}$. For example, if $\mathcal{L}$ is the class of all lattices, then, a partial algebra $\mathcal{A}$ is a partial $\mathcal{L}$–algebra (or simply, partial lattice) if $\mathcal{A}$ is a relative subalgebra (or relative sublattice) of some lattices.

Definition 5.1. *Let $\mathcal{K}$ be a class of algebras and let $\mathcal{A}$ be a partial algebra. The algebra $FK(\mathcal{A})$ is called the algebra freely generated by the partial algebra $\mathcal{A}$ over $\mathcal{K}$ if the following conditions are satisfied:*

(i) *$FK(\mathcal{A})$ is generated by A' and there exists an isomorphism $\chi : A' \to A$ between $\mathcal{A}'$ and $\mathcal{A}$, where $\mathcal{A}'$ is a relative subalgebra of $FK(\mathcal{A})$;*

(ii) *If φ is a homorphism of $\mathcal{A}$ into $\mathcal{C} \in \mathcal{K}$, then there exists a homorphism ψ of $FK(\mathcal{A})$ into $\mathcal{C}$ such that ψ is an extension of $\chi\varphi$.*

It is not difficult to prove that $FK(\mathcal{A})$ is unique up to isomorphism and, if $\mathcal{A}$ is an algebra from $\mathcal{K}$, then $FK(\mathcal{A}) \cong \mathcal{A}$. Also, it is well known that if $\mathcal{K}$ is an equational class, then $FK(\mathcal{A})$ exists if $\mathcal{A}$ is (isomorphic to) a relative subalgebra of an algebra $\mathcal{B}$ in $\mathcal{K}$. In other words, in the case of equational classes $\mathcal{K}$, $FK(\mathcal{A})$ exists if $\mathcal{A}$ is a partial $\mathcal{K}$–algebra.

For example, if $\mathcal{A}$ is a partial lattice, then $FL(\mathcal{A})$ always exists. It is well known (see [22]) these lattices (of the form $FL(\mathcal{A})$) are the lattices that can be described by finitely many generators and finitely many relations.

Proposition 5.2 ([6]). *Let $\mathcal{K} = Mod(\Sigma)$ be a variety, $\mathcal{A}$ a partial algebra. Then,*

$$FK(\mathcal{A}) \cong \mathcal{P}_{\mathcal{K}}(A, \Delta(\mathcal{A})).$$

Proof. See [6]. □

Let $\mathcal{K}$ be a class of algebras in a language $\mathcal{L}$, and let $\mathcal{A}$ be a partial $\mathcal{K}$–algebra. The *problem of partial $\mathcal{K}$–algebra $\mathcal{A}$* asks if there is an algorithm to determine for any identity $p \approx q \in Eq(\mathcal{L} \cup G)$, with no variables, whether or not $FK(\mathcal{A}) \models p \approx q$.

The *problem of partial $\mathcal{K}$–algebras* asks if there is an uniform algorithm which for any finite partial $\mathcal{K}$–algebra $\mathcal{A}$, and any identity $p \approx q \in Eq(\mathcal{L} \cup G)$, with no variables, decides whether or not $FK(\mathcal{A}) \models p \approx q$.

Proposition 5.3 ([6]). *Let $\mathcal{K}$ be a variety in a language $\mathcal{L}$. If $\mathcal{K}$ has a uniformly solvable word problem, then the problem of partial $\mathcal{K}$–algebras is solvable too.*

Proof. Let $\mathcal{A}$ be a finite partial $\mathcal{K}$–algebra, $p \approx q \in Eq(\mathcal{L} \cup G)$, with no variables. Then, because of Proposition 5.2., $FK(\mathcal{A}) \cong \mathcal{P}_{\mathcal{K}}(A, \Delta(\mathcal{A}))$, so that

$$FK(\mathcal{A}) \models p \approx q \quad iff \quad \mathcal{P}_{\mathcal{K}}(A, \Delta(\mathcal{A})) \models p \approx q.$$

Hence, directly from the algorithm for the solution of the word problem, we obtain an algorithm for the solution of the problem of partial algebras. □

Denote by $|t|$ the length of a term t (i.e. the number of symbols in t). We can formulate two rules:

(α) If a set of identities I contains an identity of the form $p \approx q$, where p and q are terms $|p| = |q| = 1$, then we take out this identity from the set I and in all the other identities we replace the symbol q by p.

(β) If a set of identities I contains some identities of the form $t \approx t_1$, $t \approx t_2$, where $t_1 \neq t_2$, then from I we take out the identity $t \approx t_2$ and in all the other identities we replace the symbol t_2 by t_1.

Let I be a set of identities. Denote by $\alpha(I)$ the set of identities which appear from I, if the rule (α) is applied, and by $\beta(I)$ if the rule (β) is applied.

We say that the set of identities I is *α–pure* if $\alpha(I) = I$. Analogously, I is *β–pure* if $\beta(I) = I$. Obviously, if I is a finite set of identities, then there are natural numbers, m, n such that the set $\alpha^n(I)$ is α–pure and set $\beta^m(I)$ is β–pure.

Definition 5.4 ([6]). *Let $\mathcal{K}$ be a variety in a language $\mathcal{L}$ and (A, R) some finite presentation in $\mathcal{K}$. Then,*

(1) *If t is a term in $\mathcal{L}$, then by $Sub(t)$ we denote the set of all the subterms of t.*

(2) $Sub(R) = \bigcup\{Sub(t) \mid (\exists s)(s \approx t \in R \vee t \approx s \in R)\}$.

(3) $A' = \{C_\sigma \mid \sigma \in Sub(R)\} \cup A$.

(4) *Define the mapping* $\varphi : Sub(R) - Eq(\mathcal{L} \cup A')$ *in the following way:*

 (i) *If* $|t| = 1$, *then* $\varphi(t)$ *is* $t \approx C_t$*;*

 (ii) *If* $t = f(t_1, t_2, \dots, t_n)$, *where* f *is an* n*–ary function symbol and* $t_1, t_2, \dots, t_n$ *are terms, then* $\varphi(t)$ *is* $t = f(C_{t_1}, C_{t_2}, \dots, C_{t_n}) \approx C_t$.

(5) *Define the set* R' *as*

$$R' = \varphi[Sub(R)] \cup \{C_p \approx C_q \mid |p| = 1 \text{ and } p \approx q \in R\} \cup$$

$$\cup \{f(C_{t_1}, C_{t_2}, \dots, C_{t_n}) \approx C_q \mid p = f(t_1, t_2, \dots, t_n) \text{ and } p \approx q \in R\},$$

where $\varphi[Sub(R)] = \{\varphi(t) \mid t \in Sub(R)\}$.

Note that if $t \in Sub(R)$ and $|t| = 1$, then $t \in A$ or t is a contant in $\mathcal{L}$ and the set R' is a set of identities, in the language $\mathcal{L} \cup A'$, with no variables.

Let $\mathcal{A} = (A, R)$ be a finite presentation in a variety $\mathcal{K}$. Let n be a finite natural number such that $\alpha^n(R')$ is α–pure and m be a natural number such that $\beta^m(\alpha^n(R'))$ is β–pure. Then let $R^* = \beta^m(\alpha^n(R'))$ and A^* be the set of all these symbols from $A' \cup const(\mathcal{L})$ which appear in the identities of R^*.

Theorem 5.5 ([6]). *Let* $\mathcal{A} = (A, F)$ *be a finite presentation in a variety* $\mathcal{K} = Mod(\Sigma)$, *in a language* $\mathcal{L}$, *and let* $\mathcal{A}^*$ *be a* $\mathcal{K}$*–partial algebra. Then, if the problem of the partial algebra* $\mathcal{A}^*$ *in* $\mathcal{K}$ *is solvable, the word problem for* $\mathcal{A} = (A, R)$ *in* $\mathcal{K}$ *is solvable, too.*

Proof. See [6]. □

6. Free spectra

Let $\mathcal{V}$ be a variety of type F. The cardinality of the free algebra over n generators ($n \geq 0$) in $\mathcal{V}$ is denoted by $f_n(\mathcal{V})$. The sequence of cardinal numbers

$$f(\mathcal{V}) = \langle f_n(\mathcal{V}) \rangle_{n \geq 0} = \langle f_0(\mathcal{V}), f_1(\mathcal{V}), \dots, f_n(\mathcal{V}), \dots \rangle$$

is called the *free spectrum* of $\mathcal{V}$.

Let $\mathcal{A} = (A, F)$ be an algebra of type F. Every term of the type F in n variables $x_1, x_2, \dots, x_n$ ($n \geq 0$) defines an n–ary operation $t : A^n \to A$ in a natural way. These operations are called *n–ary term operations*. The number of differnt n–ary term operations on $\mathcal{A}$ is denoted by $s_n(\mathcal{A})$. If $\mathcal{A}$ generates the variety $\mathcal{V}$ then obviously $f_n(\mathcal{V}) = s_n(\mathcal{A})$ for all $n \geq 0$. The investigation of free spectra of specific varieties may have started with R. Dedekind [1900]. The *Dedekind problem*, the determination of the free spectrum of the variety $\mathcal{D}$ of distributive lattices, is still open.

In group theory, the famous Burnside problem asks whether $f_n(\mathcal{G}_m)$ is always finite, where $\mathcal{G}_m$ is the variety of groups of exponent m. This was solved in the negative by S. I. Adjan and P. S. Novikov. They proved that f_2 is infinite, for instance, for $m \geq 4381$. The choice of m was improved to $m \geq 115$ for odd exponents m.

A major problem of this field is to determine *what sequences can be represented as the free spectrum* of a variety.

If $\mathcal{V}$ is a variety and all $f_n(\mathcal{V})$ are finite, then $\mathcal{V}$ is a locally finite variety. In what follows we are going to consider only locally finite varieties.

Is there a variety $\mathcal{V}$ having $f_0(\mathcal{V}) = 0$, $f_1(\mathcal{V}) = 10$ and $f_2(\mathcal{V}) = 18$? To answer this question we use the concept of s_n*–sequence* (or p_n*–sequence* in the literature). Denote

$$s(\mathcal{A}) = \langle s_0(\mathcal{A}), s_1(\mathcal{A}), \ldots, s_n(\mathcal{A}), \ldots \rangle.$$

For a nontrivial variety $\mathcal{V}$, we define s_n–sequence of $\mathcal{V}$ as the s_n–sequence of $\mathcal{F}_\mathcal{V}(\omega)$, the free algebra on ω generators in $\mathcal{V}$. $s_0(\mathcal{A})$ is the number of unary contant term operations and $s_1(\mathcal{A}) \geq 1$.

The following two formulas connect the free spectrum and the s_n–sequence for an algebra $\mathcal{A}$:

(C1)

$$f_n(\mathcal{A}) = \sum_{k=0}^{n} \binom{n}{k} s_k(\mathcal{A});$$

(C2)

$$s_n(\mathcal{A}) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_k(\mathcal{A}).$$

Back to the variety $\mathcal{V}$ with $f_0(\mathcal{V}) = 0$, $f_1(\mathcal{V}) = 10$ and $f_2(\mathcal{V}) = 18$. By formula (C2), we have $s_2 = f_0 - 2f_1 + f_2 = -2$, a contradiction. So a necessary condition for the representability of a sequence as a free spectrum of an algebra is that the corresponding s_n–sequence be nonnegative.

Let $\mathbf{S}$ be a semilattice with more than one element. Using formula (C1) we see that $f_n(\mathbf{S}) = 2^n - 1$, for all $n \geq 0$. Let $\mathcal{S}$ be the variety of semilattices then

$$s(\mathcal{S}) = \langle 0, 1, 1, \ldots, 1, \ldots \rangle.$$

Let us see some examples which explain the flavor of the field.

Proposition 6.1 ([15]). *Let $\mathcal{A}$ be an idempotent groupoid with $s_3(\mathcal{A}) < 6$. Then $\mathcal{A}$ is equivalent to a semilattice, a diagonal semigroup, a groupoid with $s_n(\mathcal{A}) = n$, or a distributive Stainer quasigroup.*

Proof. See [15]. □

Proposition 6.2. *If $\mathcal{A}$ has two commutative binary term operations, then*

(1) $s_3(\mathcal{A}) \geq 9$ *([17]);*
(2) $s_n(\mathcal{A}) \geq 3 + n!$ *for all* $n > 3$.

Proof. See [17]. □

The following result of J. Dudek seems especially attractive:

Proposition 6.3. *Let d_n be the n-th Dedekind's number, that is,* $d_n = |\mathcal{F}_\mathcal{D}(n)|$, *where $\mathcal{D}$ is the variety of ditributive lattices. For a variety $\mathcal{V}$,* $f_n(\mathcal{V}) = d_n$ *holds, for all $n \geq 0$, iff $\mathcal{V}$ is equvivalent to $\mathcal{D}$.*

Proof. See [20]. □

Among other things, J. Dudek proved the following

Theorem 6.4. *Let $(A, +, \cdot)$ be an idempotent commutative algebra of the type $(2,2)$ such that $+$ and $\cdot$ are distinct. Then*

(i) *$(A, +, \cdot)$ is a distributive lattice iff* $s_3((A, +, \cdot)) = 9$.
(ii) *If $(A, + \cdot)$ is a bisemilattice, then $(A, + \cdot)$ is a lattice iff* $s_2((A, + \cdot)) = 2$.

There is no bisemilattice $(A, + \cdot)$ for which $s_2((A, + \cdot)) = 3$.
$(A, +, \cdot)$ is a nondistributive modular lattice iff $s_3((A, + \cdot)) = 19$.

Proof. See [16], [17]. □

In a joint paper with J. Dudek we investigated so called *rectangular* groupoids

Definition 6.5. *A groupoid $(G, \cdot)$ is called rectangular (right) if it satisfies the following laws*

$$x^2 \approx x,$$

$$(xy)z \approx xz.$$

Proposition 6.6 ([13]). *For any rectangular groupoid $(G, \cdot)$, being not a semigroup, we have*

$$s_n((G, \cdot)) \geq n^2, \quad \textit{for } n \geq 3.$$

Proof. See [13]. □

This estimation is the best posible because we have

Theorem 6.7 ([13]). *Let $(G, \cdot)$ be a rectangular groupoid. Then the following conditions are equivalent*

(i) *$(G, \cdot)$ is not a semigroup and satisfies*

$$x(y(zu)) \approx x(z(yu));$$

(ii) *$s_n((G,\cdot)) = n^2$, for all n;*
(iii) *$s_4((G,\cdot)) = 16$.*

Proof. See [13]. □

In [9] we proved the following and therefore, solving the Problem 25. in [23].

Theorem 6.8 ([9]). *Let $\mathcal{V}$ be a variety of semigroups. Then $s_n(\mathcal{V}) = n^2$, for all $n \geq 0$, iff $\mathcal{V}$ is the variety of normal bands.*

Proof. See [9]. □

From Theorem 6.8. and by duing some technical calculations we were able to prove

Theorem 6.9 ([8]). *Let $\mathcal{G}$ be a groupoid. Then $S_n(\mathcal{G}) = n^2$, for all $n \geq 0$, iff one of the following conditions hold*

(i) *$\mathcal{G}$ generates the variety of normal bands;*
(ii) *$\mathcal{G}$ is not a semigroup and satisfies*

$$\begin{aligned} xx &\approx x \\ x(yz) &\approx xz \\ ((xy)z)u &\approx ((xz)y)u; \end{aligned}$$

(iii) *$\mathcal{G}$ is not a semigroup and satisfies*

$$\begin{aligned} xx &\approx x \\ (xy)z &\approx xz \\ x(y(zu)) &\approx x(z(yu)). \end{aligned}$$

Proof. See [8]. □

E. Marczewski formulated in [28] the problem of representability of s_n–sequences by algebras. He and his colleagues in Wroclaw considered many associated problems.

If one considers semigroups, the following problem can be formulated.

Problem 6.10. *Characterize s_n–sequences for the class of semigroups.*

We may start with the representability of sequences

$$s_\alpha =< 0, \alpha, \alpha, ..., \alpha, ... > \quad \alpha \in N$$

in the variety of semigroups.

Proposition 6.11. *If a semigroup $\mathbf{S}$ has $s_0(\mathbf{S}) = 0$ and $s_1(\mathbf{S}) = \alpha$, $\alpha > 0$, then the following hold:*

(i) *$x, x^2, ..., x^\alpha$ are different essentially unary term operations;*
(ii) *$\mathbf{S}$ satisfies $x^{\alpha+1} \approx x^\beta$ for some $\beta \in \{1, 2, ..., \alpha\}$;*

(iii) **S** *satisfies* $x^\gamma \approx x^\delta$ *iff* $\gamma, \delta \geq \beta$ *and* $\gamma \equiv \delta$ *(mod* $\alpha + 1 - \beta$*);*

(iv) *If* p, q *are two terms having lenghts* l_p, l_q *such that* $l_p \neq l_q$ *and* $l_p, l_q \leq \alpha$, *then* $p \neq g$. *Specially, all the terms* xy, xy^2, ..., $xy^{\alpha-1}$ *are different.*

(v) *If a semigroup* **S** *has an essentially* n*-ary term operation, then the term* $x_1x_2\cdots x_n$ *induces essentially* n*-ary term operation.*

Proof. Follows immediately. □

Proposition 6.12. *If a semigroup* **S** *satisfies*

$$xy \approx yx,\ xy^2 \approx x^2y,\ x^{\alpha+1} \approx x^\beta,\ (\alpha \geq \beta \geq 0),$$

then every nontrivial n*-ary term operation is equal to one of the following*

$$x_1x_2\cdots x_{n-1}x_n,\quad x_1x_2\cdots x_{n-1}x_n^2,\quad \ldots,\quad x_1x_2\cdots x_{n-1}x_n^\alpha.$$

Proof. Straightforward. □

It is easy to demonstrate that a semigroup has $< 0, 1, 1, ..., 1, ... >$ as the s_n sequence iff it is nontrivial semilattice. For the case $< 0, 2, 2, ..., 2, ... >$ we have the following.

Proposition 6.13. *A semigroup* **S** *has the* s_n*-sequence* $< 0, 2, 2, ..., 2, ... >$ *iff* **S** *generates the variety of semigroups determined by the identities*

$$x^3 \approx x^2,\ \ xy \approx yx,\ \ xy^2 \approx x^2y.$$

Proof. ($\rightarrow$). Let **S** be a semigroup having $< 0, 2, 2, ..., 2, ... >$ as the s_n-sequence. If **S** satisfies $x^3 \approx x$, then, because of $xy^2 = x^2y \Rightarrow x^5 = x^4$, it follows that xy, xy^2, x^2y are three different essentially binary term operations. Hence, **S** satisfies $x^3 \approx x^2$. If **S** is a non commutative semigroup, then xy, yx are only essentially binary term operations of **S**. The term xyz is essentially 3–ary (Proposition 6.12.) so that from $s_3(\mathbf{S}) = 2$ it follows that **S** satisfies $xyz = zxy = yzx$ which implies $xy^2 = y^2x = yxy$. But then $x^2y^2 = y^2x^2$ which is a contradiction since **S** does not have a commutative binary term operation. Therefore, **S** is a commutative semigroup. $s_2(\mathbf{S}) = 2$ implies that both of essentially binary term operation are commutative. Specially, $xy^2 = x^2y$. Therefore **S** belongs to the variety given by

$$x^3 \approx x^2,\ \ xy \approx yx,\ \text{ and }\ xy^2 \approx x^2y.$$

If $\mathcal{A}$ is an arbitrary semigroup from the variety above, then Proposition 6.13. implies that every essentially n–ary term operation is equal to $x_1\cdots x_{n-1}x_n$ or $x_1\cdots x_{n-1}x_n^2$. Hence, $s_2(\mathcal{A}) \leq 2$. Since $s_n(\mathbf{S}) = 2$ for all $n \geq 1$, it follows that **S** generates the variety.

($\leftarrow$). It is sufficient to prove that the free semigroup **F** in the variety $x^3 \approx x^2$, $xy \approx yx$, $xy^2 \approx x^2y$ over an infinite set of generators has

$< 0, 2, 2, ..., 2, ... >$ as the s_n–sequence. It was demonstrated above that $s_n(\mathbf{F}) \leq 2$ for all $n \geq 1$. Obviously, both of terms $x_1 \cdots x_{n-1}\ x_n$ and $x_1 \cdots x_{n-1}x_n^2$ induce essentially n–ary term operation in $\mathbf{F}$. $\mathbf{F}$ satisfies $x_1 \cdots x_{n-1}x_n \approx x_1 \cdots x_{n-1}x_n^2$ iff this identity can be deduced from the defining identities. However, we can only apply $xy \approx yx$ to $x_1 \cdots x_{n-1}x_n$ and hence obtain a permutation of it. Therefore, $s_n(\mathbf{F}) = 2$ for all $n \geq 1$ and it is obvious that $s_0(\mathbf{F}) = 0$. □

Having done some more calculations we will be able to prove the following.

Theorem 6.14. (i) *For $\alpha \geq 3$ the sequence $< 0, \alpha, \alpha, ...\alpha, ... >$ is not representable in the class of all semigroups.*

(ii) *The sequence $< 0, 1, 1, ..., 1, ... >$ is the s_n–sequence for a semigroup* **S** *iff* **S** *is a nontrivial semilattice.*

(iii) *The sequence $< 0, 2, 2, ..., 2, ... >$ is the s_n–sequence for a semigroup* **S** *iff* **S** *generates the variety determined by the identities*

$$x^3 \approx x^2, \quad xy \approx yx, \quad xy^2 \approx x^2y.$$

Proof. Follows from the considerations above.

A variety $\mathcal{V}$ is *log–linear* if it is locally finite and there exists a constant $c > 0$ such that $\log f_n(\mathcal{V}) \leq cn$ for all $n > 1$. Obviously $\mathcal{V}$ is log–linear iff the free spectrum of $\mathcal{V}$ has subexponential rate of growth, i.e. iff there exist constants $a, c > 0$ such that $f_n(\mathcal{V}) \leq ac^n$ for all $n \geq 0$.

In [10] we gave a solution of the following problem of Grätzer and Kisielewicz.

Problem 6.15 ([23], Problem 29). *Characterize log–linear varieties of semigroups. Is there any algebraic property of semigroups equivalent to (or following form) log–linearity?*

Theorem 6.16 ([10]). *For any semigroup variety $\mathcal{V}$ the following conditions are equivalent:*

(i) *$\mathcal{V}$ is log–linear;*

(ii) *$\mathcal{V}$ satisfies the identities*

$$x^{\alpha+1} \approx x^{\beta}$$
$$x_1x_2 \cdots x_m \approx x_{\sigma(1)}x_{\sigma(2)} \cdots x_{\sigma(m)},$$

for some $\alpha \geq \beta > 0$, $m > 1$, and some non–trivial permutation σ of the set $\{1, ..., m\}$;

(iii) *$\mathcal{V}$ satisfies the identities*

$$x^{\alpha+1} \approx x^{\beta}$$
$$x_1x_2 \cdots x_{i-1}x_ix_{i+1}x_{i+2} \cdots x_m \approx x_1x_2 \cdots x_{i-1}x_{i+1}x_ix_{i+2} \cdots x_m,$$

for some $\alpha \geq \beta > 0$, $m > i \geq 1$.

Proof. See [10]. □

Corollary 6.17. *Let* $\mathbf{S} = (S, \cdot)$ *be an arbitrary finite semigroup and let* n *be a natural number such that* $S^{n-1} = S^n$. $\mathbf{S}$ *generates a log–linear variety iff* $cabd = cbad$ *holds for all* $a, b \in S$ *and all* $c, d \in S^n$.

It was proved in [31] that every semigroup satisfying condition (iii) of Theorem 6.16. has a finite basis for its identities, so that every log–linear semigroup variety is finitely based. Moreover, every subvariety of a log–linear semigroup variety is log–linear, and therefore finitely based. So we have:

Corollary 6.18. *Every log–linear variety of semigroups is a hereditarily finitely based.*

However, log–linearity is not necessary condition for a semigroup variety to be finitely based, even if the variety is locally finite. The variety $\mathcal{V}$ defined by the identity $xyzx \approx x^2$ was shown in [33] to be hereditarily finitely based. On the other hand, it is easy to check that $n! \leq f_n(\mathcal{V}) \leq (n+1)^n$ for all $n \geq 1$, so that $\mathcal{V}$ is locally finite but not log–linear.

7. Decidability

Let Σ be a fixed set of identities of a given similarity type. The *elementary theory* based on Σ is the set of sentences of first–order logic which are logical consequences of Σ. An elementary, quasi–identies, equational theory is deciable iff it is a recursive set of sentences. The connections between these concepts are given in the diagram below. This diagram refers to any fixed set Σ of equations.

$$\begin{tikzcd}
& \text{Decidable elementary theory} \arrow[dl] \arrow[dr] & \\
\text{Decidable theory of quasi–identities} \arrow[d] \arrow[rr, Leftrightarrow] & & \text{Uniformly solvable word problem} \arrow[d] \\
\text{Decidable equational theory} & & \text{Solvable word problem}
\end{tikzcd}$$

In general, none of the implications above can be reversed.

It is well known ([24]) that in the case of the variety of relation algebras of Tarski every quasi–identity is equivalent to some identity. Since, in the case of relation algebras, there is an algorithm to construct, for every quasi–identity, the equivalent identity, the problem of quasi–identities is equivalent to the problem of decidability of the equational theory. It was mentioned, as a consequence of Theorem 4.3., that the word problem is unsolvable for the class of relation algebras of Tarski. Therefore, the theory of quasi–identities of relation algebras is unsolvable. Hence, we obtain, as a consequence, the well known theorem of Tarski.

Theorem 7.1. *The equational theory of the class of relation algebras is undecidable.*

Starting from the result on unsolvability of the word problem for rings, we can prove that some varieties of modules have undecidable equational theory. The main reason for that is that every ring $\mathcal{R} = (R, +, \cdot, 0)$ can be considered as an $\mathcal{R}$ module $\mathcal{M} = (R, +, \cdot, 0, (f_r)_{r \in R})$, where $f_r(x) = r \cdot x$ for every $x \in R$. Then, to every equality between two words in $\mathcal{R}$, corresponds an identity in $\mathcal{M}$, and from the unsolvability of the word problem for $\mathcal{R}$ we can prove the undecidability of equational theory for $\mathcal{M}$ (and HSP($\mathcal{M}$)).

The same idea, with some additional ones, can be applied for the class of dynamic algebras.

There are several algebraic structures which correspond to some notions from computer science. Such are Kleene and dynamic algebras. We consider Kleene algebras which are obtained from the so–called Kleene relation algebras (without inversion). Kleene relation algebra, with some base $\mathbf{U}$, is an algebra having the set of all binary relations on the set $\mathbf{U}$ as the carrier, and the fundamental operation are set–theoretical union, composition, and reflexive–transitive closure. Kleene algebra is an algebra that belongs to the variety generated by all Kleene relation algebras.

Because of the relationship between Kleene relation algebras and regular languages, it follows the the equational theory of Kleene algebras is decidable.

We proved in [5] that the word problem for the class of all Kleene algebras is unsolvable.

Dynamic algebras are algebraic counterparts of propositional dynamic logic. Roughly speaking, dynamic logic is a classical propositional logic with some modal operators $\langle x \rangle$ associated with the elements x of a Kleene algebra. We can say that the corresponding algebraic structure, dynamic algebras, are Boolean algebras with normal unary operators which are indexed by the elements of a Kleene algebra. Although the equational theory of Kleene algebras is decidable, we proved in [7] that there are infinitely many finitely generated varieties of dynamic algebras having undecidable equational theories.

Definition 7.2. *Let* $\mathcal{K} = (K, \bigvee, ;, \star)$ *be a Kleene algebra. An algebra* $\mathcal{D} = (B, \cdot, -, F_a(a \in K))$ *is a* dynamic $\mathcal{K}$ algebra *if it satisfies the following conditions:*

(1) $(B, \cdot, -)$ *is a Boolean algebra,*

(2) $F_a(0) \approx 0$,

(3) $F_a(x+y) \approx F_a(x) + F_a(y)$,
(4) $F_{a \bigvee b}(x) \approx F_a(x) + F_b(x)$,
(5) $F_{a;b} \approx F_a F_b(x)$,
(6) $x + F_a F_{a^*}(x) \leq F_{a^*}(x)$,
(7) $F_{a^*}(x) \leq x + F_{a^*}(-x \cdot F_a(x))$,

for all $a, b \in K$, $x, y \in B$.

The definition above is from the paper of B. Jónson [24].

Let $\mathbf{S}$ be a semigroup with an identity. By $\mathcal{T}(\mathbf{S})$ we denote the so–called *semigroup of left translations* of $\mathbf{S}$.

Definition 7.3. *Let* $\mathbf{S}$ *be a semigroup with an identity. By* $\Psi(\mathbf{S})$ *we denote the subalgebra of the Kleene relation algebra* $\mathcal{K}(S)$ *generated by the set* $\mathcal{T}(\mathbf{S})$. *We define the dynamic set algebra* $\mathcal{D}(\mathbf{S})$ *to be* $(\mathcal{P}S, \bigcap, -, F_a(a \in \Psi(\mathbf{S})))$.

Definition 7.4. *The* semigroup of Cejtin *is the semigroup* $\boldsymbol{C}$ *presented by* $(G(\boldsymbol{C}), R(\boldsymbol{C}))$, *where*

$G(\mathbf{C}) = \{a, b, c, d, e\}$,
$R(\mathbf{C}) = \{ac = ca, ad = da, bc = cb, bd = db, abac = adacc, eca = ac, edb = be\}$

It is well known that the semigroup of Cejtin has unsolvable word problem.

Proposition 7.5. *There is a sequence* $\mathbf{C}_0\ \mathbf{C}_1, ..., \mathbf{C}_n, ...$ *of finitely presented semigroups such taht*

(a) *all semigroups* $\mathbf{C}_i$, $(i \in N)$*have unsolvable word problems;*
(b) $HSP(\mathcal{D}(\mathbf{C}_i)) \neq HSP(\mathcal{D}(\mathbf{C}_j))$ *for all* $i \neq j$, $i, j \in N$.

Proof. See [7]. □

Theorem 7.6 ([7]). *There are infinitely many finitely generated varieties of dynamic algebras, with countably many operations, having undecidable equational theories. All these varieties are generated by representable dynamic algebras.*

Proof. See [7]. □

Corollary 7.7. *There are infinitely many finitely generated varieties od dynamic algebras with countably many operations, having uniformly unsolvable word problems.*

Theorem 7.6. does not give any information on the word problem of dynamic algebras. Therefore we can formulate

Problem 7.8. *Is the word problem for all the varieties of dynamic algebras solvable?*

Also, the following is still open

Problem 7.9. *Is there a finitely based variety of dynamic algebras having undecidable equational theory?*

Problem 7.10. *Is there finite dynamic algebra which is not finitely based?*

References

[1] Adjan, S.L., Repin, N.N.: *On exponential lower bound for class of nilpotency of Engel Lie algebras.*, Matem. Zametky 39 N3, (1986), 444–452

[2] Blok, W.J., Pigozzi, D.: *On the structure of varieties with equationally definable principal conqruences I*, Algebra Universalis 15, (1982), 195–227

[3] Church, A.: *A note on the Entscheidungspreoblem*, J. Symb. Logic 1, (1936) 40–41, 101–102

[4] Church, A.: *An unsolvable problem of elementary number theory*, Am. J. Math., 58., (1936) 345–363

[5] Crvenković, S., Madarász, Sz.R.: *On Kleene algebras*, Theoretical Compt. Sci. 108, (1993) 17–24

[6] Crvenković, S., Madarász, Sz.R.: *On a problem of partial algebras*, Rew. of Research Faculty of Sci. Math. Series 17.2, (1989) 39–55

[7] Crvenković, S., Madarász, Sz.R.: *On dynamic algebras*, Theoretical Compt. Sci. 134, (1994) 79–86

[8] Crvenković, S., Ruškuc, N.: *On groupoids having n^2 essentially n–ary polynomials*, Rew. of Research Faculty of Sci. Math. Series 23, (1993) 287–312

[9] Crvenković, S., Ruškuc, N.: *On semigroups having n^2 essentially n–ary polynomials*, Algebra Universalis 30, (1993) 269–271

[10] Crvenković, S., Ruškuc, N.: *Log–linear varieties of semigroups*, Algebra Universalis

[11] Crvenković, S., Delić, D.: *A variety with locally solvable but globally unsolvable word problem*, Algebra Universalis (to appear)

[12] Crvenković, S., Delić, D.: *Different levels of word problems for some varieties*, Rew. of Research Faculty of Sci. Math. Series (to appear)

[13] Crvenković, S., Dudek, J.: *Rectangular Groupoids*, Czechoslovak Mathematical Jour. 35(110)(1985) 405–414

[14] Dudek, J.: *Number of algebraic operation in idempotent groupoids*, Colloq. Math. 21, (1970) 169–177

[15] Dudek, J.: *On non–associative groupoids*, Colloq. Math. 36, (1976) 23–25

[16] Dudek, J.: *On bisemilattices II*, Demonstration Math. 15, (1982) 465–475

[17] Dudek, J.: *A characterization of distributive lattices*, in "Contibution to lattice theory (Szeged 1980)", Colloq. Math. Soc. János Bolyai, 33 North–Holland, Amsterdam–New York, pp. 325–335

[18] Dudek, J.: *On binary polynomials in idenpotent commutative groupoids*, Fund. Math. 120, (1984) 187–191

[19] Dudek, J.: *A polynomial characterization of some idempotent algebras*, Acta. Sci. Math. (Szeged) 50, (1986), 167–191

[20] Dudek, J.: *Dedekind's number characterize disttibutive lattices*, Algebra Universalis, (to appear)

[21] Evans, T.: *The word problem for abstarct algebras*, J. London Math. Soc. 26, (1951), 64–71

[22] Grätzer, G.: "Universal Algebra" Second edition, Springer–Verlag, New York–Heidelberg–Berlin (1979)

[23] Grätzer, G., Kisielewicz, A.: *A survey of some open problems on p_n–sequences and free spectra of algebras and varities*, in Universall Algebra aind Quasigroup Theory (eds. A. Romanowska and J.D.H. Smith), Haldermann Verlag, Berlin, (1992), 57–88

[24] Jónsson, B.: *The theory of binary relations*, in Algebraic Logic, Colloq. Math. Sci. János Bolyai, Vol 54, North–Holland, Amsterdam (1988), 245–292

[25] Kogalovskii, S.R.: *On some simple remarks about unsolvability*, (Russian), XII naučno tehničeskaja konferencija po rabotam, vipulnenim v 1965 godu, Sekcija matematika, Ivanovo, (1966), 3–5

[26] Madarász, Sz.,R., Crvenković, S.: Relacione algebre, Matematički institut SANU, Beograd, (1992)

[27] Maljcev, A.I.: The Metamathematics of Algebraic Systems, Collected papers: 1936–1967, translated, edited and provided by supplementary notes by B.F. Wells III, North–Holland, Amsterdam–London, (1971)

[28] Marczewski, E.: *Indepence in abstract algebras. Results and Problems*, Colloq. Math. 14, (1966), 169–188

[29] Markov, A.A.: *On the representation of recursive functions*, Izv. An SSSR, serija matem. 13, No.4, (1949) 417–424

[30] A. Mekler, E. Nelson, S. Shelah: *A variety with solvable, but not uniformly solvable, word problems*, Proc. London Math. Soc. 66, (1993) 225–256

[31] Novikov, P.S.: *On the algorithmic unsolvability of the problem of identity of words in the theory of groups*, (Russian) Trudi matem. in–ta AN SSSR, Vol. 44, (1955), 1–44

[31] Perkins, P.: *Bases of equatonial theories of semigroups*, J. Algebra II (1969), 298–314

[33] Pollák, G.: *A class of hereditarily finitely based varieties of semigroups*, in Algebraic Theory of Semigroups (Proc. Sixth Algebraic Conf., Szeged 1976), Colloq. Math. Soc. János Bolyai 20 (1979), 433–445

[34] Post, E.: *A variant of a recursively unsolvable problem*, Bull Amare. Math. Soc. 52 No 4 (1946), 264–268

[35] Turing, A. M.: *On computable numbers with an application to the Entscheidungsproblem*, Proc. London Math. Soc. (2) 42 (1937), 230–265

[36] Wells, B.: *A simple pseudorecursive variety of infinite type*, Abstracts Amer. Math. Soc. 3 (1982) 592

[37] Wells, B.: *Pseudorecursive varieties and their implications for word problems*, Ph. D. Thesis, University of California, Berkley, submited 11/82, copyright 1983, 243 pages

[38] Wells, B.: *Pseudorecursive varieties of semigrous–I* (submitted aruond 1990), Int. J. Algebra & Compt. (to appear)

Institute of Mathematics, University of Novi Sad, 21000 Novi Sad,
Trg D. Obradovića 4, Serbia, Yugoslavia

E-mail: sima@unsim.ns.ac.yu

FILOMAT (Niš) 9:3 (1995), 449–472

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

THE SECOND LARGEST EIGENVALUE OF A GRAPH (A SURVEY)

Dragoš Cvetković and Slobodan Simić

Abstract. This is a survey paper on the second largest eigenvalue λ_2 of the adjacency matrix of a graph. Among the topics presented are the graphs with small λ_2, bounds for λ_2, algebraic connectivity, graphs with good expanding properties (such as Ramanujan graphs), rapidly mixing Markov chains etc. Applications to computer science are mentioned. Recent results of the authors are included.

0. Introduction

Let G be a graph on vertices $1, 2, \dots, n$. The adjacency matrix of G is the matrix $A = [a_{ij}]_1^n$, where $a_{ij} = 1$ if vertices i and j are adjacent and $a_{ij} = 0$ otherwise. Since A is symmetric its eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ are real. Assuming that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, we also say that $\lambda_i (= \lambda_i(G))$ is the i–th eigenvalue of G ($i = 1, 2, \dots, n$). In particular, $\lambda_2(G)$ is the second largest eigenvalue of a graph G.

For general theory of graph spectra see monographs [26] and [27].

Concerning particular eigenvalues the following eigenvalues have been studied in some detail:

1^o the largest eigenvalue;
2^o the second largest eigenvalue;
3^o the smallest positive eigenvalue;
4^o the largest negative eigenvalue;
5^o the second smallest eigenvalue;
6^o the smallest eigenvalue.

For a survey on the largest eigenvalue of a graph see the paper [27] by D. Cvetković and P. Rowlinson (see also [26], the third edition, pp. 381–392). Concerning the smallest eigenvalue, particular attention has been paid to graphs with the smallest eigenvalue -2 (see [26], the third edition, pp. 378–381).

Graphs with small second largest eigenvalue have interesting structural properties. The second largest eigenvalue (in modulus) of a regular graph turned out to be an important graph invariant. This paper provides a survey of research on such graphs and on the second largest eigenvalue in general. The starting point for writing this survey was a shorter survey on the same subject given on pp. 392–394 of the third edition of [26].

1. Graphs with small λ_2

It is an elementary fact (see, for example, [26], p. 163) that for non-trivial connected graphs $\lambda_2(K_n) = -1$ $(n \geq 2)$, $\lambda_2(K_{n_1,n_2,\dots,n_k}) = 0$ $(\max(n_1, n_2, \dots, n_k) \geq 2)$ and $\lambda_2(G) > 0$ for other graphs G.

A graph property $\mathcal{P}$ is called *hereditary* if the following implication holds for any graph G: if G has property $\mathcal{P}$, then any induced subgraph of G also possesses property $\mathcal{P}$. (In this paper, when we say that "a graph G *contains* a graph H" we mean that G contains H as an induced subgraph). A graph H is *forbidden* for a property $\mathcal{P}$ if it does not have property $\mathcal{P}$. If a graph G contains (as an induced subgraph) the forbidden graph H (for a property $\mathcal{P}$), then G does not have property $\mathcal{P}$. Then H is called a *forbidden subgraph.* A forbidden subgraph H is called *minimal* if all vertex deleted subgraphs $H - i$ have property $\mathcal{P}$. Graphs having property $\mathcal{P}$ can be characterized by a collection (possibly infinite) of minimal forbidden subgraphs for property $\mathcal{P}$.

For any real a and any integer i the property expressed by the inequality $\lambda_i(G) \leq a$ is a hereditary property. This conclusion follows from the *interlacing theorem* (cf., e.g., [26], p. 19) which says that $\lambda_i(H) \leq \lambda_i(G)$ for any induced subgraph H of G.

The hereditary property of the form $\lambda_2(G) \leq a$, and in principal, the second largest eigenvalue of a graph, has been studied in some detail for the first time by L. Howes [50] and [51] in early seventies. The following characterization is taken from [51]:

Theorem 1. *Let $\mathcal{G}$ be an infinite set of graphs, then the following statements about $\mathcal{G}$ are equivalent:*

1° *There exists a real number a such that $\lambda_2(G) \leq a$ for every $G \in \mathcal{G}$.*

2° *There exists a positive integer s such that for each $G \in \mathcal{G}$ the graphs $(K_s \cup K_1) \bigtriangledown K_s$, $(sK_1 \cup K_{1,s}) \bigtriangledown K_1$, $(K_{s-1} \cup sK_1) \bigtriangledown K_1$, $K_s \cup K_{1,s}$, $2K_{1,s}$, $2K_s$ and the graphs on Fig.* 1 *(each obtained from two copies of $K_{1,s}$ by adding extra edges) are not subgraphs of G.*

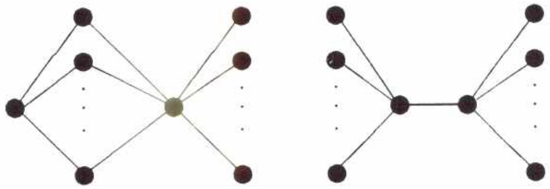


Fig. 1.

Here ∇ denotes the join of two graphs, while $\cup$ refers to union of two disjoint graphs. Notice that $G_1 \nabla G_2 = \overline{\overline{G_1} \cup \overline{G_2}}$

In the rest of this section, we shall focuss our attentoin the following values for a: $a = \frac{1}{3}$, $a = \sqrt{2} - 1$, $a = (\sqrt{5} - 1)/2$, $a = 1$ and $a = 2$.

1.1 The golden section bound

There are several results in which the (upper) bound for λ_2 does not exceed the golden section $(\sqrt{5} - 1)/2$.

It is proved in 1993 by D. Cao and Y. Hong [17] that the second largest eigenvalue of a graph G on n vertices is between 0 and $\frac{1}{3}$ if and only if $G = (n-3)K_1 \nabla (K_1 \cup K_2)$. The problem of characterizing graphs G with $\frac{1}{3} < \lambda_2(G) < (\sqrt{5} - 1)/2$ was also posed in [17]. Graphs G with $\lambda_2(G) < \sqrt{2} - 1$ are determined by M. Petrović [75]. An independent characterization of graphs with $\lambda_2 \leq \sqrt{2} - 1$ has been given by J. Li in [56]; in addition, all minimal forbidden subgraphs for the property $\lambda_2 \leq \sqrt{2} - 1$ are given there. It is proved by S. Simić [85] that the set of minimal forbidden subgraphs for the property $\lambda_2(G) < (\sqrt{5} - 1)/2$ is finite. The structure of graphs G with $\lambda_2(G) \leq (\sqrt{5} - 1)/2$ has been studied by D. Cvetković and S. Simić [29]. A part of results has been announced in [28].

We shall introduce the notation $\sigma = (\sqrt{5} - 1)/2 \approx 0.618033989$. Obviously, we have $\sigma^2 + \sigma - 1 = 0$.

Graphs having property $\lambda_2(G) \leq \sigma$ (σ-property) will be called *σ-graphs*. For convenience graphs G for which $\lambda_2(G) < \sigma$, $\lambda_2(G) = \sigma$, $\lambda_2(G) > \sigma$ will be called *σ^--graphs*, *σ^0-graphs*, *σ^+-graphs*, respectively.

The next proposition, taken from [95] (see also [9]), enables the definition of a class of graphs to which every σ^--graph belongs.

Proposition 2. *If $\overline{G}$ is a connected graph and if G has no isolated vertices, then G contains an induced subgraph equal to $2K_2$ or P_4.*

Assume now G is a σ^--graph. If $\overline{G}$ is a connected graph, then G must have at least one isolated vertex (otherwise G contains $2K_2(= E)$ or P_4 as an induced subgraph, and hence is not a σ^--graph). On the other hand, if $\overline{G}$

is a disconnected graph, then G itself is a join of at least two graphs. Since the σ-property is hereditary, it follows that G belongs to a class of graphs (here, as in [85], denoted by $\mathcal{C}$) which is defined as the smallest family of graphs that contains K_1 and is closed under adding isolated vertices (i.e., if $G \in \mathcal{C}$, then $G \cup K_1 \in \mathcal{C}$) and taking joins (i.e., if $G_1, G_2 \in \mathcal{C}$, then $G_1 \bigtriangledown G_2 \in \mathcal{C}$). An alternative way to describe graphs from the class $\mathcal{C}$ is in terms of minimal forbidden induced subgraphs. Actually, $\mathcal{C}$ is a class of graphs having no induced subgraphs equal to E $(= 2K_2)$ or P $(= P_4)$.

Clearly, any σ^--graph belongs to $\mathcal{C}$, but not vice versa.

The class $\mathcal{C}$ has been introduced and studied in [85]. Weighted rooted trees (with weights assigned to vertices) were used also in [85] in representing graphs from the class $\mathcal{C}$.

To any graph G from $\mathcal{C}$ we associate a weighted rooted tree T_G (also called *an expression tree* of G) in the following way:
if $H = (H_1 \bigtriangledown \ldots \bigtriangledown H_m) \cup nK_1$ is any subexpression of a graph G (i.e. a graph obtained by using the above rules), then a subtree T_H with a root v corresponds to H; $n(= w(v))$ is a weight of v, whereas for each i $(i = 1, \ldots, m)$ there is a vertex v_i (a son of v) representing a root of H_i.

Example. If $G = (((((K_1 \bigtriangledown K_1) \cup K_1) \bigtriangledown K_1) \bigtriangledown K_1) \bigtriangledown K_1) \cup 3K_1$, then the corresponding expression tree is depicted in Fig. 2(a). In Fig. 2(b) we represent the same graph as a set diagram (a line between two circumscribed sets of vertices denotes that each vertex inside one set is adjacent to any vertex inside the other set).

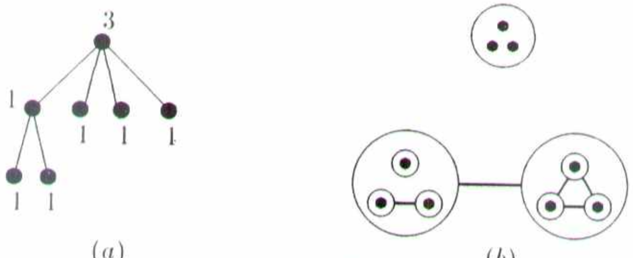


Fig. 2.

It turned out that the set of σ^--graphs falls into a finite number of structural types. These types are given in Fig. 3 by the corresponding expressioned trees.

It has been proved along the same lines in [85] that the set of minimal forbidden subgraphs for the σ^--property is finite. They all belong to $\mathcal{C}$ except for E and P_4. The whole list of these forbidden subgraphs will be described in a forth-comming paper [30].

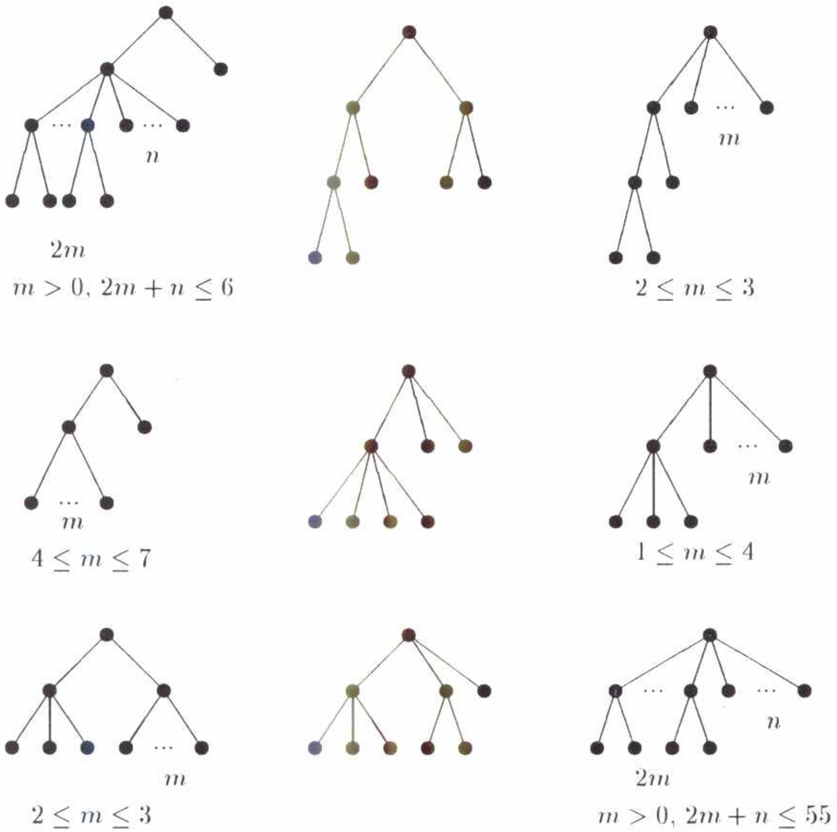


Fig. 3.

We present now main results of [29].

Theorem 3. *A σ-graph has at most one non-trivial component G for which one of the following holds:*

1^o G is a complete multipartite graph;
2^o G is an induced subgraph of C_5;
3^o G contains a triangle.

Before proceeding to describe σ-graphs mentioned in 3^o we introduce some notation.

Let G be a σ-graph with the vertex set V. Let T be a triangle in G induced by the vertices x, y, z. Next, let $A(G,T) = A$, $B(G,T) = B$, $C(G,T) = C$ be the sets of vertices outside T which are adjacent to exactly one, two, three vertices from T, respectively. Also, let G_A, G_B, G_C be the component, containing T, of the subgraph of G induced by the vertex set $V - B - C$, $V - A - C$, $V - A - B$, respectively.

Let $d(u,T)$ denote the distance of the vertex u from the triangle T, i.e. the length of the shortest path between u and a vertex from T.

σ-graphs containing triangles are now described in more detail in terms of induced subgraphs G_A, G_B, G_C.

Theorem 4. *Let G be a connected σ-graph which contains a triangle. For any triangle T of G the following holds for subgraphs G_A, G_B, G_C :*

1° *G_A is an induced subgraph of one of the graphs from Fig. 4.*

2° *For G_B one of the following holds:*

i) *G_B is an induced subgraph of graphs from Fig. 5;*

ii) *$G_B = P_4 \triangledown (H \cup K_1)$ for some σ-graph H;*

iii) *$G_B = H_1 \triangledown H_2 \triangledown H_3$ for some σ-graphs H_1, H_2, H_3.*

3° *For G_C one of the following holds:*

i) *G_C is an induced subgraph of $(K_3 \cup K_1) \triangledown H$ for some σ-graph H;*

ii) *G_C is obtained from $K_n \triangledown K_3 \triangledown H$ by adding a pendant edge to each vertex of K_n, where $n \geq 2$ and H is a σ-graph containing no induced subgraphs isomorphic to some of graphs $K_3 \cup K_1$, $K_2 \cup 3K_1$, $K_{1,2} \cup 2K_1$, $K_{2,4} \cup K_1$, $K_{3,3} \cup K_1$.*

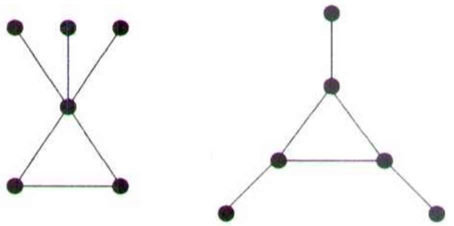


Fig. 4.

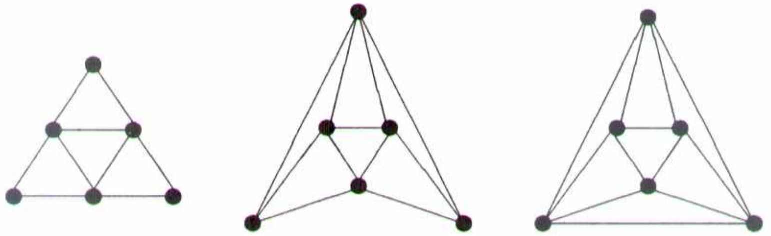


Fig. 5.

It is also proved in [29] that the set of minimal forbidden subgraphs for the σ–property is finite. The next theorem (taken from [29]) provides more details.

Theorem 5. *If H is a minimal forbidden (induced) subgraph for the $\sigma-$property, then:*

1° H is one of the graphs $E(=2K_2)$, F_1, F_2, F_3, F_4 (see Fig. 6), or
2° H belongs to the class $\mathcal{C}$.

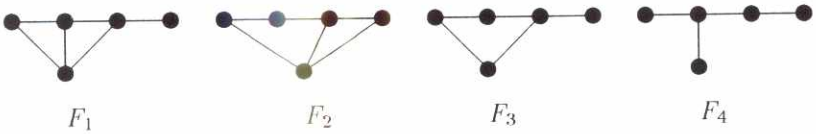


Fig. 6.

All minimal forbidden subgraphs for $\sigma-$property are not yet known. On the other hand, more can be said if we require that both, the graph and its complement are $\sigma-$graphs. Then, due to S. Simić [86], there are exactly 27 minimal forbidden subgraphs for this property. Here we rather give explicitely (following[86]) all graphs with the property in question.

Theorem 6. *G and $\overline{G}$ are both $\sigma-$graphs, if and only if either of them is one of the following graphs:*

$K_m \cup nK_1$ $(m, n \geq 0)$, $K_{2,1,1} \cup mK_1$, $K_{2,1} \cup mK_1$ $(m \geq 0)$,
$K_{3,1} \cup mK_1$ $(m \leq 3)$, $K_{2,1,1,1} \cup mK_1$ $(m \leq 2)$,
$((K_{2,1,1} \cup K_1) \bigtriangledown K_1) \cup K_1$, $((K_{2,1} \cup 2K_1) \bigtriangledown K_1) \cup K_1$,
$((K_{2,1} \cup K_1) \bigtriangledown K_1) \cup K_1$, $(K_{m,1} \cup K_1) \bigtriangledown K_n$ $(m \geq 2,\ n \geq 0)$,
$(K_{2,1,1} \cup K_1) \bigtriangledown K_m$, $(K_{2,1} \cup 2K_1) \bigtriangledown K_m$ $(m \leq 2)$,
$(K_{3,1} \cup 2K_1) \bigtriangledown K_1$, $(K_{2,1,1,1} \cup K_1) \bigtriangledown K_1$, $(((K_{2,1} \cup K_1) \bigtriangledown K_1) \cup K_1) \bigtriangledown K_1$.

1.2 Bounds equal to 1 and 2

Graphs with $\lambda_2(G) \leq 1$ have been studied in 1982 by D. Cvetković [24]. It turned out that some of these graphs are the complements of the graphs whose least eigenvalue is greater than or equal to -2. More precisely, $\lambda_n(\overline{G}) > -2$ implies $\lambda_2(G) < 1$. If $\lambda_n(\overline{G}) = -2$, then $\lambda_2(G) \leq 1$ equality holding if and only if the eigenvalue -2 of $\overline{G}$ is either non-simple or non-main (all eigenvectors are orthogonal to the vector $(1, 1, \ldots, 1)$). For other graphs G with $\lambda_2(G) \leq 1$, the complement $\overline{G}$ has exactly one eigenvalue smaller than -2. However, $\lambda_n(\overline{G}) < -2$ and $\lambda_{n-1}(\overline{G}) \geq -2$ does not imply $\lambda_2(G) \leq 1$. These results are derived by the well-known Courant-Weyl inequalities for eigenvalues of matrices. For further details see the original paper or monograph [25] (p. 11, where [Cve5] is wrongly given as [Cve7]).

A representation of graphs with $\lambda_2(G) = 1$ in the Lorentz space is given in 1983 by A. Neumaier and J. J. Seidel [72].

Bipartite graphs G with $\lambda_2(G) \leq 1$ have been characterized in 1991 by M. Petrović [74]. Three families of graphs and four particular graphs with $\lambda_2(G) \leq 1$ are constructed. It is proved that a connected bipartite graphs have the property $\lambda_2(G) \leq 1$ if and only if it is an induced subgraph of the mentioned graphs.

In particular, trees with the second largest eigenvalue less than 1 were treated by A. Neumaier [70]. More generaly, an algorithm for deciding if the second largest eigenvalue of any tree is less than some bound was also proposed by A. Neumaier.

The exact characterization of graphs with second largest eigenvalue around 1 still remains an interesting open question in spectral graph theory.

Graphs with $\lambda_2 \leq 2$ are called *reflexive* graphs [72]. Some classes of reflexive graphs are studied in [72]. In particular, trees with $\lambda_2 = 2$ are called *hyperbolic* [60]. All hyperbolic trees are known [60], [70] and [72].

2. Bounds for λ_2

Upper and lower estimates for the second largest eigenvalue of a graph under various restrictions were studied in literature (but not as extensively as for the largest eigenvalue).

The most general result concerns the connected graphs with prescribed number of vertices. According to D. Powers, for a connected graph G on n vertices the following holds

$$-1 \leq \lambda_2(G) \leq \lfloor \frac{n}{2} \rfloor - 1.$$

The upper bound is achieved, for n odd ($n = 2s+1$), if G is a graph consisting of two cliques of size s (graphs equal to K_s) bridged by a path of length 2; for n even this bound is only asymptotically sharp (see [78], or [79]; see also [48]). The lower bound is achieved if and only if G is a complete graph (see Section 1). It is interesting to note that the above (upper) estimate is proved by making use of the following more general estimate of the second largest eigenvalue in terms of the largest eigenvalue of some parts of a graph. Namely, due to D. Powers we have:

$$\lambda_2(G) \leq \max_{(G_1,G_2)} \min\{\lambda_1(G_1), \lambda_1(G_2)\},$$

where G_1 and G_2 denote the subgraphs induced by vertex sets of some bisection of (the vertex set of) a (connected) graph G. The key argument for proving this was based on partitioning the vertices of G according to sign pattern of the eigenvector corresponding to the second largest eigenvalue (see [77] for details).

If G is a connected graph on n vertices and m edges, then, due to R.C. Brigham and R.D. Dutton [13], the following inequality holds:

$$\lambda_2(G) \leq \sqrt{\frac{m(n-2)}{n}}.$$

In particular, this estimate in not too good for trees. If we assume that G is not a tree, then some refinements are possible, as shown in [79]. Then the result is expressed in terms of the estimates for the largest eigenvalue of a connected graph with a fixed number of edges (but not vertices). The latter problem is completely solved by P. Rowlinson [81] to within the graphs which realize the bounds. More precisely, as remarked in [79], then

$$\Lambda_1(\lfloor\frac{m}{2}\rfloor - 1) \leq max\{\lambda_2(G)\} \leq \Lambda_1(\lfloor\frac{m-1}{2}\rfloor),$$

where $\Lambda_1(m)$ is the maximum for the largest eigenvalue of a connected graph with m edges); thus the estimate is very tight.

In particular, for triangle-free (and bipartite) graphs some further estimates are obtained in [13].

Much better estimates for trees are known. If T is a tree with $n \geq 3$ vertices, then

$$0 \leq \lambda_2(T) \leq \sqrt{\lfloor\frac{n-2}{2}\rfloor}.$$

The upper bound was obtained by Y. Hong [47]. It is the best possible for n (≥ 3) odd (then it coincides with the bound of A. Neumaier [70] $\lambda_2(T) \leq \sqrt{\frac{n-3}{2}}$ which holds only for n odd). As remarked by D. Powers [78], with more careful analysis one can get:

$$\sqrt{\lfloor\frac{n-1}{2}\rfloor - 1} \leq max\{\lambda_2(T)\} \leq \sqrt{\lfloor\frac{n-2}{2}\rfloor},$$

i.e. the bound for n even is asimptotically sharp. The lower bound is clear from the above (it is achived for a tree isomorphic to a star, i.e. for $T = K_{1,n-1}$). Otherwise, if $T \neq K_{1,n-1}$, then $\lambda_2(T) < 1$ only for $T = S_{n-2}^2$ (here S_{n-2}^2 is the graph obtained from a star with $n-2$ arms by subdividing one arm). Also then $\lambda_2(T) = \sqrt{\frac{n-1-\sqrt{(n-3)^2+4}}{2}}$. Thus if $T \neq K_{1,n-1}, S_{n-2}^2$, then $\lambda_2(T) \geq 1$.

Star-like trees are trees homeomorphic to a star ($K_{1,s}$ for some $s \geq 3$). The second largest eigenvalue of star-like trees (with a fixed number of vertices and fixed number of arms) were studied by F. K. Bell and S. K. Simić (see [87]). We only mention here that for fixed $s \geq 4$, the trees with minimum and maximum second largest eigenvalue (on fixed number of vertices) are those as intuitively expected (i.e., those having the length of all arms as equal as possible in the former case, and those having the length of all arms but one equal to 1 in the latter case). If $s = 3$, then some interesting phenomena do occur (for detail see [87]).

Results on regular graph are given in the next section.

3. Regular graphs

There are two main reasons why regular graphs deserve special interest in this context. The first is that the largest eigenvalue of a regular graph of degree d is equal to d, so then the second largest eigenvalue becomes the dominant feature in many aspects (in particular, in spectral orderings). The second is that regular graphs allow a simple connection between the eigenvalues (of the adjacency matrix) and the eigenvalues of some other matrices associated with graphs, in particular, with the eigenvalues of the graph Laplacian (see below).

3.1 λ_2 and spectral ordering of regular graphs

The role of the second largest eigenvalue in ordering cubic graphs has been observed in 1976 by F.C. Bussemaker, S. Čobeljić, D. Cvetković and J.J. Sedel [16] (see also [26], pp. 268–269). The 621 connected cubic graphs with not more than 14 vertices, together with eigenvalues and many other data, are displayed. The sequence of eigenvalues is given in non-increasing order for each graph, and for a fixed number of vertices the graphs are ordered lexicographicaly with respect to their sequences of eigenvalues. Since the largest eigenvalue λ_1 is equal to 3 in cubic graphs, the second largest eigenvalue λ_2 determines roughly the ordering of graphs. Decreasing λ_2 shows graphs of more "round" shape (smaller diameter, higher connectivity and girth).

A partial theoretic explanation of these empirical observations was offered in 1978 by D. Cvetković [23].

Theorem 1. *Let G be a $d-$regular graph on n vertices. Let x be any vertex of G and let δ be the average vertex degree of the subgraph induced by the vertices not adjacent to x. Then we have:*

$$\delta \leq d\frac{\lambda_2^2 + \lambda_2(n-d)}{\lambda_2(n-1)+d}.$$

The same inequality (see also [25], p. 71) was derived in [6] by quite different method.

In further we shall offer some other theoretic support to these (empirical) observations (see Section 3.3).

3.2 Algebraic connectivity

For a graph G on n vertices, let $d_1, d_2, \dots, d_n$ denote the corresponding vertex degrees. The matrix $L = D - A$ with $D = [d_i \delta_{ij}]_1^n$ (δ_{ij} the Kronecker symbol) is called the Laplacian of G. The graph Laplacian is positive semidefinite and the second smallest eigenvalue of L (here denoted by $\alpha (= \alpha(G))$) is called the *algebraic connectivity* of G. It was introduced in 1973 by M. Fiedler [35].

The algebraic connectivity α of a graph (in regular case) can be expressed in terms of the second largest eigenvalue. If G is a d-regular graph, then $\alpha = d - \lambda_2$. (Thus the algebraic connectivity increases as the second largest eigenvalue becomes smaller).

Definition 1. An (n, d, ϵ)−enlarger is a d-regular graph G on n vertices with $\alpha(G) \geq \epsilon$.

The significance of enlargers lies, among others, in the fact that they enable an explicit construction of graphs with good expansion properties (such as expanders). One such construction of expanders is obtained by N. Alon and V.D. Milman [6]. For this aim we need the following definition:

Definition 2. Let $G = (V, E)$ be a graph with $V = \{v_1, \dots, v_n\}$. The extended double cover of G is a bipartite graph $H = (X, Y, F)$ with $X = \{x_1, \dots, x_n\}$, $Y = \{y_1, \dots, y_n\}$ where $F = \{x_i y_j : i = j \ or \ v_i v_j \in E\}$.

Remark. Actually, an extended double cover is a NEPS (Non-Complete Extended P-Sum, see [26], pp. 65-66) of G and K_2 in the basis $\{(0,1), (1,1)\}$ (see, also [25], p. 60).

Now the following theorem from [6] offers an explicit construction of an expander (see Definition 2 from Section 6).

Theorem 2. *Let $G = (V, E)$ be an (n, d, ϵ)−enlarger and let H be its extended double cover. Then H is a strong $(n, d+1, \delta)$−expander for*

$$\delta = \frac{4\epsilon}{d + 4\epsilon}.$$

The next theorem of N. Alon [3] points that good enlargers are in fact good magnifiers (see Definition 3 from Section 6).

Theorem 3. *Every (n, d, ϵ)−enlarger is an (n, d, δ)-magnifier, where*

$$\delta = \frac{2\epsilon}{d + 2\epsilon}.$$

It is interesting to note that the converse also holds, i.e. that every magnifier is an enlarger with some appropriate parameters (see [3], for further details).

Remark. Generally, the fact that the algebraic connectivity is relevant to expansion property of a graph (see Definition 1 from Section 6) can be also justified by the following relation (cf. [42], Lemma 5.7) given below. Namely, for any graph $G = (V, E)$ we have:

$$(\forall X \subset V)(|X| \leq \frac{1}{2} \Rightarrow |\partial X| \geq \frac{|X||V \setminus X|}{|V|}\alpha(G),$$

where $\partial X = \{y : xy \in E, x \in X\}$. A similar result is due to R.M. Tanner [92].

More information on expansion property of graphs, and other related graphs can be found in Section 6.

The work on algebraic connectivity and graph Laplacian for graphs in general (in particular, for non-regular graphs) will not be reported in this paper. For more information see papers by M. Fiedler [36], [37], [38] and also [44], [45]. Much information on graph Laplacians can be found in the book [26] and in expository papers [43], [61], [62], [66], [67].

3.3 Second largest eigenvalue in modulus

The second largest eigenvalue (in modulus) of a regular graph turned out to be an important graph invariant since it has relations with various graph invariants (such as diameter and *covering number* etc.) and graph properties (including expanding properties and convergence properties of simple random walks).

Let G be a d-regular graph, and let $\Lambda(= \Lambda(G)) = max\{|\lambda_i| : |\lambda_i| \neq d\}$. Notice that for bipartite graphs we have $\Lambda(G) = \lambda_2(G)$ (due to symmetry of the spectrum with respect to the origin).

Let G be a connected d-regular graph on n vertices. According to N. Alon and V.D. Milman [6] we have the following bound:

$$diam(G) \leq 2\lceil\sqrt{\frac{2d}{d - \Lambda}}\log_2 n\rceil. \tag{1}$$

This bound was improved by several authors, in several directions. Interesting improvements are given by B. Mohar in [65], but expressed in terms of the second smallest, and the largest eigenvalue of Laplacian matrix of any (not necessarily regular) graph. Also the following lower bound, $diam(G) \geq \dfrac{4}{n\alpha}$, valid for any (connected) graph on n vertices, can be found in [65]. Bound (1) is also improved by F.R.K. Chung [22]. For regular graphs this bounds reads:

$$diam(G) \leq \lceil \frac{\log(n-1)}{\log(\frac{d}{\Lambda})} \rceil. \tag{2}$$

For this bound it was observed in [31] that it is indeed the upper bound for covering index of a graph (i.e. it is the smallest integer c such that any pair of not necessarily distinct vertices is connected by a walk of length exactly c). Let $cover(G)$ denote the covering number of a connected graph G. As proved in [31], for any (connected) graph we have $diam(G) \leq cover(G)$ (it is also true that $cover(G) \leq diam(G) + s$ if every vertex of G is in some closed walk of odd length, at most $2s+1$; if G is a bipartite graph, then $cover(G) = \infty$). By convenient distinction between diameter and covering index we have: if G is a d-regular connected graph on n vertices and t some positive number then:

(i) if $d^m/n > \lambda^m(1-\frac{1}{n})$, then $cover(G) \leq m$ (by F.R.K. Chung, restatement of (2));

(ii) if $d^{m-1}(d+t)/n > \lambda^{m-1}|\lambda+t|(1-\frac{1}{n})$ ($\lambda \in \{\lambda_2, \ldots, \lambda_n\}$ and $t > 0$), then $diam(G) \leq m$ (by C. Delorme and P. Sóle [31]).

According to P. Sarnak [82] (see also [80]) the following estimate holds:

$$diam(G) \leq \frac{\operatorname{arccosh}(n-1)}{\operatorname{arccosh}(\frac{d}{\Lambda})}, \tag{3}$$

for any d-regular graph G on n vertices. By considering separately non-bipartite, and bipartite case, some further refinements of (3) are obtained by G. Quenell in [80]:

$$diam(G) \leq \begin{cases} \dfrac{\operatorname{arccosh}(n-1)}{\operatorname{arccosh}\left(\frac{d}{\Lambda}\right)} + 1 & G \text{ non-bipartite,} \\ \dfrac{\operatorname{arccos}(n/2-1)}{\operatorname{arccosh}\left(\frac{d}{\Lambda}\right)} + 2 & G \text{ bipartite.} \end{cases}$$

The inequality (3) can be further refined, by introducing the *injectivity radius* r of G into consideration. According to G. Quenell (see [80] for the definition of r) it holds:

$$diam(G) \leq \frac{\operatorname{arccosh}(\dfrac{n}{d(d-1)^{r-1}})}{\operatorname{arccosh}(\dfrac{d}{\Lambda})} + 2r + 1. \tag{4}$$

As also remarked in [80], the estimate (3) is better than (4) provided $\Lambda \geq 2\sqrt{d-1}$ (in other words, see below, (4) is better only for Ramanujan graphs).

Finally, let us mention that the inequality (3) has been generalized to the case of biregular graphs and regular directed graphs [31]. The authors also discuss connections to finite non-abelian simple groups, primitive association schemes, primitivity exponent of the adjacency matrix, covering radius of a linear code and Cayley graphs.

The relationship between the second largest eigenvalue in moduli of a graph and girth, was investigated by P. Solé [90]. For the graphs with small diameter we have:

$$\Lambda(G) \geq \begin{cases} \sqrt{d} & g(G) \geq 4 \\ \sqrt{2d-1} & g(G) \geq 6 \end{cases} \; ; \; G \text{ is non-bipartite},$$

$$\Lambda(G) \geq \begin{cases} \sqrt{d(n-2d)/(n-2)} & g(G) \geq 6 \\ \sigma\sqrt{d-1} & g(G) \geq 8 \end{cases} \; ; \; G \text{ is bipartite.}$$

(Here, as in Section 1, σ denotes the golden section.) For the graphs with larger diameter we have:

$$\Lambda(G) \geq \begin{cases} 2\sqrt{d-1}\cos\frac{\pi}{s+1} & G \text{ is non-bipartite,} \\ 2\sqrt{d-1}\cos\frac{2\pi}{s+1} & G \text{ is bipartite,} \end{cases}$$

where $s = \lfloor \frac{g(G)-1}{2} \rfloor$.

From the above (upper) bounds for diameter it generally follows that the diameter is expected to be smaller as λ_2 (or Λ is smaller). Thus, by these inequalities, we have at least partial explanations for the shape of cubic graphs.

We now turn to important class of graphs in this context, so called Ramanujan graphs:

Definition 3. Let G be a (conected) $d-$regular graph. Then G is called a Ramanujan graph if $\Lambda(G) \leq 2\sqrt{d-1}$.

Remark. The importance of the number $2\sqrt{d-1}$ in the above definition lies in the following lower bound due to N. Alon and R. Boppana (cf. [58];

see Proposition 4.2). Suppose $G_{d,n}$ is a d−regular (connected) graph on n vertices (d being fixed). Then for any sequence of such graphs we have:

$$\liminf_{n\to\infty} \Lambda(G_{n,d}) \geq 2\sqrt{d-1}.$$

Thus if one wants graphs with as small λ_2 as possible, the above number serves as the lower limit of what can be done. More information this kind of results can be found, for example, in [73].

Since the second largest eigenvalue is small in Ramanujan graphs, they are also good enlagers (see Definition 1 from above), and hence good magnifiers.

An infinite family of Ramanujan graphs have been constructed, for the first time, by A. Lubotzky, R. Phillips and P. Sarnak in 1988. These graphs were realised as Cayley graphs of some groups (such as, for example, group $PGL(2, F_q)$) relative to some symmetric subset (or, alternatively, as quotients of a quaternion group); see [58] for details. In particular, cubic Ramanujan graphs are treated in [20]. It is remarkable that the diameter of Ramanujan graphs cannot be too large (besides the bounds for particular Ramanujan graphs from [58], see the result of A. Nilli, given below). The girth of Ramanujan graphs is investigated in [10].

The following result of A. Nilli [73] explains some effects on λ_2 when, in fact, diameter increases. Let G be a d-regular graph, and suppose that G contains two edges the distance between which is at least $2k+2$ (the distance between two edges is the length of shortest path whose terminal vertices are the vertices of edges in question). Then we have:

$$\lambda_2(G) \geq 2\sqrt{d-1}\left(1-\frac{1}{k+1}\right)+\frac{1}{k+1}.$$

4. Rapidly mixing Markov chains

The second largest eigenvalue of graphs is of some interest in the theory of *rapidly mixing Markov chains*.

Consider a Markov chain on a finite *state space* $S_n = \{1, 2, \dots, n\}$ with *transition matrix* $P = [p_{ij}]_1^n$. Thus for any ordered pair i, j of states the quantity p_{ij} is the *transition probability* from state i to state j and is independent in the time t. The matrix P is non-negative and *stochastic*, i.e. its row sums are all equal to 1. Let π_i ($i = 1, 2, \dots, n$) be a probability distribution over S_n and suppose that $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S_n$. Then P is said to be *reversible* w.r.t. probability distribution π_i and the Markov chain is *ergodic* with the stationary distribution π_i.

As is well known, P has real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $\lambda_1 = 1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n > -1$. The rate of convergence to π_i is governed by the second largest eigenvalue in absolute value, i.e. by $\max(\lambda_2, |\lambda_n|)$. One can show that the influence of λ_n can be neglected so that the really important quantity is λ_2. A reversible Markov chain is called *rapidly mixing* if λ_2 is sufficiently small.

It is useful to identify an ergodic reversible Markov chain with a weighted undirected graph G (possibly containing loops) as follows. The vertex set is the state space S_n of the chain. If $p_{ij} \neq 0$, there is an edge in G between vertices i and j with the weight $q_{ij} = \pi_i p_{ij} = \pi_j p_{ji}$. The eigenvalues of G (i.e. of the weight matrix $Q = [q_{ij}]_1^n$) are equal to the eigenvalues of P. In this way we see that the theory of graph spectra is relevant to the problem considered. There are two immediate consequences of the above facts. Firstly, one can use the theory of graph spectra to evaluate or estimate λ_2 in Markov chains, in particular to find upper bounds for λ_2. Secondly, one can use known graphs with small λ_2 to construct rapidly mixing Markov chains.

Detailed elaboration of above ideas can be found in papers [2], [32], [33], [53], [88] and [89], just to mention a few among several papers by the same authors (D. J. Aldous, P. Diaconis, M. Jerrum, A. Sinclair). Note that rapidly mixing Markov chains are important parts in stochastic algorithms for enumeration of large combinatorial sets.

5. Miscelaneous

In this section we briefly mention other results concerning λ_2.

Let us define

$$\mu_2(G) = \lim_{d\to\infty} \inf\{\lambda_2(H) : G \subset H, d(H) > d\}.$$

A.J. Hoffman [46] proved the following result.

Theorem 1. *Let G be a graph with n vertices and with adjacency matrix A. Let Γ be the set of all $(0,1)$ matrices C with n rows and at least two columns such that every row sum of C is positive, and if C has more than two columns, no column can be deleted without destroying the property that C has positive row sums. Then*

$$\mu_2(G) = min_{C\in\Gamma}\lambda_1(A - C(J-I)^{-1}C^T).$$

It was proved by M. Doob [34] that the set of all second-largest eigenvalues is dense in the interval $(\sqrt{2+\sqrt{5}}, \infty)$. The same set has infinitely many accumulation points, but is nowhere dense in the interval $(-\infty, -1+\sqrt{2}]$. These points are described in some detail by J. Li [56].

It is proved by C. Licata and D.L. Powers [57] that the Platonic solids are *self-reproducing* in the following specific sense. We consider an eigenvalue λ (in this case $\lambda = \lambda_2$) of the graph of the solid P considered, and the corresponding eigenspace $\mathcal{E}(\lambda)$ which is of dimension k. The convex hull of a basis of $\mathcal{E}(\lambda)$ is a polytope Q. If Q is isomorphic with P, then P is called self-reproducing. It is also proved in [57] that some other polyhedra are self-reproducing.

Spectra of weighted adjacency matrices have been used by Y. C. de Verdiére to introduce a new important graph invariant in [94]. For a connected graph G we introduce the class $\mathcal{A}_G$ of matrices $A = [a_{ij}]$ for which $a_{ij} > 0$ if i and j are adjacent and $a_{ij} = 0$ otherwise. Let $\mu_1, \mu_2, \dots, \mu_m$ $(\mu_1 > \mu_2 > \cdots > \mu_m)$ be distinct eigenvalues of A with multiplicities $k_1 = 1, k_2, \dots, k_m$, respectively. Let $\mu(G) = \max k_2$, where maximum is taken over the class $\mathcal{A}_G$. For example, $\mu(K_n) = n-1$ and $\mu(K_{3,3}) = 4$. It is proved that G is planar if and only if $\mu(G) \leq 3$. It is conjectured that $\mu(G) \geq \chi(G) - 1$, where $\chi(G)$ is the chromatic number of G. The validity of this conjecture would imply the four colour theorem!

Various inequalities involving the isoperimetric number and the spectrum of graphs are provided by B. Mohar [63] and [64].

Second largest eigenvalue in random graphs is studied in [15], [39] and [40].

It is interesting to note that expanding properties in infinite graphs are related to the spectral radius of the graph [11].

6. Some applications

The topic concerning the second largest eigenvalue has many theoretical and practical applications. Its major interest stems from the fact that it is significantly related to various types of expansion (and concentration) properties of graphs. These properties, in turn, are of great practical and theoretical interest in many brances of mathematics and/or computer science (such as extremal graph theory (see, e.g., [8]), graph pebbling (see, e.g., [55]), computational complexity (see, e.g., [52]), parallel sorting algorithms (to be treated below), etc.) as well as other branches of science (like electrical engineering; some detals in connection with various networks are also included below).

We shall not attempt within this paper to go into details. Rather, we shall try to gain the importance of the topic toward various applications. The key idea is that many spectral parameters (invariants of graphs) are important link to structural properties (such as various expansion properties).

Informally, a graph has a "good" *expanding property* if each (its) vertex subset has a large neighbourhood. For bipartite graphs, more precisely, we have:

Definition 1. Let $G = (U, V, E)$ be a bipartite graph with $|U| = |V| = n$. Then G is an (n, α, β)- expanding $(0 < \alpha \leq \beta \leq n)$ if the following condition holds:

$$(\forall X \subseteq U)(|X| \geq \alpha \Rightarrow |\partial X| \geq \beta).$$

Here, for the sake of completness, we recall that $\partial X = \{y : d(y, X) = 1\}$, where d stands for the usual metric on a graph.

Bipartite graphs having good expanding properties are known as *expanders*. One of the most general definition reads as follows:

Definition 2. Let $G = (U, V, E)$ be a bipartite graph with $|U| = |V| = n$, and $|E| \leq dn$. Then G is an (n, d, δ, α)- expander $(\alpha \leq 1)$ if the follwing condition holds:

$$(\forall X \subseteq U)(|X| \leq \alpha n \Rightarrow |\partial X| \geq (1 + \delta(1 - \frac{|X|}{n}))|X|).$$

In particular, if $\alpha = \frac{1}{2}$, then G is called an (n, d, δ)-expaders and if $\alpha = 1$, then G is called a strong (n, d, δ)-expader.

In the above definition d and δ are regarded as *density* and *extension*, respectively. Notice also that the expression $(1 + \delta(1 - \frac{|X|}{n}))$ is larger as $|X|$ is smaller, which supports the fact that small subsets X, more likely, have large neighbourhood.

For non-bipartite graphs, the above definition has to be modified (since the verices are generally not distingused according to colour classes, or viewed as "input - output parts" of some system). According to [3], the non-bipartite analogon of expanders are *magnifiers*. The cooresponding definition (most frequently refering to regular graphs) reads as follows:

Definition 3. Let $G = (V, E)$ be a graph on n vertices, and maximal vertex degree d. Then G is an (n, d, δ)-magnifier if the following condition holds:

$$(\forall X \subseteq U)(|X| \leq \frac{1}{2}n \Rightarrow |\partial X| \geq \delta|X|.$$

Some examples of (good) expanders and magnifiers we have encountered in Section 3. To provide some hints on applications, we need some further definitions.

We first define two classes of graphs (having special connectivity properties and possibly small number of edges) which can be viewed as communication networks: concentrators (defined by M.S. Pinsker [76] in 1973) and superconcentrators (defined by L.G. Valiant [93] in 1975). There is an extensive literature on applications of these graphs in communication problems (a good source of references can be found, e.g., in [22]; see also [91] on construction of low complexity error-correcting codes).

Definition 4. An $(n, m)-$concentrator is a graph with n input vertices and m output vertices, $n \geq m$, having the property that, for any set of r $(\leq m)$ inputs, there exists a flow (a set of vertex-disjoint paths) that join the given inputs to some set of r outputs.

With a slight modification, we get the definition of superconcentrators.

Definition 5. An $n-$superconcentrator is a graph with n input vertices and n output vertices having the property that, for any set of r $(\leq n)$ inputs and any set of r outputs, there exists a flow that join the given inputs to given outputs.

Remark. Besides these two classes of graphs, which were firstly used in construction of various switching networks, there are many others of similar kind: for example, nonblocking networks where the partial correspondence between inputs and outputs by disjoint paths can be always extended without disturbing existing paths (see, e.g., [21] for more precise definition).

It is also worth mentioning that superconcentrators can be constructed from concentrators, but also from expanders (see, e.g., [41] and [83]). Superconcentrators, among others, are used in construction of parallel sorting networks [1].

As is well known from literature, expanding graphs within some properly choosen classes do exist. Moreover, by probabilistic arguments, one can show, with relative ease, that within many such classes almost every graph posseesses the desired property (see, for example, [12]). On the other hand, if one needs some of these graphs, there is no efficient algorithm, for a randomly choosen graph, to decide if it indeed satisfies the required properties (for example, it is known that the problem of checking if a given graph is an $(n, d, 0)-$expander is coNP-complete). So explicit constructions are desirable (but, as a rule, are very complicated). The first breakthrough was given by G.A. Margulis [59] (but without explicit estimate on expansion magnitude; only non-zero estimate is proved to exist). By a slight modification of the previous construction, O. Gabber and Z. Galil [41] have provided the estimate explicitely. Another important construction is due to N. Alon and V.D. Milman [5] (based on theory of group representations or harmonic analysis). For further constructions see [7], [4] (where finite geometries are used - points and hyperplains are the vertices of bipartite graph), [54] (bipartite graphs are obtained from affine transformations), [22] (graphs represented as k-sum are used), etc. On the other hand, it is worth noting that explicit constructions are in many circumstances poor substitute for probabilistic ones, since giving graphs with worse expanding properies then probabilistic ones.

Besides the particular graphs with good expanding properties, very frequently the (infinite) families of such graphs are more preferable.

In the rest, we give some details on sorting in rounds.

Suppose we are given n elements $x_1, \dots, x_n$ with linear order unknown to us. Our task is to determine this linear order by as few probes as possible. Each probe (or question) is a binary comparision (say, is $x_i > x_j$?). The (information) theoretic bound is, clearly, $\log_2 n!$ ($\sim n \log_2 n$). The sorting in rounds is organized as follows: In the first round we ask m_1 ($\leq m$) simultaneous questions. Having the answers, we deduce all implications and ask, in the second round, another $m_2 (\leq m)$ questions, deduce their implications, and so on. After r rounds, we need to have the unknown order. The need for such algorithms arises in structural modeling.

The sorting described above is in fact parallel sorting. Here m is a number of processors (also called the width of algorithm); r is a (parallel) time requred by algorithm (also called the depth of algorithm). The object is to minimize the size of the algorithm (equal to the number of comparisions), here denoted by $f_r(n)$.

It is known, for example that $f_1(n) = \binom{n}{2}$; $f_2(n) = O(n^{\frac{3}{2}} \log n)$ (probabilistic bound) and $f_2(n) = O(n^{\frac{7}{4}})$ (explicit construction by expanders).

Here the idea of using expanders is based on the fact that after each round enough comparisions are avoided due to good expanding properties of partial graph so far grown.

For more details see [4], [1] and [12].

References

[1] Ajtai M., Komlós J., Szemerédi, *Sorting in c log n parallel steps*, Combinatorica **3** (1988), 1–19.

[2] Aldous D. J., *The random-walk construction of uniform spanning-trees and uniform labeled trees*, SIAM J. Discrete Math. **3** (1990), no. 4, 450–465.

[3] Alon N., *Eigenvalues and expanders*, Combinatorica **6** (1986), 83–86.

[4] Alon N., *Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory*, Combinatorica **6 (3)** (1986), 207–219.

[5] Alon N., Milman V.D., *Eigenvalues, expanders and superconcentrators*, Proc. 25th Ann. IEEE Symp. on Found. of Comput. Sci. (1984,), 320–322.

[6] Alon N., Milman V.D., λ_1, *isoperimetric inequalities for graphs and superconcentrators*, J. Comb. Theory B **38** (1985), 73–88.

[7] Alon N., Galil Z., Milman V.D., *Better expanders and superconcentrators*, J. of Algorithms **8** (1987), 337–347.

[8] Beck J., *On size Ramsey number of paths, trees and circuits I*, J. Graph Theory **7** (1983), 115–129.

[9] Benzaken C., Hammer P.L., *Linear separation of dominating sets in graphs*, Annals of Discrete Math. **3** (1978), 1–10.

[10] Biggs N. L., Boshier A. G., *Note on the girth of Ramanujan graphs*, J. Comb. Theory B **49** (1990), no. 2, 190–194.

[11] Biggs N. L., Mohar B., Shawe-Taylor J., *The spectral radius of infinite graphs*, Bull. London Math. Soc. **20** (1988), 116–120.

[12] Bollobás B., *Random Graphs*, Academic Press, 1985.

[13] Brigham R. C., Dutton R. D., *Bounds on graphs spectra*, J. Combin. Theory Ser. B **37** (1984), 228–234.

[14] Brigham R. C., Dutton R. D., *Bounds on graph spectra and girth*, Ars Combinatoria **20** (1985), 91–100.

[15] Broder A., Shamir E., *On the second eigenvalue of random regular graphs*, 28th Ann. Symp. on FOCS, Los Angeles, CA, 1987, pp. 286–294.

[16] Bussemaker, F.C., Čobeljić S., Cvetković D., Seidel J.J., *Computer investigation of cubic graphs*, Technological University Eindhoven, T. H. - Report 76-WSK-01; Cubic graphs of ≤ 14 vertices, J. Comb. Theory (B) **23** (1977), 234–235.

[17] Cao D., Hong Y., *Graphs characterized by the second eigenvalue*, J. Graph Theory **17** (1993), no. 3, 325–331.

[18] Chen J. S., *Sharp bounds on the kth eigenvalue of trees*, Discrete Math. **128** (1994), no. 1–3, 61–72.

[19] Chen J. S., Cao D., *The sharp lower bound of the k-th largest eigenvalue of a forest*, J. Huazhong Univ. Sci., Tech. **18** (1990), no. 5, 1–6. (Chinese, English summary)

[20] Chiu P., *Cubic Ramanujan graphs*, Combinatorica **12** (1992), 275–285.

[21] Chung F. R. K., *On concentrators, superconcentrators, generalizers, and nonblocking networks*, The Bell System Tecn. Journal **58** (1978), no. 8, 1765–1777.

[22] Chung F. R. K., *Diameters and eigenvalues*, J. Amer. Math. Soc. **2** (1989), no. 2, 187–196.

[23] Cvetković D., *Some possible directions in further investigations of graph spectra*, Algebraic methods in graph theory, vol. I ed. L. Lovász, V. T. Sós, North-Holand, Amsterdam - Oxford - New York, 1981, pp. 47–67.

[24] Cvetković D., *On graphs whose second largest eigenvalue does not exceed 1*, Publ. Inst. Math. (Beograd) **31** (1982), 15–20.

[25] Cvetković D., Doob M., Gutman I., Torgašev A., *Recent Results in the Theory of Graph Spectra* (1988), North Holland Amsterdam .

[26] Cvetković D., Doob M., Sachs H., *Spectra of graphs - Theory and application*, Deutscher Verlag der Wissenschaften - Academic Press, Berlin - New York, 1980; III edition, Johann Ambrosius Barth, Heidelberg - Leipzig, 1995.

[27] Cvetković D., Rowlinson P., *The largest eigenvalue of a graph - a survey*, Linear and Multilinear Algebra **28** (1990), 3–33.

[28] Cvetković D., Simić S., *Graph theoretical results obtained by the support of the expert system "GRAPH"*, Bull. Acad. Serbe Sci. Arts, Cl. Sci. Math. Nat., Sci. Math. **107** (1994), no. 19, 19–41.

[29] Cvetković D., Simić S., *On the graphs whose second largest eigenvalue does not exceed* $(\sqrt{5}-1)/2$, Discrete Math. **138** (1995), 213–227.

[30] Cvetković D., Simić S., *Minimal graphs whose second largest eigenvalue is not less than* $(\sqrt{5}-1)/2$, in preparation.

[31] Delorme C., Solé P., *Diameter, covering index, covering radius and eigenvalues*, Eur. J. Comb. **12** (1991), no. 2, 95–108.

[32] Desai M., Rao V., *On the convergence of reversible Markov-chains*, SIAM J. Matrix Analysis and Appl. **14** (1993), no. 4, 950–966.

[33] Diaconis P., Strock D., *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probability **1** (1991), no. 1, 36–61.

[34] Doob M., *The limit points of eigenvalues of graphs*, Linear Algebra and Appl. **114/115** (1989), 659–662.

[35] Fiedler M., *Algebraic connectivity of graphs*, Czechoslovak Math. J. **23(98)** (1973), 298–305.

[36] Fiedler M., *An algebraic approach to connectivity of graphs*, Recent Advances in Graph Theory, ed. M. Fiedler, Academia Praha, Praha, 1975, pp. 193–196.

[37] Fiedler M., *Laplacian of graphs and algebraic connectivity*, Combinatorics and Graph Theory, Banach Centre Publ. **25**, PWN Polish Scientific Publ., Warshaw, (1989), 57–70.

[38] Fiedler M., *Absolute algebraic connectivity of trees*, Linear and Multilinear Algebra **26** (1990), no. 1–2, 85–106.

[39] Friedman J., *On the second eigenvalue and random walks in random d-regular graphs*, Combinatorica **11** (1991), 331–362.

[40] Friedman J., Kahn J., Szemeredy E., *On the second eigenvalue in random regular graphs*, Proc. 21st Ann. ACM Symp. Theory Comput., Seattle, 1989, pp. 587–598.

[41] Gabber O., Galil Z., *Explicit constructions of linear sized superconcentrators*, J. Comput. Systems Sci. **22** (1981), no. 3, 407–420.

[42] C.D. Godsil, *Linear algebra and combinatorics*, Handbook of combinatorics (to appear).

[43] Grone R., *On the geometry and Laplacian of a graph*, Linear Algebra and Appl. **150** (1991), 167–178.

[44] Grone R., Merris R., *Algebraic connectivity of trees*, Czechoslovak Math. J. **37** (1987), no. 4, 660–670.

[45] Grone R., Merris R., *Ordering trees by algebraic connectivity*, Graphs and Combinatorics **6** (1990), no. 3, 229–237.

[46] Hoffman A.J., *Applications of Ramsey style theorems to eigenvalues of graphs*, Combinatorics (ed. Hall M. Jr., Lint J. H. van) Part 2, Math. Centre Tracts, Amsterdam **56** (1974), 43–57.

[47] Hong Y., *The k-th largest eigenvalue of a tree*, Linear Algebra and Appl. **73** (1986), 151–155.

[48] Hong Y., *Bounds of eigenvalues of a graph*, Acta Math. Appl. Sinica (English Ser.) **4** (1988), no. 2, 165–168.

[49] Hong Y., *Sharp lower bounds on the eigenvalues of trees*, Linear Algebra and Appl. **113** (1989), 101–105.

[50] Howes L., *Ph. D. Thesis*, City Univ. of New York, 1970.

[51] Howes L., *On subdominantly bounded graphs - summary of results*, Recent Trends in Graph Theory, Proc. of the First New York City Graph Theory Conf. Held on Jun. 11-13, 1970, Capobianco M., Frechen J.B., Krolik M., Springer-Verlag, Berlin, 1971, pp. 181–183.

[52] JáJá J., *Time space tradeoffs for some algebraic problems*, Proc. 12th Annual ACM Sympos. on Theory of Computing, 1980, AMS, NY, 1980, pp. 339–350.

[53] Jerrum M., Sinclair A., *Approximating the permanent*, SIAM J. Computing **18** (1988), 1149–1178.

[54] Jimbo S., Maruoka A., *Expanders obtained from affine transformations*, Combinatorica **7** (1987), no. 4, 343–355.

[55] Lengauer T., R.E. Tarjan, *Asymptotically tight bounds on time space tradeoffs in a pebble game*, J. Assoc. Comput. Mach. **29** (1982), 1087–1130.

[56] Li J., *Ph. D. thesis*, The University of Manitoba, 1994.

[57] Licata C., Powers D. L., *A surprising property of some regular polytopes*, Scientia (Valparaiso), Ser. A **1** (1988), 73–80.

[58] Lubotzky A., Phillips R., Sarnak P., *Ramanujan graphs*, Combinatorica **8** (1988), 261–277.

[59] Margulis G. A., *Explicit group-theoretical construction of combinatorial schemes and their application to design of expanders and superconcentrators*, Problemy Peredachi Informatsii **9** (1973), 71–80, English translation: Problems of Infor. Trans, 1975, 325–332. (Russian)

[60] Maxwell G., *Hyperbolic trees*, J. Algebra **54** (1978), 46–49.

[61] Merris R., *Laplacian matrices of graphs: a survey*, Linear Algebra and Appl. **197/198** (1994), 143–176.

[62] Merris R., *A survey of graph Laplacians*, Linear and Multilinear Algebra (to appear).

[63] Mohar B., *Isoperimetric inequalities, growth and the spectrum of graphs*, Linear Algebra and Appl. **103** (1988), 119–131.

[64] Mohar B., *Isoperimetric numbers of graphs*, J. Comb. Theory B **47** (1989), 274–291.

[65] Mohar B., *Eigenvalues, diameter, and mean distance in graphs*, Graphs and Combinatorics **7** (1991), 53–64.

[66] Mohar B., *The Laplacian spectrum of graphs*, Graph Theory, Combinatorics and Applications, ed. Y. Alavi et al., Wiley, New York, 1991, 871–898.

[67] Mohar B., *Laplace eigenvalues of graphs – a survey*, Discrete Math. **109** (1992), 171–183.

[68] Mohar B., Poljak S., *Eigenvalues and the max-cut problem*, Czech Math. J. **40** (1990), 343–352.

[69] Mohar B., Poljak S., *Eigenvalues in combinatorial optimization*, Combinatorial and Graph-Theoretical Problems in Linear Algebra, ed. R. Brualdi, S. Friedland, V. Klee , Springer-Verlag, New York, 1993, 107–151.

[70] Neumaier A., *The second largest eigenvalue of a tree*, Linear Algebra and Appl. **46** (1982), 9–25.

[71] Neumaier A., *Derived eigenvalues of symmetric matrices with applications to distance geometry*, Linear Algebra and Appl. **134** (1990), 107–120.

[72] Neumaier A., Sedel J. J., *Discrete hyperbolic geometry*, Combinatorica **3(2)** (1983), 219–237.

[73] Nilli A., *On the second largest eigenvalue of a graph*, Discrete Math. **91** (1991), 207–210.

[74] Petrović M., *On graphs with exactly one eigenvalue less than* -1, J. Comb. Theory B **52** (1991), no. 1, 102–112.

[75] Petrović M., *On graphs whose second largest eigenvalue does not exceed* $\sqrt{2}-1$, Univ. Beograd, Publ. Elektrotehn. Fak., Ser. Mat. **4** (1993), 70–75.

[76] Pinsker M.S., *On the complexity of concentrators*, Proc. 7th International Teletraffic Conference, Stockholm, June 1973, 318/1–318/4.

[77] Powers D. L., *Structure of a matrix according to its second eigenvector .*, Current trends in matrix theory, ed. F. Uhlig, R. Grone , Elsevier, New York, 1987, 261–266.

[78] Powers D. L., *Graph partitioning by eigenvectors*, Linear Algebra and Appl. **101** (1988), 121–133.

[79] Powers D. L., *Bounds on graph eigenvalues*, Linear Algebra and Appl. **117** (1989), 1–6.

[80] Quenell G., *Spectral diameter estimates for k-regular graphs*, Adv. Math. **106** (1994), 122–148.

[81] Rowlinson P., *On the maximal index of graphs with prescribed number of edges*, Linear Algebra and Appl. **110** (1988), 43–53.

[82] Sarnak P., *Some applications of modular forms*, Cambridge Univ. Press, Cambridge, 1990.

[83] Shamir E., *From expanders to better superconcentrators without cascading*, preprint.

[84] Shao J. Y., *Bounds on the k-th eigenvalues of trees and forests*, Linear Algebra and Appl. **149** (1991), 19–34.

[85] Simić S., *Some notes on graphs whose second largest eigenvalue is less than* $(\sqrt{5}-1)/2$, Linear and Multilinear Algebra **31** (1995), 59–71.

[86] Simić S., *Complementary pairs of graphs with the second largest eigenvalue not exceeding* $(\sqrt{5}-1)/2$, Publ. Inst. Math. (Beograd) (to appear).

[87] Bell F. K. and Simić S. K., *A note on the second largest eigenvalue of some star-like trees* (to appear).

[88] Sinclair A., *Improved bounds for mixing rates of Markov chains and multicommodity flow*, LFCS Report Series, ECS-LFCS-91-178, University of Edinburgh, 1991.

[89] Sinclair A., Jerrum M., *Approximate counting, uniform generation and rapidly mixing Markov chains*, Information and Computation **82** (1988), no. 1, 93–133.

[90] Solé P., *The second eigenvalue of regular graphs of a given girth*, J. Comb. Theory B **56** (1992), 239–249.

[91] Tanner R. M., *A recursive approach to low complexity codes*, IEEE Trans. Inform. Theory **IT-17** (1981), 533–547.

[92] Tanner R. M., *Explicit concentrators from generalized n-gons*, SIAM J. Alg. Disc. Meth. **5** (1984), 287–293.

[93] Valliant L.G., *On nonlinear lower bounds in computational complexity*, Proc. 7th Annual ACM Symposium on Theory of Computing, Albuquerque, NM, May, 1975, pp. 45–53.

[94] Verdière Y. C. de, *Sur un novel invariant des graphes et un critère de planarité*, J. Comb. Theory B **50** (1990), 11–21.

[95] Wolk E.S., *A note on the comparability graph of a tree*, Proc. of A.M.S. **16** (1965), 17–20.

Faculty of Electrical Engineering, University of Belgrade, P.O.Box 816, 11 001 Belgrade, Serbia, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 473–481
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

SUBSYSTEMS OF PEANO ARITHMETIC AND CLASSICAL RESULTS OF NUMBER THEORY

C. Dimitracopoulos

ABSTRACT. We discuss problems and results concerning subsystems of first-order Peano Arithmetic, especially concerning the provability of basic theorems of elementary number theory and combinatorics.

Let LA denote the usual first-order language of arithmetic and PA denote Peano's axioms expressed in LA. Subsystems of PA are obtained by rextricting the induction schema or replacing it by a restriction of some other schema.

First we recall the definition of the "arithmetic hierarchy" of formulae of LA.

Definition 1: Let φ be a formula of LA and $n \geq 1$. We say that
(a) φ is Σ_0 or Π_0 or bounded if φ involves bounded quantifiers only, i.e. quantifiers of the form $\forall x \leq y$, $\exists x \leq y$, $\forall x < y$, $\exists x < y$.
(b) φ is Σ_n if φ is of the form $\vec{\exists}\,\vec{\forall}\,\ldots\,\theta$, where θ is bounded and there exist n alternations of blocks of similar quantifiers in front of θ.
(c) φ is Π_n if $\neg\varphi$ is logically equivalent to a Σ_n formula.

Now we proceed to the precise definition of the subsystems that were first studied.

Definition 2: For $n \geq 0$,
(a) $I\Sigma_n$ denotes PA with induction only for Σ_n formulae with parameters.
(b) $B\Sigma_n$ denotes $I\Sigma_0$ plus the collection schema for Σ_n formulas only, i.e. the schema

$$\forall x < z \exists y \varphi(x,y) \rightarrow \exists t \forall x < z \exists y < t \varphi(x,y),$$

where φ is any Σ_n formula with parameters.
(c) $L\Sigma_n$ denotes PA with the induction schema replaced by the least number

schema only for Σ_n formulae with parameters.
$I\Pi_n$, $B\Pi_n$, $L\Pi_n$ are defined similarly.

Remark 1. Strictly speaking, the subsystems defined above include a finite number of axioms expressing commutativity of + and ., associativity of + and ., etc.; these must be added to the usual axioms, because the amount of induction available in some of these systems is not sufficient to prove them.

Relations among the systems we have defined were proved by Paris & Kirby (see [14]) and are summarized as follows.

Theorem 1. *For all $n \geq 0$:*

$$\begin{tikzcd}
I\Sigma_{n+1} \arrow[d, Rightarrow] \\
B\Sigma_{n+1} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & B\Pi_n \\
I\Sigma_n \arrow[r, Leftrightarrow] & I\Pi_n \arrow[r, Leftrightarrow] & L\Sigma_n \arrow[r, Leftrightarrow] & L\Pi_n .
\end{tikzcd}$$

Furthermore, the converses of the vertical arrows do not hold.

In view of this theorem, the following question was asked in the 1970's:

Main Problem. *What is the weakest subsystem of PA that can serve as a basis for elementary number theory and combinatorics?*

It is not difficult to see that $I\Sigma_1$ is strong enough to serve as a basis: indeed, one can formalize usual proofs so that only induction for Σ_1 formulae is needed. But what happens with the strictly weaker systems $B\Sigma_1$ and $I\Sigma_0$? At this point we need to mention the following result, proved by Paris (see [12]) and, independently, H. Friedman.

Theorem 2. *For $n \geq 0$ and θ a Π_{n+2} sentence: $B\Sigma_{n+1} \vdash \theta \Rightarrow I\Sigma_n \vdash \theta$.*

By the previous theorem and the fact that all basic results of elementary number theory and combinatorics are formalized by Π_2 sentences, studying $I\Sigma_0$ is the same as studying $B\Sigma_1$, as far as our main problem is concerned. Unfortunately, $I\Sigma_0$ seems very weak, since the usual method of coding cannot work in it. To test its strength, the following problems were posed and still remain open:

Problem 1 (Paris). *Does $I\Sigma_0$ prove the MRDP theorem?*

By MRDP theorem we mean the following result of Matijasevič-Robinson-Davis-Putnam (see [10]), which was crucial for the negative solution of Hilbert's 10th problem:

MRDP Theorem. *For every Σ_1 formula $\varphi(\vec{x})$ we can effectively find a polynomial $p \in Z[\vec{x}, \vec{y}]$ such that*

$$N \models \forall \vec{x}[\varphi(\vec{x}) \rightarrow \exists \vec{y}(p = 0)],$$

where N denotes the standard model of PA.

Remark 2. Strictly speaking, $p = 0$ stands for $p^+ = p^-$, where $p = p^+ - p^-$ and $p^+, p^- \in N[\vec{x}, \vec{y}]$.

Since the MRDP theorem cannot be expressed as a set of sentences of LA, what is meant in problem 1 is: Can we replace $N \models$ by $I\Sigma_0 \vdash$ in the MRDP theorem?

Problem 2 (Wilkie). *Does $I\Sigma_0$ prove that the set of primes is unbounded?*

Problem 3 (Macintyre). *Does $I\Sigma_0$ prove $PHP\Sigma_0$?*

By $PHP\Sigma_0$ we denote the following schema, which formalizes the pigeonhole principle for Σ_0 maps:

$$\forall x \leq z \exists y < z\varphi(x, y) \rightarrow \exists x_1, x_2 \leq z \exists y < z(x_1 \neq x_2 \wedge \varphi(x_1, y) \wedge \varphi(x_2, y)),$$

where φ is any Σ_0 formula with parameters.

It is widely believed that the answer to all these problems is "no". Concerning Problem 1, this feeling is especially strong, in view of the following observation of A. Wilkie (see [19]):

If $I\Sigma_0$ proves the MRDP theorem, then $NP = co - NP$.

Given the difficulty of working with $I\Sigma_0$, it seemed worthwhile to consider systems strictly between $I\Sigma_0$ and $I\Sigma_1$. Such a system is $I\Sigma_0 + exp$, where exp denotes the axiom $\forall x, y \exists z(z = x^y)$. Here $z = x^y$ is a bounded formula defining the graph of the exponential function in the standard model; the existence of such a formula was first shown by J. Bennett (see [1]). $I\Sigma_0 + exp$ is strictly stronger than $I\Sigma_0$, since the latter can capture functions of polynomial growth only, by the following result of R. Parikh (see [11]).

Theorem 3. *If $I\Sigma_0 \vdash \forall x \exists y \varphi(x, y)$, where φ is Σ_0, then there exists $k \in N$ such that $I\Sigma_0 \vdash \forall x \exists y < x^k + k\ \varphi(x, y)$.*

It is also known that $I\Sigma_0 + exp$ is strictly weaker than $I\Sigma_1$; this holds since the former cannot prove $B\Sigma_1$, essentially by the proof that $I\Sigma_0$ cannot prove $B\Sigma_1$ (see [14]). However, $I\Sigma_0 + exp$ seems to be as strong as $I\Sigma_1$, as far as our main problem is concerned; the idea is that existential quantifiers, unbounded at first sight, are essentially bounded, as long as large numbers exist. For example, it is known that the answer to Problems 1 – 3 is "yes" if we replace $I\Sigma_0$ by $I\Sigma_0 + exp$, i.e. the following hold:

Theorem 4. *$I\Sigma_0 + exp$ proves the MRDP theorem.*

Theorem 5. *$I\Sigma_0 + exp$ proves $PHP\Sigma_0$.*

Theorem 6. *$I\Sigma_0 + exp$ proves that the set of primes is unbounded.*

The first of these results was proved by H. Gaifman and, independently, the author (see [6]) and the second one by the author and J. Paris (see [5]); the proofs of both are based on the fact that inside any $M \models I\Sigma_0 + exp$ one can code finite (in the sense of M), Σ_0 definable sequences of elements of M. Theorem 6 is proved by a straightforward formalization of the usual proof of Euclid's theorem.
To strengthen the belief that $I\Sigma_0 + exp$ is very strong, Ch. Cornaros and the author obtained (see [4])

Theorem 7. *$I\Sigma_0 + exp$ proves (a version of) the prime number theorem.*

The proof of this result is a modification of Selberg's proof, the two main differences being that
(a) an approximate logarithm function (previously introduced by A. Woods in [20]) is used instead of $log_e x$ and
(b) arguments involving limits have been replaced by inductive ones.

Attempts to solve Problems 2 and 3 led to the study of other systems strictly between $I\Sigma_0$ and $I\Sigma_1$. In connection with Problem 2, A. Woods showed

Theorem 8. *$I\Sigma_0 + PHP\Sigma_0$ proves Sylvester's theorem, i.e. that for any $1 \leq x \leq y$ one of $y+1, ..., y+x$ has a prime divisor $p > x$.*

Let us discuss briefly the idea of his proof. In the usual proof, by considering the largest powers of primes in the prime power decompositions of the numbers $1, \ldots, x, y+1, \ldots, y+x$, Sylvester showed that if no prime divisor of $y+1, \ldots, y+x$ exceeds x then for any function $\xi(x) \geq \pi(x)$ (= number of primes $\leq x$)

$$x!y! \geq (x+y-\xi(x))! \quad (\dagger).$$

But for sufficiently large x and a suitable choice of $\xi(x)$ (†) fails and so the result follows. Woods considered the logarithmic version of (†), using the approximate logarithm function referred to above. Then he "unravelled" (†) to obtain the underlying *comparison map*, the existence of which contradicts $PHP\Sigma_0$ (the "unravelling" was necessary, since it is not known how to define partial sums by a Σ_0 formula in $I\Sigma_0 + PHP\Sigma_0$).

Clearly, we obtain as corollaries of the previous theorem:
(a) the answer to Problem 2 is "yes" if $I\Sigma_0$ is replaced by $I\Sigma_0 + PHP\Sigma_0$
(b) if the answer to Problem 3 is "yes", then the answer to Problem 2 is "yes".
Note that $I\Sigma_0 + PHP\Sigma_0$ is strictly weaker than $I\Sigma_0 + exp$. To see this, let M be a nonstandard model of PA, $a \in M - N$ and K be the substructure of M with universe $\{x \in M : M \models x < a^n \text{ for some } n \in N\}$. Then $K \models I\Sigma_0 + PHP\Sigma_0$ (since $PHP\Sigma_0$ is Π_1 axiomatizable), but clearly $K \not\models exp$.

Another system studied by A. Woods is $I\mathcal{E}^2_*$, which is defined as follows:

Definition 3. Let $\mathcal{E}^2$ be the smallest class of (primitive recursive) functions containing $+, \cdot,$ all constant functions and closed under substitution and bounded recursion – this class was defined by A. Grzegorczyk (see [7]). $I\mathcal{E}^2_*$ is obtained from $I\Sigma_0$ if we

(a) add a new function symbol to LA, for each $f \in \mathcal{E}^2$,
(b) allow induction for $\mathcal{E}^2_*$ formulae, i.e. bounded formulae of the new language $LA(\mathcal{E}^2)$.
(c) add a defining axiom $DEF(f)$ for each new function symbol f.

Clearly, $I\Sigma_0 \subseteq I\mathcal{E}^2_*$, but it is unknown whether equality holds. It is easy to see that $I\mathcal{E}^2_*$ is contained in an extension by definitions of $I\Sigma_0 + exp$; in fact, this inclusion is strict, since $I\mathcal{E}^2_*$ can only capture functions of polynomial growth (i.e., Theorem 3 can be proved for $I\mathcal{E}^2_*$ instead of $I\Sigma_0$).

By exploiting the availability of "census functions" of $\mathcal{E}^2_*$-definable sets in $I\mathcal{E}^2_*$, i.e. the ability to count the number of elements of any $\mathcal{E}^2_*$-definable set by means of a function in $\mathcal{E}^2$, A. Woods proved (see [20])

Theorem 9. *$I\mathcal{E}^2_*$ proves $PHP\mathcal{E}^2_*$, where $PHP\mathcal{E}^2_*$ denotes the pigeonhole principle schema for $\mathcal{E}^2_*$ formulae.*

As a consequence of Theorems 8 and 9, Problem 2 has a positive answer if $I\Sigma_0$ is replaced by $I\mathcal{E}^2_*$.

Now we turn our attention to subsystems of $I\mathcal{E}^2_*$, studied by A. Berarducci & B. Intrigila (see [2]) and Ch. Cornaros (see [3]). Each one of these systems includes $I\Sigma_0$ and is included in $I\mathcal{E}^2_*$, but it is unknown whether any of these inclusions are proper.

Berarducci and Intrigila considered combinatorial principles provable in $I\mathcal{E}^2_*$; we will refer to only two, i.e. $weak - PHP\Sigma_0$ and $EQ\Sigma_0$.

Definition 4. (a) $Weak - PHP\Sigma_0$ is the following schema

$$\begin{aligned}(1+\varepsilon)z > z \wedge \quad & \forall x < (1+\varepsilon)z \ \exists y < z \ \varphi(x,y) \ \rightarrow \\ & \exists x_1, x_2 < (1+\varepsilon)z \ \exists y < z(x_1 \neq x_2 \wedge \varphi(x_1,y) \wedge \varphi(x_2,y)),\end{aligned}$$

where φ is any Σ_0 formula with parameters and $\varepsilon > 0$ is any rational number.
(b) $EQ\Sigma_0$ (equipartition principle for Σ_0 relations) tis the following schema

$\forall z$ "if $\varphi(x,y)$ defines an equivalence relation on z such that every
equivalence class
has exactly n elements, then $n \mid z$",

where φ is any Σ_0 formula and $n \in N$.

It should be noted that $weak - PHP\Sigma_0$ had been previously considered by J. Paris, A. Wilkie and A. Woods (see [14]) and that Theorem 9 clearly implies that $I\mathcal{E}^2_* \vdash weak - PHP\Sigma_0$.

Using Theorem 9, Berarducci and Intrigila obtained

Theorem 10. *$I\mathcal{E}^2_*$ proves $EQ\mathcal{E}^2_*$, where $EQ\mathcal{E}^2_*$ is as before, but considering $\mathcal{E}^2_*$ formulae instead of Σ_0 ones.*

They also showed that the following hold:

Theorem 11. *$I\Sigma_0 + weak - PHP\Sigma_0$ proves Lagrange's theorem, i.e. that every integer is the sum of four squares.*

Theorem 12. *$I\Sigma_0 + EQ\Sigma_0$ proves the "complementary conditions" of the quadratic reciprocity law, i.e. that for any odd prime p:*
(a) -1 is a quadratic residue $modp$ iff $p \equiv 1 mod 4$
(b) 2 is a quadratic residue $modp$ iff $p \equiv \pm 1 mod 8$.

Theorem 13. *$I\Sigma_0 + EQ\Sigma_0$ proves that a prime number is the sum of two squares iff it is of the form $4n + 1$.*

For the proofs they used the multiplicative property of Legendre's symbol $(\frac{x}{p})$ (p an odd prime) and some group-theoretical considerations – the usual proofs are based on Euler's criterion $(\frac{x}{p}) \equiv x^{p-1/2} modp$, but it is unknown whether this is provable in the theories considered.

Cornaros, continuing the work of Berarducci and Intrigila, proved

Theorem 14. *$I\mathcal{E}^2_*$ proves the quadratic reciprocity law, i.e. that for any odd primes p, q:*

$$(\frac{p}{q})(\frac{q}{p}) = (-1)^{(p-1)(q-1)/2}.$$

His proof is based on the usual one and exploits the the fact that $\prod_{0 \leq x \leq y, \varphi(x)} f(x) modp$ and $\sum_{0 \leq x \leq y} f(x)$ are $\mathcal{E}^2$ functions, for any $f \in \mathcal{E}^2$ and any $\mathcal{E}^2_*$ formula φ.

He also attempted to prove the following conjecture of A. Woods (see [20]).

Conjecture. *$I\Sigma_0(\pi) + DEF(\pi)$ proves that the set of primes is unbounded, where $I\Sigma_0(\pi) + DEF(\pi)$ is the subsystem of $I\mathcal{E}^2_*$ if we allow only one new function symbol π corresponding to the usual function $\pi(x)$ = number of primes $\leq x$.*

Cornaros showed that adding π and one more new function symbol to LA suffices, namely

Theorem 15. *$I\Sigma_0(\pi, K)+DEF(\pi)+DEF(K)$ proves Bertrand's postulate, where K is a new function symbol corresponding to the usual function $K(x) = \sum_{0<n\leq x} log_e n$.*

For the proof, an approximate logarithm function is used again and care is taken to define other functions involved in the usual proof, e.g. $\psi(x)$, in a $\Sigma_0(\pi, K)$ manner.

Next, we discuss problems and results concerning the system $I\Sigma_0 + \Omega_1$, where Ω_1 denotes the axiom $\forall x \exists y(y = x^{[log_2 x]})$. By Theorem 3, $I\Sigma_0$ is strictly weaker than $I\Sigma_0+\Omega_1$. To see that $I\Sigma_0+\Omega_1$ is strictly weaker than $I\Sigma_0+exp$, it suffices to consider the structure with universe $\{x \in M : M \models x < a^{[log_2 a]^n}$ for some $n \in N\}$, for an arbitrary nonstandard $M \models PA$ and $a \in M - N$.

Let us see what is known about Problems 1 – 3 if $I\Sigma_0$ is replaced by $I\Sigma_0 + \Omega_1$.

(a) The feeling is that Problem 1 again has a negative solution. Indeed, Wilkie's observation shows that if $I\Sigma_0 + \Omega_1$ proves the MRDP theorem, then $NP = co - NP$.

(b) By using ingenious coding techniques, Paris, Wilkie and Woods showed (see [15])

Theorem 16. *$I\Sigma_0 + \Omega_1$ proves $weak - PHP\Sigma_0$.*

(c) Again in [15] one finds

Theorem 17. *$I\Sigma_0 + \Omega_1$ proves that the set of of primes is unbounded.*

Actually this can be improved to

Theorem 18. *$I\Sigma_0 + \Omega_1$ proves Sylvester's theorem.*

This follows from Theorem 16 and the fact that Woods's proof of Theorem 8 really uses $weak - PHP\Sigma_0$, not $PHP\Sigma_0$.

We continue with a short discussion of a very weak subsystem of $I\Sigma_0$. This is denoted by $IOpen$ and is obtained from $I\Sigma_0$ if we allow induction for open formulae only. The study of free-variable systems was first advocated by T. Skolem (see [17]). Shepherdson obtained (see [16])

Theorem 19. *$IOpen$ does not prove any of the following:*
(a) $x^2 \neq 2y^2 \vee x = 0$
(b) $x^3 + y^3 \neq z^3 \vee xyz = 0$
(c) the set of primes is unbounded.

By part (a) of this theorem and the fact that $I\Sigma_0$ proves $\forall x, y(x^2 \neq 2y^2 \vee x = 0)$, it follows that $IOpen$ is strictly weaker than $I\Sigma_0$.

To prove this theorem, Shepherdson constructed a recursive nonstandard model M of $IOpen$, in which (a)-(c) fail, as follows:
The universe of M is the set of all polynomials of the form

$$a_p X^{p/q} + a_{p-1} X^{(p-1)/q} + \cdots + a_1 X^{1/q} + a_0,$$

where $p, q \in N$, $q > 0$, $a_p, \dots, a_1$ are real algebraic, $a_p > 0$ if $p > 0$, a_0 is an integer and is ≥ 0 if $p = 0$. Successor, addition, etc., are defined in the obvious way; by taking X to be "infinitely large", one can make M into a discretely ordered semi-ring.

Many other authors studied $IOpen$, among which A. Wilkie ([19]), L. Van den Dries ([18]) and A. Macintyre & D. Marker ([9]), obtaining very interesting results. We mention only one result from [9], namely

Theorem 20. *$IOpen$ does not prove Lagrange's theorem.*

Most proofs in [9], including the proof of the previous result, involve constructions of models by unions of chains arguments and repeated use of purely algebraic constructions.

Let us finish with a remark: Most of the systems we have defined in this paper have been studied extensively from more than one viewpoints, but we have been concerned only with results associated to the main problem stated at the beginning. For information concerning other viewpoints, we urge the interested reader to consult A. Macintyre's excellent survey of the subject ([8]).

References

[1] J. Bennett: *On Spectra*, Ph. D. thesis, Princeton University, 1962.

[2] A. Berarducci & B. Intrigila: *Combinatorial principles in elementary number theory*, Ann. Pure Appl. Logic **55** (1991), 35 – 50.

[3] Ch. Cornaros: *On Grzegorczyk Induction*, Ann. Pure Appl. Logic .. (1995), ...

[4] Ch. Cornaros & C. Dimitracopoulos: *The prime number theorem and fragments of PA*, Arch. Math. Logic **33** (1994), 265 – 281.

[5] C. Dimitracopoulos & J. Paris: *The pigeonhole principle and fragments of arithmetic*, Z. Math. Logik Grundlag. Math. **32** (1986), 73 – 80.

[6] H. Gaifman & C. Dimitracopoulos: *Fragments of Peano's arithmetic and the MRDP theorem*, Logic and Algorithmic, Monograph. Enseign. Math. **30** (1982), 187 – 206.

[7] A. Grzegorczyk: *Some classes of recursive functions*, Rozprawy Mat. IV (1953), 1 – 45.

[8] A. Macintyre: *The strength of weak systems*, Proc. of the 11th Wittgenstein Symposium, Kirchberg/Wechsel (Austria), Hölder-Pichler-Tempsky, Wien, 1987, 43 – 59.

[9] A. Macintyre & D. Marker: *Primes and their residue rings in models of open induction*, Ann. Pure Appl. Logic **43** (1989), 57 – 77.

[10] Y. V. Matijasevič: *Enumerable sets are diophantine* (Russian), Dokl. Akad. Nauk SSSR **191** (1970), 279 – 282. English translation: Soviet Math. Doklady **11** (1970), 354 – 357.

[11] R. Parikh: *Existence and feasibility in arithmetic*, J. Symbolic Logic **36** (1971), 494 – 508.

[12] J. Paris: *Some conservation results for fragments of arithmetic*, Lecture Notes in Math. **890** (1981), 251 – 262.

[13] J. Paris & C. Dimitracopoulos: *Truth definitions for Δ_0 formulae*, Logic and Algorithmic, Monograph. Enseign. Math. **30** (1982), 317 – 329.

[14] J. Paris & L. A. S. Kirby: *Σ_n-Collection schemas in arithmetic*, Logic Colloquium '77, North-Holland, 1978, 199 – 209.

[15] J. B. Paris, A. J. Wilkie & A. R. Woods: *Provability of the pigeonhole principle and the existence of infinitely many primes*, J. Symbolic Logic **53** (1988), 1235 – 1244.

[16] J. Shepherdson: *A non-standard model for a free variable fragment of number theory*, Bull. Polish Acad. Sci. Math. **12** (1964), 79 – 86.

[17] Th. Skolem: *Peano's axioms and models of arithmetic*, Mathematical interpretations of formal systems, Amsterdam, 1955, 1 – 14.

[18] L. Van den Dries: *Some model theory and number theory for models of weak systems of arithmetic*, Lecture Notes in Math. **834** (1980), 346 – 362.

[19] A. J. Wilkie: *Some results and problems on weak systems systems of arithmetic*, Logic Colloquium '77, North-Holland, 1978, 285 – 296.

[20] A. R. Woods: *Some problems in logic and number theory* and their connections, Ph. D. thesis, Manchester University, 1981.

Department of Methodology, History and Philosophy of Science,
University of Athens, 37, J. Kennedy Str., 161 21 Kaisariani, Greece.
E-mail address: cdimitr@atlas.uoa.ariadne-t.gr

FILOMAT (Niš) 9:3 (1995), 483–492
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

ON THE MAXIMAL ORDER OF CERTAIN ARITHMETIC FUNCTIONS

Aleksandar Ivić

ABSTRACT. An upper bound for $f(f(n))$ is obtained when $f(n)$ belongs to a certain class of multiplicative functions. Also the maximal and average order of $Q(n)$ and $Q(Q(n))$ are determined, where $Q(n)$ denotes the number of distinct exponents in the canonical decomposition of n.

It is well-known (see e.g. Hardy and Wright [3]) that

$$\limsup_{n\to\infty} \frac{\log d(n)\log\log n}{\log n} = \log 2 \tag{1}$$

where $d(n)$ denotes the number of divisors of n. A more difficult problem is to determine the maximal order of $d(d(n))$. In [1] P. Erdős and the author have shown that

$$\log d(d(n)) \ll \left(\frac{\log n \log_2 n}{\log_3 n}\right)^{1/2}, \tag{2}$$

where $\log_k x = \log(\log_{k-1} x)$ is the k-fold iterated natural logarithm of x, and $f(x) \ll g(x)$ (same as $f(x) = O(g(x))$) means that $\|f(x)\| \le Cg(x)$ for some $C > 0$, $g(x) > 0$, $x \ge x_0$. The upper bound in (2) is certainly close to being best possible. Namely if one takes

$$N = p_1^{p_1-1} p_2^{p_2-1} \cdots p_r^{p_r-1}, \qquad r \to \infty,$$

where p_j is the j-th prime number, then

$$d(N) = p_1p_2\cdots p_r, \quad d(d(N)) = 2^r.$$

Supported by Grant 0401B of SFS through Math. Inst. SANU.

But since from the prime number theorem (see [3]) it follows that

$$p_k = k(\log k + O(\log\log k)),$$

we have, with $\theta(x) = \sum_{p \le x} \log p$,

$$\log N = \sum_{k \le r} \log p_k - \theta(p_k) = \sum_{k} k \log^2 k + O(r^2 \log r \log\log r)$$
$$= \frac{1}{2} r^2 \log^2 r (1 + O\left(\frac{\log\log r}{\log r}\right)).$$

Therefore

$$r = \omega(N) = \frac{2(2\log N)^{1/2}}{\log_2 n}\left(1 + O\left(\frac{\log_3}{\log_2}\right)\right), \tag{3}$$

where $\omega(n)$ denotes the number of distinct prime factors of n. This gives

$$\log d(d(N)) = \frac{2\log 2(2\log N)^{1/2}}{\log_2 n}\left(1 + O\left(\frac{\log_3}{\log_2}\right)\right), \tag{4}$$

which was already known to S. Ramanujan (see [5]).

P. Erdős and I. Kátai [2] proved that for every $\varepsilon > 0$

$$\log d^{(r)}(n) \ll (\log n)^{1/\ell_r + \varepsilon}$$

and that

$$\log d^{(r)}(n) > (\log n)^{1/\ell_r - \varepsilon}$$

for infinitely many n; where is the r-fold iterated divisor function and is the r-th Fibonacci number: Their method, however, does not seem to yield any improvement of (2). ℓ_r is r-th Fibnacci number: ℓ_{-1}, ℓ_0, $\ell_r = \ell_{r-2}$ $(r \ge 1)$. Their method, however, does not seem to yield any improvement of (2).

The argument in [1] that led to (2) depended on an upper bound for

$$Q = Q(S; n) := \sum_{a_i \ge S} 1, \tag{5}$$

where $n > 1$, $1 \le S \le \log n / \log 2$ and

$$n = p_{j_1}^{a_1} p_{j_2}^{a_2} \cdots p_{j_r}^{a_r} \tag{6}$$

is the canonical decomposition of n. As one trivially has $n \ge 2^{QS}$, it follows that

$$Q \le \frac{\log n}{S \log 2} \qquad (n > 1),$$

but a slightly better bound also holds. Namely (6) yields

$$\log n \ge \sum_{a_i \ge S} a_i \log p_{j_i} \ge S \sum_{p \le p_Q} \log p = S\theta(p_Q) \ge \frac{1}{2} SQ \log Q$$

for $Q \ge Q_0$. Thus $Q \le Q_1 = Q_1(S, n)$ where $(2 \log n)/S = Q_1 \log Q_1$. If $S \le \log^A n$, $0 < A < 1$, then

$$2 \log Q_1 \ge \log Q_1 + \log\log Q_1 \ge \log 2 + (1 - A) \log\log n \gg \log\log n,$$

hence $logQ_1 \gg \log\log n$, which gives

$$(7) \qquad Q(S, n) \ll \frac{\log n}{S \log\log n} \qquad (1 \le S \le log^A n\, 0 < A < 1).$$

If $a(n)$ denotes the number of non-isomorphic abelian (i.e. commutative) groups with n elements, then $a(n)$ is a multiplicative function (meaning $a(mn) = a(m)a(n)$ if m, n are coprime natural numbers) and $a(pk) = P(k)$, where $P(k)$ is the number of partitions of k. It was shown in [1] that with n elements, then $a(n)$ is a multiplicative function (meaning $a(mn) = a(m)a(n)$ if m, n are coprime natural numbers) and $a(p^k) = P(k)$, where $P(k)$ is the number of partitions of k. It was shown in [1] that

$$(8) \quad \omega(a(n)) \ll (\log n)^{3/4}(\log_2 n)^{-8}, \log a(a(n)) \ll (\log n)7/8(log_2 n)^{-C}$$

with $B = 11/8, C = 19/16$. In what follows a variation of the method developed in [1] will be used to prove a general result for iterates of certain arithmetic functions, which in the case of the function a(n) yields the slightly better values B = 7/4, C = 11/8 in (8). Perhaps the correct values of the exponents of the logarithms in (8) are both 1/2 (they cannot be smaller than 1/2). If true, this conjecture seems difficult to prove. [1] will be used to prove a general result for iterates of certain arithmetic functions, which in the case of the function $a(n)$ yields the slightly better values $B = 7/4$, $C = 11/8$ in (8). Perhaps the correct values of the exponents of the logarithms in (8) are both 1/2(they cannot be smaller than 1/2). If true, this conjecture seems difficult to prove.

The functions $a(n)$ and $d(n)$ belong to the class of arithmetic functions F, which contains all multiplicative, prime-independent functions $f(n) : N \to N$ such that

$$f(p^k) = g(k),\ g(k) \le e^{Ak^c} \qquad (0 < c < 1,\ A > 0) \tag{9}$$

for all integers $k \ge 1$ and primes p, where $g(k) \in N$. As we have $d(p^k) = k+1$ (9) holds in this case for any $c > 0$, and in the case of $a(n)$ it holds with $c = 1/2$, since $P(k) \le e^{A\sqrt{k}}$ (see [5]). A simple proof that

$$\limsup_{n \to \infty} \frac{\log f(n) \log\log n}{\log n} = \max_{k \ge 1} (f(2^k))^{1/k} \tag{10}$$

if $f(n) \in F$ was given by P. Shiu [6]. We shall be interested here in the maximal order of $f(f(n))$ when $f(n) \in F$. Lack of information about the arithmetic structure of $g(k)$ makes this, in general, quite a difficult problem. Even in the relatively simple case of $d(n)$ the existing bounds (2) and (3) are of a different order of magnitude. We shall prove an upper bound result, contained in

Theorem 1. *If $f(n) \in F$ and c is given by (9), then*

$$\log f(f(n)) \ll (\log n)^{11+2c-c^2} (\log_2 n)^{(c^2-3)/2}. \tag{11}$$

Proof. We shall prove first that

$$\omega(f(n)) \ll (\log n)^{(c+1)/2} (\log_2 n)^{-(c+3)/2}, \tag{12}$$

which seems to be of independent interest. Let the a_j's denote the distinct exponents in the canonical decomposition of $n (n > 1)$. Since

$$\omega(mn) \le \omega(m) + \omega(n),\ \omega(n^k) = \omega(n),\ \omega(n) \ll \frac{\log n}{\log\log n},$$

we have, for suitaqble integers $\beta_i \geq 1$,

$$\begin{aligned}\omega(f(n)) &= \omega\left(\prod_{a_i<S} g^{\beta_i}(a_i) \prod_{a_i\geq S} G^{\beta_i}(a_i)\right) \leq \sum_{a_i<S} \omega(g(a_i)) + \sum_{a_i\geq S} \omega(g(a_i)) \\ &\ll \frac{S^c}{\log S} \sum_{a_i<S} 1 + \sum a_i \geq S \frac{a_i^c}{\log a_i} \\ &\ll \frac{S^{c+1}}{\log S} + \sum_{j=0}^{O(\log\log n)} \sum_{2^j S\leq a_i\leq 2^{j+1} S} \frac{a_i^c}{\log a_i} + \sum_{a_i\geq(\log n)^{(3+c)/4}} \frac{a_i^c}{\log a_i} \\ &\ll \frac{S^{c+1}}{\log S} + \sum_{j=0}^{O(\log\log n)} \frac{2^{jc} S^c}{\log S} Q(2^j S, n) + (\log n)^c Q((\log n)^{(3+c)/4}, n) \\ &\ll \frac{S^{c+1}}{\log S} + \sum_{j=0}^{O(\log\log n)} \frac{2^{jc} S^c}{\log S} \cdot \frac{\log n}{2^j S \log\log n} + (logn)^{(3+c)/4} \\ &\ll \frac{S^{c+1}}{\log S} + \frac{S^{c-1}\log n}{\log S \log\log n} + (logn)^{(3+c)/4},\end{aligned}$$

where summation is over j such that $2^{j+1} S \geq (\log n)^{(3c+1)/4}$, and where we used (7). Now the choice

$$S = \left(\frac{\log n}{\log\log n}\right)^{1/2}$$

gives (12), since $0 < c < 1$. To obtain (11) from (12) note that, if (6) holds (the exponents now do not have to be distinct), then by Hölder's inequality and (7) it follows that ($\Omega(n)$ is the number of all prime divisors of n)

$$\log f(n) \leq A \sum_{i=1}^{r} a_i^c \leq A(\Omega(n))^c(\omega(n))^{1-c}. \tag{13}$$

In (13) we replace n by $f(n)$, use (12),(10) and the fact that $\Omega(n) \leq \log n/\log 2$ for all $n \geq 1$. We obtain

$$\log f(f(n)) \ll \left(\frac{\log n}{\log\log n}\right)^c ((\log n)^{(c+1)/2}(\log\log n)^{-(c+1)/2})^{1-c},$$

which gives then (11).This ends the proof of Theorem 1.

We recall that $a(n) \in F$ with $c = 1/2$, so that (12) and (11) yield $B = 7/4$ and $C = 11/8$ in (8), as already mentioned.

It follows from (10) that (11) gives a non-trivial upper bound for $log f(f(n))$. However,Theorem 1 certainly does not resolve the problem of the maximal order of $log f(f(n))$, whose solution requires additional information on the function $g(k)$ in (9). To see that $f(f(n))$ may assume both large and very small values infinitely often if $f(n) \in F$, we present the following two examples.

Example 1. Let $f_1(n) \in F$ with $f_1(p^k) = g_1(k)$, $g_1(1) = g_1(2) = 2$ $g_1(k) = [e^{k^c}]$ for $k \geq 3$ and a fixed c such that $0 < c < 1$, where $[x]$ denotes the integer part of x. Then if$n = (p_1 p_2 \cdots p_K)^2$ ($K \to \infty$) we have

$$f_1(n) = 2^K,\ f_1(f_1(n)) = [e^{K^c}],\ \log n = 2\theta(p_K) \sim 2K \log K.$$

Thus for infinitely many n we have

$$\log f_1(f_1(n)) \gg \left(\frac{\log n}{\log\log n}\right)^c. \tag{14}$$

By construction the constant c in (14) is, for $f_1(n)$, the same as the one appearing in (9). If we compare the bounds in (11) and (14) for $log f_1(f_1(n))$ it is hard to tell which one lies closer to the true order of magnitude of $log f_1(f_1(n))$. Although $g_1(k)$ in this example is of simple form, its arithmetic structure is obscure, and for this reason the problem is a hard one.

Example 2. Let $f_2(n) \in F$ with

$$f_2(p^k) = \begin{cases} 1 & k \neq 2 \\ 2 & k = 2. \end{cases}$$

In the previous example the function $f_1(f_1(n))$ exhibited large values, but in this case we clearly have

$$\liminf_{n\to\infty} f_2(f_2(n)) = 1,\ \limsup_{n\to\infty} f_2(f_2(n)) = 2,$$

since $f_2(f_2(n))$ equals either 1 or 2. Here, at least, the problem of the maximal order of $f_2(f_2(n))$ is solved. Note, however, that $f_2(n)$ itself takes large values, since by (10) one has

$$\limsup_{n\to\infty} \frac{\log f_2(n) \log\log n}{\log n} = \frac{\log 2}{2}.$$

Related to the functions $\omega(n)$, $\Omega(n)$ is the function $Q(n)$, which for $n > 1$ we define as the number of *distinct* exponents a_j in the canonical decompositon (6) of n, and for convenience we set $Q(1) = 1$. Note that the function $Q(n)$ is neither multiplicative nor additive. We shall determine the maximal and average order of $Q(n)$ and $Q(Q(n))$. The results on the maximal order are contained in

Theorem 2. *For $n \geq n_o$ we have*

$$Q(n) \leq 2\left(\frac{\log n}{\log_2 n}\right)^{1/2}\left(1 + O\left(\frac{\log_3 n}{\log_2 n}\right)\right), \tag{15}$$

and equality holds in (15) *for infinitely many n. We also have*

$$Q(Q(n)) \leq \left(\frac{2\log_2 n}{\log_3 n}\right)^{1/2}\left(1 + O\left(\frac{\log_4 n}{\log_3 n}\right)\right), \tag{16}$$

and equality holds in (16) *for infinitely many n.*

Proof. Take

$$n = p_1^1 p_2^2 \cdots p_K^K, \quad K \to \infty. \tag{17}$$

Then

$$K = \omega(n) = Q(n), \quad Q(Q(n)) = Q(K). \tag{18}$$

But from (17) we have

$$\log n = \sum_{j \leq K} j \log p_j = \sum_{j \leq K} j(\log j + O(\log_2 j)) \tag{19}$$

$$= \frac{1}{2}K^2 \log K + O(K^2 \log_2 K),$$

which gives

$$Q(n) = 2\left(\frac{\log n}{\log_2 n}\right)^{1/2}\left(1 + O\left(\frac{\log_3 n}{\log_2 n}\right)\right)$$

for n given by (17), that is, for infinitely many n. From (18) we have

$$Q(Q(n)) = Q(K) = 2\left(\frac{\log K}{\log_2 K}\right)^{1/2}\left(1 + O\left(\frac{\log_3 K}{\log_4 K}\right)\right)$$

for infinitely many K of the form

$$K = p_1^1 p_2^2 \cdots p_r^r, \ r \to \infty. \tag{20}$$

But from (19) it follows that

$$\log K = \frac{1}{2}\log_2 n + O(\log_3 n), \ \log\log K = \log_3 n + O(1).$$

Inserting those values in the expression for $Q(Q(n))$ it follows that equality holds in (16) if n is given by (17) and $K$by (20).

To obtain an upper bound in (15) note that if $(j_1, j_2, \dots, j_Q)$ is any permutation of $(1, 2, \dots, Q)$ and $1 \le a_1 < \dots < a_Q$ $(Q = Q(n))$ are the distinct exponents in the canonical decomposition of n, then

$$a_i \ge i \qquad (i = 1, \dots, Q).$$

Thus we have, for some permutation $(j_1, j_2, \dots, j_Q)$ of $(1, 2, \dots, Q)$,

$$\begin{aligned}\log n &\ge \sum_{i=1}^{Q} a_i \log p_{j_i} \ge \sum_{i=1}^{Q} a_i \log p_{Q-i+1} \ge \sum_{i=1}^{Q} i \log p_{Q-i+1}\\ &= \sum_{i=1}^{Q} (Q-i+1) \log p_i = \sum_{i=1}^{Q} (Q-i+1)(\log i + O(\log_2 i))\\ &= Q \sum_{i=1}^{Q} \log i - \sum_{i=1}^{Q} i \log i + O(Q^2 \log_2 Q) = \frac{1}{2} Q^2 (\log Q + O(\log_2 Q)).\end{aligned}$$

The above expression is similar to (19) and easily implies the upper bound in (15). Since the right-hand side of (15) is an increasing function of n for $n \ge n_1$ we have

$$Q(Q(n)) \le 2 \left(\frac{\log Q(n))}{\log_2 Q(n)} \right)^{1/2} \left(1 + O\left(\frac{\log_3 Q(n)}{\log_2 Q(n)} \right) \right),$$

and if apply (15) to the right-hand side of the last inequality, we obtain (16). This completes the proof of Theorem 2.

To investigate the average order of $Q(n)$ and $Q(Q(n))$ we shall use the approach developed by G. Tenenbaum and the author [4]. Therein an s-function $f(n)$ was defined as an arithmetic function for which $f(n) = f(s(n))$, where $s(n)$ denotes the squarefull part of n (s is called squarefull if $s = 1$ or if $p^2 \mid s$ whenever $p \mid s$, p a prime). Thus $a(n)$ and $\Omega(n) - \omega(n)$ are both s-functions, the former being multiplicative and the atter additive. Now $Q(n)$ is neither multiplicative nor additive, but it turns out that it is "nearly" an s-function. Every n can be uniquely written as $n = qs$, $(q, s) = 1$, where $q = q(n)$ is squarefree (meaning that it is either 1 or a product of distinct primes) and $s = s(n)$ is squarefull. But then

$$Q(n) = \begin{cases} 1 + Q(s(n)) & \text{if } q(n) > 1, \\ Q(s(n)) & \text{if } q(n) = 1. \end{cases} \tag{21}$$

Therefore

$$\sum_{n\le x} Q(n) = \sum_{s\le x}(1+Q(s)) \sum_{1<n\le x/s,(q,s)=1} 1 + \sum_{s\le x} Q(s).$$

We evaluate the sum over q by (1.4) and (1.5) of [4], noting that $\sum_{s\le x} 1 \ll \sqrt{x}$. We obtain

$$\sum_{n\le x} Q(n) = \sum_{s\le x}(1+Q(s))\times \tag{22}$$
$$\times\left\{\frac{6x}{\pi^2 s}\prod_{p|s}(1+p^{-1})^{-1} + O(B(s)s^{-1/2}x^{1/2}\log x\right\}$$

with

$$B(n) = \prod_{p|n}(1+p^{-1/2}.$$

To estimate the error term in (22) we use (15) and

$$\sum_{s\le x} B(s)s^{-1/2} \le \prod_{p\le x}(1+B(p)\sum_{m=2}^{\infty} p^{-m/2}) \ll \log x.$$

In a similar way we may evaluate the summatory function of $Q(Q(n))$. The expression will be similar to (22), only instead of $1+Q(s)$ we shall have $Q(1+Q(s))$. We obtain

Theorem 3. *We have*

$$\sum_{n\le x} Q(n) = Dx + O(x^{1/2}\log^{5/2} x(\log_2 x)^{-1/2}),$$

$$D = \frac{6}{\pi^2}\sum_{s=1}^{\infty}\frac{1+Q(s)}{s}\prod_{p|s}(1+p^{-1})^{-1},$$

$$\sum_{n\le x} Q(Q(n)) = Ex + O(x^{1/2}\log^2 x(\log_2 x)^{1/2}(\log_3 x)^{-1/2}),$$

$$E = \frac{6}{\pi^2}\sum_{s=1}^{\infty}\frac{Q(1+Q(s))}{s}\prod_{p|s}(1+p^{-1})^{-1}.$$

It may be noted that by similar arguments one also obtains

$$\sum_{n\le x, Q(n)=k} 1 = d_k x + O(x^{1/2}\log^2 x), \tag{23}$$

where the so-called "local density" d_k is given by

$$d_k = \frac{6}{\pi^2} \prod_{s=1, Q(s)=k-1}^{\infty} \frac{1}{s} \prod_{p|s} (1+p^{-1})^{-1} \qquad (k \geq 2),$$

and $d_1 = 6\pi^{-2}$ (since $Q(n) = 1$ if n is a power of a squarefree number). The error term in (23) is uniform in k, and each $d_k > 0$, since for any given $k > 1$ the equation $Q(s) = k - 1$ has a solution in s, namely

$$s = p_1^2 p_2^3 \cdots p_{k-1}^k.$$

References

[1] P. Erdös and Aleksandar Ivić, *On the iterates of the enumerating function of finite Abelian groups*, Bull. XCIC Acad. Serbe des Sciences et des Arts, Sci. Math. **17** (1989), 13–22.

[2] P. Erdös and I. Kátai, *On the growth of $d_k(n)$*, Fibonacci Quarterly **7** (1969), 267–247.

[3] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, (4th. ed.), Oxford, 1960.

[4] A. Ivić and G. Tenenbaum, *Local densities over integers free of large prime factors*, Quart. J. Math. **37** (1986), no. 2, 401–417.

[5] S. Ramanujan, *Collected Papers*, Chelsea, New York, 1962.

[6] P. Shiu, *The maximum orders of multiplicative functions*, Quart. J. Math. (2) **31** (1980), 217–252.

Katedra Matematike RGF-a, Univerziteta u Beogradu, Đušina 7, 11000 Beograd

FILOMAT (Niš) 9:3 (1995), 493–516
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

ИТЕРАТИВНЫЕ АЛГЕБРЫ АВТОМАТОВ

В. Б. Кудрявцев

Резюме. В работе приводятся основные результаты по проблемам выразимости и полноты для итеративных алгебр автоматов, полученые за последние более, чем тридцать лет, т.е. за период возникновенияи становления теории автоматов. Описание свойств функциональных систем автоматов ведется для модельных систем, упорядоченных по мере нарастания их сложности. Сначала рассматриваются автоматы без памяти, т.е. функции k-значной логики, затем автоматы с ограниченной памятью, т.е. указанные функции с задержками, и в заключение — конечные автоматы, т.е. автоматные функции.

Введение

Понятие автомата относится к числу важнейших в математике. Оно возникло на стыке разных ее разделов, а также в технике, биологии и других областях. Содержательно автомат представляет собой устройство с входными и выходными каналами. На его входы последовательно поступает информация, которая перерабатывается им с учетом строения этой последовательности и выдается через выходные каналы. Эти устройства могут допускать соединение их каналов между собой. Отображение входных последовательностей в выходные называют автоматной функцией, а возможность получения новых таких отображений за счет соединения автоматов приводит к алгебре автоматных функций.

Сами автоматы и их алгебры начали исследоваться в тридцатые годы текущего столетия, но особенно активно, — в период 50-х годов.

Основополагающую роль здесь сыграли работы А. Тьюринга, К. Шеннона, Э. Мура, С. Клини и других авторов знаменитого сборника "Автоматы" [1]. Последующие работы по изучению алгебр автоматов велись под большим влиянием известной статьи С.В. Яблонского [2] по теории функций k-значной логики. Такие функции могут рассматриваться как автоматы без памяти, к которыми применяются операции суперпозиции. Возникшие для таких функций постановки задач о выразимости, полноте, базисах, решетке замкнутых классов и другие, а также развитый аппарат сохранения предикатов как ключевой для решения этих задач, оказались весьма действенными и для алгебр автоматов, называемых далее функциональными системами. При этом под выразимостью понимается возможность получения функций одного множества через другое с помощью заданных операций, а под полнотой — выразимость всех функций через заданные. В работе изучение итеративных алгебр автоматов осуществляется на ряде модельных объектов, начиная с автоматов без памяти, т.е. функций k-значной логики, затем для автоматов с ограниченной памятью, т.е. функции с временными задержками, далее для конечных автоматов, т.е. автоматных функций общего вида. В качестве операций выступают суперпозиции, а в последнем случае — еще и обратная связь.

Для автоматов без памяти приводятся фундаментальные результаты Э. Поста о строении решетки замкнутых классов булевских функции, знакомство с которыми сегодня затруднено в связи с библиографической редкостью книг [3,4], в которых они содержатся. Затем приводятся наиболее существенные результаты для функций k-значной логики. Их основу составляет подход, развитый А.В. Кузнецовым и С.В. Яблонским и опирающийся на понятие предполного класса. Для конечно-порожденных систем таких функций семейство предполных классов образует критериальную систему; другими словами, произвольное множество является полным точно тогда, когда не является подмножеством ни одного предполного класса. Множество этих предполных классов оказалось конечным и из их характеризации вытекает алгоритмическая разрешимость задачи о полноте. На этом пути С.В. Яблонский путем явного описания всех предполных классов была решена задача о полноте для функций трехзначной логики, а вместе с А.В. Кузнецовым найдены отдельные семейства предполных классов для произвольной конечной значности. Затем усилиями многих исследователей [5–9] последовательно были открыты новые такие семейства, а заключительные построения провел И.Розенберг [10].

Для автоматов с ограниченой памятью приводятся решения задач о полноте и выразимости, а также задачи о слабых вариантах этих подстановок [11]. Под автоматом такого рода понимается пара (f,t), где f — функция k-значной логики, а t — время ее вычисления. Слабая полнота означает возможность получения из исходных пар с помощью суперпозиций любой функции хоть с какой-нибудь задержкой. Подробно рассматривается случай функций двузначной логики. В качестве аппарата решения здесь также используются предполные классы. Отличие этого случая от автоматов без памяти состоит в том, что семейство предполных классов оказалось счетным. Вместе с тем, задача о слабой полноте остается алгоритмически разрешимой.

Другим обобщением автоматов без памяти является класс линейных автоматов с операциями суперпозиции и обратной связи. Для этого класса ситуация оказывается похожей на случай автоматов с ограниченной памятью. Также удается описание всех предполных классов, которых оказывается счетное число, откуда тем не менее извлекается алгоритм распознавания полноты конечных систем автоматов [12].

Переход к общему случаю автоматов доставляет уже континуальность множества предполных классов [13] и алгоритмическую неразрешимость задачи о полноте [14]. Поэтому актуальными становятся поиски путей, связанные как с ослаблением свойств полноты, так и, наоборот, обогащением этого понятия.

Первое направление реализуется путем рассмотрения задачи о r-полноте и A-полноте, состоящих соответственно в проверке порождения всех отображений на словах длины r, а также таких отображений при любом фиксированном r. Основными результатами здесь являются явное описание всех r- и A-предполных классов и алгоритмическая неразрешимость задачи об A-полноте [15].

Второе направление реализуется путем расслоения всех конечных систем автоматов, исследуемых на полноту, на типы. В один тип относятся все такие системы, которые содержат заданный класс Поста автоматов без памяти. Основными результатами является явное указание границы отделимости алгоритмически разрешимых случаев типов систем автоматов на диаграмме Поста. Эта же граница оказывается верной и для случая A-полноты [16].

Наряду с продвижениями в решении основных задач по проблематике, связанной с выразимостью и полнотой, в обзоре обозначаются те направления, которые еще разработаны слабо или недостаточно. Изложение ведется только для модельных случаев функциональных систем автоматов. Общие построения, осуществленные автором в [17], здесь не затрагиваются.

1. Основные понятия и задачи

Пусть $N = \{1,2,\ldots\}$, $N_0 = \{0\} \cup N$, $N_1 = N\backslash\{1\}$, для h из N пологаем $N_1^h = \{1,2,\ldots,h\}$. Рассмотрим некоторое множество M и отображение $\omega : M^n \to M$, где M^n является n-ой декартовой степенью множества M и $n \in N$. Пусть P_M — множество всех таких отображений ω при любых указанных n и $\Omega \subseteq P_M$.

Рассмотрим универсальную алгебру (у.а) $\mathcal{M} = (M,\Omega)$, в которой M называется носителем, Ω — классом операций. С каждым подмножеством $\overline{M} \subseteq M$ свяжем последовательность множеств $\overline{M}^{(i)}$, $i \in N$, таким образом.

Положим $\overline{M}^{(1)} = \overline{M}$. Далее, множество $\overline{M}^{(i+1)}$ состоит из свех таких элементов m из M, для которых найдутся ω из Ω и $m_1, m_2, \ldots, m_n$ из $\overline{M}^{(i)}$, что $m = \omega(m_1, m_2, \ldots, m_n)$. Обозначим через $I_\Omega(\overline{M})$ множество $\cup_{i=1}^{\infty} \overline{M}^{(i)}$. Нетрудно видеть, что I_Ω является оператором замикания на множестве $\mathcal{B}(M)$ образованном всеми подмножествами множества M. Тем самым для I_Ω всегда выполнены условия $I_\omega(\overline{M}) \supseteq \overline{M}$, $I_\omega(I_\omega(\overline{M})) = I_\omega(\overline{M})$, а также если $\overline{M} \supseteq \overline{\overline{M}}$, то $I_\omega(\overline{M}) \supseteq I_\omega(\overline{\overline{M}})$. Множество $I_\Omega(\overline{M})$ называется замыканием множества $\overline{M}$, а само $\overline{M}$ — порождающим множеством для $I_\Omega(\overline{M})$. Множество $\overline{M}$ называется замкнутым, если $\overline{M} = I_\Omega(\overline{M})$.

Пусть $\Sigma(\mathcal{M})$ — множество всех замкнутых подмножеств в $\mathcal{M}$. Говорят, что $\overline{M}$ выразимо через $\overline{\overline{M}}$, если $\overline{M} \subseteq I_\omega(\overline{\overline{M}})$. Множество $\overline{M}$ называется полным, если $I_\omega(\overline{M}) = M$. Полное множество называется базисом, если любое его собственное подмножество не является полным.

Основными проблемами для M, которые будут интересовать нас, являются проблемы выразимости и полноты, а также базисов, решетки замкнутых классов, модификации их и некоторые примыкающие к ним вопросы.

Под проблемой выразимости понимается указание всех пар $(\overline{M}, \overline{\overline{M}})$ таких, что $\overline{M}$ выразимо через $\overline{\overline{M}}$; под проблемой полноты — указание всех полных подмножеств; под проблемой базисов — описание всех базисов, если они существуют; под проблемой решетки — построение решетки всех замкнутых классов и нахождение ее свойств.

Знание решетки $\Sigma(\mathcal{M})$ дает решение проблем выразимости и полноты. Так, выразимость $\overline{M}$ через $\overline{\overline{M}}$ означает проверку $I_\omega(\overline{M}) \subseteq I_\omega(\overline{\overline{M}})$. Для решения проблемы полноты используется следующая схема. Систему $\Sigma' \subseteq \Sigma(\mathcal{M})$ назовем критериальной (k-системой), если любое множество $\overline{M}$ полно точно тогда, когда для любого $\overline{\overline{M}}$ из Σ' выполнено $\overline{M} \not\subseteq \overline{\overline{M}}$. Ясно, что если $\Sigma(\mathcal{M})\backslash\{M\} \neq \emptyset$, что далее предпологается, то $\Sigma(\mathcal{M})\backslash\{M\}$ является k-системой. Нетрудно видеть, что дуальные атомы решетки $\Sigma(\mathcal{M})$, называемые также предполными классами, входят в любую k-систему. Пусть $\Sigma_\pi(\mathcal{M})$ — множество всех предполных классов и $\Sigma_{\overline{\pi}}(\mathcal{M})$ — множество всех классов из $\Sigma(\mathcal{M})$, не являющихся подмножествами ни одного предполного класса из $\Sigma_\pi(\mathcal{M})$. Нетрудно убедиться в следующем утверждении.

Предложение 1.1. *Множество* $\Sigma_\pi(\mathcal{M}) \cup \Sigma_{\overline{\pi}}(\mathcal{M})$ *образует k-систему в у.а. $\mathcal{M}$.*

Особый интерес вызывает ситуация, когда $\Sigma_{\overline{\pi}}(\mathcal{M})$ является пустым множеством, так как в этом случае система $\Sigma_\pi(\mathcal{M})$ образует k-систему, что означает сведение задачи о полноте к описанию всех предполных классов. Отметим один важный случай такого рода. У.а. $\mathcal{M}$ называется конечно-порожденной, если существует конечное подмножество $M' \subseteq M$, которое является полным.

Известно [18] следующуе утверждение.

Предложение 2.1. *Если у.а. $\mathcal{M}$ является конечно-порожденной, то $\Sigma_\pi(\mathcal{M})$ образует k-систему.*

Отметим, что в общем случае это утверждение не является обратимым. Рассмотрим теперь случай, когда множество M состоит из функций. Он будет для нас основным. При этом, у. а. $\mathcal{M}$ будет называтся функциональной системой (ф.с.).

Пусть E — некоторое множество и функция f имеет вид $f : E^n \to E$, где $n \in N$. Пусть $U = \{u_1, u_2, \dots\}$ — алфавит переменных u_i со значениями в E, $i \in N$. Для записи функции f будем пользоваться выражением $f(u_{i_1}, \dots, u_{i_n})$. Класс всех таких функций обозначим через P_E. Во избежение сложных индексов у переменых u_i будем использовать для их обозначения метасимволы x, y, z, возможно, с индексами.

Следуя А.И. Мальцев [19], введем в P_E унарные операции η, τ, Δ, ∇, которые определяются следующим образом:

$$(\eta f)(x_1, x_2, \dots, x_n) = f(x_2, x_3, \dots, x_n, x_1);$$
$$(\tau f)(x_1, x_2, \dots, x_n) = f(x_2, x_1, x_3, \dots, x_n);$$
$$(\Delta f)(x_1, x_2, \dots, x_{n-1}) = f(x_1, x_1, x_2, \dots, x_{n-1}), \quad \text{если } n > 1;$$
$$(\eta f) = (\tau f) = (\Delta f) = f, \quad \text{если } n = 1;$$
$$(\nabla f)(x_1, x_2, \dots, x_n, x_{n+1}) = f(x_2, x_3, \dots, x_{n+1});$$

Форма этих операций уточняет операции из [13].

Введем в P_E а бинарную операцию $*$ так. Для функций $f(x_1, x_2, \ldots, x_n)$ и $g(x_{n+1}, x_{n+2}, \ldots, x_{n+m})$ полагаем

$$(f * g)(x_2, x_3, .., x_n, x_{n+1}, .., x_{n+m}) = f(g(x_{n+1}, x_{n+2}, .., x_{n+m}), x_2, .., x_n).$$

Описанные операции называются, соответственно, сдвигом, транспозицией, отождествлением, расширением, и подстановкой, а в совокупности — операциями суперпозиции. Множество этих операций обозначим Ω_c. Пусть $M \subseteq P_E$ и $I_{\Omega_c}(M) = M$, тогда ф.с. $\mathcal{M} = (M, \Omega_c)$ называется итеративной ф.с. Поста.

2. Функции l-значной логики

Функции из P_E называются функциями l-значной логики, если

$$E = E_l = \{0, 1, 2, \ldots, l-1\}, \quad l \geq 2.$$

В этом случае вместо P_E употребляется символ P_l. Ф.с. $\mathcal{P}_l = (P_l, \Omega_c)$ считается одной из основных моделей итеративных ф.с. Поста (для краткости: п.ф.с.), на изучении которой формировались проблематика и методы теории ф.с. Если $\mathcal{M} = (M, \Omega_c)$ и $M \subseteq P_l$, то $\mathcal{M}$ называем п.ф.с. рода l. Коротко изложим основные итоги изучения $\mathcal{P}_l$, которые будут важны нам для рассмотрения ф.с. автоматных функций.

Для P_l Э. Постом [3] дано полное решение упомянутых проблем о полноте, выразимости, базисах и решетке замкнутых классов. Опишем эту решетку, сохраняя его обозначения.

Рассморим множество Q классов

$$C_i, A_i, D_j, L_r, O_s, S_t, P_t, F_v^m, F_v^\infty,$$

где $i = 1, 2, 3, 4$; $j = 1, 2, 3$; $r = 1, 2, 3, 4, 5$; $s = 1, 2, \ldots, 9$; $t = 1, 3, 5, 6$; $v = 1, 2, \ldots, 8$; $m = 1, 2, \ldots$.

Функции из P_2 называются функциями алгебры логики (ф.а.л.). Класс P_2 Э. Пост обозначает через C_1. Класс C_2 содержит все ф.а.л. f, такие что $f(0, 0, \ldots, 0) = 0$; C_3 — все ф.а.л., такие, что $f(1, 1, \ldots, 1) = 1$; $C_4 = C_2 \cup C_3$. Говорят, что ф.а.л. f является монотоной, если всегда из неравенства $a_i \leq b_i$ при всех $i = 1, 2, \ldots, n$ следует, что $f(a_1, \ldots, a_n) \leq f(b_1, \ldots, b_n)$. Класс A_1 состоит из всех монотонных ф.а.л.; $A_2 = C_2 \cap A_1$; $A_3 = C_3 \cap A_1$; $A_4 = A_2 \cap A_3$. Класс D_3 состоит из всех ф.а.л. f таких, что $f(x_1, \ldots, x_n) = \overline{f}(\overline{x_1}, \ldots, \overline{x_n})$

где ф.а.л. $\overline{x}$ называется отрицанием и задается так: $\overline{0} = 1$, $\overline{1} = 0$; $D_1 = C_4 \cap D_3$; $D_2 = A_1 \cap D_3$. Класс L_1 состоит из всех ф.а.л. $f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \cdots + x_n + \alpha \pmod 2$; $L_2 = C_2 \cap L_1$; $L_3 = C_3 \cap L_1$; $L_4 = L_2 \cap L_3$; $L_5 = D_3 \cap L_1$. Класс O_9 состоит из всех ф.а.л., существенно зависящих не более, чем от одного переменного; $O_8 = A_1 \cap O_9$; $O_4 = D_3 \cap O_9$; $O_5 = C_2 \cap O_9$; $O_6 = C_3 \cap O_9$; $O_1 = C_5 \cap O_6$; $O_7 = \{0, 1\}$; $O_2 = O_5 \cap O_7$; $O_3 = O_6 \cap O_7$. Класс S_6 состоит из всех ф.а.л. вида $x_1 \vee x_2 \vee \cdots \vee x_n$ и констант; $S_3 = C_2 \cap S_6$; $S_5 = C_3 \cap S_6$; $S_1 = S_3 \cap S_5$. Класс P_6 состоит из всех ф.а.л. вида $x_1 \& x_2 \& \dots \& x_n$ и констант; $P_5 = C_2 \cap P_6$; $P_3 = C_3 \cap P_6$; $P_1 = P_3 \cap P_5$. Говорят, что ф.а.л. удовлетворяют условию a^μ, $\mu \in N_1$, если любые μ наборов, на которых она равна 0, имеют общую координату 0. Аналогично с заменой 0 на 1 определяется свойство A^μ. Класс F_4^μ состоит из всех ф.а.л. со свойством a^μ; $F_1^\mu = C_4 \cap F_4^\mu$; $F_3^\mu = A_1 \cap F_4^\mu$; $F_2^\mu = F_1^\mu \cap F_3^\mu$. Класс F_8^μ состоит из всех ф.а.л. со свойством A^μ; $F_5^\mu = C_4 \cap F_8^\mu$; $F_7^\mu = A_3 \cap F_8^\mu$; $F_6^\mu = F_5^\mu \cap F_7^\mu$. Ф.а.л. удовлетворяет свойству a^∞, если все наборы, на которых она равна 0, имеют общую координату 0. Аналогично с заменой 0 на 1 вводится свойство A^∞. Класс F_4^∞ состоит из всех ф.а.л. со свойством a^∞; $F_1^\infty = C_4 \cap F_4^\infty$; $F_3^\infty = A_1 \cap F_4^\infty$; $F_2^\infty = F_1^\infty \cap F_3^\infty$. Класс F_8^∞ состоит из всех ф.а.л. со свойством A^∞; $F_5^\infty = C_4 \cap F_8^\infty$; $F_7^\infty = A_3 \cap F_8^\infty$; $F_6^\infty = F_5^\infty \cap F_7^\infty$.

Теорем 2.1. [3] *Для п. ф. с. $\mathcal{P}_2$ справедливо:*

а) множество всех замкнутых классов в $\mathcal{P}_2$ счетно и совпадает с множеством Q.

б) классы из Q образуют решетку по включению, приведенную на рис. 1;

в) каждый замкнутый класс в $\mathcal{P}_2$ имеет базис и мощность его всегда не более, чем 4;

г) проблемы полноты и выразимости для п.ф.с. рода 2 применительно к конечным множествам ф.а.л. алгоритмически разрешимы.

Свойства п.ф.с. рода l при $l > 2$ оказались много сложнее, как это будет следовать из приводимых ниже утверждений.

Обозначим через $P_l^{(n)}$ множество всех функций из P_l, зависящих не более, чем от n переменных $u_1, u_2, \dots, u_n$.

Ясно, что число $p_l^{(n)}$ функций в $P_l^{(n)}$ равно $\sum_{i=1}^n C_n^i l^{l^i}$. Пусть S_l^n — множество всех функций из $P_l^{(n)}$, каждая из которых при некотором i, $i = 1, 2, \dots, n$, равна u_i. Если $M \subseteq P_l$ и M конечно, то через $\alpha(M)$ обозначим наибольшее число переменных у функций из M.

Для конечно-порожденной п.ф.с. $\mathcal{M}$ рода l пусть $\alpha(\mathcal{M})$ — наименьшее такое число α', что для некоторого $M' \subseteq M$ выполнено

$I_{\Omega_c}(M') = M$ и $\alpha(M') = \alpha'$. Назовем непустое множество $M' \subseteq P_l^{\alpha(M)} \cap M$ R-множеством в $\mathcal{M}$, если $I_{\Omega_c}(M') \cap P_l^{\alpha(M)} = M'$ и $M' \neq P_l^{\alpha(M)} \cap M'$, обозначив его через $R^{\alpha(M)}$. Пусть $\mathcal{R}^{\alpha(M)} = (R^{\alpha(M)} \cup S_l^{\alpha(M)}, R^{\alpha(M)})$. Будем говорить, что функция $f(x_1, x_2, \ldots, x_n)$ из P_l сохраняет $\mathcal{R}^{\alpha(M)})$, если для любого набора функций $g_1, g_2, \ldots, g_n$ из $R^{\alpha(M)} \cup S_l^{\alpha(M)}$ будет выполнено $f(g_1, g_2, \ldots, g_n) \in R^{\alpha(M)}$. Класс всех функций из M, сохраняющих $\mathcal{R}^{\alpha(M)}$, обозначим через $U(\mathcal{R}^{\alpha(M)})$. Назовем R-множество $R^{\alpha(M)})$ максимальным, если не существует такого R-множества $R_1^{\alpha(M)}$, что $U(\mathcal{R}_1^{\alpha(M)}) \supset U(\mathcal{R}^{\alpha(M)})$. Пусть $\mathbf{R}(M)$ — множество всех максимальных R-множеств, $\mathbf{R}_k(M)$ — множество всех пар $\mathcal{R}^{\alpha(M)}$, для которых $R^{\alpha(M)} \in \mathbf{R}(M)$ и $U(\mathbf{R}(M))$ — множество всех классов сохранения элементов из $\mathbf{R}(M)$. Назовем п.ф.с. $\mathcal{M}$ тривиальной, если $M = I_{\Omega_c}(S_k^1)$ или $M = I_{\Omega_c}(\{c(x)\})$, где $c(x) = c$, $c \in E_l$. Мощность множества A обозначим через $|A|$.

Теорем 2.2. [17] *Если п. ф. с. $\mathcal{M} = (M, \Omega_c)$ рода l нетривиальная и конечно-порожденная, то имеет место:*

а) $U(\mathbf{R}(\mathcal{M})) = \Sigma_\pi(\mathcal{M})$.

б) $|U(\mathbf{R}(M))| \leq 2^{P_l(\alpha(M))}$.

в) $\mathbf{R}(M)$ *строится эффективно.*

Эта теорема является развитием утверждения А.В. Кузнецова из [2].

Следствие. *Проблема полноты для конечно-порожденных п.ф.с. рода l алгоритмически разрешима для любого l.*

Теорема 2.3. [17] *Проблема выразимости длй конечных множеств конечно-порожденной п.ф.с. рода l алгоритмически разрешима для любого l.*

Теорема 2.4. [20] *Для каждого $l \geq 3$ существует п.ф.с. рода l, для которых выполнено:*

а) п.ф.с. имеет счетный базис;

б) п.ф.с. имеет базис заданной конечной мощности;

в) п.ф.с. не имеет базиса.

Следствие. [20] *Для каждого $l \geq 3$ решетка замкнутых классов в п.ф.с. $\mathcal{P}_l$ континуальна.*

Как установлено в [2], п.ф.с. $\mathcal{P}_l$ является конечно-порожденной, поэтому для нее справедливы теоремы 1 и 2, а для множества $U(\mathbf{R}(\mathcal{P}_l))$ найдено его явное описание [10]

Резкое отличие свойств $\mathcal{P}_l$ при $l = 2$ и $l > 2$ привело к рассмотрению разных вариаций основных задач для п.ф.с. таких, как исследование на полноту систем функций с заданными свойствами, например, систем Слупецкого, содержащих все одноместные функции; изучение строения фрагментов решетки замкнутых классов и др. Кроме того, изучались обобщения $\mathcal{P}_l$ в виде п.ф.с. неоднородных функций, то есть зависящих от разных групп переменных, области определения которых различны [17], а также функций, переменные которых, как и сами функции, принимают счетное число значений.

Начато рассмотрение декартовых степеней таких обобщений и других случаев [22]. На этих направлениях останавливаться не будем, а рассмотрим обобщение функций из P_k, возникающие за счет учета времени, требующегося для их вычисления.

3. Функции с задержками

Пусть $f(x_1, x_2, \ldots, x_n) \in P_l$ и $t \in N_0$. Пару $(f(x_1, x_2, \ldots, x_n), t)$ называем функцией f с задержкой t и множество всех таких пар обозначаем через $\tilde{P}_l$. Распространим на $\tilde{P}_l$ операции η, τ, Δ и ∇, полагая, если μ — любая из них, что $\mu((f,t)) = (\mu(f), t)$. Введем еще одну операцию $*_{\mathrm{c}}$, называемую синхронной подстановкой, полагая для пар $(f,t), (f_1, t'), (f_2, t'), \ldots, (f_n, t')$, в которых множества переменных у функций $f_1, f_2, \ldots, f_n$ попарно не пересекаются, выполненным соотношение

$$(f,t) *_{\mathrm{c}} ((f_1, t'), (f_2, t'), \ldots, (f_n, t')) = (f(f_1, f_2, \ldots, f_n), t + t').$$

Множество операций η, τ, Δ, ∇, $*_{\mathrm{c}}$ обозначим через Ω_{cc} и назовем операциями синхронной суперпозиции. Пусть $M \subseteq \tilde{P}_l$ и $J_{\Omega_{\mathrm{cc}}}(M) = M$, тогда ф.с. $\mathcal{M} = (M, \Omega_{\mathrm{cc}})$ называем итеративной ф.с. Поста функций с задержками рода l (п.ф.с.ф.з.).

Коротко изложим основные итоги изучения этих ф.с. [11, 17].

Теорема 3.1. *В конечно-порожденной п.ф.с.ф.з. $\mathcal{M}$ рода l, $l \in N_1$, множество $\Sigma_\pi(\mathcal{M})$ конечно и строится эффективно для любого l.*

Теорема 3.2. *Для конечно-порожденной п.ф.с.ф.з. рода l проблемы полноты и выразимости алгоритмически разрешимы для любого l.*

Теорема 3.3. *Для каждого l из N_1 существуют п.ф.с.ф.з. рода l, для которых выполнено:*

а) имеется счетный базис;

б) имеется конечный базис заданной мощности;

в) не имеется базиса.

Следствие. *Решетка замкнутых классов в $\tilde{P}_l$ континуальна для всех l.*

Примером конечно-порожденной п.ф.с.ф.з. является $\tilde{\mathcal{P}}_l = (\tilde{\mathcal{P}}_l, \Omega_{cc})$.

Для $\tilde{\mathcal{P}}_l$ теорема 3.1. уточняется следующим образом. Обозначим через $M^{(1)}$ множество всех функций f таких, что при некотором t выполнено $(f,t) \in M$.

Теорема 3.4. *Множество $M \subseteq \tilde{P}_l$ является полным в $\tilde{\mathcal{P}}_l$ точно тогда, когда $J_{\Omega_c}(M^{(1)}) = \tilde{P}_l$ и $M \supseteq \{(f,1)\}$, где $f \notin E_l$.*

Пусть $\mathcal{M} = (M, \Omega_{cc})$ и $M' \subseteq M$. Говорят, что M' является слабо полным в $\mathcal{M}$, если для всякой функции f из $M^{(1)}$ найдется t, такое что $(f,t) \in J_{\Omega_{cc}}(M')$. Множество M' называется слабо предполным, если оно не слабо полное, но превращается в таковое при добавлении любой пары из $M \backslash M'$. Класс всех таких множеств обозначим через $\Sigma_{c\pi}(\mathcal{M})$. Класс $K \subseteq \Sigma(\mathcal{M})$ называем слабо критериальным, если множество M' является слабо полным точно тогда, когда для любого T из $\Sigma(\mathcal{M})$ выполнено $M' \not\supseteq T$. Очевидно, для слабо критериальной системы K всегда выполнено $K \supseteq \Sigma_{c\pi}(\mathcal{M})$.

Теорема 3.5. *Для конечно-порожденной п.ф.с.ф.з. $\mathcal{M} = (M, \Omega_{cc})$ рода l выполнены положения:*

а) множество $\Sigma_{c\pi}(\mathcal{M})$ конечно или счетно;

б) множество $\Sigma_{c\pi}(\mathcal{M})$ слабо критериально;

в) множество $\Sigma_{c\pi}(\mathcal{M})$ строится эффективно.

Тхеорем 3.6. *Для конечно-порожденных п.ф.с.ф.з. типа l проблема слабой полноты алгоритмически разрешима для любого l.*

Явное описание множества $\Sigma_{c\pi}(\mathcal{M})$ получено пока для $l = 2, 3$ [11, 22]. Приведем здесь описание случая $l = 2$.

Пусть $M \subseteq P_2$ и $M \in \{C_2, C_3, D_3, A_1, L_1\}$, обозначим через $\tilde{M}$ множество всех пар (f,t) таких, что $f \in M$ и $t \in N_0$.

Функция $f(x_1, x_2, \dots, x_n)$ из P_2 называется α-, β-, γ- или δ- функцией, если $f(x_1, x_2, \dots, x_1)$ равна x, 1, 0 или $\overline{x}$ соответственно. Пусть A, B, Γ, D суть классы всех α-, β-, γ- или δ- функций соответственно.

Обозначим через $\tilde{C}$ множество всех пар вида $(f, t+1), (\varphi, t+1)$, $(\psi, 0)$, где $f \in B$, $\varphi \in \Gamma$, $\psi \in A$, $t \in N_0$.

Пусть $i \in \{0,1\}$, обозначим через E_i множество всех пар вида $(f,0)$, $(\bar{i}, t+1)$, (i,t), где $f \in C_{i+2}$, $t \in N_0$.

Назовем функцию f четной, если выполнено

$$f(x_1, x_2, \dots, x_n) = f(\overline{x}_1, \overline{x}_2, \dots, \overline{x}_n).$$

Пусть Y — множество всех четных функций.

Пусть $\tilde{H}$ множество всех пар вида $(f, 0)$, $(\varphi, t+1)$, где $\varphi \in Y$, $f \in S$, $t \in N_0$.

Для каждого r из N_0 обозначим через $\tilde{Z}_r$ множество всех пар вида

$$(f, (2t+1)2^r), (\varphi, (2t+1)2^s), (\psi, 0),$$

где $f \in D$, $\varphi \in A$, $\psi \in A$, $t \in N_0$, $s \in N_0 \backslash \{0, 1, 2, \dots, r\}$.

Для каждого r из N_0 обозначим через $\tilde{W}_r$ множество всех пар вида

$$(f, (2t+1)2^r), (0, t), (1, t), (\varphi, (2t+1)2^s), (\psi, 0),$$

где $\overline{f} \in M$, $\varphi \in M$, $\psi \in M$, $t \in N_0$, $s \in N_0 \backslash \{0, 1, 2, \dots, r\}$.

Пусть $\tilde{W} = \{\tilde{C}_2, \tilde{C}_3, \tilde{D}_3, \tilde{A}_1, \tilde{L}_1, \tilde{C}, \tilde{E}_0, \tilde{E}_1, \tilde{H}, \tilde{Z}_0, \tilde{Z}_1, \dots, \tilde{W}_0, \tilde{W}_1, \dots\}$.

Теорема 3.7. [11] *Имеет место равснство* $\Sigma_{c\pi}(\tilde{\mathcal{P}}_2) = \tilde{W}$.

Заметим, что явное описание $\Sigma_{c\pi}(\tilde{\mathcal{P}}_1) = \tilde{W}$ для любого l имеется пока лишь в виде отдельных семейств слабо предполных классов [17, 23].

Содержательный интерес представляют другие модификации проблемы полноты для п.ф.с.ф.з., рассмотренные в работе [17].

4. Автоматные функции

Рассмотренное в [3] обобщение функций l-значной логики до функций с задержками является промежуточным при переходе к классу автоматных функций, свойства которого выглядят логически существенно сложнее, чем у функций с задержками. Для введения понятия автоматной функции потребуются вспомогательные обозначения и определения.

Пусть C — конечное или счетное множество, которое называем алфавитом. Последовательность букв из C называем словом, если она конечна, и сверхсловом, если она бесконечна. Класс всех таких слов обозначим через C^*, а сверхслов — через C^ω. Пусть $\overline{C} = C^* \cup C^\omega$ и $\xi \in \overline{C}$. Слово, образованное первыми r буквами из ξ, называем префиксом для ξ и обозначаем через $\xi]_r$. Пусть A и B — алфавиты и $f : A^* — B^*$. Если $\xi = c(1)c(2)\dots c(r)$, то r называем длиной слова ξ и обозначаем через $||\xi||$. Пусть A и B — алфавиты и $f : A^* \to B^*$. Функцию f называем детерминированной (д.функцией), если для любого ξ из A^* справедливо $||f(\xi)|| = ||\xi||$, а для любых ξ_1 и ξ_2 из A^* и любого i такого, что $1 \leq i \leq \min(|\xi_1|, |\xi_2|)$, если $\xi_1]_i = \xi_2]_i$, то $\xi(\xi)_1]_i = \xi(\xi)_2]_i$. Известно, что д.функция f может быть рекуррентно задана с помощью так называемых канонических уравнений вида

$$\text{(1)}\qquad \begin{cases} q(1)=q_0,\\ q(t+1)=\varphi(q(t),a(t)), & t=1,2,\dots\\ b(t)=\psi(q(t),a(t)),\end{cases}$$

где параметар q называется состоянием для f и принимает значения из алфавита Q. Эта рекуррентность определяется так. Если $\alpha\in A^*$, $\beta\in B^*$, $\kappa\in Q$ и $\alpha=a(1)a(2)\dots a(r)$, $\beta=b(1)b(2)\dots b(r)$, $\kappa=q(1)q(2)\dots q(r)$, то при $f(\alpha)=\beta$ слово β индуктивно вычисляется по α следующим образом:

а) $b(1)=\psi(q(1),a(1))$ где $q(1)=q_0$;

б) если при $1\leq t\leq r-1$ вычислено $q(1)=q_0$, то $q(t+1)=\varphi(q(t),a(t))$ и $b(t)=\psi(q(t),a(t))$.

Часто предполагают, что алфавиты A и B являются декартовыми степенями E_l, то-есть $A=(E_l)^n$ и $B=(E_l)^m$, где $n,m,p\in N$. В этом случае от рассмотрения одноместной д.функции $f(x)$ вида $f:((E_l)^n)^*\to((E_l)^m)^*$ удобно перейти к n-местной д.функции $f(x_1,x_2,\dots,x_n)$ вида $f':((E_l)^*)^n\to((E_l)^*)^m$ следующим способом.

Пусть $\zeta^n\in(E_k^*)$ и $f(\zeta^n)=\zeta'^m$, где

$$\zeta^n=e^n(1)e^n(2)\dots e^n(r)\quad\text{и}\quad\zeta'^m=e'^m(1)e'^m(2)\dots e'^m(r);$$

при этом

$$e^n(t)=(e_1(t),e_2(t),\dots,e_n(t))\quad\text{и}\quad e'^m(t)=(e'_1(t),e'_2(t),\dots,e'_m(t)),$$

где $1\leq t\leq r$. Пусть

$$\zeta_i=e_i(1)e_i(2)\dots e_i(r)\quad\text{и}\quad\zeta'_j=e'_j(1)e'_j(2)\dots e'_j(r),$$

где $1\leq i\leq n$ и $1\leq j\leq m$.

Теперь полагаем $f'(\zeta_1,\zeta_2,\dots,\zeta_n)=(\zeta'_1,\zeta'_2,\dots,\zeta'_n)$. Функция f' получается из f фактически только за счет перехода от рассмотрения матриц, образованных вектор-буквами (строками) слов ζ^n и, соответственно, слов ζ'^m, к их представлению в виде столбцов. Канонические уравнения (2) для f' получаются из (1) заменой там всех параметров на соответствующие векторные значения, т.е.

$$\text{(2)}\quad\begin{cases} q(1)=q_0,\\ q(t+1)=\varphi(q(t),e_1(t),\dots,e_n(t)), & t=1,2,\dots,j=1,2,\dots,m.\\ b_j(t)=\psi_j(q(t),e_1(t),\dots,e_n(t)),\end{cases}$$

Функцию f' считаем интерпретацией для f и называем автоматной (а.функцией).

Параметры n и m называем, соответственно, местностью и размерностью а.функции, а мощность множества значений параметра q — числом ее состояний. Содержательным толкованием а.функции

$$f'(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_m)$$

является функционирование технического устройства F вида, указанного на рисунке 2. Здесь входные стрелки обозначены буквами x_i, а выходные — буквами y_i. Считается, что F функционирует в дискретные моменты времены $t = 1, 2, \ldots$. В эти моменты каждые вход x_i и выход y_i могут принимать значения из E_l; само же устройство может находиться в состояниях, кодируемых значениями из Q, называемых также памятью для F. По набору значений входов и состоянию в момент t устройства F по правилам (2) определяют значения его выходов и состояние в момент $t+1$.

Обозначим через $P^{n,m}_{a,l}$ класс всех а.функций с заданными параметрами n и m из N. Пусть $P_{a,l} = \bigcup\limits_{n,m\geq 1} P^{n,m}_{a,l}$.

Распространим на $P_{a,l}$ операции $\eta, \tau, \Delta, \nabla$, а также введем другие операции.

Пусть $f'(x_1, x_2, \ldots, x_n) = (y_1, \ldots, y_j, \ldots, y_m)$, тогда

$$(\pi_j f')(x_1, x_2, \ldots, x_n) = f'_j(x_1, x_2, \ldots, x_n),$$

где $f'_j(x_1, x_2, \ldots, x_n)$ совпадает со значением y_j у $f'(x_1, x_2, \ldots, x_n)$.

Пусть $f''(x_{n+1}, x_{n+2}, \ldots, x_{n+v}) = (y_{m+1}, \ldots, y_{m+w})$, тогда

$$(f' \sigma f'')(x_1, x_2, \ldots, x_n, x_{n+1}, x_{n+2}, \ldots, x_{n+v}) = (f'(x_1, x_2, \ldots, x_n)),$$
$$f''(x_{n+1}, x_{n+2}, \ldots, x_{n+v})) = (y_1, y_2, \ldots, y_m, y_{m+1}, y_{m+2}, \ldots, y_{m+w}).$$

Пусть u такое, что $m + 1 \leq u \leq m + u$, тогда

$$(f' \underset{u}{*} f'')(x_2, x_3, \ldots, x_n, x_{n+1}, x_{n+2}, \ldots, x_{n+v}) = (z_1, z_2, \ldots, z_{m+w-1}),$$

где $z_j = f'_j * f''_u$ при $j = 1, 2, \ldots, m$ и $z_{j'} = f''_{j'}(x_{n+1}, x_{n+2}, \ldots, x_{n+v})$ при $j' = m + 2, m + 3, \ldots, m + t$.

Операции π и σ называются, соответственно, проектированием и объединением, а операция вида $\underset{u}{*}$ является распространением операции $*_c$ одномерного случая функций на вектор-функции и попрежнему называется подстановкой.

В совокупности операции η, τ, Δ, ∇, π, σ и $*$ называем операциями расширенной суперпозиции и обозначим через $\Sigma_{\text{рс}}$.

Введем еще одну операцию над а.функциями, которую назовем обратной связью и обозначим через $\mathbf{O}$.

Говорят, что а.функция f' имеет тип $i-j$, если для f' найдется система вида (2) такая, что функция $\psi_j(q, e_1, e_2, \dots, e_n)$ фиктивно зависит от значений e_i. Пусть f' такая а.функция, рассмотрим функцию вида

$$(\mathbf{O}_j^i f')(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = (y_1, y_2, \dots, y_{j-1}, y_{j+1}, \dots, y_m),$$

которая определяется так. Пусть заданы слова вида $\xi_l = e_l(1)e_l(2) \dots e_l(r)$, где $l = 1, 2, \dots, i-1, i+1, \dots, n$. Тогда с помощью (2) по набору

$$(e_1(1), e_2(1), \dots, e_{i-1}(1), e_{i+1}(1), \dots e_n(1))$$

можно вычислить значение $b_j(1)$. Подставим теперь в (2) вместо $e_i(1)$ значение $b_j(1)$, после чего можно вычислить наборы $(q(2))$ и $(b_1(1), \dots, b_m(1))$. Далее, так же можно по набору

$$(e_1(2), e_2(2), \dots, e_{i-1}(2), e_{i+1}(2), \dots, e_n(2))$$

вычислить значение $b_j(2)$. Снова, подставив значение $b_i(2)$ вместо $e_j(2)$ в (2), можно вычислить наборы $(q(3))$ и $(b_1(2), b_2(2), \dots, b_m(2))$ и т.д.

Теперь пологаем

$$(\mathbf{O}_j^i f')(\zeta_1, \zeta_2, \dots, \zeta_{i-1}, \zeta_{i+2}, \dots, \zeta_n) = (\zeta'_1, \zeta'_2, \dots, \zeta'_{i-1}, \zeta'_{i+2}, \dots, \zeta'_m),$$

где $\zeta'_{l'} = b_{l'}(1)b_{l'}(2) \dots b_{l'}(r)$ и $l' = 1, 2, \dots, j-1, j+1, \dots, m$. Положим $\Omega_{\text{рс},\mathbf{O}} = \Omega_{\text{рс}} \cup \{O\}$. Класс операций $\Omega_{\text{с},\mathbf{O}}$ называется композицией. Говорят, что а.функция f' из $P_{a,l}$ является конечно-автоматной (к.а. функция), если алфавит Q в некоторой системе (2), задающей эту функцию, конечен. Класс всех к.а.функций с параметрами n и m обозначим через $P_{a,l,\text{к}}^{n,m}$. Полагаем далее, что $P_{a,l,\text{к}} = \cup_{n,m \geq 1} P_{a,l,\text{к}}^{n,m}$. Говорят, что а.функция является истинностной (и.а.функция), если в системе (2), задающий ее, алфавит Q — одноэлементный. Обозначим через $P_{a,l,\text{и}}^{n,m}$ и $P_{a,l,\text{и}}^{n,m}$ классы всех истинностных а.функций и, соответственно, истинностных к.а.функций (и.к.а.функция) с параметрами n и m. Положим

$$P_{a,l,\text{и}} = \cup_{n,m \geq 1} P_{a,\zeta,\text{и}}^{n,m} \quad \text{и} \quad P_{a,l,\text{к},\text{и}} = \bigcup P_{a,l,\text{к},\text{и}}^{n,m}.$$

Содержательно истинностные а.функции интерпретируются с одной стороны как функционирование устройства F без памяти, а с другой стороны могут считаться реализациями функций из P_l с учетом времени t, которое пробегает значения $1, 2, \ldots,$ и каждый из которых зависимость значения функции от значений переменных одна и та же.

Таким образом, п.ф.с. (P_l, Ω_{c}) при расширении в них объекта P_l до множества вектор-функций l-значной логики и, соответственно, операций — до $\Omega_{\text{рс}}$, фактически приводят к п.ф.с. $\mathcal{P}_{a,l,\text{и}} = (P_{a,l,\text{и}}, \Omega_{\text{рс}})$.

Заметим также, что функция с задержкой интерпретируется как функционирование такого устройства F с n входами и одним выходом, значение $b(t)$ которого при некоторых τ из N_0 и $f(x_1, x_2, \ldots, x_n)$ из P_l в каждый момент $t \geq \tau + 1$ определяется таким соотношением

$$b(t) = f(a_1(t-\tau), a_2(t-\tau), \ldots, a_n(t-\tau))$$

которое, очевидно, может быть описано системой вида (2).

Пусть $M \subseteq P_{a,l}$, тогда при $J_{\Omega_{\text{рс}}}(M) = M$ функциональная система $\mathcal{M} = (M, \Omega_{\text{рс}})$ и при $J_{\Omega_{\text{рс},\mathbf{O}}}(M) = M$ ф.с. $M = (\mathcal{M}, \Omega_{\text{рс},\mathbf{O}})$ называются итеративными ф.с. Поста автоматных функций (п.ф.с.а.ф.). Примерами таких ф.с. являются ф.с. вида $\mathcal{P}_{a,l,\text{и}}$. При заданном l из N_1 назовем основными п.ф.с.а.ф. следующие

$$(3) \qquad \begin{array}{ll} \mathcal{P}_{a,l,\text{к}} = (P_{a,l,\text{к}}, \Omega_{\text{рс}}), & \mathcal{P}^*_{a,l,\text{к}} = (P_{a,l,\text{к}}, \Omega_{\text{рс},\mathbf{O}}) \\ \mathcal{P}_{a,l} = (P_{a,l}, \Omega_{\text{рс}}), & \mathcal{P}^*_{a,l} = (P_{a,l}. \Omega_{\text{рс},\mathbf{O}}) \end{array}$$

Изложим главные результаты по нашей проблематике для п.ф.с.а.ф.

Теорема 4.1. **[24]** *Для любого l из N_1 среди основных п.ф.с.а.ф. конечно-порожденной является только ф.с. $\mathcal{P}^*_{a,l,\text{к}}$.*

Теорема 4.2. **[13]** *Для любых l из N и m из N_1 в $\mathcal{P}^*_{a,l,\text{к}}$ существует счетное множество базисов мощности m.*

А.функция, образующая полную систему в $\mathcal{M}$, называется Шефферовой. Дальнейшее упрощение базиса достигается за счет минимизации числа переменных, размерности и числа состояний у Шефферовой а.функции. Следующее утверждение дает окончательный ответ о виде простейших в этом смысле Шефферовых а.функций.

Теорема 4.3. **[25]** *Для любого l из N_1 в $\mathcal{P}^*_{a,l,\text{к}}$ существуют Шефферовые одномерные а.функции от двух переменных и с двумя состояниями.*

Для других основных п.ф.с.а.ф. ответ по проблеме базисов дает такое утверждение.

Теорема 4.4. **[24]** *Для любого l из N_1 в $\mathcal{P}_{a,l}$ и $\mathcal{P}^*_{a,l}$ базисов не существует, в $\mathcal{P}_{a,l,\text{к}}$ существует счетный базис, а также полная система, не содержащая в себе базиса.*

По проблеме полноты ситуацию описывают следующие утверждения.

Теорема 4.5. **[24, 26]** *Для основных п.ф.с.а.ф. множество $\Sigma_\pi(\mathcal{M})$ образует k-систему точно при $\mathcal{M} = \mathcal{P}^*_{a,l,\text{к}}$ для любого l из N_1.*

Теорема 4.6. **[13]** *Множество $\Sigma_\pi(\mathcal{M})$ континуально при $\mathcal{M} \in \{\mathcal{P}_{a,l,\text{к}}, \mathcal{P}^*_{a,l,\text{к}}\}$ и гиперконтинуально при $\mathcal{M} \in \{\mathcal{P}_{a,l}, \mathcal{P}^*_{a,l}\}$ для любого l из N_1.*

В качестве следствия заключаем, что указанными в теореме 4.6. мощностями обладают соответствующие решетки замкнутых классов в основных п.ф.с.а.ф.

Назовем систему $\Sigma' \subseteq \Sigma(\mathcal{P}^*_{a,l,\text{к}})$ k-критериальной, если всякое конечное множество $M \subseteq \mathcal{P}_{a,l,\text{к}}$, является полным точно тогда, когда для любого множества $K \in \Sigma'$ выполнено $M \not\subseteq K$.

Теорема 4.7. **[13]** *В $\mathcal{P}^*_{a,l,\text{к}}$ существуют счетные k-критериальные системы вида $\Sigma' \subseteq \Sigma_\pi(\mathcal{P}^*_{a,l,\text{к}})$ для любого l из N_1.*

Отметим, что в общем случае задание а.функций из $\mathcal{P}_{a,l}$ не является эффективным, поэтому проблема выразимости и полноты может ставится лишь для эффективно задаваемых систем.

Теорема 4.8. **[14]** *Проблема выразимости для эффективно задаваемых конечных систем а.функций в основных п.ф.с.а.ф. и проблема полноты в $\mathcal{P}_{a,l}$ алгоритмически не разрешимы для любого l из N_1.*

Таким образом, расширение функциональных возможностей а.функций по отношению к функциям l-значной логики и функциям с задержками резко усложняет решение интересующих нас проблем для алгебр а.функций. Изучение природы этой сложности осуществлялось в разных направлениях.

Мы остановимся здесь на задаче о приближенной полноте и на задаче о полноте специальной обогащенных систем а.функций.

Первая из этих задач имеет две модификации — задача о r-полноте ($r \in N$) и задача об аппроксимационной полноте (A-полноте), которым посвящается следующий параграф. Обратим внимание на специальные ф.с. $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ и $\mathcal{P}_4$, которые являются подалгебрами соответствующих основных п.ф.с.а.ф. из (3). Каждая из них состоит из всех одноместных и одномерных а.функцийи носителей указанных основных п.ф.с., а в качестве операций в них выступают те же операции, что и в соответствующих п.ф.с.а.ф, кроме операций σ и π. Как установлено в [32,24], они не имеют базисов. Кроме того, $\mathcal{P}_1$ содержит подалгебру $\mathcal{P}'_1$ всех взаимнооднозначных отображений, которая является группой с операцией подстановки, моделируя частью себя группу Бернсайдовского типа [23], то-есть такую конечно порожденную группу, в которой порядки элементов конечны, но в совокупности не ограничены. Открытыми остаются вопросы о наличии базисов в $\mathcal{P}'_1$, а также алгоритмические постановки о разрешимости конечности порядков ее элементов и выразимости этих элементов через другие. В заключение отметим, что теоремы 4.1, 4.2, 4.5, 4.6 и упомянутые здесь факты об одноместных алгебрах остаются справедливыми и в случае расширения значения l до счетного в ф.с. функций, которые тем самым обобщают а.функции.

5. Условия r-полноты и A-полноты для а.функций

Говорят, что а.функции f и g r-эквивалентны, если они совпадают на всех входных словах длины r (обозначение: $f\,r\,g$), и A-эквивалентны, если $f\,r\,g$ для любого r из N.

На множестве $B(P_{a,l})$ введем отношение, полагая для $M,\ M' \subseteq P^*_{a,l}$ выполненным $M \Delta_r M'$, если для всякой функции f из M найдется g из M', что $f\,r\,g$. Ясно, что это отношение образует предпорядок, а, значит, может быть представлено как отношение частичного порядка на классах эквивалентности, включающих в себя все такие элементы M и M', для которых выполнены соотношения $M\,\Delta_r\,M'$ и $M'\,\Delta_r\,M$; что обозначаем $M\,r\,M'$, а сами элементы называем r-эквивалентными.

На $B(P_{a,l})$ введем еще одно отношение, полагая для $M, M' \subseteq P^*_{a,l}$ выполненным $M\,\Delta_A\,M'$, если для всякого r из N имеет место $M\,r\,M'$. Это отношение также является предпорядком и для представителей M и M' его класса эквивалентности, когда тем самым выполнено $M\,\Delta_A\,M'$ и $M'\,\Delta_A\,M$, пишем $M\,A\,M'$; а сами представители называем A-эквивалентными.

Теорема 5.1. [15] *Для любых $M \subseteq P_{a,l}$ и $r \in N$ выполнено*

$$J_{\Omega_{\mathrm{pc}}}(M)\,r\,J_{\Omega_{\mathrm{pc},\mathbf{O}}}(M) \quad \text{и} \quad J_{\Omega_{\mathrm{pc}}}(M)\,A\,J_{\Omega_{\mathrm{pc},\mathbf{O}}}(M).$$

Таким образом, действия операторов $J_{\Omega_{\mathrm{pc}}}(M)$ и $J_{\Omega_{\mathrm{pc},\mathbf{O}}}(M)$ с точностью до r- и A-эквивалентностей совпадают, а тем самым в этом смысле операция обратной связи **O** оказывается моделируемой операциями расширенной суперпозиции, чем мы в дальнейшем будем пользоватся.

Пусть $M, M' \subseteq P_{a,l}$. Говорят, что M является r-выразимым через M', если $M \,\Delta_r\, J_{\Omega_{\text{pc}}}(M')$ и A-выразимым через M'', если $M \,\Delta_A\, J_{\Omega_{\text{pc}}}(M'')$.

Теорема 5.2. *Для эффективно задаваемых конечных множеств M, $M' \subseteq P_{a,l}$, отношение $M \,\Delta_r\, J_{\Omega_{\text{pc}}}(M')$ алгоритмически разрешимо для любого r из N.*

Теорема 5.3. [27] *Для конечных множеств $M, M' \subseteq P_{a,l,\text{к}}$ отношение $M \,\Delta_A\, J_{\Omega_{\text{pc}}}(M')$ алгоритмически не разрешимо.*

Пусть $M \subseteq P_{a,l}$ и $M \,A\, J_{\Omega_{\text{pc}}}(M')$. Назовем множество $M' \subseteq M$ r-полным в M, если $J_{\Omega_{\text{pc}}}(M') \,r\, M$ и A-полным, если $J_{\Omega_{\text{pc}}}(M') \,A\, M$.

Теорема 5.4. *Если $M \subseteq P_{a,l}$, $M \,A\, J_{\Omega_{\text{pc}}}(M)$, M-разрешимо, в M есть конечное A-полное подмножество и $r \in N$, то существует алгоритм, устанавливающий по любому конечному разрешимому подмножеству $M' \subseteq M$, является ли оно r-полным в M.*

На самом деле эта теорема следует из теоремы 2.2, в чем убеждаемся так. Пусть $f \in P_{a,l}$ и $r \in N$. Рассмотрим множество E_l^r в предположении, что его элементы кодируются словами длины r в алфавите E_l. Тогда рассматривая функцию f из $P_{a,l}$ только на словах длины r, можно считать ее принадлежащей P_l^r. Таким образом от рассмотрения а.функций мы перешли к функциям l^r-значной логики. Остается заметить, что операции расширенной суперпозиции в вопросах выразимости и полноты фактически редуцируются к операциям суперпозиции. Далее, из теоремы 4.9 и соотношения $P_{a,l,\text{к}} \,A\, P_{a,l}$ вытекает такое утверждение.

Теорема 5.5. *Условия r-полноты и A-полноты соответственно совпадают для всех основных п.ф.с.а.ф.*

Отметим существенное отличие понятий полноты и A-полноты, которое дается следующим предложением.

Теорема 5.6. *В каждой из основных п.ф.с.а.ф. существуют конечные A-полные системы, а также счетные A-полные системы, не содержащие конечных A-полных подсистем.*

Отличие же понятий r-полноты и A-полноты доставляет такое утверждение.

Теорема 5.7. [27] *Если $M \subseteq P_{a,l,\text{к}}$ и M — конечно, то не существует алгоритма, устанавливающего по M, является ли оно A-полным в $P_{a,l,\text{к}}$.*

В то же время существенная связь понятий r- и A-полноты имеется и проявляется прежде всего в подходе к решению задач о r- и A-полноте в терминах предполных классов.

Если $M \subseteq P_{a,l}$, $M\, A\, J_{\Omega_{\rm pc}}(M)$ и $M' \subseteq M$, то называем множество M' r-предполным в M, если оно не r-полно в M, но для любой функции f из $M \backslash M'$ множество $M' \cup \{f\}$ является r-полным в M. Аналогично вводиться понятие A-предполного множества. Пусть $\Sigma_{\pi,r}(M)$ и $\Sigma_{\pi,A}(M)$ суть множества всех r-предполных и A-предполных множеств в M соответственно.

Теорема 5.8. *Если $M \subseteq P_{a,l}$ и $M\, A\, J_{\Omega_{\rm pc}}(M)$, то*

$$\Sigma_{\pi,A}(M) = \bigcup_{r \geq 1} \Sigma_{\pi,r}(M)$$

.

Теорема 5.9. *Если $\lambda \in \{r, A\}$, $r \in N$, $M \subseteq P_{a,l}$, $M\, A\, J_{\Omega_{\rm ec}}(M)$, в M есть конечное λ-полное подмножество и $M' \subseteq M$, то M' является λ-полным в M точно тогда, когда для любого $K \in \Sigma_{\pi,\lambda}(M)$ выполнено $M' \not\subseteq K$.*

Это утверждение с учетом теорем 4.14 и 4.16 сводит решение задач о r- и A-полноте в основных п.ф.с.а.ф. к описанию множества $\Sigma_{\pi,r}(P_{a,l})$ которое было получено в [15]. Приведем это описание.

Пусть $t \in N$, обозначим через E_l^t множество всех слов $\varepsilon = e(1)e(2) \ldots e(t)$ длины t в алфавите E_l. При $h \in N$ и $T = (t_1, t_2, \ldots, t_h)$, где $t_i \in N$ при $t \in N_1^h$, положим $E_l^T = E_l^{t_1} \times E_l^{t_2} \times \cdots \times E_l^{t_h}$. Пусть $\rho(y_1, y_2, \ldots, y_h)$ — h-местный предикат, аргументы y_i которого принимают значения из $E_l^{t_i}$. Как и выше, пусть ρ_1 и ρ_0 суть соответственно множества истинных и ложных наборов значений переменных для ρ. Говорим, что а.функция $f(x_1, x_2, \ldots, x_n)$ из $P_{a,l}$ сохраняет ρ, если из истинности каждого элемента строки

$$\rho(\alpha_1^1, \alpha_2^1, \ldots, \alpha_h^1), \rho(\alpha_1^2, \alpha_2^2, \ldots, \alpha_h^2), \ldots, \rho(\alpha_1^n, \alpha_2^n, \ldots, \alpha_h^n)$$

вытекает истинность выражения

$$\rho(f(\alpha_1^1, \alpha_1^2, \ldots, \alpha_1^n), f(\alpha_2^1, \alpha_2^2, \ldots, \alpha_2^n), \ldots f(\alpha_h^1, \alpha_h^2, \ldots, \alpha_h^n)).$$

Класс всех таких а.функций обозначим через $U_a(\rho)$.

Введем функцию $\nu : E_l^* \times E_l^* \to N_0$. Пусть $\varepsilon_1 = E_l^{t_1}$, $\varepsilon_2 = E_l^{t_2}$, $t = \min\{t_1, t_2\}$. Тогда:

$$\nu(\varepsilon_1, \varepsilon_2) = \begin{cases} 0, \text{если } e_1(1) = e_2(1), \ldots, e_1(t) = e_2(t), \\ i, \text{если } 1 \leq i \leq t-1 \text{ и } \varepsilon_1(1) = \varepsilon_2(1), ..., \varepsilon_1(t-i) = \varepsilon_2(t-i), \\ \quad \text{но } e_1(t-i+1) \neq e_2(t-i+1), \\ t, \text{если } \varepsilon_1(1) \neq \varepsilon_2(1). \end{cases}$$

На множестве E_l^T определим отношение предпорядка $\leq$.

Пусть $A = (\alpha_1, \alpha_2, \ldots, \alpha_h)$ и $A' = (\alpha'_1, \alpha'_2, \ldots, \alpha'_h)$ — элементы из E_l^T. Пишем $A' \leq A$, если из включения $i, j \in N_1^h$, следует $\nu(\alpha'_i, \alpha'_j) \leq \nu(\alpha_i, \alpha_j)$.

Пусть $t' = \max\{t_1, t_2, \ldots, t_h\}$, $h \leq l^t$ и $t' \geq 2$; если $l = 2$, то полагаем $h = 2$. Пусть в A при $h \geq 2$ и условии, что $i \neq j$, выполнено $\nu(a_i, a_j) \neq 0$. Множество всех A', что $A' \leq A$, назовем ν-множеством, задаваемым элементом A, и обозначаем через ξ. Такое ξ разбиваем на два подмножества $\xi^{(m)}$, состоящее из всех максимальных по $\leq$ элементов, и $\xi^{\underline{m}}$ включающее остальные элементы из ξ. Так, при $h = 1$ имеем $\xi^{\underline{m}} = \emptyset$. Ясно, что значение $\nu(a_i, a_j)$ не зависит от выбора A из $\xi^{(m)}$, поэтому вместо $\nu(a_i, a_j)$ пишем $\nu(i, j)$.

Для $l \geq 2$ и $t \geq 1$ укажем семь семейств предикатов.

Пусть $h \geq 1$, $T = (t_1, t_2, \ldots, t_h)$ и $\xi \subseteq E_l^T$. Подстановку γ чисел $1, 2, \ldots, h$ назовем ξ-подстановкой, если для $(\alpha_1, \alpha_2, \ldots, \alpha_h)$ из ξ выполнено

$$(\alpha_{\gamma(1)}, \alpha_{\gamma(2)}, \ldots, \alpha_{\gamma(h)}) \in \xi.$$

Пусть $s \geq 1$, назовем множество

$$\{(\alpha_1^1, \alpha_2^1, \ldots, \alpha_h^1), (\alpha_1^2, \alpha_2^2, \ldots, \alpha_h^2), \ldots, (\alpha_1^s, \alpha_2^s, \ldots, \alpha_h^s)\},$$

элементов из $\xi^{(m)}$ ξ-совместимым, если существует совокупность ξ-подстановок $\gamma_1, \gamma_2, \ldots, \gamma_s$ такая, что для любых q и p из N_1^s, а также i и j из N_1^h выполнено $\nu_\xi(i, j) \leq \nu(\alpha^q_{\gamma_q}(i), \alpha^p_{\gamma_p}(j))$.

Пусть $\rho(y_1, y_2, \ldots, y_h)$ — предикат, для которого $\rho_1 \subseteq \xi$. Назовем ρ ξ-рефлексивным, если $\xi^{(m)} \subseteq \rho_1$, и ξ-симметричным, если для $(\alpha_1, \alpha_2, \ldots, \alpha_h)$ из ρ_1 и ξ-подстановки γ всегда выполнено $(\alpha_{\gamma(1)}, \alpha_{\gamma(2)}, \ldots, \alpha_{\gamma(h)}) \in \rho_1$.

ξ-рефлексивный и ξ-симметричный предикат ρ называем ξ-элементарным, если множество $\xi \backslash \rho_1$ является ξ-совместимым. Для такого ρ при $A \in \xi \backslash \rho_1$ и $i \in N_1^h$ определим подмножества $C^i_\rho(A)$, $Q^i_\rho(A)$ и $\varepsilon^i_\rho(A)$ множества E_l так:

а) $a \in C^i_\rho(A)$ точно тогда, когда найдется $\alpha'_i \in E_l^{t_i}$, что $\nu(\alpha_i, \alpha'_i) \leq 1$, $\alpha'_i(t_i) = a$, а любой элемент $(\alpha'_1, \alpha'_2, \ldots, \alpha'_h)$ из ξ содержится в ρ_1;

б) $b \in Q^i_\rho(A)$ точно тогда, когда в $\xi \backslash \rho_1$ найдется элемент

$$(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, \alpha''_i, \alpha_{i+1}, \ldots, \alpha_h),$$

такой, что $\nu(\alpha_i, \alpha''_i) \leq 1$ и $\alpha''(t_i) = b$;

в) множество $\varepsilon^i_\rho(A)$ совпадает с $C^i_\rho(A)$, если $C^i_\rho(A) \neq \emptyset$, и с $Q^i_\rho(A)$ в противном случае.

Пусть $n \geq 1$, и $R = \{\rho^1, \rho^2, \ldots, \rho^n\}$ — произвольная система ξ-элементарных предикатов. Называем R T-совместимой, если для любых $\sigma \in N_1^n$, $i \in N_1^h$ и $A \in \xi\backslash\rho_1^\sigma$ выполнено $Q_\rho^i(A) \neq E_l$. Называем R W-совместимой, если для всех σ, $\sigma' \in N_1^n$, $i \in N_1^h$, $A \in \xi\backslash\rho_1^\sigma$, $A' \in \xi\backslash\rho_1^\sigma$ множества $C_{\rho^\sigma}^i(A)$ и $C_{\rho^{\sigma'}}^i(A)$ одновременно либо пусты, либо непусты, причем при $C_{\rho^\sigma}^i(A) = \emptyset$ для всякого $b \in E_l$ существует $j \in N_1^h$ такое, что $t_j = t_i$, $\nu_\xi(i, j) \leq 1$, $a \in Q_{\rho^\sigma}^j(A)$.

Пусть $n \geq 1$, $\sigma_i \in N_1^q$, $A^i = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_n^i)$, $A^i \in \xi\backslash\rho_1^{\sigma_i}$ и $j_i \in N_1^h$ при $i \in N_1^n$. Тогда, если $\sigma_i \neq \sigma_i'$ при $i \neq i'$ и $t_{j_i} = t_{j_{i'}}$, а также $\nu(\alpha_{j_1}^1, \alpha_{j_{i''}}^2) \leq 1$ для всех $i, i' \in N_1^q$ и $i'' \in N_1^n\backslash\{1\}$, и выполнено

$$\bigcap_{i=1}^{n} \varepsilon_{\rho_1^{\sigma_i}}^{j_i}(A^i) = \emptyset,$$

то систему R назовем Q-совместимой.

Пусть $n \geq 2$, $i \in N_1^n$, $\sigma_j \in N_1^q$ и $A^i \in \xi\backslash\rho_1^{\sigma_i}$. Положим, что для всех $j, j' \in N_1^n$ при $j \neq j'$ выполнено $\nu(\alpha_1^j, \alpha_1^{j'}) \neq 1$. Пусть для всех $i_j, i_{j'} \in N_1^h$ выполнено $t_{i_j} = t_{i_{j'}}$, и при $j > 1$ справедливо $\nu(\alpha_{i_1}^1, \alpha_{i_j}) \leq 1$, а также имеет место

$$\bigcap_{j=1}^{n} \varepsilon_{\rho^{\sigma_j}}^{i_j}(A^j) = \emptyset.$$

Предположим, что при этих условиях существуют $l_v \in N_1^n$, $v \in N_1^q$, что $i_{l_v} \in N_1^2$, множества A^{l_v} являются ξ-совместимыми и

$$\bigcap_{v=1}^{q} \varepsilon_{\rho^{\sigma_{l_v}}}^{i_{l_v}}(A^{lv}) = \emptyset.$$

Тогда говорим, что система R является D-совместимой.

Семейство предикатов $Z_l(r)$. Это семейство не пусто при $l \geq 2$ и $r \geq 1$. Предикат ρ входит в $Z_l(r)$ точно тогда, когда $\rho_1 \subset \xi$, $\xi \subseteq E_l^T$, $T = (t_1, t_2, \ldots, t_h)$, $h \geq 1$, $\max\{t_1, t_2, \ldots, t_h\} \leq r$ для некоторого $m \geq 1$ выполнено $\rho_1 = \cap_{i=1}^m \rho_1^i$, где ρ^i является ξ-елементарным предикатом, а сами ρ^i образуют T-, W-, Q-совместимую систему, а для любых $j \in N_1^h$ и $A \in \xi\backslash\rho_1$ множество $C_\rho^i(A)$ непусто.

Семейство предикатов $J_l(r)$. Оно не пусто при $r \geq 1$ и $l > 2$, а также при $r \geq 2$ и $l = 2$. Предикат ρ входит в $J_l(r)$ точно тогда, когда $\rho_1 \subseteq \xi \subseteq E_l^T$, $T = (t_1, t_2, \ldots, t_h)$, $h \geq 3$, $\max\{t_1, t_2, \ldots, t_h\} \leq r$, для некоторого $m \geq 1$ выполнено $\rho_1 = \cap_{i=1}^m \rho_1^i$, где ρ^i является ξ-елементарным, а сами ρ^i образуют T-,W-,Q-совместимую систему;

существуют числа $i, j, l \in N_1^h$, $i \neq j$, $i \neq l$, $j \neq l$, что для A из $\xi \backslash \rho_1^1$ множество $C_\rho^u(A)$ пусто при $u \in \{i, j, l\}$.

Семейство предикатов $D_l(r)$. Оно не пусто при $r \geq 1$ если $l > 2$, и при $r \geq 2$, если $l = 2$. Предикат ρ входит в $D_l(r)$ точно тогда, когда, $\rho_1 \subset \xi \subseteq E_l^T$, $T = (t_1, t_2, \ldots, t_h)$, $h \geq 2$, $\max\{t_1, \ldots, t_h\} \leq r$ для некоторого $m \geq 1$ выполнено $\rho_1 = \cap_{i=1}^m \rho_1^i$, где ρ^i является ξ-элементарным, а сами ρ^i образуют T-,W- и D-совместимую систему; для любого $A \in \xi \backslash \rho$ множества $C_\rho^1(A)$ и $C_\rho^2(A)$ пустые; если $h \geq 3$, то для i из N_1^h выполнено $C_\rho^i(A) \neq \emptyset$; если $h = 2$, то $\rho_1 \cap \xi^{(M)} \neq \emptyset$.

Пусть далее $l \geq 2$, $t \geq 1$, $T = (t, t)$, $\xi_t \subseteq E_l^T$, ξ_t есть ν-подмножество и $\nu_{\xi t} = (1, 2) = 1$.

Семейство предикатов $M_l(r)$. оно не пустое при $l \geq 2$ и $r \geq 1$. Предикат ρ входит в $M_l(r)$ точно тогда, когда $\rho_1 \subset \xi_t$, $t \leq r$, ρ_1 совпадает с отношением частичного порядка, определенном на E_l^T и имеющим точно l^{t-1} минимальных и l^{t-1} максимальных елементов.

Семейство предикатов $S_l(r)$. Оно не пустое при $l \geq 2$ и $r \geq 1$. Предикат ρ входит в $S_l(r)$ точно тогда, когда $\rho_1 \subset \xi_t$, $t \leq r$, существует подстановка σ_ρ на E_l^T, разлагающая в производение циклов одинаковой простой длины $p \geq 2$, график которой совпадает с ρ_1, то есть если $a \in E_l^T$, то $(a, \sigma_\rho(a)) \in \rho_1$ и если $(a_1, a_2) \in \rho_1$, то $a_2 = \sigma_\rho(a_1)$.

Пусть $t \geq 1$, $\tilde{\Phi}_t$ — класс всех отображений φ множества E_l^T в множество подстановок на E_l. Значение φ на a обозначаем φ_a.

Пусть $\Phi_t \subseteq \tilde{\Phi}_t$ и Φ_t состоит из всех φ таких, что $\varphi_a = \varphi_{a'}$ при $\nu(a, a') \leq 1$. Положим $h \in \{3, 4\}$, $T_h = \{t, t, \ldots, t\}$, $K_t^h \subseteq E_l^{T_h}$ и K_t^h состоит из всех элементов $(a_1, a_2, \ldots, a_h)$ таких, что при $i, j \in N_1^h$ выполнено $\nu(a_i, a_j) \leq 1$.

Пусть $l = p^m$, p — простое, $m \geq 1$, $G =< E_l, + >$ — абелева группа, в которой каждый ненулевой элемент имеет простой порядок p.

При $p \neq 2$ пусть $l_p \in N_1^{p-1}$ и $2l_p = 1 (\mathrm{mod}\ p)$.

Семейство предикатов $L_l(r)$. Оно не пусто только при $l = p^m$, p — простое, $m \geq 1$, $r \geq 1$. Предикат ρ входит в $L_l(r)$ точно тогда, когда для некоторого φ из Φ_t, $t \leq r$, справедливо следующее:

а) Пусть $k = p^m$, $p > 2$. Тогда $\rho_1 \subset K_t^3$ и (a_1, a_2, a_3) из K_t^3 входит в ρ_1, если только $\varphi_{a_1}(a_3(t)) = l_p(\varphi_{a_1}(a_1(t)) + \varphi_{a_1}(a_2(t)))$.

б) Пусть $k = 2^m$. Тогда $\rho_1 \subset K_t^h$ и (a_1, a_2, a_3, a_4) из K_t^4 входит в ρ_1, если только

$$\varphi_{a_1}(a_1(t)) + \varphi_{a_1}(a_2(t)) = \varphi_{a_1}(a_3(t)) + \varphi_{a_1}(a_4(t)).$$

Отметим, что указанные семейства при $r = 1$ соответственно совпадают с известными семействами для P_l из параграфа 2.

Пусть $t \geq 2$, $T = (t, t)$, $\xi_t^{\sim}$ есть ν-подмножество E_l^T такое, что $\nu_{\xi_t^{\sim}}(1, 2) = 2$.

Семейство предикатов $V_l(r)$. Оно не пусто при $l \geq 2$, $r \geq 2$. Предикат ρ входит в $V_l(r)$ точно тогда, когда $\rho_1 \subset \xi_t^{\sim}$, $t \leq r$, и выполнено: $(a_1, a_2) \in \xi_t^{\sim(m)} \cap \rho_1$ точно тогда, когда $a_1(t) = a_2(t)$; существует φ из Φ_t такое, что включение $(a_1, a_2) \in \xi_t^{\sim(M)} \cap \rho_1$ эквивалентно существованию α из E_l такому, что $a_1(t) = \varphi_{a_1}(\alpha)$ и $a_2(t) = \varphi_{a_2}(t)$.

Пусть

$$W_l(r) = Z_l(r) \cup J_l(r) \cup D_l(r) \cup M_l(r) \cup S_l(r) \cup L_l(r) \cup V_l(r)$$

и $U(W_l(r))$ — множество классов а.функций, сохраняющих предикаты из $W_l(r)$.

Теорема 5.10. [15] *Имеет место равенство* $\Sigma_{\pi,r}(P_{a,l}) = U(W_l(r))$.

Список литературы

[1] *Автоматы*, Сборник статей под редакцией Маккарти и Шеннона, М., ИЛ., 1956.

[2] Яаблонский С.В, *Функциональные построения в k-значной логике*, В кн.: Труды математ. института им. В.А. Стеклова **Т. 51** (1958), Изд-во АН СССР, 5–42.

[3] Post. E, *Two-valued iterative systems of mathematical logic* (1941), Prinston.

[4] Яблонский С.В. Гаврилов Г.П., Кудрявцев В.Б, *Функции алгебры логики и классы Поста*, М., Наука, 1966.

[5] Ло Чжу-Кай, *Предполные классы, определяемые k-арными отношениями в k-значной логике*, Acta Sci. natur. Univ. Jilinesis **3** (1964).

[6] Ло Чжу-Кай, Лю Сюй Хуа, *Предполные классы, определяемые бинарными отношениями в многозначной логике*, Acta Sci. natur. Univ. Jilinesis **4** (1964).

[7] Захарова Е.Ю, *Критерий полноты системы функций из Пк*, Проблемы кибернетики (1967), no. 18, М., Наука, 5–10.

[8] Мартынюк В.В, *Исследование некоторых классов функций в многозначных логиках*, Проблемы кибернетики (1960), no. 3, М., Наука, 49–60.

[9] Пан Юн-Цзе, *Один разрешающий метод для отыскания всех предполных классов в многозначной логике*, Acta Sci. natur. Univ. Jilinesis **2** (1964).

[10] Rosenberg J, *La structure des functions de plusiers variables sur un ensemble fini*, Comptes Rendus Acad. Sci. Paris **260** (1965), 3817-3819.

[11] Кудрявцев В.Б, *Теорема полноты для одного класса автоматов без обратных связей*, Проблемы кибернетики **8** (1962), М., Наука, 91–115.

[12] Часовских А.А, *О полноте в классе линейных автоматов*, Математические вопросы кибернетики **3** (1995), М., Наука, 140–166.

[13] Кудрявцев В.Б, *О мощностях множеств предполных множеств некоторых функциональных систем, связанных с автоматами*, Проблемы кибернетики **13** (1965), М., Наука, 45–74.

[14] Кратко М.И, *Алгоритмическая неразрешимость проблемы распознавания полноты для конечных автоматов* **t.155** (1964), no. 1, ДАН СССР, 35–37.

[15] Буевич В.А, *О τ-полноте в классе детерминированных функций* **t. 326** (1992), no. 3, ДАН СССР, 399–404.

[16] Бабин Д.Н, *Неразрешимость полноты и A-полноты автоматных функций с истинностной системой типа F^∞, S, P, O* **t. 6** (1995), no. 1, Дискретня математика.

[17] Кудрявцев В.Б, *Функциональные сыстемы* (1982), М., Издательство МГУ, 159.

[18] Кон П, *Универсальная алгебра* (1968), М., Мир.

[19] Мальцев А.И, *Итеративные алгебры и многобразие Поста*, Алгебра и логика, семинар, vol. t. 5, Изд-во СО АНСССР, Новосибирск, 1966.

[20] Янов Ю.И., Мучник А.А, *О существовании k-значных замкнутых классов, не имеющих базиса* **t.127** (1959), no. 1, ДАН СССР, 144–146.

[21] Захарова Е.Ю., Кудрявцев В.Б., Яблонский С.В, *О предполных классах в k-значных логиках* **t. 186** (1969), no. 3, ДАН СССР, 509–512.

[22] Rosenberg I., Hikata T, *Completeness for uniformly delayed circuits* (1983), Proceedings of the 13th International Symposium on Multiple-Valued Logic, Kyoto, Japan, May 23-25, 1–9.

[23] Алешин С.В, *Конечные автоматы и проблема Бернсайда о периодических группах* **3** (1972), М., Мат. заметки, 319–328.

[24] Кудрявцев В.Б., Алешин С.В., Подколзин А.С, *Введение в теорию автоматов* (1985), М., Наука, 320.

[25] Ветренникова Е.В, *Один простой пример универсальной о.д.функции*, Дискретный анализ, Новосибирск, Институт математики СО АНСССР, 1983, pp. 5–11.

[26] Бабин Д.Н, *О суперпозициях ограничено-детерминированных функций*, Мат. заметки, vol. t. 47, 1990, pp. 3–10.

[27] Буевич В.А, *Об алгоритмической неразрешимости распознавания A-полноты для ограничено-детерминированных функций*, М., Мат. заметки, vol. 6, 1972, pp. 687–697.

[28] Slupecki J, *Kriterium petnosci wielowartosciowych systemow logiki zdan*, Comptes rendus des scanes de la Societe des Sciences et des Letters de Varsovie **111, T. 32** (1939), 102–109.

[29] Бабин Д.Н, *О полноте двухместных о.д.-функций относительно суперпозиций* **4, t. 1** (1989), Дискретная Математика, 86–91.

[30] Буевицх В.А, *Разрешимость проблемы Слупецкого для автоматов* **4, t. 4** (1995), Дискретная Математика.

[31] Бабин Д.Н, *Разрешимый случай задачи о полноте автоматных функций* **4, t. 4** (1992), Дискретная Математика, 41–55.

[32] Алешин С.В, *Об отсутствии базисов в некоторых классах инициальных автоматов*, Проблемы кибернетики, vol. 22, М., Наука, 1970, pp. 67–75.

Механико-математечиский факультет, МГУ, Москва, Россия

FILOMAT (Niš) 9:3 (1995), 517–542
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

SOME NONSTANDARD TYPES OF ORTHOGONALITY (A SURVEY)

Gradimir V. Milovanović

ABSTRACT. This survey is devoted to some nonstandard types of orthogonal polynomials in the complex plane. Under suitable integrability conditions on w, we consider polynomials orthogonal on a circular arc with respect to a non-Hermitian complex inner product as well as Geronimus' version of orthogonality on a contour in the complex plane. Also, we introduce a class of polynomials orthogonal on some selected radial rays in the complex plane. In both of cases we investigate their existence and uniqueness, recurrence relations, representations and connections with standard polynomials orthogonal on the real line. We also give an introduction to the general theory of orthogonality on the real line and the unit circle. Zero distributions of nonstandard types of orthogonal polynomials are considered.

1. Introduction

The orthogonal systems play an important role in many branches of mathematics, physics and other applied and computational sciences. Especially, orthogonal polynomial systems appear in the Gaussian quadrature processes, the least square approximation of functions, differential and difference equations, Fourier series, etc.

In this survey we mainly consider some classes of nonstandard orthogonal polynomials. The paper is organized as follows. In Section 2 we discuss two standard types of orthogonal polynomials – *polynomials orthogonal on the real line* and *polynomials orthogonal on the unit circle*. The most important properties of such polynomials are presented. Under suitable integrability conditions on a weight function, in Section 3 we consider polynomials orthogonal on the semicircle with respect to a complex-valued inner product.

This work was supported in part by the Serbian Scientific Foundation, grant number 0401F.

A generalization of such nonstandard orthogonal polynomials on a circular arc in the complex plane is treated in Section 4. Geronimus' version of orthogonality on a contour in the complex plane for polynomials orthogonal on the semicircle or on a circular arc is considered in Section 5. Sections 6 and 7 are devoted to a new class of orthogonal polynomials on some selected radial rays in the complex plane. We investigate the existence and uniqueness, recurrence relations, representations and the connection with standard polynomials orthogonal on the real line. Also, the distribution of zeros of such polynomials is included.

2. Standard types of orthogonal polynomials

A standard type of orthogonality is one on the real line with respect to a given non-negative measure $d\lambda(t)$. Namely, let $\lambda\colon \mathbb{R} \to \mathbb{R}$ be a fixed non-decreasing function with infinitely many points of increase for which all moments $\mu_k = \int_{\mathbb{R}} t^k \, d\lambda(t)$, $k = 0, 1, \dots,$ exist and are finite. Then the improper Stieltjes integral $\int_{\mathbb{R}} P(t)\, d\lambda(t)$ exists for every polynomial P. By the application of the Lebesgue-Stieltjes integral $\int_{\mathbb{R}} f(t)\, d\lambda(t)$ to characteristic functions of sets, the function λ engenders a Lebesgue-Stieltjes measure $d\lambda(t)$, which is known also as *m-distribution* (cf. Freud [10]). Moreover, if $t \mapsto \lambda(t)$ is an absolutely continuous function, then we say that $\lambda'(t) = w(t)$ is a *weight function*. In that case, the measure $d\lambda$ can be express as $d\lambda(t) = w(t)\, dt$, where the weight function $t \mapsto w(t)$ is a non-negative and measurable in Lebesgue's sense for which all moments exists and $\mu_0 = \int_{\mathbb{R}} w(t)\, dt > 0$.

In the general case the function λ can be written in the form $\lambda = \lambda_{\mathrm{ac}} + \lambda_{\mathrm{s}} + \lambda_{\mathrm{j}}$, where λ_{ac} is absolutely continuous, λ_{s} is singular, and λ_{j} is a jump function.

The set of points of increase of $t \mapsto \lambda(t)$, so-called the *support of the measure*, we denote by $\operatorname{supp}(d\lambda)$. It is always an infinite and closed set. If $\operatorname{supp}(d\lambda)$ is bounded, then the smallest closed interval containing $\operatorname{supp}(d\lambda)$ we will denote by $\Delta(d\lambda)$.

Using the measure $d\lambda(t)$ we can define the inner product (f, g), by

$$(2.1) \qquad (f,g) = \int_{\mathbb{R}} f(t)\overline{g(t)}\, d\lambda(t) \qquad \bigl(f, g \in L^2(\mathbb{R}) \equiv L^2(\mathbb{R}; d\lambda)\bigr),$$

and consider a system of (monic) orthogonal polynomials $\{p_k(t)\}$ such that

$$p_k(t) = t^k + \text{terms of lower degree} \quad (k = 0, 1, \dots),$$
$$(p_k, p_n) = 0, \quad k \neq n, \quad (p_n, p_n) = \|p_n\|^2 > 0.$$

For any m-distribution $d\lambda(t)$ there exists a unique system of polynomials $\{p_k(t)\}$.

A general property of the inner product (2.1) that $(tf, g) = (f, tg)$ provides the three-term recurrence relation for the (monic) orthogonal polynomials $p_k(t)$,

$$\begin{aligned} (2.2) \qquad & p_{k+1}(t) = (t - a_k)p_k(t) - b_k p_{k-1}(t), \qquad k = 0, 1, 2, \dots, \\ & p_0(t) = 1, \quad p_{-1}(t) = 0. \end{aligned}$$

The recursion coefficients can be expressed in terms of inner product (cf. Milovanović, Mitrinović, Rassias [31, p. 33])

$$a_k = \frac{(tp_k, p_k)}{(p_k, p_k)} \quad (k \geq 0), \quad b_k = \frac{(p_k, p_k)}{(p_{k-1}, p_{k-1})} \quad (k \geq 1).$$

The coefficient b_0, which multiplies $p_{-1}(t) = 0$ in three-term recurrence relation may be arbitrary. Sometimes, it is convenient to define it by $b_0 = \mu_0 = \int_{\mathbb{R}} d\lambda(t)$. Then the norm of p_n can be express in the form

$$\|p_n\| = \sqrt{(p_n, p_n)} = \sqrt{b_0 b_1 \cdots b_n}\,.$$

An interesting and very important property of polynomials $p_n(t)$, $n \geq 1$, is the distribution of zeros. Namely, all zeros of $p_n(t)$ are real and distinct and are located in the interior of the interval $\Delta(d\lambda)$. Also, the zeros of $p_n(t)$ and $p_{n+1}(t)$ interlace, i.e.,

$$\tau_k^{(n+1)} < \tau_k^{(n)} < \tau_{k+1}^{(n+1)} \qquad (k = 1, \dots, n;\ n \in \mathbb{N}),$$

where $\tau_k^{(n)}$, $k = 1, \dots, n$, denote the zeros of $p_n(t)$ in an increasing order

$$\tau_1^{(n)} < \tau_2^{(n)} < \cdots < \tau_n^{(n)}.$$

It is easy to prove that the zeros $\tau_k^{(n)}$ of $p_n(t)$ are the same as the eigenvalues of the following tridiagonal matrix

$$J_n = J_n(d\lambda) = \begin{bmatrix} a_0 & \sqrt{b_1} & & & \mathrm{O} \\ \sqrt{b_1} & a_1 & \sqrt{b_2} & & \\ & \sqrt{b_2} & a_2 & \ddots & \\ & & \ddots & \ddots & \sqrt{b_{n-1}} \\ \mathrm{O} & & & \sqrt{b_{n-1}} & a_{n-1} \end{bmatrix},$$

which is known as the *Jacobi matrix*. Also, the monic polynomial $p_n(t)$ can be expressed in the following determinant form

$$\hat{\pi}_n(t) = \det(tI_n - J_n),$$

where I_n is the identity matrix of the order n.

Suppose that $\Delta(d\lambda) = [a, b]$. Since every interval (a, b) can be transformed by a linear transformation to one of following intervals: $(-1, 1)$, $(0, +\infty)$, $(-\infty, +\infty)$, we can restrict the consideration (without loss of generality) only to these three intervals. A very important class of orthogonal polynomials on an interval of orthogonality $(a, b) \in \mathbb{R}$ is constituted by so-called the *classical orthogonal polynomials*. Their weight functions $w(t)$ satisfy the differential equation

$$\frac{d}{dt}(A(t)w(t)) = B(t)w(t),$$

where

$$A(t) = \begin{cases} 1 - t^2, & \text{if } (a, b) = (-1, 1), \\ t, & \text{if } (a, b) = (0, +\infty), \\ 1, & \text{if } (a, b) = (-\infty, +\infty), \end{cases}$$

and $B(t)$ is a polynomial of the first degree.

The classical orthogonal polynomials $\{Q_k\}$ on (a, b) can be specificated as the *Jacobi polynomials* $P_k^{(\alpha,\beta)}(t)$ $(\alpha, \beta > -1)$ on $(-1, 1)$, the *generalized Laguerre polynomials* $L_k^s(t)$ $(s > -1)$ on $(0, +\infty)$, and finally as the *Hermite polynomials* $H_k(t)$ on $(-\infty, +\infty)$, with the weight functions

$$t \mapsto (1-t)^\alpha(1+t)^\beta, \qquad t \mapsto t^s e^{-t}, \qquad t \mapsto e^{-t^2} \qquad (\alpha, \beta, s > -1),$$

respectively. These polynomials have many nice particular properties (cf. [9], [25], [29], [31], [40], [43]). Some characterizations of the classical polynomials were given in [2–3], [5–6], [9], [20], [23].

There are several classes of orthogonal polynomials which are in certain sense close to the classical orthogonal polynomials, so-called *semi-classical* polynomials.

In many applications of orthogonal polynomials it is very important to know the recursion coefficients a_k and b_k. If $d\lambda(t)$ is one of the classical measures, then a_k and b_k are known explicitly. Furthermore, there are certain non-classical cases when we know also these coefficients. For example, we mention here the *generalized Gegenbauer weight* $w(t) = |t|^\mu(1-t^2)^\alpha$, $\mu, \alpha > -1$, on $[-1, 1]$ (see Lascenov [22] and Chihara [9, pp. 155–156]), the *hyperbolic weight* $w(t) = 1/\cosh t$ on $(-\infty, +\infty)$ ([9, pp. 191–193]), and the *logistic weight* $w(t) = e^{-t}/(1+e^{-t})^2$ on $(-\infty, +\infty)$.

A system of orthogonal polynomials for which the recursion coefficients are not known explicitly will be said to be *strong non-classical* orthogonal polynomials. In such cases there are a few known approaches to compute the first n coefficients a_k, b_k, $k = 0, 1, \dots, n-1$. Furthermore, for such a purpose there is the package ORTHPOL developed by Gautschi [12]. These coefficients then allow us to compute all orthogonal polynomials of degree $\le n$ by a straightforward application of the three-term recurrence relation (2.2).

Another type of orthogonality is *orthogonality on the unit circle*. The polynomials orthogonal on the unit circle with respect to a given weight function have been introduced and studied by Szegő [41–43] and Smirnov [37–38]. A more general case was considered by Achieser and Kreĭn [1], Geronimus [16–17], P. Nevai [35–36], Alfaro and Marcellán [4], Marcellán and Sansigre [24], etc. These polynomials are linked with many questions in the theory of time series, digital filters, statistics, image processing, scattering theory, control theory and so on.

The inner product is defined by

$$(f, g) = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})}\, d\mu(\theta),$$

where $d\mu(\theta)$ is a finite positive measure on the interval $[0, 2\pi]$ whose support is an infinite set. In that case there is a unique system of (monic) orthogonal polynomials $\{\phi_k\}_{k\in\mathbb{N}_0}$. If $\theta \mapsto \mu(\theta)$ is an absolutely continuous function on $[0, 2\pi]$, then we say that $\mu'(\theta) = w(\theta)$ is a *weight function*.

The monic orthogonal polynomials $\{\phi_k\}$ on the unit circle $|z| = 1$ satisfy the recurrence relations

$$\phi_{k+1}(z) = z\phi_k(z) + \phi_{k+1}(0)\phi_k^*(z), \quad \phi_{k+1}^*(z) = \phi_k^*(z) + \overline{\phi_{k+1}(0)}z\phi_k^*(z),$$

for $k = 0, 1, \dots$, where $\phi_k^*(z) = z^k\bar{\phi}_k(1/z)$.

As we can see these recurrence relations are not three-term relations like (2.2). The values $\phi_k(0)$ are called reflection parameters or Szegő parameters. Defining a sequence of parameters $\{a_k\}$ by $a_k = -\overline{\phi_{k+1}(0)}$, $k = 0, 1, \dots$, Geronimus [18, Chapter VIII] derived the following three-term recurrence relations:

$$\bar{a}_{k-1}\phi_{k+1}(z) = (\bar{a}_{k-1}z + \bar{a}_k)\phi_k(z) - \bar{a}_k z(1 - |a_{k-1}|^2)\phi_{k-1}(z),$$

$$a_{k-1}\phi_{k+1}^*(z) = (a_{k-1}z + a_k)\phi_k^*(z) - a_k z(1 - |a_{k-1}|^2)\phi_{k-1}^*(z),$$

where $k \in \mathbf{N}$ and $\phi_0(z) = 1$, $\phi_1(z) = z - \bar{a}_0$.

3. Orthogonality on the semicircle

Polynomials orthogonal on the semicircle

$$\Gamma_0 = \{z \in \mathbb{C} \,|\, z = e^{i\theta},\ 0 \le \theta \le \pi\}$$

have been introduced by Gautschi and Milovanović [14–15]. The inner product is given by

$$(f, g) = \int_\Gamma f(z)g(z)(iz)^{-1}\,dz,$$

where Γ is the semicircle $\Gamma = \{z \in \mathbb{C} \mid z = e^{i\theta}, 0 \le \theta \le \pi\}$. Alternatively,

$$(f, g) = \int_0^\pi f(e^{i\theta})g(e^{i\theta})\,d\theta.$$

This inner product is not Hermitian, but the corresponding (monic) orthogonal polynomials $\{\pi_k\}$ exist uniquely and satisfy a three-term recurrence relation of the form

$$\begin{aligned}&\pi_{k+1}(z) = (z - i\alpha_k)\pi_k(z) - \beta_k\pi_{k-1}(z), \qquad k = 0, 1, 2, \ldots,\\ &\pi_{-1}(z) = 0, \quad \pi_0(z) = 1.\end{aligned}$$

Notice that the inner product possesses the property $(zf, g) = (f, zg)$.

The general case of complex polynomials orthogonal with respect to a *complex weight function* was considered by Gautschi, Landau and Milovanović [13]. Namely, let $w\colon (-1, 1) \mapsto \mathbb{R}_+$ be a weight function which can be extended to a function $w(z)$ holomorphic in the half disc

$$D_+ = \{z \in \mathbb{C} \,|\, |z| < 1, \operatorname{Im} z > 0\},$$

and

$$(3.1) \qquad (f, g) = \int_\Gamma f(z)g(z)w(z)(iz)^{-1}\,dz = \int_0^\pi f(e^{i\theta})g(e^{i\theta})w(e^{i\theta})\,d\theta.$$

Together with (3.1) consider the inner product

$$(3.2) \qquad [f, g] = \int_{-1}^1 f(x)\overline{g(x)}w(x)\,dx,$$

which is positive definite and therefore generates a unique set of real (monic) orthogonal polynomials $\{p_k\}$:

$$[p_k, p_m] = 0 \quad \text{for} \quad k \ne m \quad \text{and} \quad [p_k, p_m] > 0 \quad \text{for} \quad k = m \quad (k, m \in \mathbb{N}_0).$$

On the other hand, the inner product (3.1) is not Hermitian; the second factor g is not conjugated and the integration is not with respect to the measure $|w(e^{i\theta})|\,d\theta$. The existence of corresponding orthogonal polynomials, therefore, is not guaranteed.

We call a system of complex polynomials $\{\pi_k\}$ *orthogonal on the semicircle* if

$$(\pi_k, \pi_m) = 0 \quad \text{for } k \neq m \quad \text{and} \quad (\pi_k, \pi_m) > 0 \quad \text{for } k = m \quad (k, m \in \mathbb{N}_0).$$

where we assume that π_k is monic of degree k.

The existence of the orthogonal polynomials $\{\pi_k\}$ can be established assuming only that

$$\text{(3.3)} \qquad \operatorname{Re}(1,1) = \operatorname{Re} \int_0^\pi w(e^{i\theta})\,d\theta \neq 0.$$

Assume that the weight function w is positive on $(-1,1)$, holomorphic in D_+ and such that the integrals in (3.1) and (3.2) exist for smooth f and g (possibly) as improper integrals. We also assume that the condition (3.3) is satisfied.

Let C_ε, $\varepsilon > 0$, denote the boundary of D_+ with small circular parts of radius ε and centres at ± 1 spared out and let $\mathcal{P}$ be the set of all algebraic polynomials. Further, let Γ_ε and $C_{\varepsilon,\pm 1}$ be the circular parts of C_ε with radii 1 and ε, respectively.

Then, using Cauchy's theorem and assuming that w is such that for all $g \in \mathcal{P}$,

$$\text{(3.4)} \qquad \lim_{\varepsilon \to 0} \int_{C_{\varepsilon,\pm 1}} g(z) w(z)\,dz = 0.$$

we obtain

$$\text{(3.5)} \quad 0 = \int_C g(z)w(z)\,dz = \int_\Gamma g(z)w(z)\,dz + \int_{-1}^1 g(x)w(x)\,dx, \quad g \in \mathcal{P}.$$

The (monic, real) polynomials $\{p_k\}$, orthogonal with respect to the inner product (3.2), as well as the associated polynomials of the second kind,

$$q_k(z) = \int_{-1}^1 \frac{p_k(z) - p_k(x)}{z - x} w(x)\,dx \qquad (k = 0, 1, 2, \dots),$$

are known to satisfy a three-term recurrence relation of the form

$$y_{k+1} = (z - a_k)y_k - b_k y_{k-1} \qquad (k = 0, 1, 2, \dots), \tag{3.6}$$

where

$$y_{-1} = 0,\ y_0 = 1 \ \text{ for } \{p_k\} \qquad \text{and} \qquad y_{-1} = -1,\ y_0 = 0 \ \text{ for } \{q_k\}. \tag{3.7}$$

Denote by m_k and μ_k the moments associated with the inner products (3.1) and (3.2), respectively,

$$\mu_k = (z^k, 1), \qquad m_k = [x^k, 1], \qquad k \geq 0,$$

where, in view of (3.7), $b_0 = m_0$.

Gautschi, Landau and Milovanović [13] proved the following result:

Theorem 3.1. *Let w be a weight function, positive on $(-1, 1)$, holomorphic in $D_+ = \{z \in \mathbb{C} \mid |z| < 1,\ \operatorname{Im} z > 0\}$, and such that (3.4) is satisfied and the integrals in (3.5) exist (possibly) as improper integrals. Assume in addition that*

$$\operatorname{Re}(1, 1) = \operatorname{Re} \int_0^\pi w(e^{i\theta})\, d\theta \neq 0.$$

Then there exists a unique system of (monic, complex) orthogonal polynomials $\{\pi_k\}$ relative to the inner product (3.1). Denoting by $\{p_k\}$ the (monic, real) orthogonal polynomials relative to the inner product (3.2), we have

$$\pi_k(z) = p_k(z) - i\theta_{k-1} p_{k-1}(z) \qquad (k = 0, 1, 2, \dots), \tag{3.8}$$

where

$$\theta_{k-1} = \frac{\mu_0 p_k(0) + i q_k(0)}{i\mu_0 p_{k-1}(0) - q_{k-1}(0)} \qquad (k = 0, 1, 2, \dots). \tag{3.9}$$

Alternatively,

$$\theta_k = i a_k + \frac{b_k}{\theta_{k-1}} \quad (k = 0, 1, 2, \dots); \qquad \theta_{-1} = \mu_0, \tag{3.10}$$

where a_k, b_k are the recursion coefficients in (3.6) and $\mu_0 = (1, 1)$. In particular, all θ_k are real (in fact, positive) if $a_k = 0$ for all $k \geq 0$. Finally,

$$(\pi_k, \pi_k) = \theta_{k-1}[p_{k-1}, p_{k-1}] \neq 0 \quad (k = 1, 2, \dots), \qquad (\pi_0, \pi_0) = \mu_0. \tag{3.11}$$

As we can see, relation (3.8), with (3.9), gives a connection between orthogonal polynomials on the semicircle and the standard polynomials orthogonal on $[-1,1]$ with respect to the same weight function w. The norms of these polynomials are in relation (3.11).

In the sequel we assume that condition (3.3) is satisfied, so that the orthogonal polynomials $\{\pi_k\}$ exist. Since $(zf,g)=(f,zg)$, it is known that they must satisfy a three-term recurrence relation

$$\begin{aligned}&\pi_{k+1}(z)=(z-i\alpha_k)\pi_k(z)-\beta_k\pi_{k-1}(z), \qquad k=0,1,2,\dots,\\ &\pi_{-1}(z)=0, \quad \pi_0(z)=1.\end{aligned} \tag{3.12}$$

Using the representation (3.8), we can find a connection between the coefficients in (3.12) and the corresponding coefficients in the three-term recurrence relation (3.6) for polynomials $\{p_k\}$ (see [13]):

Theorem 3.2. *Under the assumption* (3.3), *the (monic, complex) polynomials* $\{\pi_k\}$ *orthogonal with respect to the inner product* (3.1) *satisfy the recurrence relation* (3.12), *where the coefficients* α_k, β_k *are given by*

$$\alpha_k=\theta_k-\theta_{k-1}-ia_k, \qquad \beta_k=\frac{\theta_{k-1}}{\theta_{k-2}}b_{k-1}=\theta_{k-1}(\theta_{k-1}-ia_{k-1}),$$

for $k\ge 1$ *and* $\alpha_0=\theta_0-ia_0$, *with the* θ_k *defined in Theorem* 3.1.

Alternatively, the coefficients α_k can be expressed in the form

$$\alpha_k=-\theta_{k-1}+\frac{b_k}{\theta_{k-1}}, \quad k\ge 1, \qquad \alpha_0=\frac{b_0}{\theta_{-1}}=\frac{m_0}{\mu_0}.$$

It is interesting to consider the zero distribution of polynomials $\pi_n(z)$. From (3.12) it follows that the zeros of $\pi_n(z)$ are the eigenvalues of the (complex, tridiagonal) matrix

$$J_n=\begin{bmatrix} i\alpha_0 & 1 & & & O\\ \beta_1 & i\alpha_1 & 1 & & \\ & \beta_2 & i\alpha_2 & \ddots & \\ & & \ddots & \ddots & 1\\ O & & & \beta_{n-1} & i\alpha_{n-1}\end{bmatrix}, \tag{3.13}$$

where α_k and β_k are given in Theorem 3.2.

If the weight w is symmetric, i.e.,

$$w(-z)=w(z), \qquad w(0)>0, \tag{3.14}$$

then $\mu_0 = (1, 1) = \pi w(0) > 0$, $a_k = 0$, $\theta_k > 0$, for all $k \geq 0$, and

$$\alpha_0 = \theta_0, \quad \alpha_k = \theta_k - \theta_{k-1}, \quad \beta_k = \theta_{k-1}^2, \quad k \geq 1.$$

In that case J_n can be transformed into a real nonsymmetric tridiagonal matrix

$$A_n = -iD_n^{-1} J_n D_n = \begin{bmatrix} \alpha_0 & \theta_0 & & & O \\ -\theta_0 & \alpha_1 & \theta_1 & & \\ & -\theta_1 & \alpha_2 & \ddots & \\ & & \ddots & \ddots & \theta_{n-2} \\ O & & & -\theta_{n-2} & \alpha_{n-1} \end{bmatrix},$$

where $D_n = \operatorname{diag}(1, i\theta_0, i^2\theta_0\theta_1, i^3\theta_0\theta_1\theta_2, \dots) \in \mathbb{C}^{n \times n}$. The eigenvalues η_ν, $\nu = 1, \dots, n$, of A_n can be calculated using the EISPACK subroutine HQR (see [39]). Then all the zeros ζ_ν, $\nu = 1, \dots, n$, of $\pi_n(z)$ are given by $\zeta_\nu = i\eta_\nu$, $\nu = 1, \dots, n$.

In [13] we proved the following result for a symmetric weight (3.14):

Theorem 3.3. *All zeros of π_n are located symmetrically with respect to the imaginary axis and contained in $D_+ = \{z \in \mathbb{C} \mid |z| < 1, \ \operatorname{Im} z > 0\}$, with the possible exception of a single (simple) zero on the positive imaginary axis.*

If we define the half strip $S_+ = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0, \ -\xi_n \leq \operatorname{Re} z \leq \xi_n\}$, where ξ_n is the largest zero of the real polynomial p_n, then we can prove that all zeros of π_n are also in S_+ (see [13] and [15]). Thus, all zeros are contained in $D_+ \cap S_+$.

For the Gegenbauer weight $w(z) = (1 - z^2)^{\lambda - 1/2}$, $\lambda > -1/2$, the exceptional case from Theorem 3.3 can only arise if $n = 1$ and $-1/2 < \lambda \leq 0$. Likewise, no exceptional cases seem to occur for Jacobi weights $w(z) = (1 - z)^\alpha (1 + z)^\beta$, $\alpha, \beta > -1$, if $n \geq 2$, as was observed by several numerical computations (see [13]). However, in a general case, Gautschi [11] exhibited symmetric functions w for which $\pi_n(\,\cdot\,; w)$, for arbitrary fixed n, has a zero iy with $y \geq 1$.

Some applications of these polynomials in numerical integration and numerical differentiation can be found in [8], [26–28].

4. Orthogonality on a circular arc

A generalization of polynomials orthogonal on the semicircle was given by M.G. de Bruin [7] for the circular arc

$$\Gamma_R = \{z \in \mathbb{C} \mid z = -iR + e^{i\theta}\sqrt{R^2 + 1}, \ \varphi \leq \theta \leq \pi - \varphi, \ \tan\varphi = R\}.$$

He considered the polynomials $\{\pi_k^R\}$ orthogonal on Γ_R with respect to the complex inner product

$$(4.1)\qquad (f,g)=\int_{\varphi}^{\pi-\varphi} f_1(\theta)g_1(\theta)w_1(\theta)\,d\theta,$$

where $\varphi\in(0,\pi/2)$, and for $f(z)$ the function $f_1(\theta)$ is defined by

$$f_1(\theta)=f\big(-iR+e^{i\theta}\sqrt{R^2+1}\big),\qquad R=\tan\varphi.$$

Alternatively, the inner product (4.1) can be expressed in the form

$$(4.2)\qquad (f,g)=\int_{\Gamma_R} f(z)g(z)w(z)(iz-R)^{-1}\,dz.$$

Under suitable integrability conditions on the weight function w, which is positive on $(-1,1)$ and is holomorphic in the moon-shaped region

$$M_+=\left\{z\in\mathbb{C}\ \middle|\ |z+iR|<\sqrt{R^2+1},\ \operatorname{Im} z>0\right\},$$

where $R>0$, the polynomials $\{\pi_k^R\}$ orthogonal on the circular arc Γ_R with respect to the complex inner product (4.1) always exist and have similar properties like polynomials orthogonal on the semicircle.

For $R=0$ the arc Γ_R reduces to the semicircle Γ, and polynomials $\{\pi_k^R\}$ to $\{\pi_k\}$. It is easy to prove that the condition

$$\operatorname{Re}\int_{\Gamma_R} w(z)(iz-R)^{-1}\,dz=\operatorname{Re}\int_{\varphi}^{\pi-\varphi} w_1(\theta)\,d\theta\neq 0$$

is automatically satisfied for $R>0$ in contrast to the case $R=0$ (see condition (3.3)).

Quite analogous results to Theorems 3.1–3.4 were proved by de Bruin [7]. For example, for polynomials $\{\pi_k\}$ (the upper index R is omitted) equalities (3.8) and (3.11), as well as the three-term recurrence relation (3.12) hold, where now the θ_k is given by

$$\theta_k=-R+ia_k+\frac{b_k}{\theta_{k-1}}\quad (k=0,1,2,\dots);\qquad \theta_{-1}=\mu_0,$$

instead of (3.10). Also, for the symmetric weight, $w(z)=w(-z)$, all zeros of π_n are contained in M_+ with the possible exception of just one simple zero situated on the positive imaginary axis.

Let $\{\pi_n\}$ be the set of polynomials orthogonal on the circular arc Γ_R, with respect to the inner product (4.1), i.e., (4.2). Milovanović and Rajković [33] introduced the polynomials $\{\pi_n^*\}$ orthogonal on the symmetric down circular arc Γ_R^* with respect to the inner product defined by

$$(f,g)^* = \int_{\Gamma_R^*} f(z)g(z)w(z)(iz+R)^{-1}\,dz, \tag{4.3}$$

where $\Gamma_R^* = \{z \in \mathbb{C}\,|\, z = iR + e^{-i\theta}\sqrt{R^2+1},\ \varphi \le \theta \le \pi - \varphi,\ \tan\varphi = R\}$. Such polynomials are called *dual orthogonal polynomials* with respect to polynomials $\{\pi_n\}$.

Let M be a lentil-shaped region with the boundary $\partial M = \Gamma_R \cup \Gamma_R^*$, i.e.,

$$M = \{z \in \mathbb{C}\,|\,|z \pm iR| < \sqrt{R^2+1}\},$$

where $R > 0$.

We assume that w is a weight function, positive on $(-1,1)$, holomorphic in M, and such that the integrals in (4.2), (4.3), and (3.2) exist for smooth functions f and g (possibly) as improper integrals. Under the same additional conditions on w and f, like previous, we have

$$0 = \int_\Gamma f(z)w(z)\,dz + \int_{-1}^{1} f(x)w(x)\,dx,$$

where $\Gamma = \Gamma_R$ or Γ_R^*. Then both systems of the orthogonal polynomials $\{\pi_n\}$ and $\{\pi_n^*\}$ exist uniquely.

The inner products in (4.2) and (4.3) define the moment functionals

$$\mathcal{L}z^k = \mu_k, \qquad \mu_k = (z^k,1) = \int_{\Gamma_R} z^k w(z)(iz-R)^{-1}\,dz$$

and

$$\mathcal{L}^* z^k = \mu_k^*, \qquad \mu_k^* = (z^k,1)^* = \int_{\Gamma_R^*} z^k w(z)(iz+R)^{-1}\,dz,$$

respectively. Using the moment determinants, we can express the (monic) polynomials π_k and π_k^* as

$$\pi_k(z) = \frac{1}{\Delta_k}\begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & & \mu_{k+1} \\ \vdots & & & \\ \mu_{k-1} & \mu_k & & \mu_{2k-1} \\ 1 & z & & z^k \end{vmatrix}$$

and

$$\pi_k^*(z) = \frac{1}{\Delta_k^*}\begin{vmatrix} \mu_0^* & \mu_1^* & \cdots & \mu_k^* \\ \mu_1^* & \mu_2^* & & \mu_{k+1}^* \\ \vdots & & & \\ \mu_{k-1}^* & \mu_k^* & & \mu_{2k-1}^* \\ 1 & z & & z^k \end{vmatrix},$$

where

$$\Delta_k = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_{k-1} \\ \mu_1 & \mu_2 & & \mu_k \\ \vdots & & & \\ \mu_{k-1} & \mu_k & & \mu_{2k-2} \end{vmatrix}, \qquad \Delta_k^* = \begin{vmatrix} \mu_0^* & \mu_1^* & \cdots & \mu_{k-1}^* \\ \mu_1^* & \mu_2^* & & \mu_k^* \\ \vdots & & & \\ \mu_{k-1}^* & \mu_k^* & & \mu_{2k-2}^* \end{vmatrix}.$$

We can prove that $\pi_k^*(\overline{z}) = \overline{\pi_k(z)}$, as well as the relation

$$\pi_k^*(z) = p_k(z) - i\theta_{k-1}^* p_{k-1}(z), \qquad k = 0, 1, 2, \ldots,$$

where

$$\theta_{k-1}^* = \frac{(\pi_k^*, \pi_k^*)^*}{[p_{k-1}, p_{k-1}]}, \quad k = 1, 2, \ldots, \qquad \theta_{-1}^* = \mu_0^*.$$

Here, $\theta_{k-1}^* = -\overline{\theta}_{k-1}$, where θ_{k-1} is the corresponding coefficient in the polynomial π_k.

Also, the following theorem holds:

Theorem 4.1. *The dual (monic) orthogonal polynomials $\{\pi_k^*\}$ satisfy the three-term recurrence relation*

$$\begin{gathered} \pi_{k+1}^*(z) = (z - i\alpha_k^*)\pi_k^*(z) - \beta_k^*\pi_{k-1}^*(z), \qquad k = 0, 1, 2, \ldots, \\ \pi_{-1}^*(z) = 0, \quad \pi_0^*(z) = 1, \end{gathered}$$

with $\alpha_k^* = -\overline{\alpha}_k$ and $\beta_k^* = \overline{\beta}_k$, where α_k and β_k are the coefficients in the corresponding recurrence relation for the polynomials $\{\pi_k\}$.

Using dual polynomials we can give a very short proof that $\theta_{k-1} > 0$ $(k \geq 0)$ for a symmetric weight $w(z) = w(-z)$. Namely, since $(\pi_k, \pi_k) = \theta_{k-1}[p_{k-1}, p_{k-1}]$ it is enough to prove that $(\pi_k, \pi_k) > 0$. In this *symmetric case*, θ_{k-1} is real and we have $\theta_{k-1}^* = -\theta_{k-1}$ and

$$(\pi_k, \pi_k) = (\pi_k, \pi_k^*) = \int_{\Gamma_R} G(z)w(z)(iz - R)^{-1}\, dz = -\int_{-1}^{1} G(x)\frac{w(x)}{ix - R}\, dx,$$

where $G(z) = p_k(z)^2 + \theta_{k-1}^2 p_{k-1}(z)^2$. Then

$$(\pi_k, \pi_k) = R \int_{-1}^{1} G(x) \frac{w(x)}{R^2 + x^2}\, dx + i \int_{-1}^{1} xG(x) \frac{w(x)}{R^2 + x^2}\, dx.$$

Since $x \mapsto G(x)$ is an even positive function, the second integral on the right-hand side vanishes and $(\pi_k, \pi_k) > 0$.

One complicated proof of the previous result was given in [7].

5. Geronimus' version of orthogonality

In the paper [21], J. W. Jayne considered the Geronimus' concept of orthogonality for recursively generated polynomials. Ya. L. Geronimus proved that a sequence of polynomials $\{p_k\}$, which is orthogonal on a finite interval on real line, is also orthogonal in the sense that there is a weight function $z \mapsto \chi(z)$ having one or more singularities inside a simple curve C and such that

$$(5.1) \qquad \langle p_k, p_m \rangle = \frac{1}{2\pi i} \oint_C p_k(z) p_m(z) \chi(z)\, dz = \begin{cases} 0, & k \neq m, \\ h_m, & k = m. \end{cases}$$

Following Geronimus [19] and Jayne [21], Milovanović and Rajković [32] determined such a complex weight function $z \mapsto \chi(z)$, for (monic) polynomials $\{\pi_k\}$ orthogonal on the semicircle Γ, and also for the corresponding polynomials $\{\pi_k^R\}$ orthogonal on the circular arc Γ_R $(R > 0)$.

Denoting by C any positively oriented simple closed contour surrounding some circle $|z| = r > 1$, we assume that

$$(5.2) \qquad \chi(z) = \sum_{k=1}^{\infty} \omega_k z^{-k}, \quad \omega_1 = 1,$$

for $|z| > 1$, and express z^n as a linear combination of the monic polynomials π_m, $m = 0, 1, \dots, n$, which are orthogonal on the semicircle Γ, with respect to the inner product (3.1). Thus,

$$(5.3) \qquad z^n = \sum_{m=0}^{n} \gamma_{n,m} \pi_m(z),$$

where $(z^n, \pi_m) = \gamma_{n,m}(\pi_m, \pi_m)$, $m = 0, 1, \dots, n$. Using the inner product (5.1) and the representation (5.2), we obtain

$$\langle z^n, 1 \rangle = \frac{1}{2\pi i} \oint_C z^n \chi(z)\, dz = \frac{1}{2\pi i} \oint_C \sum_{k=1}^{\infty} \omega_k z^{n-k}\, dz = \omega_{n+1}.$$

On the other hand, because of (5.3) and the orthogonality condition (5.1), we find

$$\langle z^n, 1\rangle = \langle \sum_{m=0}^{n} \gamma_{n,m}\pi_m(z), 1\rangle = \sum_{m=0}^{n} \gamma_{n,m}\langle \pi_m, 1\rangle,$$

i.e., $\langle z^n, 1\rangle = \gamma_{n,0}\langle \pi_0, \pi_0\rangle = \gamma_{n,0}h_0$. Thus, we have $w_{n+1} = \gamma_{n,0}h_0 = \gamma_{n,0}$, because $h_0 = \omega_1 = 1$.

Finally, using the moments $\mu_n = (z^n, 1)$, we obtain $\omega_{n+1} = \mu_n/\mu_0$, $n \geq 0$, and

$$\chi(z) = \frac{1}{\mu_0}\sum_{k=1}^{\infty} \mu_{k-1}z^{-k}, \qquad |z| > 1, \tag{5.4}$$

where we need the convergence of this series for $|z| > r > 1$.

Suppose that w be a weight function, nonnegative on $(-1, 1)$, holomorphic in $D_+ = \{z \in \mathbb{C} \,|\, |z| < 1, \operatorname{Im} z > 0\}$, integrable over ∂D_+, and such that (3.3) is satisfied. Then the moments μ_k can be expressed in the form

$$\mu_0 = \int_\Gamma w(z)(iz)^{-1}\,dz = \frac{1}{i}\left(i\pi w(0) - \text{ v.p. }\int_{-1}^{1} \frac{w(x)}{x}\,dx\right)$$

and

$$\mu_k = \int_\Gamma z^k w(z)(iz)^{-1}\,dz = i\int_{-1}^{1} x^{k-1}w(x)\,dx, \quad k \geq 1.$$

These moments are included in the series (5.4).

Supposing that the weight function w has such moments μ_k, which provide the convergence of the series (5.4), for all z outside some circle $|z| = r > 1$ lying interior to C, Milovanović and Rajković [32] proved:

Theorem 5.1. *The monic polynomials* $\{\pi_k\}$, *which are orthogonal on the semicircle* Γ *with respect to the inner product* (3.1), *are also orthogonal in the sense of* (5.1), *where*

$$\chi(z) = \frac{1}{z}\left(1 + \frac{i}{\mu_0}\int_{-1}^{1} \frac{w(x)}{z - x}\,dx\right), \qquad |z| > r > 1,$$

and

$$\mu_0 = \pi w(0) + i\,\text{v.p.}\int_{-1}^{1} \frac{w(x)}{x}\,dx.$$

In Gegenbauer case they obtained the following result:

Corollary 5.2. *Let $w(z) = (1-z^2)^{\lambda-1/2}$, $\lambda > -1/2$. The monic polynomials $\{\pi_k\}$, which are orthogonal on the unit semicircle with respect to the inner product* (3.1), *are also orthogonal in the sense of* (5.1), *where*

$$\chi(z) = \frac{1}{z} + \frac{i}{\sqrt{\pi}\, z^2} \cdot \frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(\lambda+1)}\, F\Big(1, \frac{1}{2}, \lambda+1; \frac{1}{z^2}\Big),$$

where F is the Gauss hypergeometric series and Γ is the gamma function.

In Legendre case ($\lambda = 1/2$) we have

$$\chi(z) = \frac{1}{z} + \frac{i}{\pi z} \log \frac{z+1}{z-1},$$

where the interval from -1 to 1 on the real axis is considered as a branch cut.

The corresponding complex weight for polynomials $\{\pi_k^R\}$ ($R > 0$) orthogonal on the circular arc Γ_R was also derived in [32] in the form

$$\chi(z) = \frac{1}{\mu_0} \int_{-1}^{1} \frac{(R+ix)w(x)}{(R^2+x^2)(z-x)}\, dx, \qquad |z| > r > 1,$$

where

$$\mu_0 = \int_{-1}^{1} \frac{R+ix}{R^2+x^2}\, w(x)\, dx.$$

6. Orthogonality on the radial rays in the complex plane

In this section we start with a new type of nonstandard orthogonality on some radial rays in the complex plane. Suppose that we have M points in the complex plane, $z_s = a_s e^{i\varphi_s} \in \mathbb{C}$, $s = 0, 1, \ldots, M-1$, with different arguments φ_s. Some of a_s (or all) can be ∞. The case $M = 5$ is shown in Fig. 6.1. We can define an inner product on these radial rays ℓ_s in the complex plane which connect the origin $z = 0$ and the points z_s, $s = 0, 1, \ldots, M-1$. Namely,

$$(f, g) = \sum_{s=0}^{M-1} e^{-i\varphi_s} \int_{\ell_s} f(z)\overline{g(z)}\, |w(z)|\, dz,$$

where $z \mapsto w(z)$ is a suitable function (complex weight).

Since, this product can be expressed in the form

$$(f, g) = \sum_{s=0}^{M-1} \int_0^{a_s} f(xe^{i\varphi_s})\overline{g(xe^{i\varphi_s})}\, \big|w(xe^{i\varphi_s})\big|\, dx,$$

we see that

$$(f, f) = \sum_{s=0}^{M-1} \int_0^{a_s} |f(xe^{i\varphi_s})|^2 \, |w(xe^{i\varphi_s})| \, dx > 0,$$

except when $f(z) = 0$.

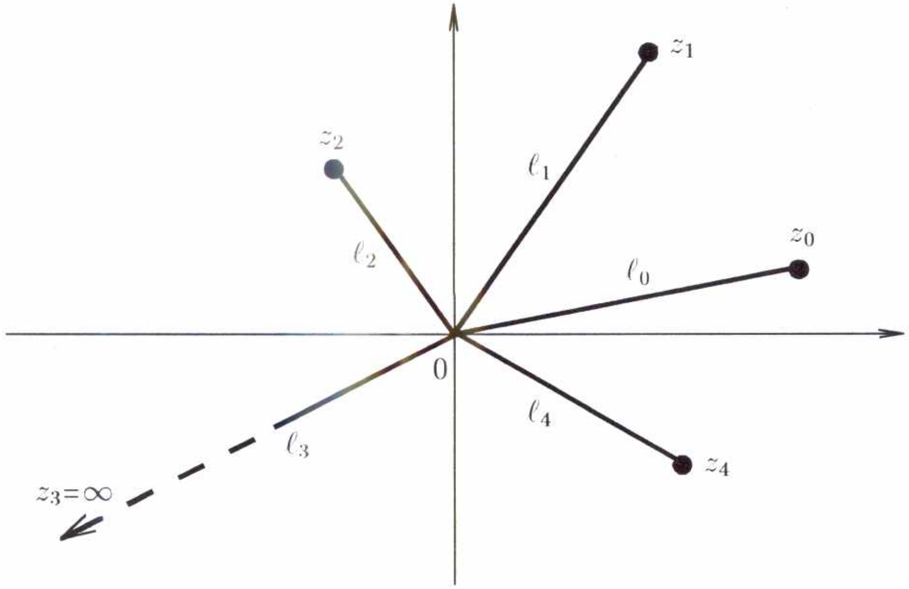


Fig. 6.1

We will consider here only the case when M is an even number and $\varphi_s = \pi s/m$, $s = 0, 1, \dots, 2m-1$. Thus, let $m \in \mathbb{N}$ and $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2m-1}$ be $(2m)$th roots of unity, i.e., $\varepsilon_s = \exp(i\pi s/m)$, $s = 0, 1, \dots, 2m-1$. We will study orthogonal polynomials relative to the inner product

$$(6.1) \qquad (f, g) = \sum_{s=0}^{2m-1} \varepsilon_s^{-1} \int_{\ell_s} f(z)\overline{g(\bar{z})}|w(z)| \, dz.$$

Suppose that $a_s = 1$ for each s and let $z \mapsto w(z)$ be a holomorphic function such that

$$|w(x\varepsilon_s)| = w(x), \quad s = 0, 1, \dots, 2m-1,$$

and $x \mapsto w(x)$ be a weight function on $(0, 1)$ (nonnegative on $(0, 1)$ and $\int_0^1 w(x)\, dx > 0$). Then, (6.1) can be written in the form

$$(6.2) \qquad (f, g) = \int_0^1 \left(\sum_{s=0}^{2m-1} f(x\varepsilon_s)\overline{g(x\varepsilon_s)} \right) w(x) \, dx.$$

In the case $m = 1$, (6.2) becomes

$$(f,g) = \int_{-1}^{1} f(x)\overline{g(x)}w(x)\,dx,$$

so we have the standard case of polynomials orthogonal on $(-1,1)$ with respect to the weight function $x \mapsto w(x)$.

The inner product (6.2) has the following property:

Lemma 6.1. $(z^m f, g) = (f, z^m g)$.

Proof. Since $\varepsilon_s^m = \varepsilon_s^{-m} = (-1)^s$ we have

$$\begin{aligned}(z^m f, g) &= \int_0^1 \biggl(\sum_{s=0}^{2m-1} x^m\varepsilon_s^m f(x\varepsilon_s)\overline{g(x\varepsilon_s)}\biggr) w(x)\,dx\\ &= \int_0^1 \biggl(\sum_{s=0}^{2m-1} f(x\varepsilon_s)\overline{x^m\varepsilon_s^m g(x\varepsilon_s)}\biggr) w(x)\,dx\\ &= (f, z^m g). \quad \square\end{aligned}$$

The moments are given by

$$\mu_{p,q} = (z^p, z^q) = \biggl(\sum_{s=0}^{2m-1} \varepsilon_s^{p-q}\biggr) \int_0^1 x^{p+q} w(x)\,dx, \quad p,q \ge 0. \tag{6.3}$$

If $p = 2mn + \nu$, $n = [p/(2m)]$, and $0 \le \nu \le 2m-1$, it is easy to verify that

$$\sum_{s=0}^{2m-1} \varepsilon_s^p = \sum_{s=0}^{2m-1} \varepsilon_s^\nu = \begin{cases} 2m & \text{if } \nu = 0,\\ 0 & \text{if } 1 \le \nu \le 2m-1.\end{cases}$$

Thus, $\mu_{p,q}$ in (6.3) is different from zero only if $p \equiv q \pmod{2m}$; otherwise $\mu_{p,q} = 0$. Using the moment determinants

$$\Delta_0 = 1, \quad \Delta_N = \begin{vmatrix} \mu_{00} & \mu_{10} & \cdots & \mu_{N-1,0}\\ \mu_{01} & \mu_{11} & \cdots & \mu_{N-1,1}\\ \vdots & & & \\ \mu_{0,N-1} & \mu_{1,N-1} & \cdots & \mu_{N-1,N-1}\end{vmatrix}, \quad N \ge 1,$$

we can prove the following existence result for the (monic) orthogonal polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$ with respect to the inner product (6.2) (see Milovanović [30]):

Theorem 6.2. *If $\Delta_N > 0$ for all $N \geq 1$ the monic polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$, orthogonal with respect to the inner product (6.2), exist uniquely.*

It is well known that an orthogonal sequence of polynomials satisfies a three-term recurrence relation if the inner product has the property $(zf, g) = (f, zg)$. In our case the corresponding property is given by $(z^m f, g) = (f, z^m g)$ (see Lemma 6.1) and the following result holds:

Theorem 6.3. *Let the inner product $(\cdot,\cdot)$ be given by (6.2) and let the corresponding system of monic orthogonal polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$ exist. They satisfy the recurrence relation*

$$(6.4)\qquad \begin{gathered}\pi_{N+m}(z) = z^m \pi_N(z) - b_N \pi_{N-m}(z), \quad N \geq m,\\ \pi_N(z) = z^N, \quad N = 0, 1, \ldots, 2m-1,\end{gathered}$$

where

$$(6.5)\qquad b_N = \frac{(\pi_N, z^m \pi_{N-m})}{(\pi_{N-m}, \pi_{N-m})} = \frac{\|\pi_N\|^2}{\|\pi_{N-m}\|^2}.$$

In a simple case when $m = 2$ and $w(x) = 1$, i.e., when the inner product $(\cdot,\cdot)$ is given by

$$(6.6)\qquad (f, g) = \int_0^1 \left[f(x)\overline{g(x)} + f(ix)\overline{g(ix)} + f(-x)\overline{g(-x)} + f(-ix)\overline{g(-ix)} \right] dx,$$

we can calculate directly the coefficient b_N in the recurrence relation (6.4). The moments are given by

$$\mu_{p,q} = (z^p, z^q) = \begin{cases} \dfrac{4}{p+q+1}, & p \equiv q \pmod 4, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if $p = 4i + \nu$ and $q = 4j + \nu$, $\nu \in \{0, 1, 2, 3\}$, we have

$$\mu_{4i+\nu, 4j+\nu} = \frac{4}{4(i+j) + 2\nu + 1}, \qquad i, j \geq 0.$$

Our purpose is to evaluate the moment determinants

$$\Delta_N = \begin{vmatrix} \mu_{00} & \mu_{10} & \cdots & \mu_{N-1,0} \\ \mu_{01} & \mu_{11} & \cdots & \mu_{N-1,1} \\ \vdots & & & \\ \mu_{0,N-1} & \mu_{1,N-1} & \cdots & \mu_{N-1,N-1} \end{vmatrix}, \quad N \geq 1.$$

In order to make it, for every $k \in \mathbb{N}$, we define the determinants

$$C_k = \begin{vmatrix} \mu_{00} & 0 & \mu_{40} & 0 & \cdots & \\ 0 & \mu_{22} & 0 & \mu_{62} & & \\ \mu_{04} & 0 & \mu_{44} & 0 & & \\ 0 & \mu_{26} & 0 & \mu_{66} & & \\ \vdots & & & & \ddots & \\ & & & & & \mu_{2k-2,2k-2} \end{vmatrix},$$

$$D_k = \begin{vmatrix} \mu_{11} & 0 & \mu_{51} & 0 & \cdots & \\ 0 & \mu_{33} & 0 & \mu_{73} & & \\ \mu_{15} & 0 & \mu_{55} & 0 & & \\ 0 & \mu_{37} & 0 & \mu_{77} & & \\ \vdots & & & & \ddots & \\ & & & & & \mu_{2k-1,2k-1} \end{vmatrix},$$

which can be expressed in terms of the determinants $E_0^{(\nu)} = 1$ and

$$E_n^{(\nu)} = \begin{vmatrix} \mu_{\nu,\nu} & \mu_{4+\nu,\nu} & \cdots & \mu_{4(n-1)+\nu,\nu} \\ \mu_{\nu,4+\nu} & \mu_{4+\nu,4+\nu} & \cdots & \mu_{4(n-1)+\nu,4+\nu} \\ \vdots & & & \\ \mu_{\nu,4(n-1)+\nu} & \mu_{4+\nu,4(n-1)+\nu} & \cdots & \mu_{4(n-1)+\nu,4(n-1)+\nu} \end{vmatrix},$$

where $\nu = 0, 1, 2, 3$.

Interpreting these determinants in terms of Hilbert-type determinants and using Cauchy's formula (see Muir [34, p. 345])

$$\det \left[\frac{1}{a_i + b_j} \right]_{i,j=1}^{n} = \frac{\prod\limits_{i>j=1}^{n} (a_i - a_j)(b_i - b_j)}{\prod\limits_{i,j=1}^{n} (a_i + b_j)}$$

with $a_i = 4i$ and $b_j = 4j + 2\nu - 7$, we obtain (see [30])

$$E_n^{(\nu)} = 4^{n^2} \frac{(0!1!\cdots(n-1)!)^2}{\prod\limits_{i=0}^{n-1} \prod\limits_{j=0}^{n-1} (4i + 4j + 2\nu + 1)}, \quad n \geq 1.$$

Also, we can prove that

$$C_k = E_{k/2}^{(0)} E_{k/2}^{(2)}, \quad k(\text{even}) \geq 2; \quad C_k = E_{(k+1)/2}^{(0)} E_{(k-1)/2}^{(2)}, \quad k(\text{odd}) \geq 1,$$

as well as

$$D_k = E^{(1)}_{k/2}E^{(3)}_{k/2},\quad k(\text{even}) \geq 2;\qquad D_k = E^{(1)}_{(k+1)/2}E^{(3)}_{(k-1)/2},\quad k(\text{odd}) \geq 1.$$

Using the same techniques we find that

$$\Delta_{2k} = C_k D_k \quad \text{and} \quad \Delta_{2k+1} = C_{k+1} D_k.$$

Combining these equalities we obtain:

Lemma 6.4. *We have*

$$\begin{aligned}\Delta_{4n} &= E^{(0)}_n E^{(1)}_n E^{(2)}_n E^{(3)}_n,\\ \Delta_{4n+1} &= E^{(0)}_{n+1} E^{(1)}_n E^{(2)}_n E^{(3)}_n,\\ \Delta_{4n+2} &= E^{(0)}_{n+1} E^{(1)}_{n+1} E^{(2)}_n E^{(3)}_n,\\ \Delta_{4n+3} &= E^{(0)}_{n+1} E^{(1)}_{n+1} E^{(2)}_{n+1} E^{(3)}_n.\end{aligned}$$

We note, first of all, that $\Delta_N > 0$ for all $N \geq 1$, and therefore, the orthogonal polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$ with respect to the inner product (6.6) exist uniquely, and

$$(\pi_N, \pi_N) = \|\pi_N\|^2 = \frac{\Delta_{N+1}}{\Delta_N} > 0.$$

Theorem 6.5. *The (monic) polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$, orthogonal with respect to the inner product (6.6), satisfy the recurrence relation*

$$\begin{aligned}&\pi_{N+2}(z) = z^2\pi_N(z) - b_N\pi_{N-2}(z),\quad N \geq 2,\\ &\pi_N(z) = z^N,\quad N = 0, 1, 2, 3,\end{aligned}\tag{6.7}$$

where

$$b_{4n+\nu} = \begin{cases} \dfrac{16n^2}{(8n+2\nu-3)(8n+2\nu+1)} & \text{if } \nu = 0, 1,\\[2ex] \dfrac{(4n+2\nu-3)^2}{(8n+2\nu-3)(8n+2\nu+1)} & \text{if } \nu = 2, 3.\end{cases}\tag{6.8}$$

Proof. Because of (6.5), the coefficients b_N can be expressed in the form

$$b_N = \frac{\|\pi_N\|^2}{\|\pi_{N-2}\|^2} = \frac{\Delta_{N+1}}{\Delta_N}\cdot\frac{\Delta_{N-2}}{\Delta_{N-1}},\quad N \geq 2.$$

In order to find these quotients we need a quotient of the determinants $E_n^{(\nu)}$. According to the previous equalities we get

$$\frac{E_{n+1}^{(\nu)}}{E_n^{(\nu)}} = \frac{4}{8n+2\nu+1}\left(\prod_{k=n}^{2n-1}\frac{4(k-n+1)}{4k+2\nu+1}\right)^2, \quad n \geq 1,$$

and $E_1^{(\nu)}/E_0^{(\nu)} = 4/(2\nu+1)$.

Then, for $\nu = 0, 1$ we find

$$b_{4n+\nu} = \frac{\Delta_{4n+\nu+1}/\Delta_{4n+\nu}}{\Delta_{4(n-1)+\nu+3}/\Delta_{4(n-1)+\nu+2}} = \frac{E_{n+1}^{(\nu)}/E_n^{(\nu)}}{E_n^{(\nu+2)}/E_{n-1}^{(\nu+2)}},$$

i.e.,

$$b_{4n+\nu} = \frac{16n^2}{(8n+2\nu-3)(8n+2\nu+1)}.$$

Similarly, for $\nu = 2, 3$, we have

$$b_{4n+\nu} = \frac{\Delta_{4n+\nu+1}/\Delta_{4n+\nu}}{\Delta_{4n+\nu-1}/\Delta_{4n+\nu-2}} = \frac{E_{n+1}^{(\nu)}/E_n^{(\nu)}}{E_{n+1}^{(\nu-2)}/E_n^{(\nu-2)}},$$

i.e.,

$$b_{4n+\nu} = \frac{(4n+2\nu-3)^2}{(8n+2\nu-3)(8n+2\nu+1)}. \quad \square$$

From (6.8) we conclude that

$$b_N \to \frac{1}{4} \quad \text{as } N \to +\infty,$$

just like in Szegő's theory for orthogonal polynomials on the interval $(-1, 1)$.

Since

$$\|\pi_N\|^2 = \begin{cases} b_N b_{N-2} \cdots b_2 \|\pi_0\|^2, & N \text{ even}, \\ b_N b_{N-2} \cdots b_3 \|\pi_1\|^2, & N \text{ odd}, \end{cases}$$

and $\|\pi_0\|^2 = \Delta_1/\Delta_0 = \mu_{00} = 4$ $(\Delta_0 \equiv 1)$,

$$\|\pi_1\|^2 = \Delta_2/\Delta_1 = \mu_{00}\mu_{11}/\mu_{00} = \mu_{11} = 4/3,$$

we can define $b_0 = 4$, $b_1 = 4/3$, so that (6.7) holds for every $N \geq 0$, where

$$\pi_{-2}(z) = \pi_{-1}(z) = 0, \ \pi_0(z) = 1, \ \pi_1(z) = z.$$

Finally, we can determine the norms of the polynomials $\{\pi_N(z)\}$. Let $N = 4n + \nu$, $n = [N/4]$, $0 \le \nu \le 3$. Since

$$\|\pi_N\|^2 = \frac{\Delta_{N+1}}{\Delta_N} = \frac{\Delta_{4n+\nu+1}}{\Delta_{4n+\nu}} = \frac{E_{n+1}^{(\nu)}}{E_n^{(\nu)}},$$

we have

$$\|\pi_N\|^2 = \frac{4}{2N+1}, \quad 0 \le N \le 3,$$

$$\|\pi_N\|^2 = \|\pi_{4n+\nu}\|^2 = \frac{4}{8n+2\nu+1}\left(\prod_{k=n}^{2n-1} \frac{4(k-n+1)}{4k+2\nu+1}\right)^2, \quad N \ge 4.$$

7. A representation of $\pi_N(z)$ and zeros

In this section we again consider the general case of the inner product (6.2) for which the corresponding system of the monic orthogonal polynomials $\{\pi_N(z)\}_{N=0}^{+\infty}$ exists and satisfies the recurrence relation (6.4). Based on this recurrence relation, we can conclude and easily prove that $\pi_N(z)$ are incomplete polynomials with the following representation (see Milovanović [30]):

Theorem 7.1. *The polynomials from Theorem 6.3 can be expressed in the form*

$$\pi_{2mn+\nu}(z) = z^\nu q_n^{(\nu)}(z^{2m}), \quad \nu = 0, 1, \ldots, 2m-1; \ n = 0, 1, \ldots, \tag{7.1}$$

where $q_n^{(\nu)}(t)$, $\nu = 0, 1, \ldots, 2m-1$, are monic polynomials of exact degree n, which satisfy the three-term recurrence relation

$$\begin{aligned} q_{n+1}^{(\nu)}(t) &= (t - a_n^{(\nu)})q_n^{(\nu)}(t) - b_n^{(\nu)} q_{n-1}^{(\nu)}(t), \quad n = 0, 1, \ldots, \\ q_0^{(\nu)}(t) &= 1, \quad q_{-1}^{(\nu)}(t) = 0. \end{aligned} \tag{7.2}$$

The recursion coefficients $a_n^{(\nu)}$ and $b_n^{(\nu)}$ are given in terms of the b-coefficients as

$$a_n^{(\nu)} = b_N + b_{N+m}, \quad b_n^{(\nu)} = b_{N-m} b_N, \quad N = 2mn + \nu.$$

The three-term recurrence relation (7.2) shows that the monic polynomial systems $\{q_n^{(\nu)}(t)\}_{n=0}^{+\infty}$, $\nu = 0, 2, \ldots, 2m-1$, are orthogonal. The following theorem gives this orthogonality:

Theorem 7.2. *Let $x \mapsto w(x)$ be a weight function in the inner product* (6.2) *which guarantees the existence of the polynomials $\pi_N(z)$, i.e., $q_n^{(\nu)}(t)$, $\nu = 0, 1, \ldots, 2m-1$, determined by* (7.1). *For any $\nu \in \{0, 1, \ldots, 2m-1\}$, the sequence of polynomials $\{q_n^{(\nu)}(t)\}_{n=0}^{+\infty}$ is orthogonal on $(0,1)$ with respect to the weight function $t \mapsto w_\nu(t) = t^{(2\nu+1-2m)/2m} w(t^{1/2m})$.*

As we can see the question of the existence of the polynomials $\pi_N(z)$ is reduced to the existence of polynomials $q_n^{(\nu)}(t)$, orthogonal on $(0,1)$ with respect to the weight function $w_\nu(t)$, for every $\nu = 0, 1, \ldots, 2m-1$.

The next result gives the zero distribution of the polynomials $\pi_N(z)$ (see [30]):

Theorem 7.3. *Let $N = 2mn + \nu$, $n = [N/2m]$, $\nu \in \{0, 1, \ldots, 2m-1\}$. All zeros of the polynomial $\pi_N(z)$ are simple and located symmetrically on the radial rays l_s, $s = 0, 1, \ldots, 2m-1$, with the possible exception of a multiple zero of order ν at the origin $z = 0$.*

At the end we mention that an analogue of the Jacobi polynomials and the corresponding problem with the generalized Laguerre polynomials were treated in [30].

References

[1] N.I. Achieser and M.G. Kreĭn, *On some problems in the moment theory*, GONTI, Har'kov, 1938.

[2] J. Aczél, *Eine Bemerkung über die Charakterisierung der "klassischen" Orthogonalpolynome*, Acta Math. Acad. Sci. Hung. **4** (1953), 315–321.

[3] R.P. Agarwal and G.V. Milovanović, *A characterization of the classical orthogonal polynomials*, Progress in Approximation Theory (P. Nevai and A. Pinkus, eds.), Academic Press, New York, 1991, pp. 1–4.

[4] M. Alfaro and F. Marcellán, *Recent trends in orthogonal polynomials on the unit circle*, IMACS Annals on Computing and Applied Mathematics, Vol. 9: Orthogonal Polynomials and Their Applications (C. Brezinski, L. Gori and A. Ronveaux, eds.), IMACS, Baltzer, Basel, 1991, pp. 3–14.

[5] W.A. Al-Salam, *Characterization theorems for orthogonal polynomials*, Orthogonal Polynomials – Theory and Practice (P. Nevai, ed.), NATO ASI Series, Series C; Mathematical and Physical Sciences, Vol. 294, Kluwer, Dordrecht, 1990, pp. 1–24.

[6] G.E. Andrews and R. Askey, *Classical orthogonal polynomials*, Polynômes Orthogonaux et Applications (C. Brezinski, A. Draux, A.P. Magnus, P. Maroni, A. Ronveaux, eds.), Lect. Notes Math. No. 1171, vol. 1985, Springer Verlag, Berlin – Heidelberg – New York, pp. 36–62.

[7] M.G. de Bruin, *Polynomials orthogonal on a circular arc*, J. Comput. Appl. Math. **31** (1990), 253–266.

[8] F. Calio', M. Frontini and G.V. Milovanović, *Numerical differentiation of analytic functions using quadratures on the semicircle*, Comput. Math. Appl. **22** (1991), 99–106.

[9] T.S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York, 1978.

[10] G. Freud, *Orthogonal Polynomials*, Akadémiai Kiadó/Pergamon, Budapest, 1971.

[11] W. Gautschi, *On the zeros of polynomials orthogonal on the semicircle*, SIAM J. Math. Anal. **20** (1989), 738–743.

[12] W. Gautschi, *Algorithm 726: ORTHPOL – a package of routines for generating orthogonal polynomials and Gauss-type quadrature rules*, ACM Trans. Math. Software **20** (1994), 21–62.

[13] W. Gautschi, H. Landau, and G.V. Milovanović, *Polynomials orthogonal on the semicircle, II*, Constr. Approx. **3** (1987), 389–404.

[14] W. Gautschi and G.V. Milovanović, *Polynomials orthogonal on the semicircle*, Rend. Sem. Mat. Univ. Politec. Torino (Special Functions: Theory and Computation) (1985), 179–185.

[15] W. Gautschi and G.V. Milovanović, *Polynomials orthogonal on the semicircle*, J. Approx. Theory **46** (1986), 230–250.

[16] Ya.L. Geronimus, *Polynomials orthogonal on a circle and their applications*, Zap. Nauč.-issled. Inst. Mat. Mech. HMO **19** (1948), 35–120. (Russian)

[17] Ya.L. Geronimus, *On some properties of generalized orthogonal polynomials*, Mat. Sb. **9 (51): 1** (1941), 121–135. (Russian)

[18] Ya.L. Geronimus, *Polynomials orthogonal on a circle and interval*, Gos. Izd. Fiz.-Mat. Lit., Moscow, 1958. (Russian)

[19] Ya.L. Geronimus, *Orthogonal polynomials*, Amer. Math. Soc. Transl. **108** (1977), 37–130.

[20] A. Guessab and G.V. Milovanović, *Weighted L^2-analogues of Bernstein's inequality and classical orthogonal polynomials*, J. Math. Anal. Appl. **182** (1994), 244-249.

[21] J.W Jayne, *Recursively generated polynomials and Geronimus' version of orthogonality on a contour*, SIAM J. Math. Anal. **19** (1988), 676–686.

[22] R.V. Lascenov, *On a class of orthogonal polynomials*, Učen. Zap. Leningrad. Gos. Ped. Inst. **89** (1953), 191–206. (Russian)

[23] P. Lesky, *Die Charakterisierung der klassischen orthogonalen Polynome durch Sturm-Liouvillesche Differentialgleichungen*, Arch. Rat. Mech. Anal. **10** (1962), 341–351.

[24] F. Marcellán and G. Sansigre, *Orthogonal polynomials on the unit circle: symmetrization and quadratic decomposition*, J. Approx. Theory **65** (1991), 109–119.

[25] G.V. Milovanović, *Numerical Analysis*, Vol. I, Naučna Knjiga, Belgrade, 1985.

[26] G.V. Milovanović, *Some applications of the polynomials orthogonal on the semicircle*, Numerical Methods, (Miskolc, 1986), Colloq. Math. Soc. János Bolyai, Vol. 50, North-Holland, Amsterdam – New York, 1987, pp. 625–634.

[27] G. V. Milovanović, *Complex orthogonality on the semicircle with respect to Gegenbauer weight: Theory and applications*, Topics in Mathematical Analysis (Th.M. Rassias, ed.), World Scientific Publ., Singapore, 1989, pp. 695–722.

[28] G.V. Milovanović, *On polynomials orthogonal on the semicircle and applications*, J. Comput. Appl. Math. **49** (1993), 193–199.

[29] G.V. Milovanović, *Orthogonal polynomial systems and some applications*, Inner Product Spaces (Th.M. Rassias, ed.) (to appear).

[30] G.V. Milovanović, *A class of orthogonal polynomials on the radial rays in the complex plane* (to appear).

[31] G.V. Milovanović, D.S. Mitrinović and Th.M. Rassias, *Topics in Polynomials: Extremal Problems, Inequalities, Zeros*, World Scientific, Singapore – New Jersey – London – Hong Kong, 1994.

[32] G.V. Milovanović and P.M. Rajković, *Geronimus concept of orthogonality for polynomials orthogonal on a circular arc*, Rend. di Matematica, Serie VII (Roma) **10** (1990), 383–390.

[33] G.V. Milovanović and P.M. Rajković, *On polynomials orthogonal on a circular arc*, J. Comput. Appl. Math. **51** (1994), 1–13.

[34] T. Muir, *The Theory of Determinants*, Vol. I, MacMillan & Co., London, 1906.

[35] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions, A case study*, J. Approx. Theory **48** (1986), 3–167.

[36] P. Nevai, *Orthogonal polynomials, measures and recurrence relations on the unit circle*, Trans. Amer. Math. Soc. **300** (1987), 175–189.

[37] I. Smirnov, *Sur la thèorie des polynomes orthogonaux à une variable complèxe*, Ž. Leningrad. Fiz.-Mat. Ob. **2** (1928), 155–179.

[38] I. Smirnov, *Sur les valeurs limites des fonctions regulières à l'intérieur d'un cercle*, Ž. Leningrad. Fiz.-Mat. Ob. **2** (1928), 22–37.

[39] B.T. Smith et al., *Matrix Eigensystem Routines – EISPACK Guide*, Lect. Notes Comp. Sci. Vol. 6, Springer-Verlag, New-York, 1974.

[40] P.K. Suetin, *Classical Orthogonal Polynomials*, Nauka, Moscow, 1976. (Russian)

[41] G. Szegő, *Beiträge zur Theorie der Toeplitzschen Formen*, Math. Z. **6** (1920), 167–202.

[42] G. Szegő, *Beiträge zur Theorie der Toeplitzschen Formen. II*, Math. Z. **9** (1921), 167–190.

[43] G. Szegő, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ., vol. 23, 4th ed., Amer. Math. Soc., Providence, R. I., 1975.

Faculty of Electronic Engineering, Department of Mathematics, P.O. Box 73, 18000 Niš, Yugoslavia

E-mail address: grade@efnis.elfak.ni.ac.yu

FILOMAT (Niš) 9:3 (1995), 543-577
Algebra, Logic & Discrete Mathematics
Niš, April 14-16, 1995.

INVITED LECTURE

ITERATIVE METHODS FOR BOUNDING THE INVERSE OF A MATRIX (A SURVEY)

Miodrag S. Petković

ABSTRACT. The aim of this paper is to give a survey of iterative methods for bounding the inverse of a point or interval matrix. These methods are based on the generalized Schulz's method and developed in interval arithmetic. The interest in bounding roundoff errors in matrix computations has come from the impossibility of exact representation of elements of matrices in those cases when numbers are represented in the computer by strings of bits of finite length or elements were experimentally determined by measurement which leads to the uncertainty in initial data. A posed problem can be usefully solved by interval analysis, a new powerful tool of applied mathematics. A detailed study of the basic inclusion method and its modifications, including the convergence features, conditions for a safe convergence, the monotonicity property, the choice of initial inclusion matrices and a number of remarks concerning a practical realization, were presented. A special attention is devoted to the construction of efficient methods for the inclusion of the inverse of a matrix.

1. Introduction

The demands of the computer age at the beginning of the sixties years with its "finite" arithmetic dictate the need for a structure which has come to be called *interval analysis* or later *interval mathematics* - a new, growing, and fruitful branch of applied mathematics. "*Although interval analysis is in a sense just a new language for inequalities, it is very powerful language and is one that has direct applicability to the important problem of significance in large computations*" (R.D. Richtmeyer, Math. Comput. 22 (1968), p. 221). The starting point for the application of interval analysis, described for the first time by Moore [21], is the desire in numerical mathematics to be able to implement algorithms on digital computers capturing all the roundoff errors

This work is supported by the Science Fund of Serbia under Grant No. 0401

automatically and therefore to calculate strict errors automatically. Interval arithmetic is powerful tool for bounding a result of some computation or a solution of an equation so that interval methods are often called *self-validiting algorithms.*

Anyone using a computer, whether in engineering design, physical sciences, technical disciplines, or whatever has surely inquired about the effect of rounding error and propagated error due to uncertain initial data or uncertain values of parameters in mathematical models. A standard question should be "*what is the error in the obtained results?*". Numerical algorithms using interval arithmetic supply techniques for keeping track of errors and provide the machine computation of rigorous error bounds on approximate solutions or results.

The application of interval mathematics to computing has several objectives: to provide computer algorithms for finding sets containing unknown solutions; to make these sets as small as possible; and to do all this as efficiently as possible. Towards these objectives, set-to-set mappings replace point-to-point mappings, and set inclusions replace approximate equalities.

The purpose of this paper is to present iterative methods for bounding the inverse of a matrix. The interest in bounding roundoff errors in matrix computations has come from the impossibility of exact representation of elements of matrices in some cases since numbers are represented in the computer by strings of bits of fixed, *finite* length. Besides, there are elements which are experimentally determined by measurement which leads to the *uncertainty* in initial data and it is only known that their values belong to some intervals. Finally, nearly all numerical computation is carried out with "fixed-precision", approximate arithmetic. In the commonly used approach, one assumes that the worst possible roundoff error occurs in each numerical step. One then determines (or bounds) how these errors can accumulate as the computation proceeds. This procedure is usually called *ordinary method for error bounding* and the abbreviation $\mathcal{OM}$ is used to refer to it. The second approach uses interval arithmetic (abbreviated as $\mathcal{IA}$) which has the advantage of an automatic control of rounding errors and, at the same time, an inclusion of the exact result of computation. For this reason, the main subject of this paper is concerned with iterative methods which use $\mathcal{IA}$ for bounding errors in matrix inversion.

In Section 2 we will give the basic matrix operations needed for the construction and analysis of iterative algorithms for the inclusion of real or interval matrices. A general approach to the problem of the inversion of matrices is described in Section 3. The two basic interval iterative methods, based on the generalized Schulz's method, are considered in Section 4. Conditions for the monotonicity of interval sequence of inclusion matrices are the subject of Section 5. In Section 6 we study the problem of finding a suitable initial matrix which insures the convergence of the presented interval algorithms. Efficient iterative methods for bounding the inverse of a matrix, which combine the efficiency of floating-point arithmetic and the control of accuracy of results by interval arithmetic, are presented in Section 7. A special attention is devoted to the choice of parameters which define the most efficient inclusion algorithm. Finally, in Section 8, we describe an iterative method for the inclusion of an *interval* matrix. Throughout this paper several numerical examples are given to illustrate presented methods as well as difficulties which appear in solving the studied problem.

The presented study is a two-way bridge between linear algebra and computing. Its aim is to encourage mathematicians to look further to computing as a source of challenging new problems, and researchers in computing to turn more frequently to contemporary mathematics in their day-to-day use of the digital machine.

2. Interval matrix operations

A subset of the set of real numbers $\mathbb{R}$ of the form

$$A = [a_1, a_2] = \{x \mid a_1 \leq x \leq a_2,\ a_1, a_2 \in \mathbb{R}\}$$

is called a closed *real interval*. The set of all closed real intervals will be denoted by $I(\mathbb{R})$. If $a_2 = a_1$ then the interval $A = [a_1, a_2]$ degenerates to the real number a_1 and A is called a *point interval*. The basic operations and properties in the set $I(\mathbb{R})$ are described in the book [3, Ch. 1 and 2]. Real intervals will be denoted by capital letters.

A *real interval matrix* is a matrix whose elements are real intervals. Since we deal in this paper only with real intervals and real interval matrices, we will use the shorter terms *interval* and *interval matrix*. The set of $m \times n$ matrices over the real numbers is denoted by $M_{mn}(\mathbb{R})$ and the set of $m \times n$ matrices over the real intervals by $M_{mn}(I(\mathbb{R}))$. An interval matrix whose all components are point intervals is called a *point matrix*. Point matrices (elements from $M_{mn}(\mathbb{R})$) will be denoted by capital letters $A, B, C, \ldots$, while interval matrices (elements from $M_{mn}(I(\mathbb{R}))$) by capital letters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ in **bold**. Interval matrices are represented, as is customary for real or complex matrices, by their components in the form $\mathbf{A} = (A_{ij})$.

Definition 1. Two $m \times n$ interval matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ are equal if and only if there is equality between all corresponding components of the matrices, that is, $\mathbf{A} = \mathbf{B} \ \Leftrightarrow\ A_{ij} = B_{ij}$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$.

A partial ordering on the set of interval matrices $M_{mn}(I(\mathbb{R}))$ is introduced by

Definition 2. Let $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ be two $m \times n$ interval matrices. Then

$$\mathbf{A} \subseteq \mathbf{B} \;\Leftrightarrow\; A_{ij} \subseteq B_{ij} \quad (i = 1, \dots, m,\; j = 1, \dots, n).$$

In particular, if $A = (a_{ij})$ is a point matrix, then we write $A \in \mathbf{B}$. Each interval matrix may be regarded as a set of point matrices.

In the following we give a short review of the basic operations between interval matrices which formally correspond to the operations on point matrices.

Definition 3. For two $m \times n$ matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ interval matrix addition and subtraction are defined by

$$\mathbf{A} \pm \mathbf{B} := (A_{ij} \pm B_{ij}).$$

Definition 4. Let $\mathbf{A} \in M_{mr}(I(\mathbb{R}))$ and $\mathbf{B} \in M_{rn}(I(\mathbb{R}))$. An interval matrix computation is defined by

$$\mathbf{AB} := \left(\sum_{k=1}^{r} A_{ik} B_{kj} \right).$$

Definition 5. If $\mathbf{A} = (A_{ij})$ is an interval matrix and X an interval, then

$$X\mathbf{A} = \mathbf{A}X := (XA_{ij}).$$

It is easy to prove that

$$\mathbf{A} \pm \mathbf{B} = \{A \pm B \mid A \in \mathbf{A}, B \in \mathbf{B}\},$$

while

$$\mathbf{AB} \supseteq \{AB \mid A \in \mathbf{A}, B \in \mathbf{B}\}.$$

In the following theorem the basic properties of the introduced operations are given (see [3, Ch. 10]):

Theorem 1. *If* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are interval matrices, then*

$$\begin{array}{ll}
\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} & \textit{(commutativity)}, \\
\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C} & \textit{(associativity)}, \\
\mathbf{A} + O = O + \mathbf{A} = \mathbf{A} & (O - \textit{zero matrix}), \\
\mathbf{A}I = I\mathbf{A} = \mathbf{A} & (I - \textit{unit matrix}), \\
\left.\begin{array}{l} (\mathbf{A} + \mathbf{B})\mathbf{C} \subseteq \mathbf{AC} + \mathbf{BC} \\ \mathbf{C}(\mathbf{A} + \mathbf{B}) \subseteq \mathbf{CA} + \mathbf{CB} \end{array}\right. & \textit{(subdistributivity)}, \\
(\mathbf{A} + \mathbf{B})C = \mathbf{A}C + \mathbf{B}C, & \\
C(\mathbf{A} + \mathbf{B}) = C\mathbf{A} + C\mathbf{B}, & \\
\mathbf{A}(BC) \subseteq (\mathbf{A}B)C. &
\end{array}$$

Let us note that the associative low is not, in general, valid for interval matrices. This low is not valid even if two of three matrices are point matrices (the last property in the above theorem).

The inclusion isotonicity property for the matrix operations is given in the following theorem ([3, Ch. 10]):

Theorem 2. *Let* $\mathbf{A}_k, \mathbf{B}_k$ $(k = 1, 2)$ *be interval matrices and* X *and* Y *real intervals. If* $* \in \{+, -, \cdot\}$ *is one of matrix operations then the conditions*

$$\mathbf{A}_k \subseteq \mathbf{B}_k \ (k = 1, 2) \quad \textit{and} \quad X \subseteq Y$$

imply $\mathbf{A}_1 * \mathbf{A}_2 \subseteq \mathbf{B}_1 * \mathbf{B}_2$ *and* $X\mathbf{A}_k \subseteq Y\mathbf{B}_k$.

In particular, from Theorem 2 we obtain

$$A \in \mathbf{A},\ B \in \mathbf{B} \ \Rightarrow\ A + B \in \mathbf{A} + \mathbf{B},$$
$$\lambda \in X,\ A \in \mathbf{A} \ \Rightarrow\ \lambda A \in X\mathbf{A} \quad (\lambda \in \mathbb{R}).$$

Definition 6. Matrix norm of an interval matrix $\mathbf{A}$ is defined by

$$\| \mathbf{A} \| := \max_{A \in \mathbf{A}} \| A \|,$$

where $\| \cdot \|$ is an arbitrary monotone norm.

Thus, the norm of an interval matrix is an extension of the norm of a point matrix and directly depends on the type of this norm. Most frequently, we use "maximum row-sum" norm $\| \cdot \|_\infty$,

$$\| \mathbf{A} \|_\infty := \max_{A \in \mathbf{A}} \| A \|_\infty = \max_i \sum_j |A_{ij}| \tag{2.1}$$

and "maximum column-sum" norm $\| \cdot \|_1$,

$$\| \mathbf{A} \|_1 := \max_{A \in \mathbf{A}} \| A \|_1 = \max_j \sum_i |A_{ij}|. \tag{2.2}$$

Both norms are monotonic and multiplicative, that is (omitting subscript indices),

$$\mathbf{B} \subseteq \mathbf{A} \ \Rightarrow \| \mathbf{B} \| \leq \| \mathbf{A} \| \quad \text{and} \quad \| \mathbf{AB} \| \leq \| \mathbf{A} \| \cdot \| \mathbf{B} \| .$$

In the sequel, we will omit the subscript indices (indicating the type of norm) and assume that the used matrix norm is monotonic and multiplicative. The application of some specific norm will be accented.

Before introducing the concept of width, absolute value and midpoint for interval matrices, we recall to the corresponding definitions for a given real interval $X = [a, b]$:

$$d(X) = b - a \quad (\mathit{width});$$
$$|X| = \max(|a|, |b|) \quad (\mathit{absolute\ value});$$
$$m(X) = \frac{a+b}{2} \quad (\mathit{midpoint}).$$

Definition 7. For an interval matrix $\mathbf{A} = (A_{ij})$ the following point matrices are associated:

a) the width matrix $d(\mathbf{A}) := (d(A_{ij}))$;
b) the absolute value matrix $|\mathbf{A}| := (|A_{ij}|)$;
c) the midpoint matrix $m(\mathbf{A}) := (m(A_{ij}))$.

The matrix $d(\mathbf{A})$ and $|\mathbf{A}|$ have nonnegative components. The elements of the midpoint matrix $m(\mathbf{A})$ are real numbers which are equal to the midpoints of the corresponding (interval) components of the interval matrix $\mathbf{A}$, so that $m(\mathbf{A}) \in \mathbf{A}$.

Definition 8. A sequence of interval matrices $\{\mathbf{A}_k\}$ is monotonically nonincreasing if $\mathbf{A}_0 \supseteq \mathbf{A}_1 \supseteq \mathbf{A}_2 \supseteq \cdots$, and monotonically nondecreasing if $\mathbf{A}_0 \subseteq \mathbf{A}_1 \subseteq \mathbf{A}_2 \subseteq \cdots$.

Definition 9. The intersection of two interval matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ of the same type is defined as

$$\mathbf{A} \cap \mathbf{B} := (A_{ij} \cap B_{ij}).$$

It is easy to see that the intersection of interval matrices has the property

$$\mathbf{A} \subseteq \mathbf{C},\ \mathbf{B} \subseteq \mathbf{D} \ \Rightarrow\ \mathbf{A} \cap \mathbf{B} \subseteq \mathbf{C} \cap \mathbf{D} \quad (\mathit{inclusion\ isotonicity}).$$

Definition 10. Let $X = (x_{ij})$ and $Y = (y_{ij})$ be point matrices from $M_{mn}(\mathbb{R})$. Then

$$X \le Y \ \Leftrightarrow\ x_{ij} \le y_{ij} \ (i = 1, \dots, m;\ j = 1, \dots, n)$$

defines the relation of partial ordering "$\le$" in $M_{mn}(\mathbb{R})$.

Using Definition 10 the following properties for real matrices, introduced in Definition 7, can be proved ([3, Ch. 10]):

Theorem 3. *If* $\mathbf{A} = (A_{ij})$ *and* $\mathbf{B} = (B_{ij})$ *are interval matrices of the same type, then*

$$\begin{aligned}
&(1)\quad \mathbf{A} \subseteq \mathbf{B} \;\Rightarrow\; d(\mathbf{A}) \le d(\mathbf{B}),\\
&(2)\quad \mathbf{A} \subseteq \mathbf{B} \;\Rightarrow\; |\mathbf{A}| \le |\mathbf{B}|,\\
&(3)\quad d(\mathbf{A} \pm \mathbf{B}) = d(\mathbf{A}) + d(\mathbf{B}),\\
&(4)\quad |\mathbf{A} + \mathbf{B}| \le |\mathbf{A}| + |\mathbf{B}|,\\
&(5)\quad |\lambda \mathbf{A}| = |\mathbf{A}\lambda| = |\lambda||\mathbf{A}| \quad (\lambda \in \mathbb{R}),\\
&(6)\quad |\mathbf{A}\mathbf{B}| \le |\mathbf{A}||\mathbf{B}|,\\
&(7)\quad d(\mathbf{A}\mathbf{B}) \le d(\mathbf{A})|\mathbf{B}| + |\mathbf{A}|d(\mathbf{B}),\\
&(8)\quad d(\mathbf{A}\mathbf{B}) \ge |\mathbf{A}|d(\mathbf{B}), \quad d(\mathbf{A}\mathbf{B}) \ge d(\mathbf{A})|\mathbf{B}|,\\
&(9)\quad d(A\mathbf{B}) = |A|d(\mathbf{B}), \quad d(\mathbf{B}A) = d(\mathbf{B})|A|,\\
&(10)\quad O \in \mathbf{A} \;\Rightarrow\; |\mathbf{A}| \le d(\mathbf{A}) \le 2|\mathbf{A}|,\\
&(11)\quad m(\mathbf{A} \pm \mathbf{B}) = m(\mathbf{A}) \pm m(\mathbf{B}),\\
&(12)\quad m(C\mathbf{A}) = Cm(\mathbf{A}), \quad m(\mathbf{A}C) = m(\mathbf{A})C,\\
&(13)\quad m(C) = C.
\end{aligned}$$

3. Problems of bounding the inverse of a matrix

In this section we will consider the problem of bounding the inverse of a matrix in the presence of rounding errors applying digital computers with the arithmetic of limited precision as well as uncertain data in elements of a given matrix.

First, we point out some more general problems in matrix inversion. Let B be an exact matrix whose elements can be exactly represented in arithmetic of finite (say, double) precision in a computer. Let $A = (a_{ij})$ be a matrix whose elements are subject to error. Suppose we know only that a_{ij} $(i, j = 1, 2, \ldots, n)$ lies in the real interval $[\underline{a}_{ij}, \overline{a}_{ij}]$, where $\underline{a}_{ij}$ and $\overline{a}_{ij}$ can be exactly represented in double precision.

Problem 1. Compute B^{-1} (approximately) and bound the errors resulting from roundoff.

Problem 2. For a given matrix $A = (a_{ij})$ with $a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}]$ for $i, j = 1, 2, \ldots, n$, compute A^{-1} (approximately) and bound both the errors resulting from roundoff and the errors from possible errors in A itself.

Problem 3. Define the set

$$(A^I)^{-1} = \{A^{-1} | \, a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}], \; A^{-1}A = I\}.$$

Compute $(A^I)^{-1}$ approximately and bound the errors due to roundoff.

Problem 4. Find $(A^I)^{-1}$ exactly.

Problem 1 can be easily solved by $\mathcal{OM}$ or by inverting B using $\mathcal{IA}$. Moreover, by use of arithmetic of sufficiently high precision, arbitrary accuracy with arbitrary sharp bounds can be obtained.

Using $\mathcal{IA}$, *Problem 2* can be solved as easily as *Problem 1*. Using $\mathcal{OM}$, only slightly more effort is required to solve *Problem 2* than *Problem 1*.

$\mathcal{OM}$ obviously cannot solve *Problem 4* and cannot solve *Problem 3* except in a very crude sense. It can be shown (see [6]) that *Problem 4* cannot be solved using $\mathcal{IA}$, even if infinite precision arithmetic is used. An approximate solution of arbitrary high accuracy can be obtained but the amount of work quickly becomes prohibitive.

Hence, we direct our attention to *Problem 3* which can be solved using $\mathcal{IA}$. Two approaches for solving this problem by $\mathcal{IA}$ have been developed in the literature: hyperpower method [4], [6] and Alefeld-Herzberger's modification of generalized Schulz' method [1].

The hyperpower method is defined by a matrix-valued fuction $\Phi(A, X)$ for real $n \times n$ matrix X in the range of the real $n \times n$ matrices, where A is a given matrix whose inverse A^{-1} has to be found. By means of the iteration

$$X^{(k+1)} = \Phi(A, X^{(k)}), \quad \text{given } X^{(0)}, \; k \geq 0,$$

we get an iterative method which generates a sequence $\{X^{(k)}\}$ of matrices. Following Altman [4] we call this iterative method a *hyperpower method* for A^{-1} of order $p > 1$ if and only if the equation

$$I - AX^{(k+1)} = (I - AX^{(k)})^p, \quad k \geq 0$$

is filfilled. If the initial matrix $X^{(0)}$ is chosen so that $\rho(I - AX^{(0)}) < 1$ (ρ denotes the spectral radius), then the sequence of matrices $\{X^{(k)}\}$ converges to the inverse A^{-1} of the matrix A with the order of convergence p. Using suitable error-bounds for the hyperpower method it is possible to derive inclusion set for A^{-1}. Further improvements can be attained using interval Schulz-Herzberger's method in the final step, as it was proposed in [16] and [17]. Let us note that Herzberger presented in [10] a class of iterative methods for inverting a linear bounded operator in a Banach space, which can be considered as a kind of hyperpower method.

The second method uses an iterative procedure to bound the inverse not only for a point matrix but also for an interval matrix. As mentioned above, this method was introduced by Alefeld and Herzberger [1] and analysed later in their books [2] and [3]. It is based on the generalized Schulz's method for point matrices and realized in real interval arithmetic. Since a number of outstanding results concerning improvements and modifications of this methods, including detailed studies of many convergence properties and behaviours, and a practical realization, were given by Prof. J. Herzberger throughout about twenty papers, it is quite natural that the mentioned methods and their modifications are referred to as Schulz-Herzberger's methods, or the S-H methods, for brevity. A survey of these interval methods will be given in the following sections.

Before we present iterative methods of Schulz-Herzberger's type, we give an example to illustrate difficulties appearing in bounding inverse matrices.

Example 1. Let us consider the interval matrix

$$\mathbf{A} = \begin{bmatrix} 1 & [0.999995, 1.000005] \\ 2 & 1 \end{bmatrix}$$

and the point matrix

$$C(x) = \begin{bmatrix} 1 & x \\ 2 & 1 \end{bmatrix}, \quad x \in X = [0.999995, 1.000005].$$

Let $\mathbf{A}^{-1} = (A'_{ij})$ and $C(x)^{-1} = (c'_{ij}(x))$ be the inverse matrices of $\mathbf{A}$ and $C(x)$, respectively. Then $A'_{ij} = \{c'_{ij}(x) | x \in X\}$. Let us determine, for instance, the component A'_{12} of the inverse matrix $\mathbf{A}^{-1}$. First, we have $c'_{12} = x/(2x-1)$. For $x \in X = [0.999995, 1.000005]$ the component $c'_{12}(x)$ is a monotone function so that the endpoints of the interval X yield the extreme values (minimum and maximum) of $c'_{12}(x)$. According to this, using 10 significant digits, we obtain $A'_{12} = [0.9999950000, 1.0000050000]$.

On the other hand, using interval arithmetic of infinite precision and the rounding of results to 10 digits to find A'_{12}, we calculate

$$\frac{X}{2X-1} = [0.9999850001, 1.0000150001],$$

which differs from the exact result given above by A'_{12}.

4. Interval versions of Schulz's method

Let $p > 1$ be a fixed natural number and I the unit matrix. If A is a given nonsingular point matrix and $X^{(0)}$ is an initial matrix such that

$\| I - AX^{(0)} \| < 1$, then for finding the inverse of A the generalized iterative method of Schulz of the order p

$$(4.1) \qquad X^{(k+1)} = X^{(k)} \sum_{r=0}^{p-1} (I - AX^{(k)})^r \quad (k = 0, 1, \dots)$$

can be applied (see [4], [26], [27], [32]). In particular, for $p = 2$, one obtains Schulz's method of the second order for calculating the inverse matrix [31]

$$(4.2) \qquad X^{(k+1)} = X^{(k)}(2I - AX^{(k)}) \quad (k = 0, 1, \dots) .$$

Let $\mathbf{X}$ be an interval matrix containing the inverse matrix A^{-1} of a given nonsingular matrix A, and let $X \in \mathbf{X}$ (for example, $X = m(\mathbf{X})$). For $B = I - AX$ we have the identity

$$I^{p-1} - B^{p-1} = (I - B)(I + B + B^2 + \cdots + B^{p-2}) = AX \sum_{r=0}^{p-2} B^r,$$

that is, after multiplying by A^{-1},

$$A^{-1} - A^{-1}(I - AX)^{p-1} = X \sum_{r=0}^{p-2} (I - AX)^r.$$

Hence, since $A^{-1} \in \mathbf{X}$,

$$(4.3) \quad A^{-1} = X \sum_{r=0}^{p-2} (I - AX)^r + A^{-1}(I - AX)^{p-1}$$
$$\in X \sum_{r=0}^{p-2} (I - AX)^r + \mathbf{X}(I - AX)^{p-1}.$$

The last relation suggests the following iterative interval version of (4.1) for the inclusion of the matrix A :

$$(4.4) \; \mathbf{X}^{(k+1)} = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} \Big(I - Am(\mathbf{X}^{(k)})\Big)^r + \mathbf{X}^{(k)} \Big(I - Am(\mathbf{X}^{(k)})\Big)^{p-1},$$

$(k = 0, 1, \dots)$, assuming that the initial matrix $\mathbf{X}^{(0)}$ contains A^{-1}.

The properties of the inclusion iterative method (4.4) are given in the following theorem ([3, Ch. 18]):

Theorem 4. *Let A be a nonsingular $n\times n$ matrix and $\mathbf{X}^{(0)}$ an $n\times n$ interval matrix such that $A^{-1}\in\mathbf{X}^{(0)}$. A sequence $\{\mathbf{X}^{(k)}\}$ of interval matrices is calculated according to* (4.4). *Then*

(4a) *each matrix $\mathbf{X}^{(k)}$ $(k\geq 0)$ contains A^{-1};*

(4b) *the sequence $\{\mathbf{X}^{(k)}\}$ converges to A^{-1} if and only if the spectral radius $\rho(I-Am(X^{(0)}))$ is smaller than 1;*

(4c) *using a matrix norm $\|\cdot\|$ the sequence $\{d(\mathbf{X}^{(k)})\}$ satisfies*

$$\| d(\mathbf{X}^{(k+1)}) \|\leq \gamma \| d(\mathbf{X}^{(k)}) \|^p, \quad \gamma\geq 0,$$

that is, the order of convergence of the method (4.4) *is at least p.*

Proof. Of (4a): Setting $\mathbf{X}^{(k)}=\mathbf{X}$ and $m(\mathbf{X}^{(k)})=X$ in (4.3) and taking into account the iterative formula (4.4), we obtain

$$A^{-1}\in m(\mathbf{X}^{(k)})\sum_{r=0}^{p-2}\Big(I-Am(\mathbf{X}^{(k)})\Big)^r+\mathbf{X}^{(k)}\Big(I-Am(\mathbf{X}^{(k)})\Big)^{p-1}=\mathbf{X}^{(k+1)}.$$

Since, in addition, $A^{-1}\in\mathbf{X}^{(0)}$, the proof of (4a) follows by complete induction.

Of (4b): Using the rules from Theorem 3 for the midpoint matrices, the midpoint mapping in the iterative procedure (4.4) gives the following iterative formula for the sequence $\{m(\mathbf{X}^{(k)})\}$:

$$m(\mathbf{X}^{(k+1)})=m(\mathbf{X}^{(k)})\sum_{r=0}^{p-1}\Big(I-Am(\mathbf{X}^{(k)})\Big)^r.$$

This is a generalization of Schulz's iterative procedure given also by (4.1). Multiplying both sides of this equation by A one obtains

$$\begin{aligned}Am(\mathbf{X}^{(k+1)})&=\Big(I-(I-Am(\mathbf{X}^{(k)}))\Big)\sum_{r=0}^{p-1}\Big(I-Am(\mathbf{X}^{(k)})\Big)^r\\&=I-(I-Am(\mathbf{X}^{(k)}))^p,\end{aligned}$$

or

$$I-Am(\mathbf{X}^{(k+1)})=(I-Am(\mathbf{X}^{(k)}))^p=(I-Am(\mathbf{X}^{(0)}))^{p^{k+1}}.$$

Hence, there follows

$$\lim_{k\to\infty}m(\mathbf{X}^{(k)})=A^{-1}\Leftrightarrow\lim_{k\to\infty}(I-Am(\mathbf{X}^{(0)}))^p=O\Leftrightarrow\rho\big(I-Am(\mathbf{X}^{(0)})\big)<1.$$

Let us show that the sequence $\{\mathbf{X}^{(k)}\}$ converges to A^{-1} if and only if the sequence of midpoint matrices $\{m(\mathbf{X}^{(k)})\}$ converges to A^{-1}. This follows from the consideration of the sequence $\{d(\mathbf{X}^{(k)})\}$ of the width matrices which satisfy

$$d(\mathbf{X}^{(k+1)}) = d(\mathbf{X}^{(k)})|(I - Am(\mathbf{X}^{(k)}))^{p-1}|$$

(see the properties (3) and (9) of Theorem 3). If $\lim_{k\to\infty} m(\mathbf{X}^{(k)}) = A^{-1}$, then the last relation implies that $\lim_{k\to\infty} d(\mathbf{X}^{(k)}) = O$. Conversely, using the continuity of m and (13) of Theorem 3 it follows trivially that $\lim_{k\to\infty} \mathbf{X}^{(k)} = A^{-1}$ implies $\lim_{k\to\infty} m(\mathbf{X}^{(k)}) = A^{-1}$. Since it was shown above that the condition $\rho(I - Am(\mathbf{X}^{(0)})) < 1$ was necessary and sufficient for the convergence of $\{m(\mathbf{X}^{(k)})\}$, it follows that (4b) is valid.

Of (4c): First we estimate

$$\begin{aligned} d(\mathbf{X}^{(k+1)}) &= d(\mathbf{X}^{(k)})|(I - Am(\mathbf{X}^{(k)}))^{p-1}| \\ &= d(\mathbf{X}^{(k)})|(AA^{-1} - Am(\mathbf{X}^{(k)}))^{p-1}| \\ &\leq d(\mathbf{X}^{(k)})(|A||A^{-1} - m(\mathbf{X}^{(k)})|)^{p-1}| \\ &\leq d(\mathbf{X}^{(k)})2^{-(p-1)}(|A|d(\mathbf{X}^{(k)}))^{p-1}. \end{aligned}$$

Using a monotonic and multiplicative matrix norm $\| \cdot \|$ and the last relation, we get

$$\| d(\mathbf{X}^{(k+1)}) \| \leq 2^{-(p-1)} \| A \|^{p-1} \| d(\mathbf{X}^{(k)}) \|^{p}$$

Since the inequality

$$\| B \| \gamma_1 \leq \| B \| \leq \gamma_2 \| B \|, \quad \gamma_1 > 0,\ \gamma_2 > 0,$$

is valid for every matrix norm $\| \cdot \|$, from this inequality we get

$$\| d(\mathbf{X}^{(k+1)}) \| \gamma_1 \leq 2^{-(p-1)} \gamma_2^{(p-1)} \| A \|^{p-1} \gamma_2^{p} \| d(\mathbf{X}^{(k)}) \|^{p},$$

which proves (4c). □

Remark 1. From the proof given above we see that the assertion of the theorem is also valid even if $\mathbf{X}^{(0)}$ is an arbitrary interval matrix not necessarily containing A^{-1}. In that case we will not have the inclusion $A^{-1} \in \mathbf{X}^{(k)}$ in general. We observe that the criterion (4b) depended only on the midpoint matrix $m(\mathbf{X}^{(0)})$ of the given inclusion matrix $\mathbf{X}^{(0)}$, while the width $d(\mathbf{X}^{(0)})$ can be arbitrary. For this reason, taking $m(\mathbf{X}^{(0)})$ to be an approximation to A^{-1} (but so that the condition (4b) holds) and choosing the elements of the matrix $\mathbf{X}^{(0)}$ to be large enough so that the enclosure of A^{-1} by $\mathbf{X}^{(0)}$ be ensured, we can provide not only the convergence of the method (4.4) but also the inclusion $A^{-1} \in \mathbf{X}^{(k)}$ $(k = 1, 2, \dots)$.

Example 2. The S-H method (4.4) for $p = 2$ was applied for the inclusion of the inverse of the point matrix

$$A = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \\ \frac{3}{10} & \frac{9}{10} \end{bmatrix}.$$

The initial inclusion matrix was constructed according to the procedure (6.2) given in Section 6. Thus, with $a = 1/(1- \parallel I - A \parallel) = \frac{5}{3}$, for the initial matrix $\mathbf{X}^{(0)}$ we choose

$$\mathbf{X}^{(0)} = \begin{bmatrix} [-a, 2+a] & [-a, a] \\ [-a, a] & [-a, 2+a] \end{bmatrix} = \begin{bmatrix} [-\frac{5}{3}, \frac{11}{3}] & [-\frac{5}{3}, \frac{5}{3}] \\ [-\frac{5}{3}, \frac{5}{3}] & [-\frac{5}{3}, \frac{11}{3}] \end{bmatrix}.$$

In this way we ensure that $A^{-1} \in \mathbf{X}^{(0)}$ holds. Besides, we have $\rho(I - Am(\mathbf{X}^{(0)})) = \rho(I - A) = 0.8 < 1$, which provides the convergence of the iterative procedure (Theorem 4). The first four iterations give the following inclusion interval matrices (using arithmetic with 7 significant digits):

$$\mathbf{X}^{(1)} = \begin{bmatrix} [0.1666666, 2.2333316] & [-0.8999999, 0.4999999] \\ [-1.4333324, 0.8333329] & [0.5000000, 1.6999988] \end{bmatrix},$$

$$\mathbf{X}^{(2)} = \begin{bmatrix} [1.1716651, 1.5043325] & [-0.3969995, -0.1749999] \\ [-0.5963338, -0.2616657] & [1.0849990, 1.3049983] \end{bmatrix},$$

$$\mathbf{X}^{(3)} = \begin{bmatrix} [1.3587207, 1.3672409] & -0.3054331, -0.2997532] \\ [-0.4581502, -0.4496299] & [1.2088432, 1.2145233] \end{bmatrix},$$

$$\mathbf{X}^{(4)} = \begin{bmatrix} [1.3636322, 1.3636379] & [-0.3030319, -0.3030281] \\ [-0.4545477, -0.4545422] & [1.2121181, 1.2121219] \end{bmatrix}.$$

The applied iterative methods converges quadratically starting from the third iteration. Besides, in each iteration step we have

$$A^{-1} = \begin{bmatrix} \frac{15}{11} & -\frac{10}{33} \\ -\frac{15}{11} & \frac{40}{33} \end{bmatrix} = \begin{bmatrix} 1.363636336\ldots & -0.30303030\ldots \\ -0.45454545\ldots & 1.21212121\ldots \end{bmatrix} \in \mathbf{X}^{(k)}.$$

The sequence of the matrices produced by (4.4) always contains A^{-1} according to (4a) and, thus, it seems natural to form the intersection of the new inclusion matrix $\mathbf{X}^{(k+1)}$ and the former matrix $\mathbf{X}^{(k)}$ in order to decrease the resulting matrix, which leads to the iterative method

(4.5)
$$\begin{cases} \mathbf{Y}^{(k+1)} = m(\mathbf{X}^{(k)})\sum_{r=0}^{p-2}(I - Am(\mathbf{X}^{(k)}))^r + \mathbf{X}^{(k)}(I - Am(\mathbf{X}^{(k)}))^{p-1}, \\ \mathbf{X}^{(k+1)} = \mathbf{Y}^{(k+1)} \cap \mathbf{X}^{(k)}, \quad (k = 0, 1, \dots). \end{cases}$$

Using this iteration procedure one obtains a monotonic sequence

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \cdots$$

of inclusions for A^{-1}. The following numerical example does show, however, that the convergence criterion (4b) is not sufficient for convergence in general.

Example 3. ([3, Ch. 18]) We choose $p = 2$ and the matrices

$$A = \begin{bmatrix} 0.4 & 0.6 \\ -0.6 & 0.4 \end{bmatrix}, \quad \mathbf{X}^{(0)} = \begin{bmatrix} [-2, 4] & [-3, 3] \\ [-3, 3] & [-2, 4] \end{bmatrix},$$

which implies that $m(\mathbf{X}^{(0)}) = I$. We obtain

$$I - Am(\mathbf{X}^{(0)}) = \begin{bmatrix} 0.6 & -0.6 \\ 0.6 & 0.6 \end{bmatrix}$$

and calculate $\rho(I - Am(\mathbf{X}^{(0)})) = 0.6\sqrt{2} \approx 0.85 < 1$. Therefore, the procedure (4.4) converges to A^{-1} using this interval matrix. Applying (4.5) we find

$$\mathbf{Y}^{(1)} = m(\mathbf{X}^{(0)}) + \mathbf{X}^{(0)}(I - Am(\mathbf{X}^{(0)})) = \begin{bmatrix} [-2, 5.2] & [-4.2, 3] \\ [-3, 4.2] & [-2, 5.2] \end{bmatrix},$$

which implies that $\mathbf{X}^{(1)} = \mathbf{X}^{(0)}$. The sequence of matrices generated by (4.5) therefore does not converge to A^{-1} in contrast to the sequence computed by (4.4).

A convergence statement for the iteration (4.5) is contained in the following theorem.

Theorem 5. *Let A be a nonsingular $n \times n$ matrix and $\mathbf{X}^{(0)}$ an $n \times n$ interval matrix for which $A^{-1} \in \mathbf{X}^{(0)}$. If the sequence of matrices $\{\mathbf{X}^{(k)}\}$ is produced by* (4.5), *then*

(5a) *each matrix $\mathbf{X}^{(k)}$, $k \geq 0$, contains A^{-1};*

(5b) *if the inequality* $\rho(|I - AX|) < 1$ *is satisfied for all* $X \in \mathbf{X}^{(0)}$, *then the sequence* $\{\mathbf{X}^{(k)}\}$ *converges toward* A^{-1};

(5c) *the sequence* $\{d(\mathbf{X}^{(k)})\}$ *is bounded as follows:*

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \gamma' \| d(\mathbf{X}^{(k)}) \|^p, \quad \gamma' \geq 0,$$

that is, the order of convergence of the iterative process (4.5) *is at least* p.

Proof. Of (5a): As in the proof of (4a) of Theorem 4 we first show that $A^{-1} \in \mathbf{Y}^{(k+1)}$, from which follows immediately that $A^{-1} \in \mathbf{X}^{(k+1)}$ since $A^{-1} \in \mathbf{X}^{(k)}$.

Of (5b): We shall use the fact that every sequence $\{\mathbf{X}^{(k)}\}$, for which $\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \cdots$ holds, converges to an interval matrix $\mathbf{X} = (X_{ij})$, where

$$X_{ij} = \bigcap_{k=0}^{+\infty} X_{ij}^{(k)} \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

(see [3, Corollary 8 in Ch. 10]). Therefore, the sequence $\{\mathbf{X}^{(k)}\}$ obtained by (4.5) always converges to an interval matrix $\mathbf{X}$. We now show that under the assumptions of the theorem we must necessarily have $d(\mathbf{X}) = O$. We define

$$\mathbf{Y} = m(\mathbf{X}) \sum_{r=0}^{p-2} (I - Am(\mathbf{X}))^r + \mathbf{X}(I - Am(\mathbf{X}))^{p-1}$$

and obtain $\mathbf{X} = (X_{ij} \cap Y_{ij}) \subseteq \mathbf{Y}$ from (4.5). By (1) of Theorem 3 we get $d(\mathbf{X}) \leq d(\mathbf{Y})$. For $d(\mathbf{X})$ we obtain from (4.5)

$$d(\mathbf{X})|I - Am(\mathbf{X})|^{p-1} \geq d(\mathbf{X})|(I - Am(\mathbf{X}))^{p-1}| = d(\mathbf{Y}) \geq d(\mathbf{X}),$$

which implies that

$$d(\mathbf{X})(I - |I - Am(\mathbf{X})|^{p-1}) \leq O.$$

The assumption $\rho(|I - Am(\mathbf{X})|) < 1$ implies the existence of $(I - |I - Am(\mathbf{X})|^{p-1})^{-1}$. It can be shown that this inverse is also nonnegative. From this it follows that $d(\mathbf{X}) \leq O$, and hence $d(\mathbf{X}) = O$. Taking into account (5a) we obtain $\mathbf{X} = A^{-1}$.

Of (5c): As in the proof of (4c) one first derives the inequality

$$\| d(\mathbf{Y}^{(k+1)}) \| \leq \gamma \| d(\mathbf{X}^{(k)}) \|^p$$

for a monotonic and multiplicative matrix norm $\|\cdot\|$. From this it follows that the inequality

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \| d(\mathbf{Y}^{(k+1)}) \| \leq \gamma \| d(\mathbf{X}^{(k)}) \|^{p}$$

is valid since $\mathbf{X}^{(k+1)} \subseteq \mathbf{Y}^{(k+1)}$ as well as using (1) of Theorem 3 and the monotonicity of the norm $\|\cdot\|$. Analogous to the proof of (4c) we use the norm equivalence theorem to prove the final statement. □

5. Monotonicity of Schulz-Herzberger's method

J.W. Schmidt has proved in [28] that the inclusion $\mathbf{X}^{(1)} \subseteq \mathbf{X}^{(0)}$ is a necessary and sufficient condition for the monotonicity of the interval Schulz's method

$$\mathbf{X}^{(k+1)} = m(\mathbf{X}^{(k)}) + \mathbf{X}^{(k)}\Big(I - Am(\mathbf{X}^{(k)})\Big). \tag{5.1}$$

Starting from the above inclusion J. Herzberger has derived in [7] the necessary and sufficient condition which is of practical importance. Furthermore, using Schmidt's remark (given without a proof) that the inclusion $\mathbf{X}^{(1)} \subseteq \mathbf{X}^{(0)}$ is also necessary and sufficient for the monotonicity of the higher-order method (4.4) (see [28]), J. Herzberger has considered in [9] the monotonicity of (4.4).

The aim of this section is to give a useful sufficient condition for the monotonicity of the S-H method (4.4). Our consideration reduces to Herzberger's results [7] concerning the iterative method (5.1), which can be generalized for the method (4.4).

Lemma 1. *Let $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \ldots$ be the sequence of interval matrices produced by the iterative formula* (4.4) *and let $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$. If the inequality*

$$2\Big|m(\mathbf{X}^{(k)})\Big(I - Am(\mathbf{X}^{(k)})\Big)\Big| \leq d(\mathbf{X}^{(k)})(I - |I - Am(\mathbf{X}^{(k)})|) \tag{5.2}$$

is valid for $k = 0$, then it holds for each $k = 0, 1, 2, \ldots$.

Proof. For brevity, let us introduce the notations

$$\mathbf{C}_k = I - Am(\mathbf{X}^{(k)}), \quad \mathbf{B}_k = |\mathbf{C}_k|.$$

From (4.4) we find the midpoint matrix $m(\mathbf{X}^{(k+1)})$ and the width matrix $d(\mathbf{X}^{(k+1)})$,

$$m(\mathbf{X}^{(k+1)}) = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-1} \mathbf{C}_k^{r}, \tag{5.3}$$

$$d(\mathbf{X}^{(k+1)}) = d(\mathbf{X}^{(k)})|\mathbf{C}_k^{p-1}|. \tag{5.4}$$

Using inequalities

$$|\mathbf{XY}| \le |\mathbf{X}||\mathbf{Y}|, \quad |\mathbf{X}+\mathbf{Y}| \le |\mathbf{X}| + |\mathbf{Y}|$$

for the absolute value matrices, in the special case of the point matrices we obtain

$$(5.5) \qquad |\mathbf{C}_k^r| \le |\mathbf{C}_k|^r = \mathbf{B}_k^r,$$

$$(5.6) \qquad \left|\sum_{r=0}^{p-1} \mathbf{C}_k^r\right| \le \sum_{r=0}^{p-1} |\mathbf{C}_k|^r = \sum_{r=0}^{p-1} \mathbf{B}_k^r.$$

Starting from (5.3), we find

$$Am(\mathbf{X}^{(k+1)}) = \Big(I - \Big(I - Am(\mathbf{X}^{(k)})\Big)\Big) \sum_{r=0}^{p-1} \mathbf{C}_k^r = I - \mathbf{C}_k^p,$$

wherefrom

$$(5.7) \qquad \mathbf{C}_{k+1} = I - Am(\mathbf{X}^{(k+1)}) = \mathbf{C}_k^p = \mathbf{C}_0^{p^{k+1}}.$$

Since $\rho(|I - Am(\mathbf{X}^{(0)})|) = \rho(\mathbf{B}_0) < 1$ implies $\rho(\mathbf{B}_0^\nu) < 1$ $(\nu > 1)$, we have

$$\rho(|\mathbf{C}_k|) = \rho(|\mathbf{C}_0^{p^k}|) \le \rho(\mathbf{B}_0^{p^k}) < 1,$$

that is

$$(5.8) \qquad \rho(\mathbf{B}_0) < 1 \quad \text{implies} \quad \rho(\mathbf{B}_k) < 1, \quad k = 0, 1, \ldots .$$

Furthermore, because of $\rho(\mathbf{B}_k) < 1$ there exists the inverse matrix $(I - \mathbf{B}_k)^{-1} \ge O$ and the following identity is valid

$$\sum_{r=0}^{p-1} \mathbf{B}_k^r = (I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p).$$

We shall now prove that the inequality (5.2), where $\mathbf{X}^{(k)}$ is given by (4.4), is valid for each $k = 1, 2, \ldots$ if

$$(5.9) \qquad 2|m(\mathbf{X}^{(0)})\mathbf{C}_0| \le d(\mathbf{X}^{(0)})(I - \mathbf{B}_0)$$

(the inequality (5.2) for $k = 0$) holds.

Let us rewrite (5.2) in a (shorter) form

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k) \tag{5.10}$$

and assume that this inequality holds for some index $k \ge 0$. Multiplying both sides of (5.10) by $(I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}|$, one obtains

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k|(I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}|$$

or

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k|\sum_{r=0}^{p-1}\mathbf{B}_k^r|\mathbf{C}_k^{p-1}| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^k)|\mathbf{C}_k^{p-1}|. \tag{5.11}$$

Using inequalities

$$\mathbf{B}_k^p = |\mathbf{C}_k|^p \ge |\mathbf{C}_k^{p-1}||\mathbf{C}_k| \ge |\mathbf{C}_k^p|,$$

we find

$$\begin{aligned}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}| &\le (I - |\mathbf{C}_k^{p-1}||\mathbf{C}_k|)|\mathbf{C}_k^{p-1}| \\ &\le |\mathbf{C}_k^{p-1}| - |\mathbf{C}_k^{p-1}||\mathbf{C}_k^p| = |\mathbf{C}_k^{p-1}|(I - |\mathbf{C}_k^p|) \\ &= |\mathbf{C}_k^{p-1}|(I - \mathbf{B}_{k+1}).\end{aligned}$$

According to (5.6) and the last inequality, from (5.11) we obtain

$$2\Big|m(\mathbf{X}^{(k)})(\sum_{r=0}^{p-1}\mathbf{C}_k^r)\mathbf{C}_k^p\Big| \le d(\mathbf{X}^{(k)})|\mathbf{C}_k^{p-1}|(I - \mathbf{B}_{k+1}). \tag{5.12}$$

Taking into account formulas (5.3), (5.4) and (5.7), the inequality (5.12) becomes

$$2|m(\mathbf{X}^{(k+1)})\mathbf{C}_{k+1}| \le d(\mathbf{X}^{(k+1)})(I - \mathbf{B}_{k+1}).$$

This proves (5.10) (that is, (5.2)) by complete induction since (5.9) holds as the assumption of Lemma 1. □

Theorem 6. *Let $A^{-1} \in \mathbf{X}^{(0)}$ and $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$. Then the generalized interval method* (4.4) *converges to A^{-1}, where $A^{-1} \in \mathbf{X}^{(k)}$ $(k = 0, 1, \dots)$, and if*

$$2|m(\mathbf{X}^{(0)})(I - Am(\mathbf{X}^{(0)}))| \le d(\mathbf{X}^{(0)})(I - |I - Am(\mathbf{X}^{(0)})|) \tag{5.13}$$

holds, then the method (4.4) *is monotone.*

Proof. First, we observe that under the given assumption, there follows that (4.4) converges because

$$\rho(|I - Am(\mathbf{X}^{(0)})|) < 1 \quad \text{implies} \quad \rho(I - Am(\mathbf{X}^{(0)})) < 1.$$

The inclusion $A^{-1} \in \mathbf{X}^{(k)}$ for each $k \geq 0$ has been proved in Theorem 4.

Under the condition (5.13) of Theorem 6 (and Lemma 1, too) the inequality

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \leq d(\mathbf{X}^{(k)})(I - \mathbf{B}_k)$$

holds for each $k \geq 0$. Multiplying both sides of the last inequality by

$$\sum_{r=0}^{p-2} \mathbf{B}_k^r = (I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^{p-1}) \geq O,$$

we obtain

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \sum_{r=0}^{p-2} \mathbf{B}_k^r \leq d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^{p-1}). \tag{5.14}$$

Since

$$\Big| m(\mathbf{X}^{(k)}) \sum_{r=1}^{p-1} \mathbf{C}_k^r \Big| \leq |m(\mathbf{X}^{(k)})\mathbf{C}_k| \sum_{r=0}^{p-2} \mathbf{B}_k^r$$

and

$$I - \mathbf{B}_k^{p-1} \leq I - |\mathbf{C}_k^{p-1}|,$$

from (5.14) we obtain

$$2\Big| m(\mathbf{X}^{(k)}) \sum_{r=1}^{p-1} \mathbf{C}_k^r \Big| \leq d(\mathbf{X}^{(k)})(I - |\mathbf{C}_k^{p-1}|)$$

or

$$\Big| m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-1} \mathbf{C}_k^r - m(\mathbf{X}^{(k)}) \Big| \leq \frac{1}{2}\big(d(\mathbf{X}^{(k)}) - d(\mathbf{X}^{(k)})|\mathbf{C}_k^{p-1}|\big).$$

Finally, according to the formulas (5.3) and (5.4) for the matrices $m(\mathbf{X}^{(k+1)})$ and $d(\mathbf{X}^{(k+1)})$, the last inequality becomes

$$|m(\mathbf{X}^{(k+1)}) - m(\mathbf{X}^{(k)})| \leq \frac{1}{2}\big(d(\mathbf{X}^{(k)}) - d(\mathbf{X}^{(k+1)})\big). \tag{5.15}$$

But, the inequality (5.15) is necessary and sufficient for the inclusion

$$\mathbf{X}^{(k+1)} \subseteq \mathbf{X}^{(k)}. \tag{5.16}$$

Therefore, if the condition (5.13) is satisfied, then the inclusion (5.16) holds for each $k \geq 0$, which means that the generalized iterative method (4.4) is monotone. This completes the proof of the theorem. □

Remark 2. The condition (5.13) can be rewritten in the form

$$2|m(\mathbf{X}^{(0)})(I - Am(\mathbf{X}^{(0)}))|(I - |I - Am(\mathbf{X}^{(0)})|)^{-1} \leq d(\mathbf{X}^{(0)}). \tag{5.17}$$

Since this condition depends only on the given matrix A and the initial approximation $m(\mathbf{X}^{(0)})$ for A^{-1}, the matrix $d(\mathbf{X}^{(0)}) \geq O$ can always be chosen so that (5.17) is satisfied. Since the convergence condition $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$ does not depend on the width matrix $d(\mathbf{X}^{(0)})$, this matrix can be taken so that

(i) an initial interval matrix $\mathbf{X}^{(0)}$ safely includes A^{-1} and
(ii) the monotonicity of the iterative method (4.4) is provided.

We observe that (5.13) coincides with the corresponding condition obtained for the interval Schulz's method (5.1). Since the construction of the proof of the assertion which gives a sufficient condition for the monotonicity of (5.1) is directly based on the relation (5.13) (see [7, Theorem 2]), for the higher-order interval method (4.4) ($p > 1$) we immediately have the following theorem:

Theorem 7. *Let $\|I - Am(\mathbf{X}^{(0)})\| < 1$ ($\|\cdot\|$ the column-sum norm), then the method (4.4) converges to A^{-1}. In addition, this method is monotone if the following is valid*

$$d(X_{ij}^{(0)}) = h \geq \frac{2 \cdot \max\limits_{i,j} |m(X_{ij}^{(0)})|}{1 - \|I - Am(\mathbf{X}^{(0)})\|} \quad \textit{for } i \neq j, \quad d(X_{ii}^{(0)}) \geq h. \tag{5.18}$$

Theorem 7 gives a sufficient condition for the monotonicity of the generalized interval method (4.4). Under the given assumptions of this theorem it is always possible to choose the width matrix $d(\mathbf{X}^{(0)})$ in such a way that the method (4.4) is monotone. A detailed description of the construction of the initial including matrix $\mathbf{X}^{(0)}$ which guarantees for $A^{-1} \in \mathbf{X}^{(0)}$ is given in the next section.

6. Construction of the initial inclusion matrix

The convergence criterion (5b) in Theorem 5 depends on the width of the inclusion matrix $\mathbf{X}^{(0)}$ for A^{-1}, which is not a case with the criterion (4b) in Theorem 4. Nevertheless, it is not difficult to find a relation between these criteria. For instance, if an interval matrix $\mathbf{X}^{(0)}$ satisfies the inequality $\| I - Am(\mathbf{X}^{(0)}) \| < 1$, for a monotonic and multiplicative norm $\| \cdot \|$, then we have that

$$\| d(\mathbf{X}^{(0)}) \| < \alpha = 2(1- \| I - Am(\mathbf{X}^{(0)}) \|)/ \| A \| \tag{6.1}$$

is a sufficient criterion for the statement that $\| I - AX \| < 1$ for all $X \in \mathbf{X}^{(0)}$. To construct a suitable interval matrix $\mathbf{X}^{(0)}$ let us assume that A may be represented as $A = I - B$ with $\| B \| < 1$. The choice $m(\mathbf{X}^{(0)}) := I$ gives

$$\| I - Am(\mathbf{X}^{(0)}) \| = \| B \| < 1,$$

and, according to the criterion (4b), the inclusion method (4.4) is convergent for every interval matrix $\mathbf{X}^{(0)}$ for which $m(\mathbf{X}^{(0)}) = I$. In order to insure the inclusion $A^{-1} \in \mathbf{X}^{(0)}$ we consider the equation $AX = (I - B)X = I$ or $X = BX + I$. In regard to this there follows (using a multiplicative matrix norm) that

$$\| X \| \leq a := \frac{1}{1- \| B \|},$$

wherefrom (using the row-sum or the column-sum norm)

$$-a \leq x_{ij} \leq a \quad (1 \leq i, j \leq n)$$

for all the elements of $X = (x_{ij})$. For the matrix $\mathbf{X}^{(0)} = (X_{ij})$ defined by

$$X_{ij}^{(0)} = \begin{cases} [-a, a] & \text{for } i \neq j \\ [-a, 2 + a] & \text{for } i = j, \end{cases} \tag{6.2}$$

we have $A^{-1} \in \mathbf{X}^{(0)}$ and also $m(\mathbf{X}^{(0)}) = I$. By virtue of Theorem 4 the iterative method converges to A^{-1}.

From the above consideration, we see that the iterative method (4.4) requires weaker convergence conditions compared to (4.5). For this reason, it is convenient to start with the method (4.4) as soon as the sufficient condition (6.1) is fulfilled provided $\| I - Am(\mathbf{X}^{(0)}) \| < 1$ and then to continue with the method (4.5). Such a combined process has been described in details by Alefeld and Herzberger [1].

The sufficient condition (6.1) can be weakened, which is the subject of the following assertion:

Theorem 8. *If $\mathbf{X}^{(k)}$ is an inclusion matrix for A^{-1}, then*

$$\| d(\mathbf{X}^{(k)}) \| < \beta = \frac{2}{\| A \|} \tag{6.3}$$

is a sufficient condition for the convergence of (4.5) to A^{-1}.

Proof. Applying the width operator d to the iterative formula (4.5), we obtain

$$d(\mathbf{X}^{(k+1)}) \le d\Big(m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} \Big(I - Am(\mathbf{X}^{(k)}) \Big) + \mathbf{X}^{(k)} \Big(I - Am(\mathbf{X}^{(k)}) \Big)^{p-1} \Big)$$
$$\le d(\mathbf{X}^{(k)}) 2^{-p+1} \Big(|A| d(\mathbf{X}^{(k)}) \Big)^{p-1}.$$

Using a monotonic and multiplicative matrix norm, we get

$$\| d(\mathbf{X}^{(k+1)}) \| \le \left(\frac{\| A \|}{2} \right)^{p-1} \| d(\mathbf{X}^{(k)}) \|^{p},$$

which proves that (6.3) is sufficient for $\| d(\mathbf{X}^{(k+1)}) \| \to 0$, and whence, $\mathbf{X}^{(k+1)} \to A^{-1}$. □

Remark 3. Comparing the numbers α and β appearing in (6.1) and (6.3) we infer that $\alpha < \beta$, which means that the condition (6.3) is weaker than (6.1). Furthermore, β is considerably simpler to calculate and has the same value for all the matrices $\mathbf{X}^{(k)}$. Finally, the criterion (6.3) from Theorem 6 is even considerably less restrictive than that of Theorem 5, as it was shown in [8].

The result given in the following theorem provides a better inclusion for A^{-1} compared with (6.2).

Theorem 9. *For the initial inclusion matrix $\widehat{\mathbf{X}}^{(0)}$ defined by*

$$\widehat{\mathbf{X}}^{(0)} = I + ([-c, c]) \quad \textit{with} \quad c = \frac{\| B \|}{1 - \| B \|}$$

we have $A^{-1} \in \widehat{\mathbf{X}}^{(0)}$ and the iterative process (4.4) converges to A^{-1} ($\| \cdot \|$ row-sum or colomn-sum norm).

Proof. Starting from the obvious equalities

$$A^{-1} - I = (I - B)^{-1} - I = (I - B)^{-1} B$$

and using a multiplicative matrix norm $\| \cdot \|$ and the well-known inequality

$$\| (I-B)^{-1} \| \leq \frac{1}{1- \| B \|},$$

we obtain

$$\| A^{-1} - I \| \leq \| (I-B)^{-1} \| \| B \| \leq \frac{\| B \|}{1- \| B \|}.$$

In this way the inclusion $A^{-1} \in \hat{\mathbf{X}}^{(0)}$ is proved. Further, since

$$\| I - Am(\hat{\mathbf{X}}^{(0)}) \| = \| B \| < 1,$$

the iterative method (4.4) converges to A^{-1} (see Theorem 4). □

Remark 4. The computation of $\mathbf{X}^{(0)}$ and $\hat{\mathbf{X}}^{(0)}$ requires the same amount of work but we have $\hat{\mathbf{X}}^{(0)} \subset \mathbf{X}^{(0)}$.

For nonsingular matrices A which do not have the same property as in the previous, some other approach which uses Theorem 7 has to be applied for constructing a starting matrix for (4.4) with $A^{-1} \in \mathbf{X}^{(0)}$. Namely, according to Remark 1, the iterative method (4.4) converges to A^{-1} even if $\mathbf{X}^{(0)}$ does not include A^{-1}. But the construction (5.18) guarantees the monotonicity of the interval matrices produced by (4.4),

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \dots ,$$

and thus we necessarily have $A^{-1} \in \mathbf{X}^{(0)}$.

7. Combined Schulz-type methods

In this section we describe a general approach to the construction of new methods of *Schulz's type* for improving bounds for the inverse A^{-1} of a given $n \times n$ nonsingular matrix A. These methods, proposed by J. Herzberger and Lj. Petković [18], [19], possess a great computational efficiency.

It is well known that interval evaluations are more costly than ordinary floating-point computations. For this reason, it would be advisable to apply the necessary interval computations only in a part of the algorithm. The aim of this section is to present an approach for solving this problem, which combines iterative methods in floating-point arithmetic as well as in interval arithmetic. In this way, we take advantage of comparatively small computational costs of floating-point arithmetic and the very important inclusion property of interval arithmetic (the enclosure of the exact result).

Definition 11. The mapping Φ from the set of $n \times n$-matrices onto itself is called a Schulz-type method of order $p \geq 2$ for A^{-1} if and only if for $Y = \Phi(X, A)$ the equation

$$I - AY = (I - AX)^p \tag{7.1}$$

holds true.

Remark 5. For practical computations Φ should only consist of matrix multiplications and additions.

Two the most frequently used examples of the mapping Φ are given below:

Example 4. Let $p \geq 2$, then

$$\Phi_p(X, A) = X \sum_{i=0}^{p-1} (I - AX)^i \tag{7.2}$$

defines a Schulz-type method for A^{-1} of order p.

Example 5. We can use Ostrowski's identity (see [22])

$$\begin{aligned}\Phi_5(X, A) &= X \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX) + (I - AX)^2\right)\\ &\quad \times \left(I - \frac{\sqrt{5}+1}{2}(I - AX) + (I - AX)^2\right)\\ &= X \cdot \sum_{\nu=0}^{4} (I - AX)^\nu,\end{aligned} \tag{7.3}$$

which also gives a Schulz's type method of order 5.

By means of a Schulz-type method for A^{-1} we can construct an iteration method in ordinary floating-point arithmetic as follows:

$$X^{(k+1)} = \Phi(X^{(k)}, A), \quad X^{(0)} \quad \text{given}, \quad k \geq 0. \tag{7.4}$$

The following assertion has been proved in [18]:

Theorem 10. *Let Φ be a Schulz-type method for A^{-1} of order p. Then the sequences of matrices $\{X^{(k)}\}$ produced by (7.4) have the following properties:*

(a) $X^{(k)} \to A^{-1} \Leftrightarrow \rho(I - AX^{(0)}) < 1$,

(b) *if the method (7.4) is convergent, then its order of convergence is at least p.*

Let $\mathbf{X}^{(0,0)} \ni A^{-1}$ be an initial inclusion for A^{-1} and $\Phi(X, A)$ define a Schulz-type method of order p for A^{-1}. Then for fixed integers $k \geq 0$ and $p \geq 1$ we define:

$$(7.5) \qquad \begin{aligned} X^{(n,0)} &= m(\mathbf{X}^{(n,0)}), \\ X^{(n,i)} &= \Phi(X^{(n,i-1)}, A), \quad 1 \leq i \leq k, \end{aligned}$$

(empty statement in case $k = 0$)

$$\mathbf{X}^{(n+1,0)} = X^{(n,k)} \sum_{i=0}^{r-1} (I - AX^{(n,k)})^i + \mathbf{X}^{(n,0)}(I - AX^{(n,k)})^r,$$

(*Horner-scheme evaluation in interval arithmetic*), $n \geq 0$,

and

$$(7.6) \qquad \begin{aligned} X^{(n,0)} &= m(\mathbf{X}^{(n,0)}), \\ X^{(n,i)} &= \Phi(X^{(n,i-1)}, A), \quad 1 \leq i \leq k, \end{aligned}$$

(empty statement in case k=0)

$$\mathbf{X}^{(n+1,0)} = \left\{ X^{(n,k)} \sum_{i=0}^{r-1} (I - AX^{(n,k)})^i + \mathbf{X}^{(n,0)}(I - AX^{(n,k)})^r \right\} \cap \mathbf{X}^{(n,0)},$$

(*Horner-scheme evaluation in interval arithmetic*), $n \geq 0$,

where $m(\mathbf{X}) = \big(m(X_{ij})\big)$ is the midpoint matrix.

Remark 6. For $k = 0$ we get as special cases the methods (4.4) and (4.5) discussed in Section 4.

In particular, for the fixed $n = 0$ in (7.5) and (7.6), we obtain the combined methods

$$(7.7) \qquad \begin{cases} X^{(i)} &= \Phi(X^{(i-1)}, A), \quad 1 \leq i \leq k, \\ \mathbf{X}^{(1,k)} &= X^{(k)} \sum_{i=0}^{r-1} (I - AX^{(k)})^i + \mathbf{X}^{(0)}(I - AX^{(k)})^r. \end{cases}$$

and the monotonic version

$$(7.8)\quad \begin{cases} X^{(i)} & = \Phi(X^{(i-1)}, A), \quad 1 \le i \le k, \\ \mathbf{X}^{(1,k)} & = \left\{ X^{(k)} \sum_{i=0}^{r-1} (I - AX^{(k)})^i + \mathbf{X}^{(0)}(I - AX^{(k)})^r \right\} \bigcap \mathbf{X}^{(0)}. \end{cases}$$

The combined methods (7.7) and (7.8) are, therefore, performed applying k iterations in floating-point arithmetic in order to obtain sufficiently good approximation (point matrix) $X^{(k)}$ to the inverse matrix A^{-1} and then, in the final step, the inclusion method of the order r to provide the guaranteed error bounds to A^{-1}. Such a combination is of a great interest in practice and, for this reason, it was studied extensively in the papers [19] and [25].

For the combined methods (7.5) and (7.6) the following theorem has been proved in [19].

Theorem 11. *For the methods* (7.5) *and* (7.6) *the inclusion* $A^{-1} \in \mathbf{X}^{(n,0)}$ $(n \ge 0)$ *holds.*

Theorem 12. *The sequence* $\{\mathbf{X}^{(n,0)}\}$ *obtained by the method* (7.5) *converges to* A^{-1} *if and only if* $\rho(I - AX^{(0,0)}) < 1$.

As presented in Sections 4 and 6, the convergence criterion for monotonic methods like (7.6) for which

$$\mathbf{X}^{(0,0)} \supseteq \mathbf{X}^{(1,0)} \supseteq \cdots \supseteq \mathbf{X}^{(n,0)} \ni A^{-1}$$

obviously holds, differ from those of the non-monotonic methods like (7.5). This is contained in the convergence theorem, which is quite similar to Theorem 8.

Theorem 13. *The sequence* $\{\mathbf{X}^{(n,0)}\}$ *generated by the method* (7.6) *converges to* A^{-1} *if the inequality*

$$\| d(\mathbf{X}^{(0,0)}) \| < 2/ \| A \|,$$

with a monotonic matrix norm $\| \cdot \|$, *is fulfilled.*

According to Traub [33, Appendix C] the *efficiency index* of an iterative method of order q can be defined by $q^{1/\Theta}$, where Θ is the total amount of work for one iteration step. In methods like (7.5) and (7.6) one usually measures Θ in terms of matrix multiplications and all other computational costs are considered to be negligible compared with these. If we count the computational efforts by Traub's formula we get the following results, assuming that one interval matrix multiplication costs at least about two times as much as a point matrix multiplication:

ks multiplications for the application of the Schulz-type method where **s** is the number of multiplications for the evaluation of Φ;

$\mathbf{r+1}$ interval matrix multiplications for the Horner-scheme interval evaluation or approximately $\mathbf{2(r+1)}$ point matrix multiplications.

This makes a total cost of $\mathbf{ks+2(r+1)}$ multiplications for one step of methods (7.5) or (7.6), reduced to point matrix multiplications. A Schulz-type method of the form (7.4) requires only ordinary floating point operations whereas the Horner-scheme interval evaluation has to be done completely by rounded interval operations to ensure the inclusion property of Theorem 11.

From Theorem 11 and Theorem 12 we get lower bounds for the order of convergence of our methods (7.5) and (7.6) as $q = rp^k + 1$ so that lower bounds for the efficiency index are given by

$$E(p,r,k) = (rp^k+1)^{1/(ks+2(r+1))}.$$

Before determining parameters p and r in order to establish the optimal combined method concerning the computational efficiency expressed by the efficiency index $E(p,r,k)$, we recall that the most efficient method of Schulz's type in ordinary floating-point arithmetic reads

$$X^{(k+1)} = X^{(k)} \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX^{(k)}) + (I - AX^{(k)})^2\right) \\ \times \left(I - \frac{\sqrt{5}+1}{2}(I - AX^{(k)}) + (I - AX^{(k)})^2\right),$$

which is constructed using the mapping Φ_5 given in Example 5. Namely, the number of multiplication is $\mathbf{s} = 4$ for the evaluation of $\Phi_5(X,A)$ given by (7.3) and $\mathbf{s} = p$ for $\Phi_p(X,A)$ $(p \neq 5)$ given by (7.2) when Horner-scheme evaluation is applied.

In the sequel, speaking about the function Φ_5 (the case $p=5$), we will assume the function defined by Ostrowski's identity (7.3), while in the remaining cases Φ_p $(p \neq 5)$ will denote the mapping (7.2). According to this, we define the total amount of work (expressed by point matrix multiplications) by

$$\Theta = \begin{cases} 4k + 2(r+1), & p = 5, \\ pk + 2(r+1), & p \neq 5 \end{cases}$$

(see [19]). Therefore, the lower bound of the efficiency index is given by

$$E(p,r,k)=\begin{cases}(r5^k+1)^{1/(4k+2r+2)}, & p=5,\\ (rp^k+1)^{1/(pk+2r+2)}, & p\neq 5.\end{cases} \tag{7.9}$$

The detailed procedure for finding optimal values of p and r (with respect to definition (7.9)) has been done by M. Petković and J. Herzberger in the paper [25]. This problem is of a great practical importance in applying the combined methods (7.5) and (7.6), and also (7.7) and (7.8). It leads to an optimization problem in the field of integers. First, the following theorem has been proved:

Theorem 14. *Let $r\in\{1,\dots,7\}$ and let $k\geq 1$ and p ($p\geq 2$ and $p\neq 5$) be arbitrary integers. Then*

$$E(5,r,k)>E(p,r,k). \tag{7.10}$$

As explained in [25], the restriction for r to be less than 8 is made for practical reasons. Namely, for a sufficiently great r (at least $r=16$ but usually considerably greater, even more than 100) it is possible to find $p\geq 6$ and k such that the inequality (7.10) becomes converse. But, such values of p (at least 6 iterations in floating-point arithmetic) and r (at least 16 iterations in interval arithmetic) are meaningless in practice, especially in the situation when it is easy to provide initial matrices which insures the safe convergence.

The optimal choice of the number of point iterations r has been considered in the following theorem, assuming that $p=5$.

Theorem 15. *The function $q(r):=(r5^k+1)^{1/(4k+2r+2)}$ attains its maximum on the interval $(1,2)$ for arbitrary $k\geq 1$.*

Using the result of Theorem 15 and the fact that r is an integer, we conclude that the optimal r in the combined methods can be either $r=1$ or $r=2$, depending on the number of iterative steps. A short analysis has shown that

$$E(5,2,k)>E(5,1,k)\quad\text{for } k=1(1)6$$

and

$$E(5,2,k)<E(5,1,k)\quad\text{for } k\geq 7.$$

Thus, if the number of point iterative steps k is less than 7 then $r=2$ is the optimal value, while for $k\geq 7$ the optimal value is $r=1$. However, the second case ($k\geq 7$) is only of theoretical importance due to the very fast convergence of the applied point method (of the order 5). For example, if $\| d(\mathbf{X}^{(0)}) \| = 0.8$, using the estimation

$$\| d(\mathbf{X}^{(1,k)}) \| \sim \| d(\mathbf{X}^{(0)}) \|^{5k+1}$$

for $k = 4$ we obtain even $\| d(\mathbf{X}^{(1,k)} \| \sim 10^{-61}$, which is an indicative illustration that the use of a (relatively) great number of iterative steps (say, $k > 3$) is not only meaningless but also not profitable (because of the limited precision of digital computers).

Finally, according to the previous results and discussion, in a practical realization of the combined method it should be chosen $p = 5$ and $r = 2$ (optimal for $k \leq 6$). Thus, the most efficient combined method of Schulz-type is of the form

$$(7.11) \quad X^{(n,i+1)} = X^{(n,i)} \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX^{(n,i)}) + (I - AX^{(n,i)})^2 \right) \times$$
$$\times \left(I - \frac{\sqrt{5}+1}{2}(I - AX^{(n,i)}) + (I - AX^{(n,i)})^2 \right),$$
$$i = 0, 1, \ldots, k-1 \text{ (\textit{in floating point arithmetic})}$$

$$(7.12) \quad \mathbf{X}^{(n+1,0)} = \left(\mathbf{X}^{(0)}(I - AX^{(n,k)}) + X^{(n,k)}\right) \cdot (I - AX^{(n,k)}) + X^{(n,k)}$$
$$\text{(\textit{in interval arithmetic})},$$

where $X^{(n,0)} = m(\mathbf{X}^{(n,0)})$ and $n \geq 0$ and the starting matrix $\mathbf{X}^{(0,0)}$ includes A^{-1}.

The combined methods (7.11) and (7.12) have been considered in details in [19].

Example 6. To illustrate numerically the combined method (7.11) - (7.12), we present the example taken from the paper [19], where a 9×9 nonsingular matrix A with $A = I - B$, $\| B \| < 1$, was considered. Here $\| \cdot \|$ denotes the column-sum norm and the matrix $B = (b_{ij})$ is defined by

$$b_{ij} = \begin{cases} 0.1 & i \neq j \\ 0 & i = j \end{cases}, \quad (1 \leq i, j \leq 9).$$

A starting inclusion matrix $\mathbf{X}^{(0,0)}$ is constructed according to Theorem 9, that is, $\mathbf{X}^{(0,0)} = I + ([-c, c])$, where $c = \frac{\|B\|}{1 - \|B\|}$. Evidently $m(\mathbf{X}^{(0,0)}) = I$ and the inclusion $A^{-1} \in \mathbf{X}^{(0,0)}$ holds.

For the method (7.11) - (7.12), referred to as the method (a), it was taken $k = 2$. The result of this combined method was compared to the classical optimal method, referred to as method (b), which can be defined as

$$\mathbf{Y}^{(0)} = \mathbf{X}^{(0,0)} \quad \text{and}$$

$$\mathbf{Y}^{(n+1)} = m(\mathbf{Y}^{(n)}) + (m(\mathbf{Y}^{(n)}) + \mathbf{Y}^{(n)}(I - Am(\mathbf{Y}^{(n)}))(I - Am(\mathbf{Y}^{(n)}))$$

for $n = 0, 1, \ldots$.

Let the inequality $\| d(\mathbf{X}) \| < \varepsilon = 5 \times 10^{-10}$ define the stopping criterion. The results obtained by the methods (a) and (b) are given in Table 1.

The total computational amount of work in terms of point matrix multiplications under the same assumption for interval matrix multiplications as above is as follows:

for the method (a): $2 \times (4 + 6) = 20$
for the method (b): $5 \times 6 = 30$.

It is clear that method (a) converges faster with the smaller computational efforts. Moreover, the computational efficiency of the method (a) is greater the greater is k.

n	$\| d(\mathbf{X}^{(n,0)}) \|$	$\| d(\mathbf{Y}^{(n)}) \|$
0	$7.20000000000 \times 10^1$	$7.20000000000 \times 10^1$
1	$7.73094113280 \times 10^0$	$4.60800000000 \times 10^1$
2	$1.96000000000 \times 10^{-10}$	$1.20795955200 \times 10^1$
3		$2.17606647543 \times 10^{-1}$
4		$1.27215720000 \times 10^{-6}$
5		$2.56000000000 \times 10^{-10}$

Table 1

8. Bounding the inverse of an interval matrix

Let $\mathbf{A} = (A_{ij})$ be an $n \times n$ interval matrix for which A^{-1} exists for every real matrix $A \in \mathbf{A}$ and denote $\mathbf{A}^i = \{A^{-1} | \ A \in \mathbf{A}\}$. In this section the problem of computing an interval matrix $\mathbf{X}$ with $\mathbf{A}^i \subseteq \mathbf{X}$ is considered. In many cases one can find an initial inclusion $\mathbf{X}^{(0)} \supseteq \mathbf{A}^i$, for example, by means of norm inequalities. But, in that case, the question arises how to improve $\mathbf{X}^{(0)}$ in such a way that its width $d(\mathbf{X}^{(0)}) = d((X_{ij}^{(0)})) = (d(X_{ij}^{(0)}))$ will be reduced. Theoretically, it is possible to find the interval hull of $\mathbf{A}^i$ in the form $\widehat{\mathbf{X}} = \cap\{\mathbf{X} | \ \mathbf{X} \supseteq \mathbf{A}^i\}$, but this, in general, cannot be done without an unreasonable amount of work. For this reason, we are not dealing with this kind of problem and we are looking for an improvement $\mathbf{X}^*$ for $\mathbf{X}^{(0)}$ with $\mathbf{A}^i \subseteq \mathbf{X}^* \subseteq \mathbf{X}^{(0)}$ and $d(\mathbf{X}^*) \leq d(\mathbf{X}^{(0)})$ such that at least for a monotone matrix norm $\| \cdot \|$ the strict inequality

$$\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$$

is valid. Schmidt found in [28] and [30] a monotone algorithm for the iterative improvement of $\mathbf{X}^{(0)}$. Alefeld and Herzberger suggested in [3, Ch. 18] (see, also, Section 4 of this paper) a somewhat different approach by means of interval analysis. The proposed method is closely related to the monotone version of the interval Schulz method for the iterative improvement of bounds for the inverse of a real nonsingular matrix A and it can be read in the form

$$\mathbf{X}^{(k+1)} = \{m(\mathbf{X}^{(k)}) + \mathbf{X}^{(k)}(I - Am(\mathbf{X}^{(k)}))\} \cap \mathbf{X}^{(k)}, \tag{8.1}$$

where $m(\mathbf{X})$ is the midpoint matrix of $\mathbf{X}$. A similar generalization with $\mathbf{A}_k$ and $\lim_{k\to\infty} \mathbf{A}_k = A$ instead of $\mathbf{A}$ was already used in Chapter 20 in [3] in connection with the Newton-method. In the case $\mathbf{A} = A$ one obtains the well-known interval Schulz-method. For iteration (8.1) we get immediately the following lemma:

Lemma 2. *For $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ the sequence of matrices $\{\mathbf{X}^{(k)}\}$ produced by (8.1) has the property*

$$\mathbf{A}^i \subseteq \mathbf{X}^{(k)}, \quad (k = 0, 1, \dots) .$$

Proof. Since $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$, we choose $A^{-1} \in \mathbf{X}^{(0)}$ and by the use of the inclusion property of the interval operations we find

$$A^{-1} = m(\mathbf{X}^{(0)}) + A^{-1}(I - Am(\mathbf{X}^{(0)})) \in m(\mathbf{X}^{(0)}) + \mathbf{X}^{(0)}(I - Am(\mathbf{X}^{(0)})) \subseteq \mathbf{X}^{(1)}.$$

For $k > 1$ the proof can be done analogously. □

From (8.1) there follows

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \mathbf{X}^{(3)} \supseteq \dots$$

and thus

$$\lim_{k\to\infty} \mathbf{X}^{(k)} = \mathbf{X}^*$$

is valid. But the iteration process (8.1) could already fail with $\mathbf{X}^* = \mathbf{X}^{(0)}$ especially if $d(\mathbf{A})$ is of considerable size. In that case, instead of improving $X^{(0)}$, the process starts reproducing the same disk. Therefore, a convergence analysis for (8.1) which gives sufficient conditions for

$$\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$$

has to be done in such a way that the method (8.1) yields an improved inclusion $\mathbf{X}^*$. For a given matrix $\mathbf{A}$ these sufficient conditions will impose some restrictions for $\| |\mathbf{X}^{(0)}| \|$ as well as for $\| d(\mathbf{X}^{(0)}) \|$ and so determine a class of matrices $\mathbf{X}^{(0)}$ with $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ for which method (8.1) improves $\mathbf{X}^{(0)}$. The main result is the following theorem whose proof was given in [14].

Theorem 16. *Let* $\mathbf{A}$ *be given, then the iteration process* (8.1) *converges to* $\mathbf{X}^*$ *with* $\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$ *if the matrix* $\mathbf{X}^{(0)}$ *with* $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ *fulfills the inequalities*

$$\| d(\mathbf{A}) \| < \frac{3}{\| |\mathbf{X}^{(0)}| \| (8 \cdot \| m(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \| + \frac{3}{2}} \tag{8.2}$$

and

$$\frac{4 \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|^2}{2 - \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|} < \| d(\mathbf{X}^{(0)}) \| < \frac{16}{19} \cdot \frac{1}{\| m(\mathbf{A}) \|}. \tag{8.3}$$

In addition to this, for $\mathbf{X}^*$ *the inequality*

$$\| d(\mathbf{X}^*) \| \leq \frac{2 \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|^2}{1 - \frac{1}{2} \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|} \tag{8.4}$$

holds.

Remark 7. A sufficient condition for $\| |\mathbf{X}^{(0)}| \|$ in terms of $\| m(\mathbf{A}) \|$ and $\| d(\mathbf{A}) \| > 0$ such that (8.2) is fulfilled can easily be derived as

$$\| |\mathbf{X}^{(0)}| \| < \frac{2}{\frac{\| d(\mathbf{A}) \|}{2} + \sqrt{\frac{\| d(\mathbf{A}) \|^2}{4} + \frac{32}{3} \| m(\mathbf{A}) \| \cdot \| d(\mathbf{A}) \|}}.$$

Remark 8. From (8.4) it follows that $\| d(\mathbf{X}^*) \| \to 0$ as $\| d(\mathbf{A}) \| \to 0$. Thus, the estimation (8.4) claims that for $\mathbf{A} = A$ the interval Schulz-method converges to A^{-1}. This is the reason why (8.1) can be regarded as a generalization of the Schulz-method (5.1) in the case of an interval matrix $\mathbf{A}$.

Remark 9. The condition (8.3) is more restrictive than the corresponding result for the interval Schulz-method in the case $\mathbf{A} = A$ (see [8]) where the sufficient condition for the convergence

$$\| d(\mathbf{X}^{(0)}) \| < \frac{2}{\| A \|}$$

is proved. Here, $\mathbf{X}^{(0)}$ can contain singular matrices as examples show.

Remark 10. Condition (8.3) implies that every $X \in \mathbf{X}^{(0)}$ is nonsingular. This can be seen taking $X \in \mathbf{X}^{(0)}$. Then $(m(\mathbf{A}))^{-1} \in \mathbf{X}^{(0)}$ and we have

$$\| X - (m(\mathbf{A}))^{-1} \| = \| \, |X - (m(\mathbf{A}))^{-1}| \, \| \leq \| d(\mathbf{X}^{(0)}) \| < \frac{16}{19} \cdot \frac{1}{\| m(\mathbf{A}) \|}$$
$$< \frac{1}{\| m(\mathbf{A}) \|}.$$

According to [4, Theorem 4 in Section 4] it follows that

$$(m(\mathbf{A}))^{-1} + (X - (m(\mathbf{A}))^{-1}) = X$$

is nonsingular.

As it was shown in [14], the assumption on $\| \, |\mathbf{X}^{(0)}| \, \|$ can be weakened. But this requires more complicated form of the upper bound for $\| d(\mathbf{X}^{(0)}) \|$. Both is given in

Corollary of Theorem 16. *Let* $\mathbf{A}$ *be given. Then the iteration process* (8.1) *converges to* $\mathbf{X}^*$ *with* $\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$ *if the matrix* $\mathbf{X}^{(0)}$ *with* $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ *fulfills the inequalities*

$$\| d(\mathbf{A}) \| < \frac{1}{\| \, |\mathbf{X}^{(0)}| \, \| \cdot (2 \, \| m(\mathbf{A}) \| \cdot \| \, |\mathbf{X}^{(0)}| \, \| + 1}$$

and

$$\frac{4 \cdot \| d(\mathbf{A}) \| \cdot \| \, |\mathbf{X}^{(0)}| \, \|^2}{2 - \| d(\mathbf{A}) \| \cdot \| \, |\mathbf{X}^{(0)}| \, \|} < \| d(\mathbf{X}^{(0)}) \|$$
$$< \frac{1}{\| m(\mathbf{A}) \|} \left(1 - \frac{\| d(\mathbf{A}) \| \| \, |\mathbf{X}^{(0)}| \, \|}{2} \right).$$

In addition to this, the inequality (8.4) *holds.*

In practical computations the quantity $\| d(\mathbf{A}) \|$ is of small size. The interval matrix $\mathbf{A}$ appears, for instance, because of inaccurate initial data for a real matrix A or from conversion errors which are usually not too large. Therefore, the necessary initial inclusion $\mathbf{X}^{(0)}$ for $\mathbf{A}$ can often be calculated by an application of an interval Gaussian elimination or even by norm inequalities (see [14]).

References

[1] G. Alefeld, J. Herzberger, *Matrizeninvertierung mit Fehlereffassung Elektron*, Datenverarbeit **12** (1970), 410–416.

[2] G. Alefeld, J. Herzberger, *Einführung in die Intervallrechnung*, Bibliographisches Institut AG, Mannheim, 1974.

[3] G. Alefeld, J. Herzberger, *Introduction to interval computation*, Academic Press, New York, 1983.

[4] M. Altman, *An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space*, Pacific J. Math. **10** (1960), 1107–1113.

[5] P. Foster, *Bemerkungen zum Iterationsverfahren von Schulz zur Bestimmung der Inversen einer Matrix*, Numer. Math. **12** (1968), 211–214.

[6] E. Hansen, *Interval arithmetic in matrix computations. Part I*, J. SIAM Numer. Anal. Ser. B **2** (1965), 308–320.

[7] J. Herzberger, *On the monotonicity of the interval version of Schulz's method*, Computing **38** (1987), 71–74.

[8] J. Herzberger, *Remarks on the interval version of Schulz's method*, Computing **39** (1987), 183–186.

[9] J. Herzberger, *Zur Monotonie der intervallmäßigen Schulz-Verfahren höherer Ordnung*, Z. Angew. Math. Mech. **67** (1987), 137–138.

[10] J. Herzberger, *A class of optimal iterative methods of inverting a linear bounded operator*, Numer. Funct. Anal. and Optimiz. **9** (1987), 521–533.

[11] J. Herzberger, *Ein effizienter Algorithmus zur iterativen Einschließung der inversen Matrix*, Aplikace Matematiky **32** (1987), 271–275.

[12] J. Herzberger, *Monotone Einschließungsalgorithmen für die inverse Matrix mit Hilfe von PASCAL-SC*, Angew. Informatik **30** (1988), 207–212.

[13] J. Herzberger, *Iterationsverfahren höherer Ordnung zur Einschließung der Inversen einer Matrix*, Z. Angew. Math. Mech. **69** (1989), 115–120.

[14] J. Herzberger, *On the convergence of an iterative method for bounding the inverse of an interval matrix*, Computing **41** (1989), 153–162.

[15] J. Herzberger, *Uber die Wirksamkeit eines iterationsverfahrens zur Einschließ ung der inversen einer intervallmatrix*, Z. Angew. Math. Mech. **70** (1990), 470–472.

[16] J. Herzberger, *Using error-bounds for hyperpower methods to calculate inclusions for the inverse of a matrix.*, BIT **30** (1990), 508–515.

[17] J. Herzberger, D. Bethke, *On two algorithms for bounding the inverses of an interval matrix*, Interval Computations **1** (1991), 44–53.

[18] J. Herzberger, Lj. Petković, *On the construction of efficient interval Schulz's methods for bounding the inverse matrix*, Z. Angew. Math. Mech. **71** (1991), 411–412.

[19] J. Herzberger, Lj. Petković, *Efficient iterative algorithms for bounding the inverse of a matrix*, Computing **44** (1990), 237–244.

[20] L. V. Kantorovich, G. P. Akilov, *Functional Analysis*, Pergamon Press, Oxford, 1982.

[21] R. E. Moore, *Interval analysis*, Prentice Hall, New Jersey, 1966.

[22] A. M. Ostrowski, *Sur quelques transformations de la série de Liouville-Neumann*, C.R. Acad. Sci. Paris **206** (1938), 1345–1347.

[23] Lj. D. Petković, M. S. Petković, *On the monotonicity of the higher-order Schulz's method*, Z. Angew. Math. Mech. **68** (1988), 455–456.

[24] M. S. Petković, *Introduction to interval mathematics*, Naučna knjiga, Beograd, 1989. (In Serbian)

[25] M. S. Petković, J. Herzberger, *On the efficiency of a class of combined Schulz's methods for bounding the inverse matrix*, Z. Angew. Math. Mech. **71** (1991), 181–187.

[26] W. V. Petryshyn, *On the inversion of matrices and linear operators*, Proc. Amer. Math. Soc. **16** (1965), 893–901.

[27] W. V. Petryshyn, *On generalized inverses and on the converses of $(I-\beta K)^n$ with application to iterative methods*, J. Math. Anal. Appl. **18** (1967), 417–439.

[28] J. W. Schmidt, *Einschließung inverser Elemente durch Fixpunktverfahren*, Numer. Math. **31** (1978), 313–320.

[29] J. W. Schmidt, *Monotone Eingrezung von inversen Elementen durch ein quadratisch konvergentes Verfahren ohne Durchschnittsbildung*, Z. Angew. Math. Mech. **60** (1980), 202–204.

[30] J. W. Schmidt, *Two-sided approximations of inverses, square-roots and Cholesky factors*, Computational Mathematics. Banach Center Publications **13** (1984), 483–497.

[31] G. Schulz, *Iterative Berechnung der reziproken Matrix*, Z. Angew. Math. Mech. **13** (1933), 57–59.

[32] E. Stickel, *On a class of high order methods for inverting matrices*, Z. Angew. Math. Mech. **67** (1987), 334–336.

[33] J. F. Traub, *Iterative methods for the solutions of equations*, Prentice-Hall, New Jersey, 1964.

Faculty of Electronic Engineering, Department of Mathematics, P.O. Box 73, 18000 Niš, Yugoslavia

CONTRIBUTIONS

FILOMAT (Niš) 9:3 (1995), 579-590
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON SOME 4- AND 5-DESIGNS ON $\leq$ 49 POINTS

Dragan M. Acketa and Vojislav Mudrinski

ABSTRACT. A search for those t-$(q+1,k,\lambda)$ designs is made, which arise by action of the groups $PSL(2,q)$ and $PGL(2,q)$ on the ground-set $\Omega(q) = \{0,1,...,q-1\} \cup \{\infty\}$. The search is made for $(t,k) = (4,5)$ with prime powers $q \leq 49$ and for $(t,k) \in \{(4,6),(5,6)\}$ with prime powers $q \leq 31$. The group $PSL(2,q)$ is used for $q \equiv 3 \pmod 4$ and the group $PGL(2,q)$ is used otherwise.

The search uses orbit incidence matrices determined by orbits of t-subsets and k-subsets (shortly: t-orbits and k-orbits) of the ground-set, obtained by action of the group used. An element of an orbit incidence matrix is the number of those k-sets within a k-orbit, which contain a fixed t-set (representative) of a t-orbit. Construction of orbit incidence matrices essentially uses 3-homogenicity of the groups.

The total number of distinct quadruples (t,q,k,λ) of parameters, for which t-$(q+1,k,\lambda)$ designs are constructed is equal to 75. It is guaranteed that the obtained values of λ are the only possible, which can be reached by action of the groups used, for the considered triples (t,q,k). It is assumed that most of the obtained quadruples of design parameters are new, in particular those for $q = 19, 25, 27, 31$ and 37.

1. Introduction

Let *n-set* denote a set of cardinality n. A t-(v,k,λ) *design* [5] is an incidence structure on v points, which consists of some k-sets of points (called *blocks*) without repetitions and which satisfies that each t points are contained in exactly λ blocks. $GF(q)$ is the Galois field associated to a prime power $q = p^s$.

The group $GL(2,q)$ is the group of all non-singular 2×2 matrices with elements in $GF(q)$ (= non-singular linear transformations over $(GF(q))^2$), while $SL(2,q)$ is its subgroup consisting of the matrices with determinant 1. The projective general linear group $PGL(2,q)$ and the projective special linear group $PSL(2,q)$ are obtained from $GL(2,q)$ and $SL(2,q)$ respectively, by reduction with the corresponding groups of homoteties.

Both $PGL(2,q)$ and $PSL(2,q)$ act on the common ground-set $\Omega(q) = \{0,1,...,q-1\} \cup \{\infty\}$. It is known that $PGL(2,q)$ acts 3-transitively for all q, while $PSL(2,q)$ acts 3-homogenously for $q \equiv 3 \pmod 4$ and only 2-transitively for other prime powers q. Construction of these two groups is described in [3] and [2] respectively.

The orbit incidence matrix method for searching designs, which will be referred to as "Λ-technique", introduced in [2], can be sketched as follows:

- Let be given a 3-homogenous permutation group G acting on $\Omega(q)$ and a pair (t,k) of natural numbers satisfying $4 \le t < k \le q$.
- Construct the orbits $T_1,...,T_m$ of those t-subsets of $\Omega(q)$, which include the set $\{0,1,\infty\}$. Similarly, construct the orbits $B_1,...,B_n$ of those k-subsets of $\Omega(q)$, which include the set $\{0,1,\infty\}$.
- Construct the orbit incidence matrix $\Lambda = (\lambda_{ij})$, $1 \le i \le m$, $1 \le j \le n$, where λ_{ij} denotes the number k-subsets of $\Omega(q)$ within B_j, which contain a fixed t-subset (representative) of T_i; the sum of all elements in each row of Λ is equal to
$\lambda_{\text{trivial}} = \binom{q+1-t}{k-t} = \lambda$-value of the trivial t-$(q+1,k,\lambda)$-design.
- Try to find for a *proper* subset P of the column set of Λ, which satisfies that the sum of elements within the columns of P is equal to the same constant λ for all the rows $(1 \le \lambda \le \lambda_{\text{trivial}}/2)$.
- If the subset P is found, then all the k-subsets of $\Omega(q)$, which belong to the orbits B_j corresponding to the columns of P, are the blocks of a t-$(q+1,k,\lambda)$ design. The complementary k-subsets of $\Omega(q)$ are the blocks of a t-$(q+1,k,\lambda_{\text{trivial}}-\lambda)$ design.

1.1. A comparision between the use of $PSL(2,q)$ and $PGL(2,q)$

Statement. *If a prime power q is of the form $4k+3$, then the group $PSL(2,q)$ is more suitable for looking for designs than $PGL(2,q)$.*

Namely, as already mentioned, the group $PSL(2,q)$ is 3-homogenous with the values of q of this form. Although 3-transitivity (possessed by $PGL(2,q)$) is a stronger property, it is only 3-homogenicity that matters when the application of the Λ-technique is considered. On the other hand, the group $PSL(2,q)$ is a subgroup (normal, of index 2) of $PGL(2,q)$, which implies that orbits by action of $PSL(2,q)$ are included in orbits by action of $PGL(2,q)$. "Building constituents" of the designs are k-orbits. The smaller are the constituents, the larger is the chance for making equilibrium (suitable sums of λ_{ij}'s), which leads to designs. Therefore we have the following:

Consequence. *If a prime power q is of the form $4k+3$, then each design which can be derived by Λ-technique with application of the group $PGL(2,q)$, can be also derived with application of $PSL(2,q)$.*

However, the group $PGL(2,q)$ is more suitable with other prime powers. It is always 3-transitive (and consequently 3-homogenous), while, when $PSL(2,q)$ is considered, only 2-transitivity is guaranteed.

Conclusion. *The group $PSL(2,q)$ is used for searching for designs with prime powers q of the form $4k+3$, while the group $PGL(2,q)$ is used with other prime powers q.*

2. Results

2.1. A global account of the generated designs

The computer search was performed for prime powers $q \leq 31$ with $k = 6$ and for further prime powers $q \leq 49$ with $k = 5$.

The search was successful with:

$PSL(2,q)$ and $(t,k) = (4,5)$ for $q = 47$;

$PSL(2,q)$ and $(t,k) = (4,6)$ for $q = 19$;

$PSL(2,q)$ and $(t,k) = (5,6)$ for $q = 11, 23, 27, 31$;

$PGL(2,q)$ and $(t,k) = (4,6)$ for $q = 25$;

$PGL(2,q)$ and $(t,k) = (4,5)$ for $q = 17, 32, 37$.

Note that the reported success with $(t,k) = (4,6)$ means that there was no success with $(t,k) = (5,6)$; otherwise, a 4-$(q+1, 6, \lambda_2)$ design would be a consequence of a 5-$(q+1, 6, \lambda_1)$ design, which corresponds to the same set of columns of the λ_{ij} matrix.

More precisely, the constructed t-$(q+1, k, \lambda)$ designs are summarized in the following table (the numbers of t-orbits and k-orbits by action of the group cited are denoted by m and n respectively):

t	q	k	$\lambda \leq \lambda_{\text{trivial}}/2$	λ_{trivial}	G	m	n
5	11	6	1,2	7	$PSL(2,11)$	2	6
4	17	5	4	14	$PGL(2,17)$	3	4
4	19	6	60	120	$PSL(2,19)$	5	19
5	23	6	1,2,3,4,5,6,7,8,9	19	$PSL(2,23)$	7	34
4	25	6	51,60,81,90,111	231	$PSL(2,25)$	5	28
5	27	6	2,3,4,5,6,7,8,9,10,11	23	$PSL(2,27)$	10	54
5	31	6	6,12	27	$PSL(2,31)$	15	83
4	32	5	4,5,9	29	$PSL(2,32)$	5	11
4	37	5	16	34	$PGL(2,37)$	7	15
4	47	5	8,12,16,20	44	$PSL(2,47)$	10	33

When the design complementations are taken into account, it turns out that the total number of generated designs with distinct parameters is equal to $75 = 2\cdot 2 + 2\cdot 1 + 1\cdot 1 + 2\cdot 9 + 2\cdot 5 + 2\cdot 10 + 2\cdot 2 + 2\cdot 3 + 2\cdot 1 + 2\cdot 4$. (note that $\lambda = \lambda_{\text{trivial}}/2$ for $q = 19$).

A global conclusion concerning the generated designs, obtained after a thorough examination of the generated Λ-matrices, is the following:

Statement. *The above listed values of λ (taking in addition the values complementary w.r.t. λ_{trivial} into account), are the only possible values of λ which can be reached by action of the corresponding listed groups.*

However, it is not to say that there may not exist t-(v,k,λ) designs, obtained in another manner, which have some other values of λ and the same values of t, v and k as some of the listed ones.

2.2. Detailed results of application of λ-technique

In this section are listed Λ-matrices corresponding to each one of the ten above cited groups, together with representatives of the underlying orbits and with a representative of the generated designs, for each possible quadruple of parameters. The t-orbits and k-orbits corresponding to successive rows and columns of a Λ-matrix are listed in front of it.

2.2.1. Denotations.

Λ-matrices in this section will be denoted as $\Lambda(G;t,k)$. A Λ-matrix is determined by the corresponding group G and by the values of parameters t and k; it establishes relationship between t-orbits and k-orbits by action of G.

In order to enable precise identification of s-orbits (for $s \in \{4,5,6\}$), the following data will be given in the form $(A : B; C)$, where

$A =$ the ordinal number of the coresponding orbit (= row or column of the (λ_{ij}) matrix).

$B =$ $s-2$ elements of the lexicographically the first "special" representative, apart from the compulsory elements $0,\ 1,\ \infty$.

$C =$ the number of "special" subsets (supersets of $\{0,1,\infty\}$) within the orbit.

For example, the denotation $(4 : 2,3,7; 10)$ below (that is, $A = 4$; $B = 2,3,7$; $C = 10$), used for a 6-orbit by action of $PSL(2,11)$, means that this orbit is the fourth one among the 6-orbits (corresponds to the 4th column of the λ_{ij} matrix), has the 6-subset $\{0,1,2,3,7,\infty\}$ as a representative and contains ten "special" 6-subsets.

The design(s) generated from a Λ-matrix are listed after the word "**Design(s)**". A representative design is given in ( )-brackets separately for each possible λ. Designs are denoted by the ordinal numbers of the columns belonging to the set P (cited in the description of Λ-technique); the blocks of the designs are exactly the k-sets belonging to the k-orbits corresponding to the columns of P.

Thus the denotation $(\lambda = 2 : 7, 21, 22, 30)$ after the matrix $\Lambda(PSL(2,23); 5, 6)$ means that the 6-sets of the 7th, 21st, 22nd and 30th orbit of this Λ-matrix constitute a 5-(24,6,2) design.

2.2.2. $PSL(2,11)$, $t = 5$, $k = 6$, $\lambda_{\text{trivial}} = 7$.

5-orbits: $(1 : 2, 3; 30)$ $(2 : 3, 4; 6)$

6-orbits: $(1 : 2, 3, 4; 30)$ $(2 : 2, 3, 5; 12)$ $(3 : 2, 3, 6; 10)$ $(4 : 2, 3, 7; 10)$ $(5 : 2, 3, 8; 10)$ $(6 : 2, 3, 9; 12)$

The 2 × 6 matrix $\Lambda(PSL(2,11); 5, 6)$: $\begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 5 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$

Designs: $(\lambda = 1 : 2)$ $(\lambda = 2 : 2, 6)$

2.2.3. $PGL(2,17)$, $t = 4$, $k = 5$, $\lambda_{\text{trivial}} = 14$.

4-orbits: $(1 : 2; 3)$ $(2 : 3; 6)$ $(3 : 4; 6)$

5-orbits: $(1 : 2, 3; 30)$ $(2 : 2, 5; 15)$ $(3 : 2, 6; 30)$ $(4 : 3, 7; 30)$

The 3 × 4 matrix $\Lambda(PGL(2,17); 4, 5)$: $\begin{pmatrix} 8 & 2 & 4 & 0 \\ 4 & 0 & 4 & 6 \\ 2 & 4 & 4 & 4 \end{pmatrix}$

Designs: $(\lambda = 4 : 3)$

2.2.4. $PSL(2,19)$, $t = 4$, $k = 6$, $\lambda_{\text{trivial}} = 120$.

4-orbits: $(1 : 2; 3)$ $(2 : 3; 6)$ $(3 : 4; 6)$ $(4 : 8; 1)$ $(5 : 12; 1)$

6-orbits: $(1 : 2, 3, 4; 30)$ $(2 : 2, 3, 5; 60)$ $(3 : 2, 3, 6; 60)$ $(4 : 2, 3, 7; 30)$ $(5 : 2, 3, 8; 60)$ $(6 : 2, 3, 9; 60)$ $(7 : 2, 3, 10; 30)$ $(8 : 2, 3, 11; 10)$ $(9 : 2, 3, 12; 30)$ $(10 : 2, 3, 13; 60)$ $(11 : 2, 3, 15; 30)$ $(12 : 2, 5, 6; 30)$ $(13 : 2, 5, 8; 10)$ $(14 : 2, 5, 12; 60)$ $(15 : 2, 5, 15; 30)$ $(16 : 2, 5, 16; 30)$-$(17 : 2, 6, 12; 10)$ $(18 : 2, 6, 16; 30)$ $(19 : 3, 4, 9; 20)$

The 5 × 19 matrix $\Lambda(PSL(2,19); 4, 6)$:

									1	1	1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
8	12	16	6	12	8	8	4	8	8	6	4	2	8	2	4	2	2	0
7	12	8	6	12	10	4	2	4	10	6	2	1	10	7	5	1	7	6
4	8	10	5	8	14	5	1	5	14	5	9	2	12	4	4	2	4	4
0	12	12	0	12	12	12	0	0	0	6	6	6	12	12	12	0	6	0
0	12	12	6	12	0	0	0	12	12	0	6	0	12	6	12	6	12	0

Design: $(\lambda = 60 : 4, 5, 7, 9, 10, 11, 13, 14, 15)$.

2.2.5. $PSL(2,23)$, $t = 5$, $k = 6$, $\lambda_{\text{trivial}} = 19$.

5-orbits: $(1 : 2, 3; 30)$ $(2 : 2, 5; 30)$ $(3 : 2, 6; 30)$ $(4 : 2, 8; 30)$ $(5 : 3, 4; 30)$ $(6 : 3, 7; 30)$ $(7 : 3, 14; 30)$

6-orbits: (1 : 2, 3, 4; 30) (2 : 2, 3, 5; 60) (3 : 2, 3, 6; 60) (4 : 2, 3, 7; 60) (5 : 2, 3, 8; 60) (6 : 2, 3, 9; 60) (7 : 2, 3, 10; 60) (8 : 2, 3, 11; 60) (9 : 2, 3, 12; 30) (10 : 2, 3, 13; 10) (11 : 2, 3, 14; 30) (12 : 2, 3, 15; 60) (13 : 2, 3, 18; 60) (14 : 2, 3, 19; 60) (15 : 2, 5, 6; 30) (16 : 2, 5, 7; 30) (17 : 2, 5, 8; 30) (18 : 2, 5, 10; 30) (19 : 2, 5, 11; 60) (20 : 2, 5, 14; 30) (21 : 2, 5, 15; 30) (22 : 2, 5, 17; 30) (23 : 2, 5, 18; 30) (24 : 2, 5, 19; 60) (25 : 2, 6, 8; 60) (26 : 2, 6, 10; 10) (27 : 2, 6, 14; 60) (28 : 2, 6, 19; 30) (29 : 2, 8, 14; 10) (30 : 3, 4, 9; 20) (31 : 3, 4, 11; 30) (32 : 3, 4, 16; 30) (33 : 3, 7, 10; 10) (34 : 3, 7, 21; 10)

The 7 × 34 matrix $\Lambda(PSL(2,23);5,6)$:

									1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4
2	2	2	1	1	2	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2	1	1	1	0	0	0	0	0	0	1	1	1	2	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
0	1	1	3	0	0	1	2	1	0	0	0	1	0	1	0	0	1	1	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0
0	0	1	0	1	1	1	0	0	0	1	2	0	3	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	0	0
1	1	0	0	2	1	0	1	0	0	0	1	2	0	0	0	0	1	1	1	0	0	0	1	2	0	0	0	0	2	1	1	0	0
0	1	0	1	0	1	2	1	0	0	0	1	0	1	0	0	0	0	2	0	0	0	0	2	0	0	2	1	0	0	1	1	1	1
0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	0	1	2	0	2	0	0	0	1	1	0	0

Designs: ($\lambda = 1$: 9, 20, 32) ($\lambda = 2$: 7, 21, 22, 30) ($\lambda = 3$: 2, 11, 18, 20, 27) ($\lambda = 4$: 5, 7, 10, 18, 20, 22, 27) ($\lambda = 5$: 3, 5, 10, 13, 24, 26, 27, 28) ($\lambda = 6$: 1, 5, 7, 11, 18, 19, 20, 21, 23, 27) ($\lambda = 7$: 1, 7, 10, 12, 13, 18, 19, 20, 21, 22, 23, 27) ($\lambda = 8$: 1, 6, 7, 10, 13, 16, 18, 21, 23, 24, 25, 26, 27, 28) ($\lambda = 9$: 1, 5, 7, 8, 12, 13, 14, 16, 18, 19, 22, 23, 27)

2.2.6. $PGL(2,25)$, $t = 4$, $k = 6$, $\lambda_{\text{trivial}} = 231$.

4-orbits: (1 : 2; 3) (2 : 5; 6) (3 : 6; 6) (4 : 7; 6) (5 : 8; 2)

6-orbits: (1 : 2, 3, 4; 1) (2 : 2, 3, 5; 120) (3 : 2, 5, 6; 120) (4 : 2, 5, 7; 120) (5 : 2, 5, 8; 60) (6 : 2, 5, 9; 120) (7 : 2, 5, 10; 60) (8 : 2, 5, 11; 60) (9 : 2, 5, 13; 20) (10 : 2, 5, 15; 60) (11 : 2, 5, 16; 60) (12 : 2, 5, 17; 120) (13 : 2, 5, 18; 60) (14 : 2, 5, 19; 30) (15 : 2, 5, 20; 60) (16 : 2, 5, 21; 120) (17 : 2, 5, 22; 30) (18 : 2, 5, 23; 60) (19 : 2, 5, 24; 120) (20 : 2, 6, 8; 60) (21 : 2, 6, 9; 60) (22 : 2, 6, 10; 60) (23 : 2, 6, 11; 30) (24 : 2, 6, 12; 60) (25 : 2, 6, 21; 30) (26 : 2, 10, 12; 30) (27 : 5, 7, 12; 20) (28 : 6, 7, 15; 20)

The 5 × 28 matrix $\Lambda(PGL(2,25);4,6)$ the first part:

									1	1	1	1	1
1	2	3	4	5	6	7	8	9	0	1	2	3	4
1	40	16	16	8	8	8	8	4	8	8	8	12	4
0	12	16	16	8	20	6	8	2	12	10	20	6	5
0	12	16	16	8	8	12	8	2	6	4	20	6	2
0	12	16	20	6	20	8	6	4	8	10	12	8	6
0	12	12	0	12	24	0	12	0	0	6	12	12	0

The 5 × 28 matrix $\Lambda(PGL(2,25);4,6)$ the second part:

1	1	1	1	1	2	2	2	2	2	2	2	2	2
5	6	7	8	9	0	1	2	3	4	5	6	7	8
8	8	4	8	16	4	4	8	6	8	4	4	0	0
4	8	2	4	16	12	6	10	4	10	2	6	6	0
4	20	5	10	16	6	12	10	4	10	2	6	0	6
12	20	6	10	16	6	6	2	2	2	5	0	4	4
18	24	0	6	12	12	12	12	6	12	12	3	0	0

Designs: ($\lambda = 51 : 1, 3, 14, 17, 18, 20, 23, 25, 26$) ($\lambda = 60 : 3, 9, 11, 13, 18, 21, 22, 27$) ($\lambda = 81 : 1, 3, 7, 10, 13, 14, 15, 17, 21, 22, 25, 26, 27$) ($\lambda = 90 : 3, 4, 10, 13, 15, 18, 20, 21, 22, 23$) ($\lambda = 111 : 1, 3, 5, 7, 9, 11, 12, 13, 14, 15, 17, 18, 22, 23, 25, 26, 27$)

2.2.7. $PSL(2, 27)$, $t = 5$, $k = 6$, $\lambda_{\text{trivial}} = 23$.

5-orbits: $(1 : 2, 3; 30)$ $(2 : 2, 6; 30)$ $(3 : 3, 4; 30)$ $(4 : 3, 5; 30)$ $(5 : 3, 7; 30)$ $(6 : 3, 10; 30)$ $(7 : 3, 12; 30)$ $(8 : 3, 15; 30)$ $(9 : 4, 6; 30)$ $(10 : 4, 11; 30)$

6-orbits: $(1 : 2, 3, 4; 60)$ $(2 : 2, 3, 6; 30)$ $(3 : 2, 3, 7; 30)$ $(4 : 2, 3, 8; 30)$ $(5 : 2, 3, 9; 60)$ $(6 : 2, 3, 10; 30)$ $(7 : 2, 3, 11; 60)$ $(8 : 2, 3, 12; 60)$ $(9 : 2, 3, 13; 30)$ $(10 : 2, 3, 14; 60)$ $(11 : 2, 3, 15; 30)$ $(12 : 2, 3, 17; 60)$ $(13 : 2, 3, 18; 60)$ $(14 : 2, 3, 19; 30)$ $(15 : 2, 3, 20; 60)$ $(16 : 2, 3, 21; 30)$ $(17 : 2, 3, 22; 60)$ $(18 : 2, 3, 23; 60)$ $(19 : 2, 3, 24; 30)$ $(20 : 2, 6, 7; 60)$ $(21 : 2, 6, 12; 30)$ $(22 : 2, 6, 13; 60)$ $(23 : 2, 6, 14; 60)$ $(24 : 2, 6, 18; 60)$ $(25 : 2, 6, 20; 30)$ $(26 : 2, 6, 21; 30)$ $(27 : 3, 4, 9; 20)$ $(28 : 3, 4, 10; 60)$ $(29 : 3, 4, 11; 60)$ $(30 : 3, 4, 12; 30)$ $(31 : 3, 4, 15; 60)$ $(32 : 3, 4, 16; 60)$ $(33 : 3, 4, 17; 30)$ $(34 : 3, 4, 19; 30)$ $(35 : 3, 4, 23; 60)$ $(36 : 3, 4, 26; 30)$ $(37 : 3, 5, 12; 30)$ $(38 : 3, 5, 14; 30)$ $(39 : 3, 5, 15; 60)$ $(40 : 3, 5, 17; 60)$ $(41 : 3, 5, 18; 30)$ $(42 : 3, 5, 19; 30)$ $(43 : 3, 7, 10; 60)$ $(44 : 3, 7, 11; 30)$ $(45 : 3, 7, 17; 20)$ $(46 : 3, 7, 18; 60)$ $(47 : 3, 7, 20; 30)$ $(48 : 3, 10, 12; 60)$ $(49 : 3, 10, 14; 20)$ $(50 : 3, 10, 15; 30)$ $(51 : 3, 10, 18; 30)$ $(52 : 3, 12, 15; 30)$ $(53 : 3, 12, 18; 30)$ $(54 : 4, 6, 13; 20)$

The 10×54 matrix $\Lambda(PSL(2, 27); 5, 6)$

									1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4
2	1	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0	2	1	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	2	2	1	1	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	1	0	1	1	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	2	0	0	0	1	0	2	1	0	0	0	2	0	0	1	1	2	1	1	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	1	0	2	0	2	0	0	2	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	2	0	0	0	1	1	0	0	1	1	0	1	0	1	1	0	0	0	1	0	1	1	0	1	0	0	0	1	0	1	0	1	1	0	0	1	1	0
1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	2	0	0	1	1	0	0	0	0	1	0	2	0	1	0	1	0	0	1	0	0	0	1	1	1	1	0
0	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	2	0	0	2	0	2	0	0	1	0	1	2
0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0	1	2	0	0	1	1	1	0	1	1	2	0	1	0	1	0	0

Designs: ($\lambda = 2 : 7, 22, 39, 49$) ($\lambda = 3 : 2, 9, 14, 28, 33, 40, 48$) ($\lambda = 4 : 6, 7, 11, 22, 25, 26, 30, 34, 39, 45, 54$) ($\lambda = 5 : 2, 7, 9, 14, 22, 28, 33, 39, 40, 48, 49$) ($\lambda = 6 : 3, 4, 6, 11, 16, 19, 21, 25, 26, 27, 30, 33, 34, 39, 44, 45, 47, 51, 53, 54$) ($\lambda = 7 : 2, 6, 7, 9, 11, 14, 22, 25, 26, 28, 30, 33, 34, 39, 40, 45, 48, 54) ($\lambda = 8$: 3, 4, 6, 7, 11, 16, 19, 21, 22, 25, 26, 27, 29, 32, 37, 38, 39, 43, 45, 49, 52, 54) ($\lambda = 9$: 2, 3, 4, 6, 9, 11, 14, 16, 19, 21, 25, 26, 27, 28, 30, 33, 34, 36, 39, 40, 44, 45, 47, 48, 51, 53, 54) ($\lambda = 10$: 1, 5, 8, 10, 13, 15, 17, 18, 20, 24, 30, 33, 34, 36, 38, 39, 42, 44, 45, 47, 51, 53, 54) ($\lambda = 11$: 2, 3, 4, 6, 7, 9, 11, 14, 16, 19, 21, 22, 25, 26, 28, 29, 31, 32, 35, 39, 40, 43, 46, 48, 49).

2.2.8. $PSL(2,31)$, $t = 5$, $k = 6$, $\lambda_{\text{trivial}} = 27$.

5-orbits: (1 : 2, 3; 30) (2 : 2, 5; 30) (3 : 2, 6; 30) (4 : 2, 8; 30) (5 : 2, 9; 30) (6 : 2, 18; 30) (7 : 3, 4; 30) (8 : 3, 7; 30) (9 : 3, 8; 30) (10 : 3, 10; 30) (11 : 4, 6; 30) (12 : 4, 9; 30) (13 : 5, 6; 10) (14 : 5, 7; 30) (15 : 12, 13; 6)

6-orbits: (1 : 2, 3, 4; 30) (2 : 2, 3, 5; 60) (3 : 2, 3, 6; 30) (4 : 2, 3, 7; 60) (5 : 2, 3, 8; 60) (6 : 2, 3, 9; 60) (7 : 2, 3, 10; 60) (8 : 2, 3, 11; 60) (9 : 2, 3, 13; 60) (10 : 2, 3, 14; 60) (11 : 2, 3, 15; 60) (12 : 2, 3, 16; 30) (13 : 2, 3, 17; 10) (14 : 2, 3, 18; 30) (15 : 2, 3, 19; 60) (16 : 2, 3, 20; 60) (17 : 2, 3, 21; 60) (18 : 2, 3, 22; 60) (19 : 2, 3, 24; 60) (20 : 2, 3, 25; 60) (21 : 2, 3, 26; 60) (22 : 2, 3, 27; 60) (23 : 2, 3, 28; 30) (24 : 2, 5, 6; 60) (25 : 2, 5, 7; 60) (26 : 2, 5, 8; 60) (27 : 2, 5, 9; 30) (28 : 2, 5, 10; 60) (29 : 2, 5, 12; 60) (30 : 2, 5, 13; 60) (31 : 2, 5, 14; 60) (32 : 2, 5, 15; 30) (33 : 2, 5, 18; 60) (34 : 2, 5, 19; 20) (35 : 2, 5, 21; 60) (36 : 2, 5, 23; 60) (37 : 2, 5, 24; 30) (38 : 2, 5, 25; 60) (39 : 2, 5, 27; 60) (40 : 2, 5, 28; 30) (41 : 2, 6, 7; 30) (42 : 2, 6, 8; 60) (43 : 2, 6, 9; 60) (44 : 2, 6, 10; 30) (45 : 2, 6, 12; 30) (46 : 2, 6, 18; 60) (47 : 2, 6, 21; 60) (48 : 2, 6, 23; 60) (49 : 2, 6, 26; 30) (50 : 2, 6, 27; 30) (51 : 2, 6, 28; 60) (52 : 2, 8, 10; 30) (53 : 2, 8, 13; 60) (54 : 2, 8, 18; 60) (55 : 2, 8, 21; 60) (56 : 2, 8, 26; 30) (57 : 2, 9, 13; 60) (58 : 2, 9, 21; 30) (59 : 2, 9, 27; 60) (60 : 2, 9, 28; 30) (61 : 2, 18, 21; 30) (62 : 2, 18, 26; 30) (63 : 3, 4, 9; 30) (64 : 3, 4, 10; 60) (65 : 3, 4, 11; 30) (66 : 3, 4, 12; 30) (67 : 3, 4, 15; 60) (68 : 3, 4, 23; 30) (69 : 3, 4, 24; 30) (70 : 3, 4, 25; 60) (71 : 3, 4, 26; 30) (72 : 3, 7, 8; 30) (73 : 3, 7, 15; 12) (74 : 3, 7, 20; 10) (75 : 3, 7, 23; 20) (76 : 3, 8, 12; 30) (77 : 3, 8, 14; 30) (78 : 3, 8, 18; 60) (79 : 3, 8, 22; 30) (80 : 3, 10, 18; 10) (81 : 4, 6, 17; 10) (82 : 5, 7, 23; 10) (83 : 5, 7, 29; 12)

The 15×83 matrix $\Lambda(PSL(2,31); 5, 6)$:

```
         11111111112222222222333333333344444444445555555555666666666677777777778888
12345678901234567890123456789012345678901234567890123456789012345678901234567890123
22111112111111111211111000000000000000000000000000000000000000000000000000000000000
01100000011000011011000111111121112111110000000000000000000000000000000000000000000
01121000201100000000000110001001100000001121211111100000000000000000000000000000000
00001112100000001011100102120000000101000101000100021111000000000000000000000000000
00000110110000110100001101111100110001010020000000101000211100000000000000000000000
00000000000001102100121000001000102001100000011011000111011012000000000000000000000
10001110100000001011100000000000000110000100020000001010000100111121111000000000000
00010001001000110001010000000011100000001010100100002020001000000101101011110000000
01011000010000010100110010000200000110100000000001101000001100000101000100011210000
00000020000000110010010010000102010000001000000011100011001000110000110000000221000
01001000000110000000200010001000100001000101011000100100200000110000110000020200100
00000100010000010001000002020010001000100010002100110100100000110020011100001000000
00000000000000000000000030000000000003003000000033000300000030030000003000000000000
00010101011000100000010010000100001200000200000000000100110011010110100000101000011
00000000000000000000000005000000000005500000000005000000005000000000000010000000001
```

Designs: $(\lambda = 6 : 8, 13, 15, 16, 21, 25, 26, 29, 30, 38, 45, 46, 49, 67, 71, 76, 80, 83)$ $(\lambda = 12 : 1, 2, 5, 6, 9, 10, 13, 15, 16, 22, 25, 26, 28, 29, 31, 33, 36, 37, 41, 42, 46, 47, 49, 53, 54, 55, 58, 59, 70, 73, 78, 83)$.

2.2.9. $PGL(2,32)$, $t = 4$, $k = 5$, $\lambda_{\text{trivial}} = 29$.

4-orbits: $(1 : 2; 6)$ $(2 : 4; 6)$ $(3 : 6; 6)$ $(4 : 14; 6)$ $(5 : 16; 6)$

5-orbits: $(1 : 2, 3; 15)$ $(2 : 2, 5; 60)$ $(3 : 2, 6; 60)$ $(4 : 2, 8; 60)$ $(5 : 2, 9; 60)$ $(6 : 2, 11; 60)$ $(7 : 2, 12; 15)$ $(8 : 4, 5; 15)$ $(9 : 4, 17; 60)$ $(10 : 6, 14; 15)$ $(11 : 14, 22; 15)$

The 5×11 matrix $\Lambda(PGL(2,32); 4, 5)$: $$\begin{pmatrix} 4&4&8&4&4&4&1&0&0&0&0\\ 1&4&0&8&4&4&0&4&4&0&0\\ 0&0&4&4&4&4&4&0&8&1&0\\ 0&8&4&4&4&0&0&0&4&4&1\\ 0&4&4&0&4&8&0&1&4&0&4 \end{pmatrix}$$

Designs: $(\lambda = 4 : 5)$ $(\lambda = 5 : 1, 7, 8, 10, 11)$ $(\lambda = 9 : 1, 5, 7, 8, 10, 11)$

2.2.10. $PGL(2,37)$, $t = 4$, $k = 5$, $\lambda_{\text{trivial}} = 34$.

4-orbits: $(1 : 2; 3)$ $(2 : 3; 6)$ $(3 : 4; 6)$ $(4 : 5; 6)$ $(5 : 6; 6)$ $(6 : 8; 6)$ $(7 : 11; 2)$

5-orbits: $(1 : 2, 3; 30)$ $(2 : 2, 5; 60)$ $(3 : 2, 6; 60)$ $(4 : 2, 7; 15)$ $(5 : 2, 8; 60)$ $(6 : 3, 4; 30)$ $(7 : 3, 7; 60)$ $(8 : 3, 12; 60)$ $(9 : 3, 14; 30)$ $(10 : 3, 15; 30)$ $(11 : 3, 26; 60)$ $(12 : 4, 5; 30)$ $(13 : 4, 11; 10)$ $(14 : 4, 17; 30)$ $(15 : 5, 8; 30)$

The 7×15 matrix $\Lambda(PGL(2,37); 4, 5)$:

$$\begin{pmatrix} 8&8&8&2&8&0&0&0&0&0&0&0&0&0&0\\ 4&0&4&0&0&4&8&4&4&2&4&0&0&0&0\\ 2&4&0&0&4&4&0&8&0&0&4&4&2&2&0\\ 0&8&4&0&0&2&4&0&0&4&4&4&0&0&4\\ 0&0&4&4&4&0&4&4&0&4&4&2&0&4&0\\ 0&4&4&0&4&0&0&4&6&0&4&0&0&4&4\\ 0&0&0&0&12&0&12&0&0&0&0&0&4&0&6 \end{pmatrix}$$

Design: $(\lambda = 16 : 2, 3, 7, 8, 13, 14)$

2.2.11. $PSL(2,47)$, $t = 4$, $k = 5$, $\lambda_{\text{trivial}} = 44$.

4-orbits: $(1 : 2; 3)$ $(2 : 3; 6)$ $(3 : 4; 6)$ $(4 : 5; 3)$ $(5 : 6; 6)$ $(6 : 7; 6)$ $(7 : 10; 3)$ $(8 : 11; 6)$ $(9 : 13; 3)$ $(10 : 22; 3)$

5-orbits: $(1 : 2, 3; 30)$ $(2 : 2, 5; 30)$ $(3 : 2, 6; 30)$ $(4 : 2, 7; 30)$ $(5 : 2, 8; 30)$ $(6 : 2, 10; 30)$ $(7 : 2, 12; 30)$ $(8 : 2, 13; 30)$ $(9 : 2, 14; 30)$ $(10 : 2, 16; 30)$ $(11 : 3, 4; 30)$ $(12 : 3, 7; 30)$ $(13 : 3, 8; 30)$ $(14 : 3, 11; 30)$ $(15 : 3, 12; 30)$ $(16 : 3, 13; 30)$ $(17 : 3, 14; 30)$ $(18 : 3, 15; 30)$ $(19 : 3, 17; 30)$ $(20 : 3, 19; 30)$ $(21 : 3, 20; 30)$ $(22 : 3, 22; 30)$ $(23 : 3, 26; 30)$ $(24 : 3, 39; 30)$ $(25 : 4, 9; 30)$ $(26 : 4, 13; 30)$ $(27 : 4, 19; 30)$ $(28 : 4, 20; 30)$ $(29 : 4, 21; 30)$ $(30 : 4, 27; 30)$ $(31 : 5, 8; 30)$ $(32 : 6, 10; 30)$ $(33 : 7, 11; 30)$

The 10×33 matrix $\Lambda(PSL(2,47); 5, 6)$:

									1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3
8	4	4	4	8	4	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	2	0	0	0	0	0	0	2	4	8	8	2	2	2	4	2	2	2	2	4	2	2	0	0	0	0	0	0	0	0	0
2	2	0	0	4	0	0	2	0	0	4	0	0	0	4	0	4	0	0	0	2	0	2	0	4	4	2	2	4	2	0	0	0
0	4	4	0	0	4	4	0	0	0	0	0	0	8	4	4	0	4	4	0	0	0	0	0	0	0	0	4	4	0	4	0	0
0	0	2	4	4	0	0	0	0	2	0	4	4	0	0	4	4	0	0	0	2	0	2	0	0	0	4	4	0	4	2	2	0
0	0	0	4	4	2	2	0	2	0	2	4	4	2	0	0	0	4	4	4	0	0	0	2	2	2	0	0	2	4	0	0	4
0	4	0	0	0	4	4	0	0	4	0	0	0	0	0	0	0	4	4	0	0	4	8	0	0	4	4	0	0	0	0	4	4
0	2	0	0	4	2	2	2	2	0	0	4	4	2	0	0	8	0	0	0	0	2	0	2	4	0	0	0	0	0	4	4	2
0	0	0	0	0	0	0	4	4	4	0	0	0	0	4	4	0	0	0	4	0	0	0	8	0	4	4	0	0	0	4	0	0
0	0	4	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	4	8	4	0	0	0	0	0	4	4	0	0	4	4

Designs: $(\lambda = 8 : 5, 14, 21, 23, 24)$ $(\lambda = 12 : 2, 4, 10, 14, 17, 20, 21, 26)$ $(\lambda = 16 : 3, 6, 8, 9, 11, 14, 17, 20, 23, 27, 30)$ $(\lambda = 20 : 3, 4, 5, 8, 11, 14, 18, 21, 22, 23, 24, 25, 27, 31)$

2.3. Some observations on the constructed designs

In this section we give some miscalleneous data concerning the constructed designs and the construction itself.

The designs for $q \in \{17, 32\}$ were considered in more detail in [2]; just a few data are mentioned here. The construction for $q = 17$ is due to Alltop and was described in [5], Example 8.5, pp. 186-187; Λ-technique is an improvement of the Alltop's construction. The design constructed for $q = 32$ and $\lambda = 5$ is the first member of an Alltop's infinite class of 4-designs. It is likely that all the constructed designs for $q = 32$ can arise ([8]) by action of the 4-homogeneous group $PGamaL(2, 32)$.

The designs with $q = 11$ and $q = 23$ are related to the well-known ([5]) Steiner systems $S(5, 6; 12)$ and $S(5, 6; 24)$ (that is, to the 5-$(12, 6, 1)$ design and to the 5-$(24, 6, 1)$ design). The first one of these Steiner systems is, as stated in [7], Theorem 2.26., the uniquely determined Steiner system $S(5, 6; 12)$, with the automorphism group isomorphic to the famous Mathieu 5-transitive group M_{12} of cardinality $8 \cdot 9 \cdot 10 \cdot 11 \cdot 12$.

The brute-force search over the colums of Λ-matrices was applicable on a PC-386 computer in the cases when the number n of columns was restricted to 30 ($n = 30$ required one week of computing time and each added unit to n would double the time required). The following shortcut was used for $q = 23$ and $q = 47$, where n is equal to 34 and 33 respectively:

It is observed that there exist in both cases several pairs of duplicate columns within the Λ-matrix (exactly four pairs with $q = 23, n = 34$ and three pairs with $q = 47, n = 33$). The search is performed over the reduced 30-column matrices, which are obtained from the Λ-matrices by discarding one of the columns from each duplicate pair. Such a reduction does not guarantee completeness of the search; it might happen that some of the existing designs require combinations of columns which include both columns in a duplicate pair. However, the arguments related to the specific coefficients of the two Λ-matrices show that no set of design parameters is missed in this way.

For example, the set of λ values with $q = 23$ is complete (all the values in the interval $[1,...,18 = \lambda_{\text{trivial}} - 1]$ are present). Similarly, all the elements in the first row of the Λ-matrix for $q = 47$ are divisible by 4, which implies that all the corresponding values of λ must be divisible by 4; an additional argument shows that $\lambda = 4$ is impossible.

The Λ-matrices with $q = 27$ and $q = 31$ have very large numbers of columns (54 and 83 respectively), so there is no chance for a full search. However, ad hoc designed heuristic approaches ([3]) have given designs with all the possible values of λ in these cases.

The number of successful (that is, design-corresponding) combinations of columns is very large with some of the Λ-matrices (several hundreds with $q = 19$ and $q = 47$ and several thousands with $q = 23$).

Some of the obtained parameters seem to be particularly interesting. For example, the designs constructed for $q = 37$ seem ([8]) to be the first 4-designs known on 38 points.

A design isomorphism search was performed ([6]) among the constructed 560 4-designs on 48 points for $\lambda \in \{8, 12, 16, 20\}$. Auxiliary graphs were attached to the designs so that non-isomorphism of some two attached graphs implies non-isomorphism of the corresponding designs. Global results of this search seem to be very interesting. All the equivalence classes of isomorphic attached graphs are of cardinality 2; this implies that at least one half of the total number of the constructed designs are pairwise non-isomorphic. Moreover, the unique and involutory (a product of transpositions) isomorphism maps onto each other the two graphs of each one of the equivalence classes; this means that the recognized isomorphism is a global symmetry of the whole found class of 4-designs.

Finally, it seems worth-while to try an isomorphism search for $q = 19$. It is only in this case that there exists a unique (and self-complementary) value $\lambda = 60$. Is the 4-(20,6,60) design unique up to an isomorphism within the class of designs with these parameters generated by $PSL(2, 19)$? The isomorhism search in this case might use attached hypergraphs with edges containing three vertices each.

References

[1] D. M. Acketa and V. Mudrinski, *A 4-design on 38 points*, (submitted).

[2] D. M. Acketa, V. Mudrinski and Dj. Paunić, *A search for 4-designs arising by action of $PGL(2,q)$*, Publ. Elektrotehn. fak., Univ. Beograd, Ser. Mat. **5** (1994), 13–18.

[3] D. M. Acketa and V. Mudrinski, *Some 5-designs on 28 and 32 points*, (submitted).

[4] W. O.Alltop, *An infinite class of 4-designs*, J. Comb. Th. **6** (1969), 320–322.

[4] T. Beth, D. Jungnickel and B. Lenz, *Design theory*, Bibliographisches Institut, Mannheim/Wien/Zürich, 1985.

[5] S. Dautović, D. M. Acketa and V. Mudrinski, *A graph approach to isomorphism testing of 4-(48,5,λ) designs arising from $PSL(2,47)$*, (submitted).

[6] D. Gorenstein, *Finite Simple Groups, An Introduction to Their Classification*, Plenum Press,. New York, London, 1982.

[7] Z. Janko and V. Tonchev, private communication.

Institute of Mathematics, Trg Dositeja Obradovića,
21000 Novi Sad, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 591-597
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ORDERED LINEAR RESOLUTION AS THE BASE OF THE SYSTEM FOR AUTOMATIC THEOREM PROVING

Ivana F. Berković

Abstract. U radu se opisuje uređena linearna rezolucija sa markiranim literalima i njene specifičnosti. Da bi se očuvala potpunost metode izvršena je modifikacija algoritma za određivanje rezolvente.

Na bazi modifikovane uređene linearne rezolucije sa markiranim literalima izgrađen je sistem za automatsko dokazivanje teorema. Sistem je implementiran na PC - računaru i dopušta varijabilne strategije pretraživanja. Pretpostavke i tvrđenja koje treba dokazati, zapisuju se odgovarajućim formulama predikatskog računa prvog reda. U radu se daje opis implementiranog sistema za automatsko dokazivanje teorema, prikazuju se njegove karakteristike i oblasti primene. Posebno se razmatra odnos ovakvog automatskog dokazivača teorema i Prolog-a.

1. Introduction

Automated reasoning is very important area in Artificial Intelligence, but the common sense is difficult to model in a computer. The needed knowledge is not easy to represent. Another problem is how it can deduce something from a set of facts, or how it can prove that a conclusion follows from a given set of premises. Computational logic, based on formulations by some formal-language (propositional logic, predicate logic), provides problem-solving methods.

The developing of theorem-proving can be divided in two directions. The first direction is pure automated theorem proving, which is mostly resolution-based. The other approach is non-resolution-based theorem proving or natural deduction, which includes some heuristics and user-supplied knowledge.

2. The Rule of Ordered Linear (OL) Resolution with Marked Literals

The most popular method for automatic theorem proving is the resolution method, which is discovered by J. A. Robinson in 1965 ([2], [5]). Resolution method is a syntactic method of deduction. This procedure is a general automatic method for determining if a theorem (conclusion) follows from a given set of premises (axioms). Each formula will be transformed to the clauses form. *Reduction ad absurdum* is in the basis of resolution method. Resolution rule will be applied on the set of clauses (axioms) which was expanded by negating the desired conclusion in clause form.

Since 1965., many resolution forms and techniques are developed because the pure resolution rule has been unable to handle complex problems. Also, many resolution theorem provers are created.

Ordered Linear (OL) resolution rule with marked literals ([6]) increases efficiency and doesn't disturb completeness of pure resolution rule.
The generating process of OL-resolvent from central clause (d1) and auxiliary clause (d2):

1. Redesignate variables (without common.variables in the clauses).
2. Determine universal unificator Θ for last literal of d1 and k-literal ($k = 1, 2, ...$) of d2 (if it exists for some k, else it is impossible to generate OL-resolvent for specification clauses).
3. Create resolvent with marked last literal in d1Θ and add the rest of clause d2Θ without k-literal (d1Θ and d2Θ are clauses, which were formed by universal unificator Θ applied on d1 and d2, respectively).
4. Eliminate identical non-marked literals and tautology examination (tautologies are not memorized).
5. The Shortening Operation (delete all ending marked literals).
6. The Compressing Operation (delete the last non-marked literal, which is complemented in relation to negation, with some marked literal for unificator λ).
7. Repeat steps: 5 and 6 until the empty clause is got, or the Compressing Operation is not applied on the last non-marked literal.

The final result of this process is: the forming one OL-resolvent from central clause (d1) and auxiliary clause (d2).

To preserve completeness of the OL-resolution rule with marked literals, some resolvents have to be memorized.

The rule of OL-resolution with marked literals is separated in two parts: in-resolution and pre-resolution. The steps: 1 - 5 are represented in-resolution. The steps: 6 - 7 are represented pre-resolution. Mid-resolvents are the products of in-resolution and without their memorizing, the completeness of the method can be lost. It can be illustrated by example.

Example 1.
Central clause is: $\neg R(A)$
Auxiliary clauses are:
a1 $\neg S(X, Y) \vee R(X) \vee R(Y)$
a2 $S(A, B)$
a3 $\neg R(B)$
where: X and Y are variables, A and B are constants, R and S are predicates, $\neg$ is negation.

The results without memorizing mid-resolvents are:

Taking into consideration the central clause $\neg R(A)$ and the auxiliary clause a1 by literal $R(X)$ and literal $R(Y)$, one resolvent: $/\neg R(A) \vee \neg S(A, A)$ is generated at the first level. This resolvent has not produced new resolvents and it is not possible to generate empty clause.

Not to lose the completeness of the method some resolvents must be memorized that are got during the resolution procedure.
The results with memorizing mid-resolvents are:

Three resolvents are generated at the first level:
1. $/\neg R(A) \vee \neg S(A, Y) \vee R(Y)$ from $\neg R(A)$ and a1 by literal $R(X)$
2. $/\neg R(A) \vee \neg S(X, A) \vee R(X)$ from $\neg R(A)$ and a1 by literal $R(Y)$
3. $/\neg R(A) \vee \neg S(A, A)$ from 1., or 2. with pre-resolution.

There are two resolvents generated at the second level:
4. $/\neg R(A) \vee \neg S(A, B)$ from 1. and a3 with in-resolution
5. $/\neg R(A) \vee \neg S(B, A)$ from 2. and a3 with in-resolution.

Empty clause is generated at the third level from $/\neg R(A) \vee \neg S(A, B)$ and a2. The set of clauses is contradictory.

From the point of scientific researching this example shows that some resolvents have to be memorized to preserve the completeness of the method. This modification of Ordered Linear resolution rule is served as the base for development of the system for automatic theorem proving ADT.

3. The System for Automatic Theorem Proving ADT

In our country, the first resolution theorem-prover is developed in a scope of GRAPH expert system at the Faculty of Electrical Engineering in Belgrade, ([8]).

The system ADT is based on the resolution rule. The system is developed at Technical Faculty "Mihajlo Pupin" in Zrenjanin. ADT is a system for automatic theorem proving, which is implemented on PC - computer by Pascal (Turbo Pascal ver. 6.0) programming language. The rule of Ordered Linear Resolution with marked literals presents the system base, ([6]).

ADT system differs from the other resolution-based theorem-provers which are characterized by one fixed strategy. The system permits various syntactic search strategies, ([2], [3], [5]).

The system ADT disposes three search strategies: breadth-first, depth-first and their combination. The first and the second strategy are common blind search procedures. The third blind search procedure is constructed as their combination.

In breadth-first search are the nodes starting with the root node of the search tree. They all are generated level by level. In depth-first search, a new node is generated at the next level, from the one current, and the search is continuing deeper and deeper in this way until it is forced to backtracking. In combine-search, the nodes of the search tree are generated and examined in the breadth, until the fulfilling of the level. Then the procedure is backing one level up and continues in depth with backtracking.

The system ADT permits comparisons of strategies. It is also possible to use various strategies to find the proof, especially if it can not be detected by means of other ones.

ADT is projected for scientific - researching, teaching and practical purpose. Some results of the experimental work with ADT system are described in ([3]).

There are many different possibilities for using the system in education. ADT can be used for learning the elements of theorem-proving. It allows the illustration of the Unification Algorithm or the Resolution Rule. It is also possible to use this system for experimental work in: deduction of proofs, comparison of strategies, influence of various factors on efficiency proving.

The methods of automatic theorem proving can be applied in various domains of artificial intelligence. They are applicable in fields as mathematical theorem proving, expert systems, question-answering systems, automatic programming, program verification, situational control and decision, relation data bases, logical programming, etc. It is presented in some concrete examples ([3]).

This system is incorporated in the system for automatic creating of the combinatorial disposition DEDUC ([11]), where it has presented the satisfying practical efficiency. ADT system is the basic generating mechanism in DEDUC system. DEDUC system is aimed to automated creating time-table. It is implemented on PC computer.

4. ADT system and PROLOG

Specific high-level languages have been developed for different application domains. PROLOG and LISP are the most famous programming languages in artificial intelligence.

The logical programming language PROLOG and ADT system are compared.

PROLOG is a logic-oriented language ([4], [10]), which contains a resolution-based theorem-prover. The theorem-prover in PROLOG appears with the depth-first search approach. The first-order predicate logic is the form of representation in PROLOG. Programs in PROLOG consists of axioms (clauses, facts) and a theorem to be proved (goal). The axioms are restricted in "Horn clause" form.

The first-order logic is the form of representation in ADT system, too. But, this system has not restriction in "Horn clause". It appears with clauses. The axioms are presented by auxiliary clause. The central clause is negating the theorem to be proved.

PROLOG has the negation defect. This defect is corrected in ADT system. It can be illustrated by example.

Example 2.

Program in PROLOG:

```
vegetarian(tom).
vegetarian(ivan).
vegetarian(isak).
smoker(tom).
smoker(isak).
ana_likes(X1) :  not (smoker(X1)), vegetarian(X1).
```

PROLOG-system gives unconnected answer on following questions:

```
?- ana_likes(X1).
no
?- ana_likes(ivan).
yes
```

If the last clause is now:

```
ana_likes(X1) :  vegetarian(X1), not (smoker(X1)).
```

PROLOG-system gives wrong answers on following questions:

```
?- ana_likes(X1).
X1=ivan
?- ana_likes(ivan).
yes
```

These answers are incorrect because we have not data about Ivan and smoking. We don't know is Ivan a smoker or not. The correct answer will be: "I don't know".

In both cases ADT system gives the correct answer: "I don't know". In fact, ADT system generates only one resolvent and can not complete the proof with none of the three strategies.

ADT system allows recursion using (example with family relationship, [3]) and works with structures and lists, as well as PROLOG.

5. Conclusion

Completeness and universality of the resolution method, as the base of ADT system, enables it to be applied in various domains of artificial intelligence. In the scientific researching is given an example which shows that some resolvents must be memorized to preserve the completeness of this method. The relationship between ADT system and PROLOG are emphasized. In this sense, the further development and applications of this system is possible. The system is convenient for teaching and has the practical purposes.

References

[1] Albus J.S., *Outline for a Theory of intelligence*, IEEE Transactions on Systems, Man, and Cybernetics, New York **21** (1991), no. 3, 473-509.

[2] Barr A., Cohen P.R., Feigenbaum E.A., *The Handbook of Artificial Intelligence, Vol.I,II,III*, Heuris Tech Press, W. Kaufmann, Inc., California, 1982.

[3] Berković I., *Variable searching strategies in the teaching aimed system for automatic theorem proving*, M.Sci. Thesis, Technical Faculty "M. Pupin", Zrenjanin, 1994. (Serbian)

[4] Bratko I., *Prolog programming for Artificial Intelligence*, Addison–Wesley Publ. Comp., 1986.

[5] Gevarter W.B., *Intelligent Machines*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1985.

[6] Hotomski P., Pevac I., *Mathematical and program problems of Artificial Intelligence in the field of automatic theorem proving*, Naucna knjiga, Belgrade, 1988 (1991). (Serbian)

[7] Hotomski P., *Deductive approach to automatic creating of combinatorial disposition*, Proc. of the VI Conf. on Logic and Comp. Sci. LIRA'92, Novi Sad (1992), 35–42. (Russian)

[8] Hotomski P., *Systems of Artificial Intelligence*, Technical Faculty "M. Pupin", Zrenjanin, 1994. (Serbian)

[9] Jones M., *Applications of artificial intelligence within education*, Comp.& Maths. with Appls. **11** (1985), no. 5, 517-526.

[10] Kluzniak F., Szpakowicz S., Bien J., *Prolog for programmers*, Academic Press Inc., Orlando, 1985.

[11] Madžarević T., Cvetković D., *Some supplements to the strategy for a proof directing in the theorem prover of the expert system Graph*, Proc. of SYM-OP-IS '93, Belgrade (1993), 31-34. (Serbian)

[12] Prohaska D., *A deductive method for the combinatorial disposition generating and its program algorithmisation*, M.Sci. Thesis, Technical Faculty "M. Pupin", Zrenjanin, 1993. (Serbian)

[13] Winston P. H., *Artificial Intelligence*, Addison Wesley Publsh. Comp., 1984.

University of Novi Sad, Technical Faculty "Mihajlo Pupin", 23000 Zrenjanin, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 599–602
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

HADAMARD'S INEQUALITY AND FIXED-POINT METHOD

Momčilo Bjelica

ABSTRACT. The famous inequality in matrix theory of J. Hadamard has different proofs and extensions [1, 6]. Here given proof is by the method of common fixed-point of mappings monotonic with respect to a functional, which can be applied to many, including all main inequalities [3]. Condition for equality: rows of a matrix are orthogonal or at least one of them is zero, is replaced by proportionality (appearing in numerous other inequalities) between rows of a matrix and corresponding rows of cofactors.

Theorem. *Let $A = (a_{ij})$ be a real square matrix and $|A|$ be it's determinant, then*

$$|A|^2 \le \prod_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right). \tag{1}$$

Equality in (1) *holds if and only if*

$$a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{in}a_{jn} = 0, \tag{2}$$

for each pair of different i, j, or if at least one factor on the right side of (1) *is equal to zero.*

Condition (2), *including the disjunct, can be replaced by the next one: there are numbers λ_i, μ_i, $\lambda_i^2 + \mu_i^2 \ne 0$, $1 \le i \le n$, such that*

$$\lambda_i a_{ij} + \mu_i A_{ij} = 0, \qquad 1 \le j \le n, \tag{3}$$

where A_{ij} are cofactors.

Proof. Define the space (product of n-spheres)

$$\mathcal{X} = \left\{ X = (x_{ij}) \,\middle|\, \sum_{j=1}^{n} x_{ij}^2 = \sum_{j=1}^{n} a_{ij}^2,\ 1 \le i \le n \right\} \tag{a}$$

and the functional $f : \mathcal{X} \to \mathbb{R}$

$$\text{(b)} \qquad f(X) = |X|.$$

Define mappings $F_i : \mathcal{X} \to \mathcal{X}$, $1 \le i \le n$

$$\text{(c)}\ F_i(X) = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i1} & y_{i2} & \dots & y_{in} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix}, \qquad y_{ij} = \frac{r_i}{R_i} X_{ij}, \qquad 1 \le j \le n,$$

$$r_i = \sqrt{x_{i1}^2 + x_{i2}^2 + \dots + x_{in}^2}, \qquad R_i = \sqrt{X_{i1}^2 + X_{i2}^2 + \dots + X_{in}^2}.$$

If $r_i = 0$, then $F_i(X) = X$; define the same if $R_i = 0$. The row $(y_{ij})_j$ is defined to be proportional to the corresponding row of cofactors $(X_{ij})_j$ and that $F_i(X) \in \mathcal{X}$. The mapping F_i is monotonic nondecreasing with respect to the functional f

$$\text{(d)} \qquad f(X) \le f(F_i(X)),$$

by Laplace development (d) is equivalent to

$$\begin{aligned} &x_{i1}X_{i1} + x_{i2}X_{i2} + \dots + x_{in}X_{in} \le y_{i1}X_{i1} + y_{i2}X_{i2} + \dots + y_{in}X_{in}, \\ \text{(4)} \qquad &x_{i1}X_{i1} + x_{i2}X_{i2} + \dots + x_{in}X_{in} \le r_i R_i. \end{aligned}$$

The Cauchy inequality (4) is equality [1] if and only if there are numbers λ_i, μ_i, not both 0, such that

$$\lambda_i x_{ij} + \mu_i X_{ij} = 0, \qquad 1 \le j \le n.$$

If $\lambda_i \mu_i \ne 0$, then equality in (d) holds if and only if sets $(x_{ij})_j$ and $(X_{ij})_j$ are proportional, or equivalently $x_{ij} = y_{ij}$, $1 \le j \le n$. Hence, F_i is strictly monotonic with respect to f: equality in (d) holds if and only if X is a fixed point of mapping $F_i(X) = X$, and for non-fixed points strict inequality holds. On the compact set $\mathcal{X}$ the functional f attains maximal value, and, because of strict monotonicity, it is attained on a set $\mathcal{F}$ of common fixed points of mappings F_i, $1 \le i \le n$. The set $\mathcal{F}$ is not empty, since it contains, e.g., diagonal matrix with diagonal r_i, $1 \le i \le n$. If $X \in \mathcal{F}$, $|X| \ne 0$, $X_{ij} = c_i x_{ij}$, $1 \le i, j \le n$, then

$$|X| = c_i r_i^2, \qquad 1 \le i \le n, \qquad |X|^n = \prod_{i=1}^{n} c_i r_i^2.$$

On the other hand

$$|X| = \left|\left(\frac{X_{ij}}{c_i}\right)\right| = \frac{1}{c_1c_2\dots c_n}|(X_{ij})| = \frac{1}{c_1c_2\dots c_n}||X|X^{-1}| = \frac{1}{c_1c_2\dots c_n}|X|^{n-1}$$
$$|X|^n = c_1c_2\dots c_n|X|^2,$$

so that

$$|X|^2 = \prod_{i=1}^n r_i^2 = \prod_{i=1}^n\Big(\sum_{j=1}^n a_{ij}^2\Big).$$

The equivalence between conditions (2), including the disjunct, and (3), both determining the same set of matrices on which the equality in (1) holds, follows from earlier proofs of the theorem and this one. However, we give a direct proof that (2) $\Leftrightarrow$ (3).

(2) $\Rightarrow$ (3). If $a_{ij} = 0$, $1 \le j \le n$, then $A_{kj} = 0$, $1 \le k \le n$, $k \ne i$, $1 \le j \le n$ and (3) holds. Let A be an orthogonal matrix with no one zero row, then lineal over rows of the matrix A

$$\mathcal{L}(\,\{(a_{i1}, a_{i2}, \dots, a_{in}) \mid 1 \le i \le n\}\,)$$

is n-dimensional vector space. Also

$$(a_{i1}, a_{i2}, \dots, a_{in}) \;\perp\; \mathcal{L}(\,\{a_{j1}, a_{j2}, \dots, a_{jn}) \mid 1 \le j \le n,\ j \ne i\}\,).$$

From

$$\text{(5)}\qquad a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn} = \delta_{ij}|A|,$$

follows

$$(A_{i1}, A_{i2}, \dots, A_{in}) \;\perp\; \mathcal{L}(\,\{a_{j1}, a_{j2}, \dots, a_{jn}) \mid 1 \le j \le n,\ j \ne i\}\,).$$

Vectors $(a_{i1}, a_{i2}, \dots, a_{in})$ and $(A_{i1}, A_{i2}, \dots, A_{in})$ in n-dimensional space are orthogonal to the same hyperplane ($(n-1)$-variety) and therefore they are collinear.

(3) $\Rightarrow$ (2). If in (3) some $\mu_i = 0$, then $a_{ij} = 0$, $1 \le j \le n$, that is, the disjunct in (2) holds. Hence, suppose that $\mu_i \ne 0$, $1 \le i \le n$, i. e., rows of A are not zero-vectors. If $\lambda_i \ne 0$, $1 \le i \le n$, then from (5) follows (2). Now, without lose of generality, suppose that $\lambda_k = 0$, $1 \le k \le i$; $\lambda_l \ne 0$, $i < l \le n$. From $A_{kj} = 0$, $1 \le k \le i$, $1 \le j \le n$ follows that $|A| = 0$ and that rows of A are linearly dependent. From (5) follows that rows $(a_{lj})_j$, $i < l \le n$ are orthogonal and, therefore, linearly independent. Using also

$$\mathcal{L}(\,\{(a_{k1}, a_{k2}, \dots, a_{kn}) \mid 1 \le k \le i\}\,) \;\perp\; \mathcal{L}(\,\{(a_{l1}, a_{l2}, \dots, a_{ln}) \mid i < l \le n\}\,)$$

obtains that rows $(a_{kj})_j$, $1 \leq k \leq i$ are linearly dependent. Therefore, $A_{lj} = 0$, $i < l \leq n$, $1 \leq j \leq n$, what implies that $i = n$. □

Note that there is an orbit of X

$$|X| \leq |F_1(X)| \leq |F_2(F_1(X))| \leq \cdots \leq |F_n(\dots F_2(F_1(X))\dots)|,$$

where $F_n \circ \cdots \circ F_1(\mathcal{X}) = \mathcal{F}$, what gives a direct proof of (1). Geometric interpretation of Hadamard's inequality is that the volume of a parallelepiped in n-dimensional space does not exceed product of lengths of it's edges, equality holds if the edges are orthogonal, or if length of one edge is zero. Also, mention analogy between Hadamard' inequality and generalization of Cauchy inequality, namely, one special case of Hölder's inequality

$$\left(\sum_{j=1}^{m}\left(\prod_{i=1}^{n} a_{ij}\right)\right)^{n} \leq \prod_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij}^{n}\right),$$

$|A|$ is also a sum of products of elements of matrix A.

References

[1] E. F. Beckenbach, R. Bellman, *Inequalities*, Springer—Verlag, 1983.
[2] R. Bellman, *Notes on Matrix Theory*, series of articles, Amer. Math. Monthly **60**-(1953–).
[3] M. Bjelica, *Fixed Point and Inequalities*, Ph. D. Thesis, University of Belgrade, 1990.
[4] J. Hadamard, *Resolution d'une question relative aux determinants*, Bull. Sci. Math. **2** (1893), 240–248.
[5] J. Hadamard, *The Psychology of Invention in the Mathematical Field*, Princeton University Press, 1949.
[6] J. G. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934.

University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin
23000, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 603–610
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SEMILATTICES OF WEAKLY LEFT ARCHIMEDEAN SEMIGROUPS

Stojan Bogdanović and Miroslav Ćirić

Abstract. By the well-known result of A. H. Clifford, any band of left Archimedean semigroups is a semilattice of matrices (rectangular bands) of left Archimedean semigroups. The converse of this assertion don't hold, i.e. the class of semilattices of matrices of left Archimedean semigroups is larger than the class of bands of left Archimedean semigroups. In this paper we characterize semilattices of matrices of left Archimedean semigroups, and especially matrices of left Archimedean semigroups. The obtained results generalize the some results of M. S. Putcha and L. N. Shevrin.

Bands of left (also right and two-sided) Archimedean semigroups form important classes of semigroups studied by a number of authors. General characterizations of these semigroups have been given by M. S. Putcha [16], and in the completely π-regular case by L. N. Shevrin [17]. Some new characterizations of bands of left Archimedean semigroups and of bands of nil-extensions of left simple semigroups have been given recently by the authors [6]. By the well-known result of A. H. Clifford, any band of left Archimedean semigroups is a semilattice of matrices (rectangular bands) of left Archimedean semigroups. The converse of this assertion don't hold, i.e. the class of semilattices of matrices of left Archimedean semigroups is larger than the class of bands of left Archimedean semigroups. In this paper we give a complete characterization of semigroups having a semilattice decomposition whose components are matrices of left Archimedean semigroups. Moreover, we describe such components in the general and some special cases. For the related results see [7], [12] and [13]. For more informations about semilattice-matrix decompositions of semigroups the reader is reffered

Supported by Grant 0401B of RFNS through Math. Inst. SANU.

to [10] and [11]. The obtained results generalize the above quoted results of M. S. Putcha and L. N. Shevrin.

Throughout this paper $\mathbf{Z}^+$ will denote the set of positive integers. The division relations $|$ and $\underset{l}{|}$ on a semigroup S are defined by

$$a \mid b \;\Leftrightarrow\; (\exists x, y \in S^1)\, b = xay, \qquad a \underset{l}{|} b \;\Leftrightarrow\; (\exists x \in S^1)\, b = xa,$$

and that the relations $\longrightarrow$ and $\xrightarrow{l}$ on S are defined by

$$a \longrightarrow b \;\Leftrightarrow\; (\exists n \in \mathbf{Z}^+)\, a \mid b^n, \qquad a \xrightarrow{l} b \;\Leftrightarrow\; (\exists n \in \mathbf{Z}^+)\, a \underset{l}{|} b^n.$$

The relation $\xrightarrow{r}$ on S is defined dually. For $n \in \mathbf{Z}^+$, $\xrightarrow{l}^n$ will denote the n-th power of $\xrightarrow{l}$, and $\xrightarrow{l}^\infty$ will denote the transitive closure of $\xrightarrow{l}$. For an element a of a semigroup S we define sets $\Lambda_n(a)$, $n \in \mathbf{Z}^+$, and $\Lambda(a)$ by

$$\Lambda_n(a) = \{x \in S \mid a \xrightarrow{l}^n x\}, \qquad \Lambda(a) = \{x \in S \mid a \xrightarrow{l}^\infty x\},$$

and the equivalence relations λ_n, $n \in \mathbf{Z}^+$, and λ on S by

$$a\, \lambda_n\, b \;\Leftrightarrow\; \Lambda_n(a) = \Lambda_n(b), \qquad a\, \lambda\, b \;\Leftrightarrow\; \Lambda(a) = \Lambda(b),$$

[3]. For undefined notions and notations we refer to [1], [2] and [14].

First we prove the following theorem:

Theorem 1. *The following conditions on a semigroup S are equivalent:*

(i) *λ is a matrix congruence on S;*
(ii) *λ is a right zero band congruence on S;*
(iii) $(\forall a, b, c \in S)\; abc \xrightarrow{l}^\infty ac$;
(iv) $(\forall a, b \in S)\; aba \xrightarrow{l}^\infty a$;
(v) $(\forall a, b \in S)\; ab \xrightarrow{l}^\infty b$;
(vi) *S is a disjoint union of all its principal left radicals;*
(vii) *$\xrightarrow{l}^\infty$ is a symmetric relation on S.*

Proof. (i) ⇒ (iii), (iii) ⇒ (iv) and (ii) ⇒ (i). This follows immediately.

(iv) ⇒ (v). For all $a, b \in S$, $ab \xrightarrow{l} bab$, so by (iv), $ab \xrightarrow{l}^\infty b$.

(v) ⇒ (ii). Let $a, b \in S$ such that $a\,\lambda\, b$, and $x \in S$. By (v), $\Lambda(ax) = \Lambda(x) = \Lambda(bx)$ and $\Lambda(xa) = \Lambda(a) = \Lambda(b) = \Lambda(xb)$. Therefore, λ is a congruence. Clearly, it is a right zero band congruence.

(ii) ⇒ (vi). Let S be a right zero band B of semigroups S_i, $i \in B$, which are λ-classes of S. Assume $a \in S$. Then $a \in S_i$, for some $i \in B$, and since S_i is a completely semiprime left ideal of S (Lemma 4 [3]), then $\Lambda(a) \subseteq S_i$. On the other hand, if $b \in S_i$, then $b \,\lambda\, a$, so $b \in \Lambda(b) = \Lambda(a)$, whence $S_i \subseteq \Lambda(a)$. Therefore, $\Lambda(a) = S_i$, so (vi) holds.

(vi) ⇒ (vii). Let $a, b \in S$ such that $a \xrightarrow{l}{}^{\infty} b$. Then $b \in \Lambda(a)$, whence $\Lambda(a) \cap \Lambda(b) \neq \varnothing$, so by (vi), $\Lambda(a) = \Lambda(b)$. Therefore, $b \xrightarrow{l}{}^{\infty} a$.

(vii) ⇒ (v). For all $a, b \in S$, $b \xrightarrow{l} ab$, so by (vii), $ab \xrightarrow{l}{}^{\infty} b$. □

Corollary 1. *The following conditions on a semigroup S are equivalent:*

(i) λ_n *is a matrix congruence on* S*;*

(ii) λ_n *is a right zero band congruence on* S*;*

(iii) $(\forall a, b \in S)\ \Lambda_n(a) \subseteq \Lambda_n(aba)$*;*

(iv) $(\forall a, b \in S)\ \Lambda_n(b) \subseteq \Lambda_n(ab)$*;*

(v) $\xrightarrow{l}{}^{n}$ *is a symmetric relation on* S.

Lemma 1. *Let ξ be a band congruence on a semigroup S contained in $\overset{l}{-\!-}$, where $\overset{l}{-\!-} = \xrightarrow{l} \cap \xrightarrow{l}{}^{-1}$. Then any ξ-class of S is a left Archimedean semigroup.*

Recall that a semigroup S is called *left Archimedean* if $a \xrightarrow{l} b$, for all $a, b \in S$. Here we introduce a more general notion: a semigroup S will be called *weakly left Archimedean* if $ab \xrightarrow{l} b$, for all $a, b \in S$. *Weakly right Archimedean semigroups* are defined dually. A semigroup S is *weakly t-Archimedean* (or *weakly two-sided Archimedean*) if it is both weakly left and weakly right Archimedean, i.e. if for all $a, b \in S$ there exists $n \in \mathbf{Z}^+$ such that $a^n \in abSba$.

We give the following characterization of semilattices of weakly left Archimedean semigroups:

Theorem 2. *A semigroup S is a semilattice of weakly left Archimedean semigroups if and only if*

$$a \longrightarrow b \quad \Rightarrow \quad ab \xrightarrow{l} b,$$

for all $a, b \in S$.

Proof. Let S be a semillatice Y of weakly left Archimedean semigroups S_α, $\alpha \in Y$. Assume $a, b \in S$ such that $a \longrightarrow b$. If $a \in S_\alpha$, $b \in S_\beta$, for some $\alpha, \beta \in Y$, then $\beta \leq \alpha$, whence $b, ba \in S_\beta$. Now, $b^n \in S_\beta bab \subseteq Sab$, for some $n \in \mathbf{Z}^+$, since S_β is weakly left Archimedean. Therefore, $ab \xrightarrow{l} b$.

Conversely, let for all $a, b \in S$, $a \longrightarrow b$ implies $ab \xrightarrow{l} b$. Assume $a, b \in S$. Since $a \longrightarrow ab$, then by the hypothesis, $a^2b \xrightarrow{l} ab$, i.e. $(ab)^n \in Sa^2b \subseteq Sa^2S$, for some $n \in \mathbf{Z}^+$. Now, by Theorem 1 [9], S is a semilattice Y of Archimedean semigroups S_α, $\alpha \in Y$. Further, assume $\alpha \in Y$, $a, b \in S_\alpha$. Then $a \longrightarrow b$, so by the hypothesis, $ab \xrightarrow{l} b$ in S, and by Lemma 11 (c) [3], $ab \xrightarrow{l} b$ in S_α. Therefore, S_α is weakly left Archimedean. □

Corollary 2. *A semigroup S is a semilattice of weakly t-Archimedean semigroups if and only if*

$$a \longrightarrow b \;\Rightarrow\; ab \xrightarrow{l} b \;\;\&\;\; ba \xrightarrow{r} b,$$

for all $a, b \in S$.

The components of the semilattice decomposition treated in Theorem 2 will be characterized by the next theorem. Namely, we will give a description of weakly left Archimedean semigroups.

Theorem 3. *The following conditions on a semigroup S are equivalent:*

(i) *S is weakly left Archimedean;*
(ii) *S is a matrix of left Archimedean semigroups;*
(iii) *S is a right zero band of left Archimedean semigroups;*
(iv) *$\xrightarrow{l}$ is a symmetric relation on S.*

Proof. (i) ⇒ (iv). Let $a, b \in S$ such that $a \xrightarrow{l} b$, i.e. $b^n = xa$, for some $n \in \mathbf{Z}^+$, $x \in S$. By (i), $a^m = yxa = yb^n$, for some $m \in \mathbf{Z}^+$, $y \in S$, whence $b \xrightarrow{l} a$.

(iv) ⇒ (i). This follows by the proof for (vii) ⇒ (v) of Theorem 1.

(iv) ⇒ (iii). Let $a, b, c \in S$ such that $a \xrightarrow{l} b$ and $b \xrightarrow{l} c$. By (iv), $c \xrightarrow{l} b$, so $b^n = xa = yc$, for some $n \in \mathbf{Z}^+$, $x, y \in S$. Since (iv) ⇔ (i), then there exists $m \in \mathbf{Z}^+$, $z \in S$ such that $c^m = z(yc) = zb^n = zxa \in Sa$. Therefore, $a \xrightarrow{l} c$, so $\xrightarrow{l}$ is transitive, i.e. $\xrightarrow{l} = \xrightarrow{l}{}^{\infty}$. Now, by Theorem 1, $\lambda_1 = \lambda$ is a right zero band congruence. By Lemma 1, λ_1-classes are left Archimedean semigroups.

(iii) ⇒ (ii). This follows immediately.

(ii) ⇒ (i). Let S be a matrix B of left Archimedean semigroups S_i, $i \in B$. Then for $a, b \in S$, $a, aba \in S_i$, for some $i \in B$, whence $a^n \in S_i aba \subseteq Sba$, for some $n \in \mathbf{Z}^+$. □

Recall that the relation $\xrightarrow{t}$ on a semigroup S is defined by $\xrightarrow{t} = \xrightarrow{l} \cap \xrightarrow{r}$. Now by Theorem 3 and its dual we obtain the following corollary:

Corollary 3. *The following conditions on a semigroup S are equivalent:*

(i) *S is weakly t-Archimedean;*

(ii) *S is a matrix of t-Archimedean semigroups;*

(iii) *$\xrightarrow{t}$ is a symmetric relation on S;*

(iv) *$\xrightarrow{l}$ and $\xrightarrow{r}$ are symmetric relations on S.*

By the following theorem we characterize matrices of nil-extensions of left simple semigroups.

Theorem 4. *The following conditions on a semigroup S are equivalent:*

(i) *S is weakly left Archimedean and left π-regular;*

(ii) *S is weakly left Archimedean and intra-π-regular;*

(iii) *S is a matrix of nil-extensions of left simple semigroups;*

(iv) *S is a right zero band of nil-extensions of left simple semigroups;*

(v) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n \in S(ba)^n$*;*

(vi) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n \in Sb^n a$*.*

Proof. (i) ⇒ (iv). This follows by Theorem 3 and Theorem 4.1 [15], since the components of any band decomposition of a left π-regular semigroup are also left π-regular.

(iv) ⇒ (iii). This follows immediately.

(iii) ⇒ (ii). This follows by Theorem 3, since a nil-extension of a left simple semigroup is intra-π-regular.

(ii) ⇒ (i). By Theorem 3, S is a right zero band B of left Archimedean semigroups S_i, $i \in B$. Let $a \in Intra(S)$, i.e. $a = xa^2y$, for some $x, y \in S$. Then $a = (xa)^k ay^k$, for each $k \in \mathbf{Z}^+$. Further, $a \in S_i$, for some $i \in B$, and clearly, $y \in S_i$, so $y^k = za^2$, for some $k \in \mathbf{Z}^+$, $z \in S$, since S_i is left Archimedean. Therefore, $a = (xa)^k ay^k = (xa)^k aza^2$, whence $a \in LReg(S)$, so by Theorem 1 [5], S is left π-regular.

(iv) ⇒ (vi). Let S be a right zero band B of semigroups S_i, $i \in B$, and for each $i \in B$, let S_i be a nil-extension of a left simple semigroup K_i. Since (v) ⇔ (i), then S is a nil-extension of a left completely simple semigroup K. Clearly, $K = LReg(S) = \bigcup_{i \in B} K_i$. Now, for $a, b \in S$, $a \in S_i$, $b \in S_j$, for some $i, j \in B$, and $a^n \in K_i$, $b^n \in K_j$, for some $n \in Z$, whence $b^n a \in S_i \cap K = K_i$, so $a^n \in K_i b^n a \subseteq Sb^n a$.

(vi) ⇒ (v). Assume $a, b \in S$. By (vii), there exists $n \in \mathbf{Z}^+$ such that $a^n \in S(ab)^n a \subseteq S(ba)^n$.

(v) ⇒ (i). This follows immediately. □

Let T be a semigroup of a semigroup S. A mapping φ of S onto T is a *right retraction* of S onto T if $a\varphi = a$, for each $a \in T$, and $(ab)\varphi = a(b\varphi)$, for all $a, b \in S$. *Left retractions* are defined dually. A mapping φ of S onto

T is a *retraction* of S onto T if it is a homomorphism and $a\varphi = a$, for each $a \in T$. If T is an ideal of S, then φ is a retraction of S onto T if and only if it is both left and right retraction of S onto T. An ideal extension S of a semigroup T is a *(left, right) retractive* extension of T if there exists a *(left, right) retraction* of S onto T.

Some characterizations of matrices of nil-extensions of left groups have been given by L. N. Shevrin in [17]. By the next theorem we prove that such semigroups are exactly right retractive nil-extensions of completely simple semigroups. In this way we generalize some results of S. Bogdanović and S. Milić [7], J. L. Galbiati and M. L. Veronesi [12] and A. Mărkuş [13].

Theorem 5. *The following conditions on a semigroup S are equivalent:*

(i) *S is a right retractive nil-extension of a completely simple semigroup;*

(ii) *S is weakly left Archimedean and has an idempotent;*

(iii) *S is a matrix of nil-extensions of left groups;*

(iv) *S is a right zero band of nil-extensions of left groups;*

(v) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n \in a^n S(ba)^n$;

(vi) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n \in a^n S b^n a$.

Proof. (iv) ⇒ (iii) and (iii) ⇒ (ii). This follows immediately.

(ii) ⇒ (i). By Theorem 4.1 [15]. S is a nil-extension of a simple semigroup K, so it is intra-π-regular and by Theorem 1 [5], S is left π-regular, it is a right zero band B of semigroups S_i, $i \in B$, and for each $i \in B$, S_i is a nil-extension of a left simple semigroup K_i. Further, $K = Intra(S) = LReg(S) = \bigcup_{i\in B} K_i$, by Theorem 1 [5], since the components of any band decomposition of a left π-regular semigroup are also left π-regular. Thus, K is left completely simple, so it is completely simple, since it has an idempotent. Thus, for each $i \in B$, K_i is a left group, so by Theorem VI 3.1 [1] (or Theorem 3.7 [2]), it has a right identity e_i. Define a mapping φ of S onto K by:

$$a\varphi = ae_i \qquad\qquad \text{if } a \in S_i,\ i \in B.$$

Clearly, $a\varphi = a$, for each $a \in K$. Further, for $a, b \in S$, $a \in S_i$, $b \in S_j$, for some $i, j \in B$, and $ab \in S_j$, whence $(ab)\varphi = (ab)e_j = a(be_j) = a(b\varphi)$. Therefore, φ is a right retraction of S onto K.

(i) ⇒ (vi). Let S be a right retractive nil-extension of a completely simple semigroup K, and let K be a right zero band B of left groups K_i, $i \in B$. Let $a, b \in S$. Then $a^n, b^n \in K$, for some $n \in \mathbf{Z}^+$, and $a^n \in K_i$, $b^n \in K_j$, for some $i, j \in B$. If $a\varphi \in K_l$, for some $l \in B$, since $a^{n+1} \in K_i$, then $a^{n+1} = a^{n+1}\varphi = a^n(a\varphi) \in K_iK_l \subseteq K_l$, whence $l = i$. Thus, $a\varphi \in K_i$, so $b^na = (b^na)\varphi = b^n(a\varphi) \in K_jK_i \subseteq K_i$. Therefore, $a^n, b^na \in K_i$, so by the dual of Lemma 1.1 [8], $a^n \in a^nK_ib^na \subseteq a^nSb^na$.

(vi) ⇒ (v). For $a, b \in S$ there exists $n \in \mathbf{Z}^+$ such that $a^n \in a^n S(ab)^n a = a^n S a(ba)^n \subseteq a^n S(ba)^n$.

(v) ⇒ (iv). This follows by Theorem 4. □

Corollary 4. *The following conditions on a semigroup S are equivalent:*

(i) *S is a retractive nil-extension of a completely simple semigroup;*
(ii) *S is weakly t-Archimedean and intra-π-regular;*
(iii) *S is weakly t-Archimedean and has an idempotent;*
(iv) *S is a matrix of π-groups;*
(v) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n \in (ab)^n S(ba)^n$.

Let us introduce the following notations for some classes of semigroups:

Notation	Class of semigroups	Notation	Class of semigroups
$\mathcal{LA}$	*left Archimedean*	$\mathcal{M}$	*rectangular bands*
$\mathcal{B}$	*bands*	$\mathcal{S}$	*semilattices*

and by $\mathcal{X}_1 \circ \mathcal{X}_2$ we denote the Mal'cev product of classes $\mathcal{X}_1$ and $\mathcal{X}_2$ of semigroups. Let

$$\mathcal{LA} \circ \mathcal{M}^{k+1} = (\mathcal{LA} \circ \mathcal{M}^k) \circ \mathcal{M}, \quad k \in \mathbf{Z}^+.$$

Now we can state the following:

Problem. *Describe the structure of semigroups from the following classes*

$$\mathcal{LA} \circ \mathcal{M}^{k+1}, \quad (\mathcal{LA} \circ \mathcal{M}^{k+1}) \circ \mathcal{B}, \quad (\mathcal{LA} \circ \mathcal{M}^{k+1}) \circ \mathcal{S}.$$

References

[1] S. Bogdanović, *Semigroups with a system of subsemigroups*, Inst. of Math., Novi Sad, 1985.

[2] S. Bogdanović and M. Ćirić, *Polugrupe*, Prosveta, Niš, 1993.

[3] M. Ćirić and S. Bogdanović, *Semilattice decompositions of semigroups*, Semigroup Forum (to appear).

[4] S. Bogdanović and M. Ćirić, *Semilattices of Archimedean semigroups and (completely) π-regular semigroups* I *(A survey)*, Filomat (Niš) **7** (1993), 1–40.

[5] S. Bogdanović and M. Ćirić, *A note on left regular semigroups* (to appear).

[6] S. Bogdanović and M. Ćirić, *Bands of left Archimedean semigroups* (to appear).

[7] S. Bogdanović and S. Milić, *A nil-extension of a completely simple semigroup*, Publ. Inst. Math. **36 (50)** (1984), 45–50.

[8] S. Bogdanović and B. Stamenković, *Semigroups in which S^{n+1} is a semilattice of right groups (Inflations of a semilattices of right groups)*, Note di Matematica **8** (1988), no. 1, 155–172.

[9] M. Ćirić and S. Bogdanović, *Decompositions of semigroups induced by identities*, Semigroup Forum **46** (1993), 329–346.

[10] M. Ćirić and S. Bogdanović, *Theory of greatest decompositions of semigroups (A survey)*, FILOMAT (Niš) this volume.

[11] M. Ćirić and S. Bogdanović, *Semilattice-matrix decompositions of semigroups* (to appear).

[12] J. L. Galbiati e M. L. Veronesi, *Sui semigruppi che sono un band di t-semigruppi*, Rend. Ist. Lomb, Cl. Sc. (A) **114** (1980), 217–234.

[13] A. Mărkuş, *Retract extensions of completely simple semigroups by nil-semigroups*, Mathematica **34 (57)** (1992), no. 1, 37–41.

[14] M. Petrich, *Lectures in semigroups*, Akad. Verlag, Berlin, 1977.

[15] M. S. Putcha, *Semilattice decompositions of semigroups*, Semigroup Forum **6** (1973), 12–34.

[16] M. S. Putcha, *Bands of t-archimedean semigroups*, Semigroup Forum **6** (1973), 232–239.

[17] L. N. Shevrin, *Theory of epigroups* II, Mat. Sbornik **185** (1994), no. 8, 153–176. (in Russian)

Univeristy of Niš, Faculty of Economics, Trg JNA 11, 18000 Niš, Yugoslavia
E-mail address: root@eknux.eknfak.ni.ac.yu

Univeristy of Niš, Faculty of Philosophy, Ćirila i Metodija 2, 18000 Niš, Yugoslavia
E-mail address: mciric@archimed.filfak.ni.ac.yu

FILOMAT (Niš) 9:3 (1995), 611–617
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SEMILATTICES OF HEREDITARY ARCHIMEDEAN SEMIGROUPS

Stojan Bogdanović, Miroslav Ćirić and Melanija Mitrović

ABSTRACT. In this paper we investigate semigroups whose any subsemigroup is Archimedean, called hereditary Archimedean, and semilattices of such semigroups. The obtained results generalize some results of J. L. Chrislock, M. Schutzenberger and M. V. Sapir and E. V. Suhanov.

T. E. Nordahl [14] studied commutative semigroups whose any proper ideal is a power-joined semigroup. C. S. H. Nagore [12] extended this result to quasi-commutative semigroups. Semigroups containing an idempotent and whose any subsemigroup is t-Archimedean have been studied by A. Cherubini and A. Varisco [9]. S. Bogdanović [2] studied weakly commutative semigroups whose any proper right ideal is power-joined. B. Pondeliček [15] described semigroups whose one-sided ideals are t-Archimedean semigroups. Semigroups whose any proper ideal is a power-joined semigroup have been studied by A. Nagy [13]. S. Bogdanović in [3] described semigroups whose proper (left) ideals are Archimedean (left Archimedean, t-Archimedean, power-joined) semigroups. S. Bogdanović and T. Malinović [8] studied semigroups whose any proper subsemigroup is right Archimedean (t-Archimedean). In this paper we study semigroups whose any subsemigroup is Archimedean, called hereditary Archimedean semigroups, and semilattices of such semigroups. We prove also a more general theorem concerning semigroups whose any subsemigroup is a semilattice of Archimedean semigroups. Note that semilattices of Archimedean semigroups have been studied by a number of authours. M. S. Putcha in [16] gave the first complete description of such semigroups. Another characterizations of semilattices of

Supported by Grant 0401B of RFNS through Math. Inst. SANU.

Archimedean semigroups have been given by T. Tamura [17], S. Bogdanović and M. Ćirić [4] and M. Ćirić and S. Bogdanović [10].

Throughout this paper, $Reg(S)$ $(E(S))$ will denote the set of regular (idempotent) elements of a semigroup S, and for $e \in E(S)$, G_e will denote the maximal subgroup of S with e as its identity. A semigroup S is said to be *π-regular* if for any $a \in S$, some power of a is regular.

For undefined notions and notations we refer to [1] and [5].

Recall that the *division relation* $\mid$ on a semigroup S is defined by

$$a \mid b \quad \Leftrightarrow \quad (\exists x, y \in S^1)\, b = xay,$$

and the relation $\longrightarrow$ is defined by

$$a \longrightarrow b \quad \Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, a \mid b^n.$$

Also, on a semigroup S we define the relations $\uparrow$, $\underset{l}{\uparrow}$, $\underset{r}{\uparrow}$ and $\underset{t}{\uparrow}$ by

$$\begin{aligned} a \uparrow b \quad &\Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, b^n \in \langle a, b\rangle\, a\, \langle a, b\rangle\,, \\ a \underset{l}{\uparrow} b \quad &\Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, b^n \in \langle a, b\rangle\, a, \\ a \underset{r}{\uparrow} b \quad &\Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, b^n \in a\, \langle a, b\rangle\,, \\ a \underset{t}{\uparrow} b \quad &\Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, a \underset{l}{\uparrow} b \;\&\; a \underset{r}{\uparrow} b \end{aligned}$$

Clearly, $a \underset{t}{\uparrow} b$ if and only if $b^n \in a\, \langle a, b\rangle\, a$, for some $n \in \mathbf{Z}^+$.

A semigroup S is a *hereditary Archimedean* if $a \uparrow b$ for all $a, b \in S$. By a *hereditary left Archimedean semigroup* we mean a semigroup S satisfying the condition: $a \underset{l}{\uparrow} b$, for all $a, b \in S$. A *hereditary right Archimedean semigroup* is defined dually. A semigroup S is called *hereditary t-Archimedean* if it is both hereditary left Archimedean and hereditary right Archimedean. i.e. if $a \underset{t}{\uparrow} b$ for all $a, b \in S$.

The next lemma gives an explanation why we the notion "hereditary Archimedean" is used.

Lemma 1. [5] *A semigroup S is hereditary Archimedean if and only if any subsemigroup of S is Archimedean.*

Similar assertions hold for hereditary left, right or t-Archimedean semigroups.

T. Tamura in [23] proved that the class of all semigroups which are semilattices of Archimedean semigroups is not subsemigroup closed. By the following theorem we determine the greatest subsemigroup closed subclass of this class. In other words, we describe all semigroups having the property that any its subsemigroup is a semilattice of Archimedean semigroups.

Theorem 1. *Any subsemigroup of a semigroup S is a semilattice of Archimedean semigroups if and only if for all $a, b \in S$ there exists $n \in \mathbf{Z}^+$ such that*

$$(ab)^n \in \langle a, b\rangle \, a^2 \, \langle a, b\rangle .$$

Proof. If $a, b \in S$ and $T = \langle a, b\rangle$, then by Theorem 1 [11] it follows that

$$(ab)^m \in Ta^2T = \langle a, b\rangle \, a^2 \, \langle a, b\rangle ,$$

for some $m \in \mathbf{Z}^+$.

Conversely, if T is a subsemigroup of S and $a, b \in T$, then there exists $m \in \mathbf{Z}^+$ such that

$$(ab)^m \in \langle a, b\rangle \, a^2 \, \langle a, b\rangle \subseteq Ta^2T,$$

so by Theorem 1 [11], T is a semilattice of Archimedean semigroups. □

The main result of this paper is the following theorem which characterizes semilattices of hereditary Archimedean semigroups.

Theorem 2. *The following conditions on a semigroup S are equivalent:*

(i) *S is a semilattice of hereditary Archimedean semigroups;*
(ii) $(\forall a, b \in S)\ a \longrightarrow b \ \Rightarrow\ a^2 \upharpoonright b;$
(iii) $(\forall a, b, c \in S)\ a \longrightarrow c \ \&\ b \longrightarrow c \ \Rightarrow\ ab \upharpoonright c;$
(iv) $(\forall a, b, c \in S)\ a \longrightarrow b \ \&\ b \longrightarrow c \ \Rightarrow\ a \upharpoonright c.$

Proof. (i) ⇒ (ii). Let S be a semilattice Y of hereditary Archimedean semigroups S_α, $\alpha \in Y$. Assume $a, b \in S$ such that $a \longrightarrow b$. Then $b, a^2b \in S_\alpha$, for some $\alpha \in Y$, so by the hypothesis we obtain that

$$b^n \in \left\langle b, a^2b\right\rangle a^2b \left\langle b, a^2b\right\rangle \subseteq \left\langle a^2, b\right\rangle a^2 \left\langle a^2, b\right\rangle .$$

Thus $a^2 \upharpoonright b$, so (ii) holds.

(ii) ⇒ (iii). Assume $a, b, c \in S$ such that $a \longrightarrow c \ \&\ b \longrightarrow c$. Then by Theorem 5.5 [5], $ab \longrightarrow c$. Now by (ii) it follows $(ab)^2 \upharpoonright c$, whence $ab \upharpoonright c$.

(iii) ⇒ (iv). By (iii) and Propositions 7 [22], $\longrightarrow$ is transitive. Assume $a, b, c \in S$ such that $a \longrightarrow b$ and $b \longrightarrow c$. Then $a \longrightarrow c$, so $a^2 \upharpoonright c$, by (iii), whence $a \upharpoonright c$.

(iv) ⇒ (i). By (iv), $\longrightarrow$ is transitive, so by Proposition 7 [22], S is a semilattice Y of Archimedean semigroups S_α, $\alpha \in Y$.

Assume $\alpha \in Y$ and $a, b \in S_\alpha$. Then $a \longrightarrow b$ and $b \longrightarrow b$, whence $a \upharpoonright b$, by (iv). Therefore, S_α is hereditary Archimedean. Hence, (i) holds. □

The next theorem gives a characterisation of semigroups which are chains of hereditary Archimedean semigroups.

Theorem 3. *A semigroup S is a chain of hereditary Archimedean semigroups if and only if*

$$ab \mid a \quad or \quad ab \uparrow b.$$

for all $a, b \in S$.

Proof. Let S be a chain Y of hereditary Archimedean semigroups S_α, $\alpha \in Y$. If $a \in S_\alpha$, $b \in S_\beta$, for some $\alpha, \beta \in Y$, then $a, ab \in S_\alpha$ or $b, ab \in S_\beta$, whence

$$a^n \in \langle a, ab\rangle\, ab\, \langle a, ab\rangle \qquad \text{or} \qquad b^n \in \langle b, ab\rangle\, ab\, \langle b, ab\rangle$$

for some $n \in \mathbf{Z}^+$.

Conversely, by the hypothesis and Theorem 1 [7], S is a chain Y of Archimedean semigroups $S_\alpha, \alpha \in Y$. If $\alpha \in Y$ and $a, b \in S_\alpha$, then then there exists $n \in \mathbf{Z}^+$ such that $b^n \in S_\alpha a S_\alpha$, and by Theorem 2, $a^2 \uparrow b^n$, whence $a \uparrow b$. Thus, S_α is hereditary Archimedean. Hence, S is a chain of hereditary Archimedean semigroups. □

Further we study semilattices of hereditary left Archimedean semigroups.

Theorem 4. *A semigroup S is a semilattice of hereditary left Archimedean semigroups if and only if for all $a, b \in S$,*

$$a \longrightarrow b \ \Rightarrow\ a \underset{l}{\uparrow} b.$$

Proof. Let S be a semilattice Y of hereditary left Archimedean semigroups S_α, $\alpha \in Y$. Assume $a, b \in S$ such that $a \longrightarrow b$. Since $a \in S_\alpha$, $b \in S_\beta$, for some $\alpha, \beta \in Y$, we then have that $\beta \le \alpha$, so $b, ba \in S_\beta$. Now $ba \underset{l}{\uparrow} b$, whence $a \underset{l}{\uparrow} b$, which proves the direct part of the theorem.

Conversely, by the hypothesis and Theorem 3(2) [18], S is a semilattice Y of left Archimedean semigroup S_α, $\alpha \in Y$. Assume $\alpha \in Y$ and $a, b \in S_\alpha$. Then $a \longrightarrow b$, whence $a \underset{l}{\uparrow} b$, by the hypothesis. Therefore, any S_α is hereditary left Archimedean, so S is a semilattice of hereditary left Archimedean semigroups. □

Corollary 1. *A semigroup S is a semilattice of hereditary t-Archimedean semigroups if and only if for all $a, b \in S$,*

$$a \longrightarrow b \ \Rightarrow\ a \underset{t}{\uparrow} b.$$

Now we prove a theorem which generalizes a result of J. L. Chrislock [10].

Theorem 5. *The folowing conditions on a semigroup S are equivalent:*

(i) *S is hereditary Archimedean and π-regular;*
(ii) *S is hereditary Archimedean and has a primitive idempotent;*
(iii) *S is a nil-extension of a periodic completely simple semigroup;*
(iv) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ a^n = (a^n b^n a^n)^n$.

Proof. (i) ⇒ (ii). First we prove that

$$(1) \qquad (\forall a \in S)(\forall e \in E(S))(\exists n \in \mathbf{Z}^+)\ e = (eae)^n.$$

Indeed, for $a \in S$, $e \in E(S)$, $ea \mid e$, by (i), whence $e = (ea)^n$ or $e = (ea)^n e$, for some $n \in \mathbf{Z}^+$. However, in both of cases it follows that $e = (ea)^n e = (eae)^n$. Thus (1) holds.

Further, assume $a \in S$. Let $m \in \mathbf{Z}^+$ such that $a^m \in Reg(S)$ and let x be an inverse of a^n. Then $a^m x, xa^m \in E(S)$, so by (1) we obtain that

$$a^m x = (a^m x \cdot a \cdot a^m x)^n = (a^{m+1}x)^n,$$

for some $n \in \mathbf{Z}^+$, whence

$$\begin{aligned}
a^m = a^m x a^m &= (a^{m+1}x)^n a^m = (a^{m+1}x)^{n-1} a^{m+1} x a^m = \\
&= (a^{m+1}x)^{n-1} a a^m x a^m = (a^{m+1}x)^{n-1} a^{m+1} = \\
&= (a^{m+1}x)^{n-2} a^{m+1} x a^{m+1} = (a^{m+1}x)^{n-2} a a^m x a^m a = \\
&= (a^{m+1}x)^{n-2} a a^m a = (a^{m+1}x)^{n-2} a^{m+2} = \cdots = \\
&= (a^{m+1}x)^{n-(n-1)} a^{m+(n-1)} = \\
&= a^{m+1} x a^{m+n-1} = a a^m x a^m a^{n-1} = \\
&= a a^m a^{n-1} = a^{m+n}.
\end{aligned}$$

Thus, S is periodic, and by Theorem 3.14 [5], S has a primitive idempotent.

(ii) ⇒ (iii). By Theorem 3.14 [5], S is a nil-extension of a completely simple semigroup K. But, K is hereditary Archimedean and regular, so it is periodic, by the proof of (i) ⇒ (ii).

(iii) ⇒ (iv). Assume $a, b \in S$. Then $a^k = e$ and $b^n = f$, for some $e, f \in E(S)$, $k \in \mathbf{Z}^+$. Further, $efe \in eSe = G_e$, by Lemma 3.13 [5], whence $(efe)^m = e$, for some $m \in \mathbf{Z}^+$. Now, for $n = km$ we obtain that $a^n = (a^n b^n a^n)^n$.

(iv) ⇒ (i). This follows immediately. □

Finaly we prove the following theorem which generalizes some results of M. V. Sapir and E. V. Suhanov [19] and M. Schutzenberger [20] (see L. N. Shevrin and E. V. Suhanov [21]).

Theorem 6. *The following conditions on a semigroup S are equivalent:*

(i) *S is π-regular and a semilattice of hereditary Archimedean semigroups;*

(ii) *S is a semilattice of nil-extensions of periodic completely simple semigroups;*

(iii) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ (ab)^n = (ab)^n((ba)^n(ab)^n)^n;$

(iv) $(\forall a, b \in S)(\exists n \in \mathbf{Z}^+)\ (ab)^n = ((ab)^n(ba)^n(ab)^n)^n.$

Proof. (i) ⇒ (ii). This follows immediately by Theorem 5.

(ii) ⇒ (iii) and (ii) ⇒ (iv). This follows by Theorem 5.

(iii) ⇒ (i) and (iv) ⇒ (i). This follows immediately. □

References

[1] S. Bogdanović, *Semigroups with system of subsemigroups*, Inst. of Math., Novi Sad, 1985.

[2] S. Bogdanović, *Q_r-semigroups*, Publ. Inst. Math. **29(43)** (1981), 15–21.

[3] S. Bogdanović, *Semigroups whose proper ideals are Archimedean semigroups*, Zb. rad. PMF Novi Sad **13** (1983), 289–296.

[4] S. Bogdanović and M. Ćirić, *Semigroups in which the radical of every ideal is a subsemigroup*, Zb. rad. Fil. fak. Niš, Ser. Mat. **6** (1992), 129–135.

[5] S. Bogdanović and M. Ćirić, *Polugrupe*, Prosveta, Niš, 1993.

[6] S. Bogdanović and M. Ćirić, *Semilattices of Archimedean semigroups and (completely) π-regular semigroups I (A survey)*, FILOMAT (Niš) **7** (1993), 1–40.

[7] S. Bogdanović and M. Ćirić, *Chains of Archimedean semigroups (Semiprimary semigroups)*, Indian J. Pure Appl. Math. **25** (1994), no. 3, 331–336.

[8] S. Bogdanović and T. Malinović, *Semigroups whose proper subsemigroups are (right) t-Archimedean*, Algebra and Logic, Cetinje, 1986, Inst. of Math., Novi Sad, 1988, pp. 1–14.

[9] A. Cherubini Spoletini e A. Varisco, *Sui semigruppi i cui soto semigruppi propri sono t-Archemedei*, Ist. Lombardo **112** (1978), 91–98.

[10] J. L. Chrislock, *A certain class of identities on semigroups*, Proc. Amer. Math. Soc. **21** (1969), 189–190.

[11] M. Ćirić and S. Bogdanović, *Decompositions of semigroups induced by identities*, Semigroup Forum **46** (1993), 329–346.

[12] M. Ćirić and S. Bogdanović, *Semilattice decompositions of semigroup*, Semigroup Forum (to appear).

[13] C. S. H. Nagore, *Quasi-commutative Q-semigroups*, Semigroup Forum **15** (1978), 189–193.

[14] A. Nagy, *Semigroups whose proper two-sided ideals are power-joined*, Semigroup Forum **25** (1982), 325–329.

[15] T. E. Nordahl, *Commutative semigroups whose proper subsemigroups are power-joined*, Semigroup Forum **6** (1973), 35–41.

[16] B. Pondeliček, *Semigroups whose proper one-sided ideals are t-Archimedean*, Mat. Vesnik **37 (3)** (1985), 315–321.

[17] M. S. Putcha, *Semilattice decompositions of semigroups*, Semigroup Forum **6** (1973), 12–34.

[18] M. S. Putcha, *Rings which are semilattices of Archimedean semigroups*, Semigroup Forum **23** (1981), 1–5.

[19] M. V. Sapir and E. V. Suhanov, *On varieties of periodic semigroups*, Izv. vysh. uchebn. zav, Mat. **4** (1981), 48–55. (in Russian)

[20] M. Schutzenberger, *Sur le produit de concatenation non ambigu*, Semigroup Forum **13** (1976), 47–75.

[21] L. N. Shevrin and E. V. Suhanov, *Structural aspects of the theory of varieties of semigroups*, Izv. vysh. uchebn. zav, Mat. **6** (1989), 3–39. (in Russian)

[22] T. Tamura, *On Putcha's theorem concerning semilattice of Archimedean semigroups*, Semigroup Forum **4** (1972), 255–261.

[23] T. Tamura, *Quasi-orders, generalized Archimedeaness, semilattice decompositions*, Math. Nachr. **68** (1975), 201–220.

Univeristy of Niš, Faculty of Economics, Trg JNA 11, 18000 Niš, Yugoslavia

E-mail address: root@eknux.eknfak.ni.ac.yu

Univeristy of Niš, Faculty of Philosophy, Ćirila i Metodija 2, 18000 Niš, Yugoslavia

E-mail address: mciric@archimed.filfak.ni.ac.yu

Univeristy of Niš, Faculty of Mechanical Engineering, Beogradska 14, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 619–632
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A CONNECTION BETWEEN CUT ELIMINATION AND NORMALIZATION

Mirjana Borisavljević

ABSTRACT. Sequent systems for classical and intuitionistic logic and natural deduction systems for these logics are characterized by two important theorems. Sequent systems are characterized by cut- elimination theorems, and natural deduction systems by normalization theorems. In this paper, by means of multicategories and the typed λ-calculus we exhibit some similarities and differences between cut elimination and normalization. We consider the sequent system and the natural deduction system for intuitionistic propositional logic. We define a multicategory corresponding to the sequent system. On the other hand, a typed λ-calculus corresponds to the natural-deduction system. We show how to form a typed λ-calculus out of a multicategory and vice-versa. In some kinds of multicategory, some equations necessary for cut elimination, are not necessary for normalization.

Introduction

In this paper we shall consider Gentzen's sequent system and his system of natural deduction for propositional intuitionistic logic. We shall investigate the connection between cut elimination in the sequent system and normalization in natural deduction.

In their papers Zucker and Pottinger have already described this connection, in a certain way. In this paper this will be done by linking multicategories and typed λ-calculi.

Certain multicategories will correspond to Gentzen's sequent system for intuitionistic propositional logic. Objects of these multicategories will be formulas and operations on objects will be logical connective. We shall take the arrows as proofs and operations on arrows will correspond to inference

I would like to express my gratitude to Professor Kosta Došen for many useful and important comments and advices he made on my manuscript.

rules. The arrow $\Gamma \to A$, which way have been constructed from other arrows by applying operations on arrows, will correspond to a particular derivation of the sequent $\Gamma \vdash A$.

Equations between arrows in multicategories equate arrows which correspond to derivations with the same end-sequent. On the basis of the equality $f = g$ we shall be able to transform the derivation corresponding to the arrow f into the derivation corresponding to the arrow g. Equations between arrows which we shall assume will make cut elimination possible in some multicategories; for example, in multicategories with axioms which are closed for cut. In these multicategories equations can be explained in the following way: there will be two kinds of equation. In the equations of the first kind the arrows on the left-hand side will correspond to a derivation in which cut has to be eliminated and the arrows on the right-hand side will correspond to a derivation in which cuts are eliminated or of smaller degree than the cut on the left-hand side. Equations of the second kind equate arrows which are constructed by application of the same operations on the same arrows and the only difference is in the order of application of these operations.

A typed λ-calculus will correspond to the system of natural deduction of intuitionistic propositional logic. Types of the λ-calculus will correspond to formulas and a term of type A will be considered as a derivation with the end-formula A. We shall define equations on terms that will correspond to the steps of reduction leading to proofs in normal form. We shall postulate equations on terms that will represent reductions in natural deduction such that the middle part of the proof is made of atomic formulas.

In Section 1 four kinds of multicategory will be defined; they differ in equations their arrows. In MG-multicategories with axioms which are closed for cut equations will represent steps of transformation of a derivation into a derivation where no cut appears. In the other kinds of multicategory, called MGI, MN and MNI, we shall require other equations on arrows. The equations of MN and MNI will be closer to normalization of proofs.The polynomial multicategory $\mathcal{M}[X]$ will be formed out of a multicategory a $\mathcal{M}$ in a standard way (cf.[2]). Functional completness will be proved for all those multicategories.

In Section 2 two kinds of typed λ-calculus will be defined. One kind of typed λ-calculus will have equations between terms that correspond to reduction steps in the normalization of proofs. The other kind of typed λ -calculus will have added equations by which formulas of the middle part of the proof are broken into their atomic subformulas.

In Section 3 we shall define how typed λ-calculi can be formed out of objects and arrows of multicategories and vice versa. We shall be able to form only an MN-multicategory and an MNI-multicategory out of the given typed λ-calculus. Then it will be possible to separate some equations which are needed for cut elimination in some extensions of Gentzen's sequent system, but are not needed for normalizing proofs in these systems.

1. Multicategories

Definition 1.1. A **multigraph** is made of a class of objects and a class of arrows together with two mappings: source: $\{\text{arrows}\} \rightarrow \{\text{objects}\}^*$, target: $\{\text{arrows}\} \rightarrow \{\text{objects}\}$, where $\{\text{objects}\}^*$ is the free monoid generated by the class of objects; $f : \Gamma \rightarrow A$ is an arrow, where $\Gamma = A_1 ... A_n$ is string of objects (our *arrows* are sometimes called *multiarrows*).

Definition 1.2. A **context-free recognition grammar** is a multigraph with operations on objects: $\wedge$, $\vee$ and $\Rightarrow$; special object 0. We also have operations on arrows:
structural rules:

$$\frac{f:\Gamma AB\Delta\rightarrow C}{p_{A,B}(f):\Gamma BA\Delta\rightarrow C} \qquad \text{permutation}$$

$$\frac{f:\Gamma\rightarrow C}{t_A(f):A\Gamma\rightarrow C} \qquad \text{thinning}$$

$$\frac{f:AA\Gamma\rightarrow C}{c_A(f):A\Gamma\rightarrow C} \qquad \text{contraction}$$

$$\frac{f:\Gamma\rightarrow A \quad g:A\Delta\rightarrow C}{g\langle f\rangle:\Gamma\Delta\rightarrow C} \qquad \text{cut}$$

connective rules:

MI $\dfrac{f:A\Gamma\rightarrow C}{f_p:A\wedge B\Gamma\rightarrow C}$ $\dfrac{g:B\Gamma\rightarrow C}{g_{p'}:A\wedge B\Gamma\rightarrow C}$ MII $\dfrac{f:\Gamma\rightarrow A \quad g:\Gamma\rightarrow B}{\langle f,g\rangle:\Gamma\rightarrow A\wedge B}$

MIII $\dfrac{f:A\Gamma\rightarrow C \quad g:B\Gamma\rightarrow C}{[f,g]:A\vee B\Gamma\rightarrow C}$ MIV $\dfrac{f:\Gamma\rightarrow A}{{}_k f:\Gamma\rightarrow A\vee B}$ $\dfrac{g:\Gamma\rightarrow B}{{}_{k'} g:\Gamma\rightarrow A\vee B}$

MV $\dfrac{f:\Gamma\rightarrow \quad g:B\Delta\rightarrow C}{g[f]:A\Rightarrow B\Gamma\Delta\rightarrow C}$ MVI $\dfrac{f:A\Gamma\rightarrow B}{f^*:\Gamma\rightarrow A\Rightarrow B}$

and special arrows:
MVII $1_A : A \rightarrow A$, for each object A from the class of objects.
MVIII $\square^A : 0 \rightarrow A$, for each object A from the class of objects.

Definition 1.3.

1. Let $f : \Gamma\Delta\Lambda\Theta \rightarrow A$ and $\Delta = C_1...C_n$, $\Lambda = D_1...D_m$, then $p_{\Delta,\Lambda}(f) =_{def.} p_{C_1,D_m}(...(p_{C_{n-1},D_1}(p_{C_n,D_m}...(p_{C_n,D_1}(f))...))...)$ and $p_{\Delta,\Lambda}(f) : \Gamma\Lambda\Delta\Theta \rightarrow A$.

2. Let $f : \Gamma \rightarrow A$ and $\Delta = B_1...B_n$; then $t_\Delta(f) =_{def.} t_{B_1}(...(t_{B_m}(f))..)$ and $t_\Delta(f) : \Delta\Gamma \rightarrow A$.

3. Let $f : \Delta\Delta\Gamma \rightarrow A$ and $\Delta = B_1...B_n$, $h : B_i B_{i+1}...B_n B_i B_{i+1}...B_n B_1...B_{i-1}\Gamma \rightarrow A$, $\Delta_i = B_{i+1}...B_n$, $\Delta_n = \emptyset$, $B_0 = B_{n+1} = \emptyset$, where $\emptyset$ is the empty string and $1 \leq i \leq n$; then

$c^{B_i}(h) =_{def} p_{B_i,\Delta_i\Delta_i B_1 \ldots B_{i-1}}(c_{B_1}(p_{B_{i+1},B_1}(\ldots(p_{B_n,B_i}(h))\ldots)))$ and
$c_\Delta(f) =_{def} c^{B_n}(c^{B_{n-1}}(\ldots(c^{B_1}(f))\ldots)))$ and
$c_\Delta(f) : \Delta\Gamma \rightarrow A$.

Definition 1.4. A **multicategory** is a context-free recognition grammar in which the following equations on arrows hold:

M1. $f\langle 1_A \rangle = f$, where $f : A\Gamma \rightarrow B, 1_A : A \rightarrow A$

M2. $1_B\langle g \rangle = g$, where $g : \Delta \rightarrow B$, $1_B : B \rightarrow B$

M3. $h\langle g\langle f\rangle\rangle = h\langle g\rangle\langle f\rangle$, where $h : B\Lambda \rightarrow C$, $g : A\Delta \rightarrow B$, $f : \Gamma \rightarrow A$

M4. $p_{\Lambda,\Gamma}(p_{\Gamma,B}(h\langle f\rangle)\langle g\rangle) = p_{\Lambda,A}(p_{A,B}(h)\langle g\rangle)\langle f\rangle$,

where $h : AB\Delta \rightarrow C$, $g : \Lambda \rightarrow B$, $f : \Gamma \rightarrow A$

Now we shall define four kinds of multicategory. Each of them will have five families of equations which hold for their arrows.

Definition 1.5. An **MG-multicategory** (M is for "multicategory" and G is for "Gentzen") is a multicategory in which the following families of equations hold:

I PTC-equations:

PP1. $p_{B,A}(p_{A,B}(f)) = f$, for all $f : \Gamma AB\Delta \rightarrow C$.

PP2. $p_{A,B}(p_{C,D}(f)) = p_{C,D}(p_{A,B}(f))$, for all $f : \Gamma AB\Delta CD\Lambda \rightarrow E$.

PP3. $p_{A,B}(p_{A,C}(p_{B,C}(f))) = p_{B,C}(p_{A,C}(p_{A,B}(f)))$, for all $f : \Gamma ABC\Delta \rightarrow C$.

PT1. $p_{A,B}(t_C(f)) = t_C(p_{A,B}(f))$, for all $f : \Gamma AB\Delta \rightarrow D$.

PC1. $c_C(p_{A,B}(f)) = p_{A,B}(c_C(f))$, for all $f : CC\Delta AB\Lambda \rightarrow D$.

PC2. $c_C(p_{C,C}(f)) = c_C(f)$, for all $f : CC\Delta \rightarrow D$.

TC1. $p_{C,B}(t_C(c_B(f))) = c_B(p_{C,BB}(t_C(f)))$, for all $f : BB\Delta \rightarrow D$.

TC2. $c_B(t_B(f)) = f$, for all $f : \Delta \rightarrow D$.

CC1. $c_A(p_{B,AA}(c_B(p_{AA,BB}(f)))) = p_{B,A}(c_B(p_{A,BB}(c_A(f))))$, for all $f : AABB\Delta \rightarrow D$.

Derivations in this family of equations differ only in the order of appearance of the structural rules: of contraction, permutation and thinning.

II CUT-PTC-equations:

CUTP1. $p_{B,C}(g)\langle f\rangle = p_{B,C}(g\langle f\rangle)$, for all $f : \Gamma \rightarrow A$, $g : A\Delta BC\Lambda \rightarrow D$.

CUTP2. $g\langle p_{C,D}(f)\rangle = p_{C,D}(g\langle f\rangle)$, for all $f : \Gamma CD\Delta \rightarrow A$, $g : A\Lambda \rightarrow D$.

CUTT1. $t_C(g\langle f\rangle) = p_{\Gamma,C}(p_{C,A}(t_C(g))\langle f\rangle)$, for all $f : \Gamma \rightarrow A$, $g : A\Delta \rightarrow D$.

CUTT2. $g\langle t_C(f)\rangle = t_C(g\langle f\rangle)$, for all $f : \Gamma \rightarrow A$, $g : A\Delta \rightarrow D$.

CUTT3. $t_A(g)\langle f\rangle = t_\Gamma(g)$, for all $f : \Gamma \rightarrow A$, $g : A\Delta \rightarrow D$.

CUTC1. $p_{C,A}(c_C(g))\langle f\rangle = p_{C,\Gamma}(c_C(p_{\Gamma,CC}(p_{CC,A}(g)\langle f\rangle)))$,

for all $f : \Gamma \rightarrow A$, $g : CCA\Delta \rightarrow B$.

CUTC2. $g\langle c_C(f)\rangle = c_C(g\langle f\rangle)$, for all $f : CC\Gamma \rightarrow A$, $g : A\Delta \rightarrow B$.

CUTC3. $c_A(g)\langle f\rangle = c_\Gamma(p_{\Gamma,A}(g\langle f\rangle)\langle f\rangle)$, for all $f : \Gamma \rightarrow A$, $g : AA\Delta \rightarrow B$.

Derivations in this family of equations differ in the order of appearance of cut and other structural rules. For example, in one derivation cut will appear first and permutation will follow and in the other permutation will precede cut. However, in CUTC3 we replace one cut that comes after contraction on the left by two cuts preceding contraction on the right, and in CUTT3 we replace one cut that comes after thinning on the left by zero cuts on the right.

III IMPORTANT CUTS equations:

ICUT1. $g_p\langle\langle f,h\rangle\rangle = g\langle f\rangle$, for all $g : A\Gamma \rightarrow C$, $f : \Delta \rightarrow A$, $h : \Delta \rightarrow B$.

ICUT2. $g_{p'}\langle\langle f,h\rangle\rangle = g\langle h\rangle$, for all $g : B\Gamma \rightarrow C$, $f : \Delta \rightarrow A$, $h : \Delta \rightarrow B$.

ICUT3. $\langle f,h\rangle\langle g\rangle = \langle f\langle g\rangle, h\langle g\rangle\rangle$, for all $g : \Delta \rightarrow C$, $f : C\Gamma \rightarrow A$, $h : C\Gamma \rightarrow B$.

ICUT4. $[f,h]\langle {}_k g\rangle = f\langle g\rangle$, for all $g : \Gamma \rightarrow A$, $f : A\Delta \rightarrow C$, $h : B\Delta \rightarrow C$.

ICUT5. $[f,h]\langle {}_{k'} g\rangle = h\langle g\rangle$, for all $g : \Gamma \rightarrow B$, $f : A\Delta \rightarrow C$, $h : B\Delta \rightarrow C$.

ICUT6. $f[h]\langle g^*\rangle = p_{\Lambda,\Delta}(f\langle g\rangle\langle h\rangle)$, for all $g : A\Delta \rightarrow B$, $f : B\Gamma \rightarrow C$, $h : \Lambda \rightarrow A$.

ICUT7. $h^*\langle g\rangle = (p_{\Delta,A}(p_{A,C}(h)\langle g\rangle))^*$, for all $g : \Delta \rightarrow C$, $h : AC\Gamma \rightarrow B$.

ICUT8. $f\langle \Box^A\rangle = \Box^B$, for all $\Box^A : 0 \rightarrow A, \Box^B : 0 \rightarrow B$ and $f : A \rightarrow B$.

By equations of this family the following derivations will be equated:

1. a derivation with a cut and a derivation with one or more cuts whose cut formulas are subformulas of the cut formula of the first cut;
2. derivations which differ in the order of application of a conecctive rule and a cut.

IV TROUBLESOME EQUATION: This family consists of a single equation:

T.E. $h\langle[f,g]\rangle = [h\langle f\rangle, h\langle g\rangle]$, for all $h : C\Delta \rightarrow D$, $f : A\Gamma \rightarrow C$, $g : B\Gamma \rightarrow C$.

V PTC-CUT-CON-equations:

P∧1. $p_{A,B}(g_p) = (p_{A,B}(g))_p$, for all $g : C\Delta AB\Lambda \rightarrow E$.

P∧2. $p_{A,B}(g_{p'}) = (p_{A,B}(g))_{p'}$, for all $g : D\Delta AB\Lambda \rightarrow E$.

P∧3. $p_{A,B}(\langle f,g\rangle) = \langle p_{A,B}(f), p_{A,B}(g)\rangle$, for all $f : \Gamma AB\Delta \rightarrow C$, $g : \Gamma AB\Delta \rightarrow D$.

P∨1. $p_{A,B}([f,g]) = [p_{A,B}(f), p_{A,B}(g)]$, for all $f : C\Gamma AB\Delta \rightarrow E$, $g : D\Gamma AB\Delta \rightarrow E$.

P∨2. $p_{A,B}({}_k g) = \ {}_k(p_{A,B}(g))$, for all $g : \Gamma AB\Delta \rightarrow C$.

P∨3. $p_{A,B}({}_{k'} g) = \ {}_{k'}(p_{A,B}(g))$, for all $g : \Gamma AB\Delta \rightarrow D$.

P⇒1. $p_{A,B}(g[f]) = p_{A,B}(g)[f]$, for all $g : D\Delta AB\Lambda \rightarrow E$, $f : \Gamma \rightarrow C$.

P⇒2. $p_{A,B}(g[f]) = g[p_{A,B}(f)]$, for all $g : D\Lambda \rightarrow E$, $f : \Gamma AB\Delta \rightarrow C$.

P⇒3. $p_{A,B}(g^*) = (p_{A,B}(g))^*$, for all $g : C\Gamma AB\Delta \rightarrow D$.

T∧1. $p_{C,A\wedge B}(t_C(g_p)) = (p_{C,A}(t_C(g)))_p$, for all $g : A\Gamma \rightarrow D$.

T∧2. $p_{C,A\wedge B}(t_C(g_{p'})) = (p_{C,B}(t_C(g)))_{p'}$, for all $g : B\Gamma \rightarrow D$.

T∧3. $t_C(\langle f,g\rangle) = \langle t_C(f), t_C(g)\rangle$, for all $f : \Gamma \rightarrow A$, $g : \Gamma \rightarrow B$.

T∨1. $t_C([f,g]) = p_{A\vee B,C}([p_{C,A}(t_C(f)), p_{C,B}(t_C(g))])$,

for all $f : A\Gamma \to D,\ g : B\Gamma \to D$.

T∨2. $t_C({}_k g) =_k (t_C(g))$, for all $g : \Gamma \to A$.

T∨3. $t_C({}_{k'} g) =_{k'} (t_C(g))$, for all $g : \Gamma \to B$.

$T \Rightarrow$1. $p_{C,A\Rightarrow B}(t_C(g[f])) = g[t_C(f)]$, for all $f : \Gamma \to A,\ g : B\Delta \to D$.

$T \Rightarrow$2. $p_{C,A\Rightarrow B\Gamma}(t_C(g[f])) = p_{C,B}(t_C(g))[f]$, for all $f : \Gamma \to A,\ g : B\Delta \to D$.

$T \Rightarrow$3. $t_C(g^*) = (p_{C,A}(t_C(g)))^*$, for all $g : A\Gamma \to B$.

C∧1. $c_C(p_{A\wedge B,CC}(p_{CC,A}(g))_p) = p_{A\wedge B,C}(p_{C,A}(c_C(g)))_p$,

for all $g : CCA\Delta \to D$.

C∧2. $c_C(p_{A\wedge B,CC}(p_{CC,A}(g))_{p'}) = p_{A\wedge B,C}(p_{C,A}(c_C(g)))_{p'}$,

for all $g : CCB\Delta \to D$.

C∧3. $c_C(\langle f,g\rangle) = \langle c_C(f), c_C(g)\rangle$, for all $f : CC\Gamma \to A,\ g : CC\Gamma \to B$.

C∨1. $p_{C,A\vee B}(c_C(p_{A\vee B,CC}([g,f]))) = [p_{C,A}(c_C(p_{A,CC}(g))), p_{C,B}(c_C(p_{B,CC}(f)))]$,

for all $g : ACC\Gamma \to D,\ f : BCC\Gamma \to D$.

C∨2. $c_C({}_k g) = \ {}_k(c_C(g))$, for all $g : CC\Gamma \to A$.

C∨3. $c_C({}_{k'} g) = \ {}_{k'}(c_C(g))$, for all $g : CC\Gamma \to B$.

C⇒1. $c_C(p_{A\Rightarrow B,CC}(g[f])) = p_{A\Rightarrow B,C}(g[c_C(f)])$, for all $f : CC\Gamma \to A,\ g : B\Delta \to D$.

C⇒2. $c_C(p_{a\Rightarrow B\Gamma,CC}(p_{CC,B}(g)[f])) = p_{A\Rightarrow B\Gamma,C}(p_{C,B}(c_C(g))[f])$,

for all $f : \Gamma \to A,\ g : CCB\Delta \to D$.

C⇒3. $c_C(f^*) = (p_{C,A}(c_C(p_{A,CC}(f))))^*$, for all $f : ACC\Gamma \to B$.

CUT∧1. $p_{\Gamma,A\wedge B}(p_{A\wedge B,C}(g_p)\langle f\rangle) = (p_{\Gamma,A}(p_{A2,C}(g)\langle f\rangle))_p$

$p_{\Gamma,B\wedge A}(p_{B\wedge A,C}(g_{p'})\langle f\rangle) = (p_{\Gamma,A}(p_{A,C}(g)\langle f\rangle))_{p'}$

for all $f : \Gamma \to C,\ g : AC\Delta \to D$.

CUT∧2. $g\langle f_p\rangle = (g\langle f\rangle)_p$,

$g\langle f_{p'}\rangle = (g\langle f\rangle)_{p'}$, for all $f : A\Gamma \to C,\ g : C\Delta \to D$.

CUT∨1. $p_{\Gamma,A\vee B}(p_{A\vee B,C}([f,g])\langle h\rangle) = [p_{\Gamma,A}(p_{A,C}(f)\langle h\rangle), p_{\Gamma,B}(p_{B,C}(g)\langle h\rangle)]$,

for all $f : AC\Delta \to D,\ g : BC\Delta \to D,\ h : \Gamma \to C$.

CUT∨2. ${}_k g\langle f\rangle = \ {}_k(g\langle f\rangle)$,

${}_{k'} g\langle f\rangle = \ {}_{k'}(g\langle f\rangle)$, for all $f : \Gamma \to C,\ C\Delta \to A$.

CUT⇒1. $h\langle g[f]\rangle = h\langle g\rangle[f]$, for all $f : \Gamma \to A$, $g : B\Delta \to C$, $h : C\Lambda \to D$.

CUT⇒2. $p_{A\Rightarrow B,C}(g[f])\langle h\rangle = p_{A\Rightarrow B,\Lambda}(g[f\langle h\rangle])$,

for all $h : \Lambda \to C$, $f : C\Gamma \to A$, $g : B\Delta \to D$.

CUT⇒3. $p_{A\Rightarrow B\Gamma,C}(g[f])\langle h\rangle = p_{A\Rightarrow B\Gamma,\Lambda}(p_{\Lambda,B}(p_{B,C}(g)\langle h\rangle)[f])$,

for all $h : \Lambda \to C$, $g : BC\Delta \to D$, $f : \Gamma \to A$.

Equations of this family equate derivations which differ in the order of applying connective rules and structural rules.

Definition 1.6. An **MN-multicategory** (N stands for "Natural deduction") is a multicategory in which hold all the equations that hold for MG-multicategories except that the TROUBLESOME EQUATION is replaced by the following special cases of this equation:

NE1. $1_{Cp}\langle [f,g]\rangle = [1_{Cp}\langle f\rangle, 1_{Cp}\langle g\rangle]$,

NE2. $1_{Dp'}\langle [f,g]\rangle = [1_{Dp'}\langle f\rangle, 1_{Dp'}\langle g\rangle]$, for all $f : A\Gamma \to C\wedge D$, $g : B\Gamma \to C\wedge D$.

NE3. $1_D[h_1]\langle [f,g]\rangle = [1_D[h_1]\langle f\rangle, 1_D[h_1]\langle g\rangle]$,

for all $h_1 : \Delta \to C$, $f : A\Gamma \to C\Rightarrow D$, $g : B\Gamma \to C\Rightarrow D$.

NE4. $[h_1, h_2]\langle [f,g]\rangle = [[h_1, h_2]\langle f\rangle, [h_1, h_2]\langle g\rangle]$,

for all $h_1 : C\Delta \to E$, $h_2 : D\Delta \to E$, $f : A\Gamma \to C\vee D$, $g : B\Gamma \to C\vee D$.

NE5. $\Box^C\langle [f,g]\rangle = [\Box^C\langle f\rangle, \Box^C\langle g\rangle]$, for all $f : A\Gamma \to 0$, $g : B\Gamma \to 0$.

We shall call these normalizing equations (NOR-equations) because they correspond to steps the normalizing of a derivation.

Definition 1.7. An **MGI-multicategory** is a multicategory in which hold all the equations that hold for MG-multicategories except that PTC-CUT-CON equations are replaced by the following equations:

I∧. $\langle 1_{Ap'}, 1_{Bp'}\rangle = 1_{A\wedge B}$.

I∨. $[_k 1_A, _{k'} 1_B] = 1_{A\vee B}$.

I⇒. $(p_{A\Rightarrow B,A}(1_B[1_A]))^* = 1_{A\Rightarrow B}$.

I0. $\Box^0 = 1_0$.

This family of equations will be called identity equations (ID-equations).

Definition 1.8. An **MNI-multicategory** is a multicategory in which hold all the equations that hold for MN-multicategories except that PTC-CUT-CON-equations are replaced by ID-equations.

It can be easily seen that if the TROUBLESOME EQUATION holds, then the NOR-equations hold, too. This means that each MG-multicategory is an MN-multicategory. In some multicategories the TROUBLESOME EQUATION cannot be derived from the NOR-equations and other equations. For example, let the arrow $h : C\Delta \rightarrow D$ be an axiom, where C is an atomic formula then the TROUBLESOME EQUATION cannot be derived from NE1-NE5 for the arrow h.

It can be shown that from the M1-M4 equations, the equations of the family IMPORTANT CUTS and the ID-equations, we can derive the PTC-CUT-CON-equations. According to that each MG-multicategory is an MGI-multicategory, and in the same way each MN-multicategory is an MNI-multicategory.

From now on in this text, if not stated otherwise, we shall take as multicategory $\mathcal{M}$ a multicategory of any of the four kinds previously defined.

Now we shall form in a standard way a polynomial multicategory $\mathcal{M}[X]$ out of a multicategory $\mathcal{M}$. Suppouse a multicategory $\mathcal{M}$ is given. The polynomial multicategory $\mathcal{M}[X]$ will be of the same kind as the multicategory $\mathcal{M}$ itself. In a usual way, a set of new arrows $X = \{x_A : \ \rightarrow A : A$ is an object of $\mathcal{M}$ $\}$ is added to the multicategory $\mathcal{M}$. Arrows of $\mathcal{M}[X]$ will be arrows of $\mathcal{M}$, arrows from X and arrows obtained with the help of the operations on arrows extending those assumed for $\mathcal{M}$. The arrow of $\mathcal{M}[X]$ will have the form $\varphi(x_1, ..., x_n) : \Gamma \rightarrow A$, where x_1, ..., x_n are arrows from X which can appear in the construction of the arrow the φ. Arrows x_i, $1 \leq i \leq n$ need not occur in this order in construction of φ; they can occur several times, and they need not occur explicitly. The next step is making a multicategory out of $\mathcal{M}[X]$. For this we need some equations that hold on arrows in $\mathcal{M}$. For two arrows of $\mathcal{M}[X]$ which have the some source and target consider all the equivalence relations $\equiv_Y$ for some $Y \subseteq X$ on arrows of $\mathcal{M}[X]$. Which basides some of the families of equations used for defining MG, MGI, MN and MNI-multicategories must satisfy also
X. the equations following form:
if $\varphi \equiv_Y \psi$ then $\varphi_p \equiv_Y \psi_p$, where $\varphi, \psi : A\Gamma \rightarrow C$ in $\mathcal{M}[X]$, and similar by with other operations on arrows.

The relations $\equiv_Y$ will satisfy the families equations assumed for the kind of multicategory to which $\mathcal{M}$ belongs. Then $\cong_Y$ is the smallest of the equivalence relations $\equiv_Y$ (cf. [2], p. 57). For example, let $\mathcal{M}$ be an MG-multicategory; then $\equiv_X$ satisfies the following families of equations: X, M, PTC, CUT-PTC, IMP-CUT, TR-EQ, PTC-CUT-CON. The relation $\cong_X$ is the smallest equivalence relation which satisfies these families of equations. Then $\mathcal{M}[X]$ with $\cong_X$ is a polynomial MG-multicategory.

Let $\mathcal{M}$ be a multicategory of one of the kinds defined above. The functional completness theorem will hold for the multicategory $\mathcal{M}$.

Theorem 1.1. *Let $\varphi(x_1, ..., x_n) : \Gamma \rightarrow A$ be an arbitrary arrow of multicategory $\mathcal{M}[X]$, where $\Gamma = B_1 ... B_m$ and $x_i : \ \rightarrow A_i$, $1 \leq i \leq n$. Then for an arbitrary order of formulas the A_i, $i \leq i \leq n$, for example $A_1 ... A_n$, there is* *a unique arrow $\overline{\varphi}$ in the multicategory $\mathcal{M}$, such that:*

$$\overline{\varphi} : A_1 ... A_n \Gamma \rightarrow A \quad \textit{and} \quad \overline{\varphi}\langle x_1\rangle ... \langle x_n\rangle \cong_Y \varphi(x_1, ..., x_n), \qquad (**)$$

where $Y = \{x_1, ..., x_n\}$.
The arrow $\overline{\varphi}_{A_1 ... A_n}$ will be denoted simply by $\overline{\varphi}$, and $A_1 ... A_n$ is Δ.

Proof. Only a sketch of the proof will be given here.
I We shall first show that the arrow $\overline{\varphi}$ of $\mathcal{M}$ exists for each arrow $\varphi(x_1, ..., x_n)$ of $\mathcal{M}[X]$ for which hold (**). The proof will be executed by induction on the complexity of the arrow $\varphi(x_1, ... x_n)$. The a arrow $\varphi(x_1, ... x_n)$ must have one of the following forms:
1. f, where f is an arrow of $\mathcal{M}$; 2. $x_i : \rightarrow A_i$; 3. $p_{D,C}(\psi)$; 4. $c_B(\psi)$; 5. $t_E(\psi)$;
6. ψ_p; 7. $\psi_{p'}$; 8. $\langle\psi, \xi\rangle$; 9. $[\psi, \xi]$; 10. ${}_k\psi$; 11. ${}_{k'}\psi$; 12. $\psi[\xi]$; 13. ψ^*; 14. $\psi\langle\xi\rangle$;
15. $\psi\langle g\rangle$; 16. $g\langle\psi\rangle$, where g in 15. and 16. is an arrow of $\mathcal{M}$.

Arrows ψ and ξ are of smaller complexity than arrow φ and then on the basis of inductional hypothesis there are $\overline{\psi}$, $\overline{\xi}$ in $\mathcal{M}$ for which:

$$\overline{\psi}\langle x_1\rangle ... \langle x_n\rangle \cong_Y \psi(x_1, ..., x_n) \text{ and } \overline{\xi}\langle x_1\rangle ... \langle x_n\rangle \cong_Y \xi(x_1, ..., x_n)$$

Then $\overline{\varphi}$ is:
1. $\overline{f} = t_{A_1}(...(t_{A_n}(f))...)$; 2. $\overline{x_i} = t_{A_1} ... t_{A_{i-1}}(p_{A_{i+1} ... A_n, A_i}(t_{A_{i+1} ... A_n}(1_{A_i})))$;
3. $p_{D,C}(\overline{\psi})$; 4. $p_{B,A_1 ... A_n}(c_B(p_{A_1 ... A_n, BB}(\overline{\psi})))$; 5. $p_{E,A_1 ... A_n}(t_E(\overline{\psi}))$;
6. $p_{B \wedge C,\Delta}((p_{\Delta,B}(\overline{\psi}))_p)$; 7. $p_{C \wedge B,\Delta}((p_{\Delta,B}(\overline{\psi}))_{p'})$; 8. $\langle\overline{\psi}, \overline{\xi}\rangle$;
9. $p_{B \vee C,\Delta}([p_{\Delta,B}(\overline{\psi}), p_{\Delta,C}(\overline{\xi})])$; 10. ${}_k\overline{\psi}$; 11. ${}_{k'}\overline{\psi}$;
12. $c_\Delta(p_{B \Rightarrow C,\Delta\Delta}(p_{B_1 ... B_k,\Delta}(p_{\Delta,C}(\overline{\psi})[\overline{\xi}])))$; 13. $(p_{\Delta,B}(\overline{\psi}))^*$;
14. $c_\Delta(p_{B_1 ... B_k,\Delta}(p_{\Delta,B}(\overline{\psi})\langle\overline{\xi}\rangle))$; 15. $p_{B_1 ... B_k,\Delta}(p_{\Delta,B}(\overline{\psi})\langle g\rangle)$; 16. $g\langle\overline{\psi}\rangle$;

On the basis of the equations that hold in the multicategory $\mathcal{M}[X]$ it can be easily shown that (**) holds in all cases 1-16.
II In order to show the uniqueness of the arrow $\overline{\varphi}$, we first define an equivalence relation $\simeq_Y$ in the following way:
for arrows $\psi(y_1, ... y_m)$ and $\xi(y_1, ... y_m) : \Gamma \rightarrow A$

$$\psi \simeq_Y \xi \quad \text{if and only if} \quad \overline{\psi}_\Theta = \overline{\xi}_\Theta \quad \text{in } \mathcal{M}, \qquad (\circ)$$

where $Y = y_1, ... y_m$ and $y_i : \rightarrow B_i$ and Θ is $B_1 ... B_m$ taken in any order.

Then it is proved that the relation $\simeq_Y$ is an equivalence relation $\equiv_Y$. Since $\cong_Y$ is the smallest equivalence relation we get $\psi \cong_Y \xi$ only if $\overline{\psi} = \overline{\xi}$.

Now we suppose that there are two arrows in $\mathcal{M}$, say $\overline{\varphi}$ and $\overline{\varphi}'$ such that $\overline{\varphi}$, $\overline{\varphi}' : A_1 ... A_n\Gamma \rightarrow A$ and $\overline{\varphi}\langle x_1\rangle ... \langle x_n\rangle \cong_Y \varphi$, $\overline{\varphi}'\langle x_1\rangle ... \langle x_n\rangle \cong_Y \varphi$. As $\cong_Y$ is transitive we have $\overline{\varphi}\langle x_1\rangle ... \langle x_n\rangle \cong \overline{\varphi}'\langle x_1\rangle ... \langle x_n\rangle$ and by implication above we get $\overline{\overline{\varphi}\langle x_1\rangle ... \langle x_n\rangle} = \overline{\overline{\varphi}'\langle x_1\rangle ... \langle x_n\rangle}$ in $\mathcal{M}$. Then from the part I of the proof and from equations which hold in $\mathcal{M}$ we get that $\overline{\varphi} = \overline{\varphi}'$. This means that $\overline{\varphi}$ is the unique arrow of $\mathcal{M}$ for which (**) holds. □

2. Typed λ-calculi

Definition 2.1. A typed λ-calculus ΛC is a formal theory defined by classes of types, terms of each type, and equations between terms. We shal write $t \in A$ to say that t is term of type A.

Types. The class of types contains special a type 0, and is closed under three operations: $A \wedge B$, $A \vee B$, $A \Rightarrow B$, where A, B are types.

Terms 2.1. For each type A there are countably many variables of type A: $x_i \in A$, $i = 1, 2, 3, \ldots$
2.2. If $t \in A \wedge B$ then $Lt \in A$ and $Rt \in B$.
2.3. If $u \in A$ and $t \in B$, then $\Pi(u, t) \in A \wedge B$.
2.4. If $t(x) \in C$ and $x \in A$, $s(y) \in C$, $y \in B$, $u \in A \vee B$, then $\delta_{x,y}(u, t(x), s(y)) \in C$.
2.5. If $t \in A$ and $s \in B$, then $K_B t \in A \vee B$, $K'_A s \in A \vee B$.
2.6. If $u \in A$ and $t \in A \Rightarrow B$, then $tu \in B$.
2.7. If $u(x) \in B$ and $x \in A$, then $\lambda_x u \in A \Rightarrow B$.
2.8. If $t \in 0$, then $\iota_A t \in A$.

Equations. Equations have the form $t =_X s$, where t and s have the same type A, and X is a finite set of variables such that all variables occuring freely in t and s are in X.
3.1. $=_X$ is an equivalence relation.
3.2. If $t =_Y s$ and $Y \subseteq X$, then $t =_X s$.
3.3. $=_X$ satisfies the usual substitution rules for all terms forming operations. For example:

$$\text{if } t_1 =_X t_2 \text{ and } s_1 =_X s_2 \text{ then } \Pi(t_1, s_1) =_X \Pi(t_2, s_2)$$

where $t_1, t_2 \in A$, $s_1, s_2 \in B$.

If $t(x)$ is a term of type A, $x \in B$ and s is a term of type B then $t[x/s]$ is the result of replacing an occurence of x in the term t by the term s.

Definition 2.2. A typed λ-calculus is **ΛC-ND** if the following equations on terms hold in it:

L. $L(\Pi(t, s)) =_X t$, for all $t \in A$, $s \in B$.

R. $R(\Pi(t, s)) =_X s$, for all $t \in A$, $s \in B$.

K. $\delta_{x,y}(K_A u, t(x), s(y)) =_X t[x/u]$, for all $u \in A$, $t, s \in C$, $x \in A$, $y \in B$.

K'. $\delta_{x,y}(K'_B u, t(x), s(y)) =_X s[y/u]$, for all $u \in B$, $t, s \in C$, $x \in A$, $y \in B$.

β. $(\lambda_x t)u =_X t[x/u]$, for all $u \in A$, $t \in B$, $x \in A$,

where in K, K' and β no free variable in u becomes bound in $t[x/u]$ and $s[y/u]$.

0. $t[x/\iota_A(z)] =_X \iota_B z$, for all $x \in A$, $t \in B$, $z \in 0$.

N1. $L\delta_{x,y}(u, t(x), s(y)) =_X \delta_{x,y}(u, Lt(x), Ls(y))$,

for all $u \in A \vee B$, $x \in A$, $y \in B$ $t, s \in C \wedge D$.

N2. $R\delta_{x,y}(u, t(x), s(y)) =_X \delta_{x,y}(u, Rt(x), Rs(y))$,

for all $u \in A \vee B$, $x \in A$, $y \in B$ $t, s \in C \wedge D$.

N3. $\delta_{x_1,x_2}(\delta_{x,y}(u, t(x), s(y)), v_1(x_1), v_2(x_2)) =_X$

$\delta_{x,y}(u, \delta_{x_1,x_2}(t, v_1(x_1), v_2(x_2))(x), \delta_{x_1,x_2}(s, v_1(x_1), v_2(x_2))(y)))$,

for all $u \in A \vee B$, $x \in A$, $y \in B$, $t, s \in C \vee D$, $x_1 \in C$, $x_2 \in D$, $v_1, v_2 \in E$.

N4. $\delta_{x,y}(u, t(x), s(y))v =_X \delta_{x,y}(u, tv(x), sv(y))$,

for 'all $u \in A \vee B$, $x \in A$, $y \in B$, $t, s \in C \Rightarrow D$, $v \in C$.

N5. $\iota_C(\delta_{x,y}(u, t(x), s(y)) =_X \delta_{x,y}(u, \iota_C t(x), \iota_C s(y))$,

for all $u \in A \vee B$, $x \in A$, $y \in B$, $t, s \in \Lambda$.

The terms of typed a λ-calculus code derivations of the natural deduction system for intuitionistic propositional logic. Equations L, R, K, K', β and 0 correspond to reduction steps in the normalization theorem when a derivation in which maximum formula appears is transformed into derivation without maximum formula. Equations N1, N2, N3, N4, N5 correspond to reduction steps by which a maximum segment in a derivation is eliminated. Terms on the left-hand side of all equations will correspond to derivations in which a maximum formula or a maximum segment occurs. Those derivations are transformed by reductions into derivations corresponding to terms on the right-hand side of equations.
Equations on terms in a typed λ-calculus ΛC-ND equate derivations with their normal forms.

Definition 2.3. A typed λ-calculus ΛC-ND is Λ **C-NDA** if the following equations on terms hold in it:

$\eta\wedge$. $z =_X \Pi(Lz, Rz)$, for all $z \in A \wedge B$.

$\eta\vee$. $z =_X \delta(z, K_B x, K'_A y)$, for all $z \in A \vee B$, $x \in A$, $y \in B$.

$\eta\Rightarrow$. $z =_X \lambda_x zx$, for all $z \in A \Rightarrow B$, $x \in A$.

$\eta 0$. $z =_X \iota_0 z$, for all $z \in 0$.

Equations on terms $\eta\wedge$, $\eta\vee$, $\eta \Rightarrow$ and $\eta 0$ equate derivations consisting of nonatomic formulas in their middle part with derivations whose middle part consists of atomic formulas.

3. Connection between multicategories and typed λ-calculi

In this section we show how to form a typed λ-calculus out of a multicategory $\mathcal{M}$ and how a multicategory can be formed from a typed λ-calculus $\mathcal{L}$.

First we shall define types and terms LC($\mathcal{M}$) from objects and arrows of some multicategory $\mathcal{M}$.

LCI 1. the types of LC($\mathcal{M}$) will be the objects of $\mathcal{M}$.

2. the operations on types in LC($\mathcal{M}$) will be the operations on objects of $\mathcal{M}$.

LCII Every arrow term $\varphi(x_1,...,x_n): \ \rightarrow A$ of $\mathcal{M}[X]$ is a term of type A in LC($\mathcal{M}$), and all the free variables of $\varphi(x_1,...,x_n)$ are in $X = \{x_1,..,x_n\}$.

0T. $x \in A$, $x =_{def.} x: \ \rightarrow A$.

1T. If $\varphi(x_1,...x_n): \ \rightarrow A \wedge B$, then $L\varphi =_{def} 1_{Ap}\langle\varphi\rangle$.

2T. If $\varphi(x_1,...,x_n): \ \rightarrow A \wedge B$, then $R\varphi =_{def} 1_{Bp'}\langle\varphi\rangle$.

3T. If $\varphi(x_1,...,x_n): \ \rightarrow A$, $\psi(x_1,...,x_n): \ \rightarrow B$, then $\Pi(\varphi,\psi) =_{def} \langle\varphi,\psi\rangle$.

4T. If $\varphi(z_1,...,z_k): \ \rightarrow A \vee B$, $\psi(x_1,...,x,...,x_n): \ \rightarrow C$,
$\xi(y_1,...,y,...,x_m): \ \rightarrow C$, $x: \ \rightarrow A$, $y: \ \rightarrow B$
then $\delta_{x,y}(\varphi,\psi(x),\xi(y)) =_{def} [\overline{\psi},\overline{\xi}]\langle\varphi\rangle$,
where $\overline{\psi}$ is defied with respect to $Y = \{x\}$ and $\overline{\xi}$ is defined with respect to $Y = \{y\}$.

5T. If $\varphi(x_1,...,x_n): \ \rightarrow A$, then $K_B\varphi =_{def} {}_k\varphi$.

6T. If $\varphi(x_1,...,x_n): \ \rightarrow B$, then $K'_A\varphi =_{def} {}_{k'}\varphi$.

7T. If $\varphi(x_1,...,x_n): \ \rightarrow A$, $\psi(x_1,...,x_n): \ \rightarrow A \Rightarrow B$,
then $\psi\varphi =_{def} 1_B[1_A]\langle\psi\rangle\langle\varphi\rangle$.

8T. If $\varphi(x_1,...,x_n): \ \rightarrow B$, $x: \ \rightarrow A$, then $\lambda_x\varphi =_{def} \overline{\varphi}^*$,
where $\overline{\varphi}$ is defined with respect to $Y = \{x\}$.

9T. If $\varphi(x_1,...,x_n): \ \rightarrow 0$, then $\iota_A\varphi =_{def} \Box^A\langle\varphi\rangle$.

LCEQ. If $\varphi(x_1,...,x_n), \psi(x_1,...,x_n): \ \rightarrow A$,
then $\varphi =_X \psi$ if and only if $\varphi \cong_X \psi$ in $\mathcal{M}[X]$, where $X = \{x_1,...,x_n\}$.

Now let one typed λ-calculus, $\mathcal{L}$, be given. We shall define objects and arrows of MC($\mathcal{L}$) from types and terms of the typed λ-calculus $\mathcal{L}$ in the following way:

MCI 1. the objects of MC($\mathcal{L}$) will be the types of $\mathcal{L}$.

2. the operations on objects in MC($\mathcal{L}$)will be the operrations on types of $\mathcal{L}$.

MCII An arrow $f: \ A_1...A_n \rightarrow A$ in MC($\mathcal{L}$) will be $(x_1...x_n; \varphi(x_1,...,x_n) \in A)$ where $\varphi(x_1,...,x_n)$ is the term of $\mathcal{L}$ and $x_1,...,x_n$ are all free variables of term φ, and $x_1 \in A_1$, ..., $x_n \in A_n$.

0A1. $1_A =_{def} (x; x \in A)$

0A2. $\Box^A =_{def} (x \in 0; \iota_A x \in A)$

1A. If $f = (x_1...x_n x y y_1...y_m; \varphi \in C)$, $x \in A$, $y \in B$
then $p_{A,B}(f) =_{def} (x_1...x_n y x y_1...y_m; \varphi \in C)$.

2A. If $f = (x x x_1...x_n; \varphi \in C)$, $x \in A$, then $c_A(f) =_{def} (x x_1...x_n; \varphi \in C)$.

3A. If $f = (x_1...x_n; \varphi \in X)$, $x \in A$, then $t_A(f) =_{def} (x x_1...x_n; \varphi \in C)$.

4A. If $f = (x_1...x_n; \varphi \in A)$, $g = (x y_1...y_m; \psi \in B)$, $x \in A$,
then $g\langle f\rangle =_{def} (x_1...x_n y_1...y_m; \psi[x/\varphi] \in B)$.

5A. If $f = (x x_1...x_n; \varphi \in C)$, $x \in A$, $z \in A \wedge B$,
then $f_p =_{def} (z x_1...x_n; \varphi[x/Lz] \in C)$.

6A. If $f = (y x_1...x_n; \varphi \in C$, $y \in B$, $z \in A \wedge B$,
then $f_{p'} =_{def} (z x_1...x_n; \varphi[y/Rz] \in C)$.

7A. If $f = (x_1...x_n; \varphi \in A)$, $g(x_1...x_n; \psi \in B)$,
then $\langle f,g\rangle =_{def} (x_1...x_n; \Pi(\varphi,\psi) \in A \wedge B)$.

8A. If $f = (xx_1...x_n; \varphi \in C)$, $g = (yx_1...x_n; \psi \in C)$, $x \in A$, $y \in B$ and $z \in A \wedge B$,
then $[f, g] =_{def} (zx_1...x_n; \delta_{x,y}(\varphi(x), \psi(y)) \in C)$.
9A. If $f = (x_1...x_n; \varphi \in A)$, then ${}_k f =_{def} (x_1...x_n; K_B\varphi \in A \vee B)$.
10A. If $f = (x_1...x_n; \psi \in B)$, then ${}_{k'} f =_{def} (x_1...x_n; K'_A\psi \in A \vee B)$.
11A. If $f = (x_1...x_n; \varphi \in A)$, $g = (xy_1...y_m; \psi \in C)$, $x \in B$, $y \in A$ $z \in A \Rightarrow B$,
then $g[f] =_{def} (x_1...x_n y_1...y_m; \psi[x/zy[y/\varphi]] \in C)$.
12A. If $f = (xx_1...x_n; \varphi \in B)$, then $f^* =_{def} (x_1...x_n; \lambda_x\varphi \in A \Rightarrow B)$.
MCEQ. If $f = (x_1...x_n; \varphi \in A)$, $g = (y_1...y_n; \psi \in A)$,
then $f = g$ if and only if 1. $x_i, y_i \in A_i$, $1 \leq i \leq n$, 2. $\varphi =_X \psi[\overline{y}/\overline{x}]$,
where $\overline{x} = x_1...x_n$, $\overline{y} = y_1...y_n$ and $X = \{x_1...x_n\}$.

In this way types and terms of LC($\mathcal{M}$) and objects and arrows of MC($\mathcal{L}$) are defined. It is still neccessary to show that LC($\mathcal{M}$) is a typed λ-calculus and that MC($\mathcal{L}$) is a multicategory.

It has to be shown that the corresponding equations hold for terms of LC($\mathcal{M}$). Depending on what kind of multicategory $\mathcal{M}$ belongs to, LC($\mathcal{M}$) will be LC-ND or LC-NDA typed λ-calculus. The following theorem holds:

Theorem 3.1.

1. *If $\mathcal{M}$ is an MN-multicategory, then LC($\mathcal{M}$) is a ΛC-ND typed λ-calculus.*
2. *If $\mathcal{M}$ is an MNI-multicategory, then LC($\mathcal{M}$) is a ΛC-NDA typed λ-calculus.*

Proof. In both cases we have to verify what equations hold on terms in LC($\mathcal{M}$).
1. As an example we show that the equation N1. holds:
$L\delta(u, t(x), s(y)) =_X \delta(u, Lt(x), Ls(y)$, $u : \ \to A \vee B$, $t, s : \ \to C \wedge D$, $x : \ \to A$, $y : \ \to B$.
$Lt =_{def} 1_{Cp}\langle t\rangle$, $Ls =_{def} 1_{Cp}\langle s\rangle$, $\delta_{x,y}(u, t(x), s(y)) =_{def} [\overline{t}, \overline{s}]\langle u\rangle$,
$\delta_{x,y}(u, Lt(x), Ls(y)) =_{def} [\overline{1_{Cp}\langle t\rangle}, \overline{1_{Cp}\langle s\rangle}]\langle u\rangle$, $L\delta_{x,y}(u, t(x), s(y)) =_{def}$
$1_{Cp}\langle[\overline{t}, \overline{s}]\langle u\rangle\rangle$, $1_{Cp}\langle[\overline{t}, \overline{s}]\langle u\rangle\rangle =_X 1_{Cp}\langle[\overline{t}, \overline{s}]\rangle\langle u\rangle =_X [\overline{1_{Cp}\langle t\rangle}, \overline{1_{Cp}\langle s\rangle}]\langle u\rangle$,
then $L\delta_{x,y}(u, t(x), s(y)) =_X \delta_{x,y}(u, Lt(x), Ls(y))$.

It can be proved in a similar way that all the other equations hold.

2. As an example we will show that the equation $\eta\wedge$ holds and that it depends on the equation I1: $z =_X \Pi(Lz, Rz)$.
$z : \ \to A \Rightarrow B$, $Lz =_{def} 1_{Ap}\langle z\rangle$, $Rz =_{def} 1_{Bp'}\langle z\rangle$,
$\Pi(Lz, Rz) =_{def} \langle 1_{Ap}\langle z\rangle, 1_{Bp'}\langle z\rangle\rangle$,
$\langle 1_{Ap}\langle z\rangle, 1_{Bp'}\langle z\rangle\rangle =_X \langle 1_{Ap}, 1_{Bp'}\rangle\langle z\rangle =_X 1_{A \wedge B}\langle z\rangle =_X z$,
then $z =_X \Pi(Lz, Rz)$ in LC($\mathcal{M}$).

All the other equations are verified in a similar way. □

Let $\mathcal{M}$ be a multicategory of any of the four kinds previously defined and let its axioms be closed for cut. This means that if the arrows $f : A\Delta \to B$ and $g : \Gamma \to A$ are axioms of $\mathcal{M}$ and A is atomic formula, then the arrow $f\langle g\rangle$ is an axiom in $\mathcal{M}$, too. In this multicategory equations on arrows connect a derivation in which cut occurs and a derivation in which cut is eliminated. Equations on terms connect a derivation with the normal form of this derivation. Roughly speaking, equations we need for cut elimination yield normalization.

On the other hand, let us see what we get from equations which we need for normalization. These equations will 'select' from the equations which we need for cut elimination only those really required. We shall show that from LC-ND and LC-NDA typed λ-calculi we obtain only MN-multicategories and MNI-multicategories, which are not necessarily MG and MGI-multicategories.

The following theorem shows this:

Theorem 3.2.

1. If $\mathcal{L}$ is a ΛC-ND typed λ-calculus, then MC($\mathcal{L}$) is an MN-multicategory.
2. If $\mathcal{L}$ is a ΛC-NDA typed λ-calculus, then MC($\mathcal{L}$) is an MNI-multicategory.

Proof. We have to check that the equations which must hold in MN-multicategories, respectively MNI-multicategories, also hold for arrows in MC($\mathcal{L}$).

Some remarks.

The free MG-multicategory corresponds to Gentzen's sequent system for intuitionistic propositional logic. Some MG-multicategories with axioms which are closed for cut are extensions of this system. Equations between arrows which must hold in an MG-multicategory don't make cut elimination possible in an MG-multicategory with arbitrary atomic axioms.

The situation is analogous with MN-multicategories.

We can investigate the connection between some MG-multicategories mentioned above and the corresponding MN-multicategories. This can be another way to link cut elimination and normalization, without going via typed λ-calculi.

References

[1] G. Gentzen, *The Collected Papers of Gerhard Gentzen*, North-Holland, Amsterdam, 1969.

[2] J. Lambek and Ph. J. Scott, *Introduction to Higher Order Categorical Logic*, Cambridge University Press, Cambridge, 1986.

[3] D. Prawitz, *Natural Deduction, A Proof-Theoretical Study*, Almqvist and Wiksell, Stockholm, 1965.

[4] G. Pottinger, *Normalization as a homomorphic image of cut elimination*, Annals of Mathematical Logic **12** (1977), 323-357.

[5] J. I. Zucker, *Cut elimination and normalization*, Annals of Mathematical Logic **1** (1974), 1-112.

Faculty of Transport and Traffic Engineering, University of Belgrade,
Vojvode Stepe 305, 11000 Belgrade, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 633–638
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON WEAK CONGRUENCE MODULAR LATTICES

Ivan Chajda, Branimir Šešelja* and Andreja Tepavčević*

ABSTRACT. The main result of the paper is a characterization of weak congruence modular varieties (every algebra of which has a modular lattice of weak congruences). Varieties are supposed to have a nullary operation, and every algebra a one element subalgebra. It is proved that such a variety is weak congruence modular if and only if it is polynomially equivalent to the variety of modules over a ring with unit. Some other characterizations of such varieties and of algebras in these varieties having distributive weak congruence lattices, are also given.

1. Introduction

A variety $\mathcal{V}$ whose similarity type contains a nullary operation 0 and every algebra of which has a one element subalgebra is a 0_1-*variety*. An algebra with a nullary operation 0 is 0-*regular* if $\theta = \phi$ for each $\theta, \phi \in Con\mathcal{A}$, whenever $[0]_\theta = [0]_\phi$. A variety $\mathcal{V}$ with 0 is 0-*regular* if each $\mathcal{A} \in \mathcal{V}$ has this property. A 0_1-variety which is 0-regular is a 0_1-*regular* variety. A single algebra $\mathcal{A}$ with a minimal one element subalgebra A_m is A_m-*regular* if every congruence on $\mathcal{A}$ is uniquely determined by the class containing A_m.

An algebra with 0 is *weakly coherent* if for every subalgebra $\mathcal{B}$ of $\mathcal{A}$ and each $\theta \in Con\mathcal{A}$, if $[0]_\theta \subseteq B$, then $[x]_\theta \subseteq B$ for each $x \in B$. A variety $\mathcal{V}$ is *weakly coherent* if each $\mathcal{A} \in \mathcal{V}$ has this property.

Recall that an algebra $\mathcal{A}$ is *Hamiltonian* if every subalgebra of $\mathcal{A}$ is a class of some congruence on $\mathcal{A}$. A variety $\mathcal{V}$ is *Hamiltonian* if each $\mathcal{A} \in \mathcal{V}$ has this property.

The *weak congruence lattice* $Cw\mathcal{A}$ of an algebra $\mathcal{A}$ is the lattice of all symmetric and transitive subalgebras of $\mathcal{A}^2$, i.e. the lattice of all congruences on all subalgebras of $\mathcal{A}$. The congruence lattice $Con\mathcal{A}$ of $\mathcal{A}$ is the filter

* Supported by Grant 0401B of RFNS through Math. Inst. SANU

$[\Delta)$ in that lattice generated by the diagonal relation Δ, and $Sub\mathcal{A}$ is isomorphic with the sublattice (ideal) $(\Delta]$ of all diagonal relations. Because of that isomorphism, subalgebras are usually identified with the corresponding diagonal relations; hence, $Sub\mathcal{A}$ is a sublattice of $Cw\mathcal{A}$. In addition, the congruence lattice of every subalgebra of $\mathcal{A}$ is an interval sublattice of $Cw\mathcal{A}$.

An algebra $\mathcal{A}$ has the *Congruence Intersection Property* (the *CIP*), if for all $\rho \in Con\mathcal{B}$, $\theta \in Con\mathcal{C}$, $\mathcal{B}, \mathcal{C} \in Sub\mathcal{A}$

$$\rho_A \cap \theta_A = (\rho \cap \theta)_A,$$

where ρ_A is the least congruence on $\mathcal{A}$ whose restriction to B^2 is ρ. In the lattice $Cw\mathcal{A}$ the CIP is usually expressed in the following way: for $\rho, \theta \in Cw\mathcal{A}$,

$$(\rho \wedge \theta) \vee \Delta = (\rho \vee \Delta) \wedge (\theta \vee \Delta).$$

Recall that $\mathcal{A}$ has the *Congruence Extension Property* (the *CEP*) if for every congruence ρ on a subalgebra of $\mathcal{A}$ there is a congruence on $\mathcal{A}$ collapsing ρ.

$\mathcal{A}$ has the *Strong Congruence Extension Property* (*SCEP*) if for every $\mathcal{B} \in Sub\mathcal{A}$ and $\rho \in Con\mathcal{B}$ there is $\theta \in Con\mathcal{A}$ such that $[b]_\rho = [b]_\theta$ for every $b \in B$. A variety $\mathcal{V}$ has the SCEP if every $\mathcal{A} \in \mathcal{V}$ has this property.

A variety $\mathcal{V}$ is *weak-congruence modular* (Cw-modular) if the weak congruence lattice of every $\mathcal{A} \in \mathcal{V}$ is modular.

Let $\mathcal{E}$ be some lattice identity

$$p(x_1, ..., x_n) = q(x_1, ..., x_n).$$

We say that $\mathcal{E}$ *implies modularity* if every lattice satisfying $\mathcal{E}$ is modular. Similarly, $\mathcal{E}$ *implies distributivity* if every lattice satisfying $\mathcal{E}$ is distributive.

For more details about weak coherence, 0-regularity and SCEP, see [2,3], and for some other properties of weak congruences see [7,8,9].

2. Results

It is obvious that a variety $\mathcal{V}$ with exactly one nullary operation 0 in its similarity type is a 0_1-variety if and only if the identity

$$f(0, ..., 0) = 0 \tag{1}$$

holds in $\mathcal{V}$ for every n-ary operational symbol f. Another characterization is the following.

Lemma 1. *A variety $\mathcal{V}$ with a single nullary operation 0 is a 0_1-variety if and only if there exists at most unary term g such that for every n-ary term f the identity*

$$f(g(x), \ldots, g(x)) = g(x) \tag{2}$$

holds in $\mathcal{V}$.

Proof. Let $\mathcal{V}$ be a 0_1-variety. Then (1) holds, and the term $g(x) \equiv 0$ satisfies the requirement (2).

On the other hand, if (2) holds, then by Theorem 9 in [1], every congruence on an algebra $\mathcal{A}$ in $\mathcal{V}$ has a class which is a subalgebra of $\mathcal{A}$, and since there is a constant in $\mathcal{A}$, this class is unique. The diagonal relation then provides a one element subalgebra. □

Corollary 1. *A variety $\mathcal{V}$ with 0 is a 0_1-variety if and only if every congruence on an algebra $\mathcal{A}$ in $\mathcal{V}$ has exactly one class which is a subaglebra of $\mathcal{A}$.*

Lemma 2. *(Theorem 1 in [4]) If the lattice of subalgebras of every free algebra in a variety $\mathcal{V}$ is modular, then $\mathcal{V}$ is Hamiltonian.*

Theorem 1. *Let $\mathcal{V}$ be a 0_1-regular variety. If for each $\mathcal{A} \in \mathcal{V}$ the lattice $Sub\mathcal{A}$ satisfies an identity $\mathcal{E}$ which implies modularity, then $Cw\mathcal{A}$ satisfies $\mathcal{E}$.*

Proof. If $Sub\mathcal{A}$ satisfies $\mathcal{E}$ for each $\mathcal{A} \in \mathcal{V}$, then also $Sub\mathcal{F}$ is modular for every free algebra $\mathcal{F}$ of $\mathcal{V}$. By Lemma 2, $\mathcal{V}$ is Hamiltonian. Since $\mathcal{V}$ is 0_1-regular and Hamiltonian, then also each $\mathcal{A} \in \mathcal{V}$ has this property, and by (ii), Theorem 18 in [7], $Con\mathcal{A} \cong Sub\mathcal{A}$. Hence, both $Con\mathcal{A}$ and $Sub\mathcal{A}$ satisfy $\mathcal{E}$. By (iii), Theorem 18 in [7], $\mathcal{A}$ has the CEP and the CIP. Using Theorem 3 in [8], we conclude that also $Cw\mathcal{A}$ satisfies $\mathcal{E}$ for every $\mathcal{A} \in \mathcal{V}$. □

The proof of the preceding theorem shows that the same argument (the use of Theorem 18 in [7] and Theorem 3 in [8]) can be used for a single algebra $\mathcal{A}$ in any variety satisfying $\mathcal{E}$, provided that $\mathcal{A}$ is A_m-regular (in which case it has a unique minimum one element subalgebra, but there should be no constants in the similarity type of $\mathcal{V}$). $\mathcal{E}$ may also imply distributivity.

Theorem 2. *If for each $\mathcal{A}$ in a variety $\mathcal{V}$ the lattice $Sub\mathcal{A}$ satisfies an identity $\mathcal{E}$ which implies modularity (distributivity), then the weak congruence lattice of every A_m-regular algebra in $\mathcal{V}$ also satisfies $\mathcal{E}$.*

Proof. If $\mathcal{A}$ is an A_m-regular algebra in $\mathcal{V}$, then it is Hamiltonian (by Lemma 2), has the CEP and the CIP, and thus, as above, $Sub\mathcal{A} \cong Con\mathcal{A}$. By Theorem 3 in [8]. $\mathcal{E}$ holds in $Cw\mathcal{A}$, since it is satisfied on both, $Con\mathcal{A}$ and $Sub\mathcal{A}$. □

By the definition given in Introduction, an algebra $\mathcal{A}$ is weakly coherent whenever every its subalgebra is a union of congruence classes provided it contains a congruence class containing 0. This will be used in the following theorem.

Theorem 3. *Let $\mathcal{V}$ be a Hamiltonian 0_1-variety which is weakly coherent. Then $Cw\mathcal{A}$ is Arguesian (and hence modular) for every $\mathcal{A}$ in $\mathcal{V}$.*

Proof. If $\mathcal{V}$ is weakly coherent then, by Corollary 1 in [2] and by the assumption, $\mathcal{V}$ is 0_1-regular. Since $\mathcal{V}$ is also Hamiltonian, $Sub\mathcal{A} \cong Con\mathcal{A}$ for each $\mathcal{A} \in \mathcal{V}$ directly by Theorem 18 in [7]. By Corollary 2 in [2], $\mathcal{V}$ has permutable congruences and thus, by the famous Jónsson result, $Con\mathcal{A}$ is Arguesian (end hence modular) for each $\mathcal{A} \in \mathcal{V}$. So also $Sub\mathcal{A}$ is Arguesian. By (iii), Theorem 18 in [7], $\mathcal{A}$ satisfies the CIP and the CEP. Using again Theorem 3 in [8], we get that $Cw\mathcal{A}$ is Arguesian for each $\mathcal{A} \in \mathcal{V}$. □

Thus we have obtained some sufficient conditions under which a 0_1-variety is Cw-modular. In the following, we give also the necessary conditions, and we characterize algebras in these varieties with distributive lattices of weak congruences.

First we advance some known results.

Lemma 3. *(Theorem 2.9 in [9]) For an algebra $\mathcal{A}$, $Cw\mathcal{A}$ is a modular lattice if and only if $Con\mathcal{A}$ and $Sub\mathcal{A}$ are modular and $\mathcal{A}$ has the CEP and the CIP.*

Lemma 4. *([3])*

a) *A variety $\mathcal{V}$ has the SCEP if and only if it is Hamiltonian.*
b) *An algebra $\mathcal{A}$ has the SCEP if and only if $\mathcal{A}$ is Hamiltonian and has the CEP.*

If $\mathcal{A}$ is a Hamiltonian algebra with a one element subalgebra, and each subalgebra of $\mathcal{A}$ is 0-regular, then by Theorem 18 in [7] $\mathcal{A}$ has both, CEP and CIP. This result will be used in a characterization of $0_1 - Cw$-modular varieties. For a single algebra, similar problems are solved in the following.

Proposition 1. *Let $\mathcal{A}$ be a Hamiltonian 0-regular algebra which has the CEP. Then,*

a) *$\mathcal{A}$ has the CIP; and*
b) *$\mathcal{A}$ is weakly coherent.*

Proof. a) Let $\rho,\theta \in Cw\mathcal{A}$, $\rho \in Con\mathcal{B}$, $\theta \in Con\mathcal{C}$, $\mathcal{B},\mathcal{C} \in Sub\mathcal{A}$. Then by the CEP, in the lattice $Cw\mathcal{A}$

$$((\rho \wedge \theta) \vee \Delta) \wedge (B \wedge C)^2 = \rho \wedge \theta \text{ and}$$

$$((\rho \vee \Delta) \wedge (\theta \vee \Delta)) \wedge (B \wedge C)^2 = (\rho \vee \Delta) \wedge B^2 \wedge (\theta \vee \Delta) \wedge C^2 = \rho \wedge \theta.$$

By Lemma 4 b), $\mathcal{A}$ has the SCEP, since it is Hamiltonian and has the CEP. Hence, both $(\rho \wedge \theta) \vee \Delta$ and $(\rho \vee \Delta) \wedge (\theta \vee \Delta)$ have the same blocks as $\rho \wedge \theta$. By 0-regularity then

$$(\rho \wedge \theta) \vee \Delta = (\rho \vee \Delta) \wedge (\theta \vee \Delta),$$

and the CIP holds.

b) Let $\mathcal{B} \in Sub\mathcal{A}$, $\theta \in Con\mathcal{A}$, and B contains $[0]_\theta$. Now, $(B^2 \wedge \theta) \vee \Delta$ and θ have the same blok $[0]_\theta$, since $\mathcal{A}$ has the SCEP. By 0-regularity then $(B^2 \wedge \theta) \vee \Delta = \theta$. Again by the SCEP $[0]_\theta \subseteq B$ for each $x \in B$, and $\mathcal{A}$ is weakly coherent. □

Now we can give a characterization theorem for 0_1-Cw-modular varieties.

Theorem 4. *The following are equivalent for a 0_1-variety $\mathcal{V}$:*

(i) *$\mathcal{V}$ is weak congruence modular;*

(ii) *$\mathcal{V}$ is subalgebra modular and 0-regular;*

(iii) *$\mathcal{V}$ is Hamiltonian and 0-regular;*

(iv) *$\mathcal{V}$ is Hamiltonian and weakly coherent;*

(v) *$\mathcal{V}$ is polynomially equivalent to the variety of modules over a ring with unit.*

Proof. (i)$\Longrightarrow$(v). If $\mathcal{A}$ belongs to a Cw-modular variety, then $Sub\mathcal{A}^2$ is a modular lattice and thus (see [4]) $\mathcal{A}^2$ is Hamiltonian. $Con\mathcal{A}$ is also modular and hence, by (vi), Theorem 5.5 in [5], $\mathcal{A}$ is polynomially equivalent to a module over a ring with unit.

(v)$\Longrightarrow$(i). If (v) holds and $\mathcal{A} \in \mathcal{V}$, then $\mathcal{A}$ is 0-regular and $Con\mathcal{A}$ is a modular lattice. By Theorem 1 in [4], $\mathcal{A}^2$ is Hamiltonian, and by the result of E. Kiss ([6]), $\mathcal{A}$ has the CEP. By the similar argument $\mathcal{A}^2$ also has the CEP, and by a), Proposition 1, $\mathcal{A}^2$ has the CIP. In the presence of the CEP, the CIP is hereditary for subalgebras (Corollary 3 in [8]). Hence, $\mathcal{A}$ has the CIP as well (since $\mathcal{A}$ is, up to the isomorphism, a subalgebra of $\mathcal{A}^2$). By Theorem 18 in [7], $Sub\mathcal{A} \cong Con\mathcal{A}$ and $Sub\mathcal{A}$ is also modular. By Lemma 3, $Cw\mathcal{A}$ is a modular lattice.

(ii)$\Longrightarrow$(i). By Theorem 1.

(iv)$\Longrightarrow$(i). By Theorem 3.

(iii)$\Longrightarrow$(iv). By b), Proposition 1.

(v)$\Longrightarrow$(ii). Similarly to (v)$\Longrightarrow$(i), since $\mathcal{A}$ is 0-regular, $Con\mathcal{A}$ is modular, and $Sub\mathcal{A} \cong Con\mathcal{A}$. Hence, $Sub\mathcal{A}$ is modular.

(ii)$\Longrightarrow$(iii). By Lemma 2. □

A consequence of Theorem 17 in [7] is that in the above characterized weak congruence modular varieties, congruence and subalgebra lattices of every algebra in the variety are isomorphic, and the algebra has the CEP and the CIP. Under these conditions, by already used Theorem 3 in [8], lattice identities satisfied on $Con\mathcal{A}$ and $Sub\mathcal{A}$, also hold on $Cw\mathcal{A}$. By these arguments, it is possible to discuss the weak congruence distributivity for algebras in Cw-modular varieties.

Theorem 5. *Let $\mathcal{V}$ be a 0_1-Cw-modular variety. Then, the following are equivalent for an algebra $\mathcal{A} \in \mathcal{V}$:*

(i) *$Cw\mathcal{A}$ is distributive;*
(ii) *$Sub\mathcal{A}$ is distributive;*
(iii) *$Con\mathcal{A}$ is distributive.*

Proof. If $Cw\mathcal{A}$ is distributive, then obviously (ii) and (iii) hold. On the other hand, if $Sub\mathcal{A}$ or $Con\mathcal{A}$ are distributive, then by the above argument, since the variety is weak congruence modular ($Sub\mathcal{A} \cong Con\mathcal{A}$, $\mathcal{A}$ has the CIP and the CEP), the distributivity is transfered to the weak congruence lattice. □

References

[1] B. Csákány, *Subalgebras and Congruences*, Annales Univ. Sci. Budapest, Sectio Math. **18** (1975), 37–44.

[2] I. Chajda, *Weak Coherence of Congruences*, Czech. Math. J. **40** (1991), 149–154.

[3] I. Chajda, *Some Modifications of the Congruence Extension Property*, Math. Slovaca (to appear).

[4] T. Evans and B. Ganter, *Varieties with Modular Subalgebra Lattices*, Bull. Austral. Math. Soc. **28** (1983), 247–254.

[5] H. P. Gumm, *Geometrical Methods in Congruence Modular Algebras*, Memoirs of AMS, No. 286.

[6] E. W. Kiss, *Each Hamiltonian Variety has the Congruence Extension Property*, Alg. Univ. **12** (1981), 395–398.

[7] B. Šešelja and A. Tepavčević, *Infinitely Distributive Elements in the Lattice of Weak Congruences*, General Algebra 1988, Elsevier Sci. (North. Holland), 1990, 241–253.

[8] B. Šešelja and A. Tepavčević, *Special Elements of the Lattice and Lattice Identities*, Rev of Research, Fac. of Sci. Univ. of Novi Sad, Math. Ser. **20** (1990), 21–29.

[9] G. Vojvodić and B. Šešelja, *On the Lattice of Weak Congruence Relations*, Algebra Universalis **25** (1989), 221–230.

Ivan Chajda, Department of Algebra and Geometry, Fac. Sci., Palacký University Olomouc, Tomkova 40, 77900 Olomouc, Czech Republic

Branimir Šešelja and Andreja Tepavčević, Institute of Mathematics, Fac. of Sci., University of Novi Sad, Trg D. Obradovića 4, 21000 Novi Sad, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 639-646

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A REMARK ON CONVOLUTION POLYNOMIALS

L. N. Đorđević, Đ. R. Đorđević and Z. A. Ilić

ABSTRACT. A family of polynomials $\{P_i(x), i = 0, 1, \dots, m\}$ $(m \in \mathcal{N}_0)$ of degree **i** is a convolution one if satisfies the functional equation

$$\sum_{i=0}^{m} P_{m-i}(x)P_i(y) = P_m(x+y), \tag{1}$$

for every $x, y \in \mathcal{R}$. The generalization of (1) is the functional equation

$$\sum_{i=0}^{m} P_{m-i,i}(a,p,q;x,y) = P_m(a,p+q;x+y), \tag{2}$$

where $P_{j,k}(a,p,q;x,y)$ is polynomial of degree $j+k=m$ in two variables, x and y, and a, p, q are real parameters. The n-dimensional generalization of (1) is

$$\begin{aligned}\sum_{m_1+\cdots+m_n=m} & P_{m_1,\dots,m_n}(a,p_1,\dots,p_n;x_1,\dots,x_n) \\ & = P_m(a,p_1+\cdots+p_n;x_1+\cdots+x_n)\,.\end{aligned} \tag{3}$$

1. Introduction

In the paper *Convolution Polynomials* [1] D. E. Knuth systematized the known identities regarding convolution polynomials.

The identities involving not only convolution of variables but also convolution of parameters are presented in this paper.

A family of polynomials $\{P_i(x), i = 0, 1, \dots, m\}$ $(m \in \mathcal{N}_0)$ of degree **i** is a convolution one if satisfies the functional equation (convolution condition)

$$\sum_{i=0}^{m} P_{m-i}(x)P_i(y) = P_m(x+y), \tag{1}$$

for every $x, y \in \mathcal{R}$.

The polynomials $P_m(x) = \dfrac{x^m}{m!}$ and $P_m(x) = \dfrac{(x)_m}{m!}$, which are necessary for further work, satisfy the convolution condition (1).

The generalization of (1) is a functional equation

$$\sum_{i=0}^{m} P_{m-i,i}(a,p,q,x,y) = P_m(a,p+q,x+y), \tag{2}$$

where $P_m(a,p+q,x+y)$ is the polynomial in one variable of degree m and $P_{m-i,i}$ (a,p,q,x,y) is the polynomial in two variables, x and y, of degree m, and a,p,q are real parameters. The polynomial $P_{m-i,i}(a,p,q,x,y)$ can not be factorized to two polynomials in x and y, respectively.

The equality (2) is satisfied by two families:

$$\text{(3)}\quad \left\{P_m(a,p+q,x+y) = \frac{1}{m!}G_m(a;p+q;x+y)\right\},$$
$$\left\{P_{m-i,i}(a,p,q,x,y) = \frac{1}{(m-i)!i!}G_{m-i,i}(a;p,q;x,y),\right.$$
$$\left. m=0,1,\dots,\ i=0,1,\dots,m\right\}.$$

The monic Gauss hypergeometric polynomial $G_m(a;p+q;x+y)$ in variable $(x+y)$ is defined by Gauss hypergeometric function F in the following way:

$$\text{(4a)}\quad G_m(a;p+q;x+y) := (-1)^m\frac{(p+q)_m}{(a)_m}F(a,-m;p+q;x+y)$$
$$= (-1)^m\frac{(p+q)_m}{(a)_m}\sum_{j=0}^{m}\frac{(-m)_j(a)_j}{(p+q)_j}\frac{(x+y)^j}{j!},$$

i.e.,

$$\text{(4}b\text{)}\quad G_m(a;p+q;x+y) := \sum_{j=0}^{m}(-1)^{m+j}\binom{m}{j}\frac{(p+q+j)_{m-j}}{(a+j)_{m-j}}(x+y)^j.$$

The polynomials $G_{m-i,i}(a;p,q;x,y)$ are monic Appell's hypergeometric polynomials in two variables defined by the Appell hypergeometric function F_2

$$\text{(5a)}\quad G_{m-i}(a;p,q;x,y) := (-1)^m\frac{(p)_{m-i}(q)_i}{(a)_m}F_2(a;-m+i,-i;p,q;x,y)$$
$$= (-1)^m\frac{(p)_{m-i}(q)_i}{(a)_m}\sum_{j=0}^{m-i}\sum_{k=0}^{i}\frac{(a)_{j+k}(-m+i)_j(-i)_k}{(p)_j(q)_k}\frac{x^jy^k}{j!k!},$$

i.e.,

$$\text{(5b)}\quad G_{m-i,i}(a;p,q;x,y) := \sum_{j=0}^{m-i}\sum_{k=0}^{i}(-1)^{m+j+k}\binom{m-i}{j}\binom{i}{k}\frac{(p+j)_{m-i-j}(q+k)_{i-k}}{(a+j+k)_{m-j-k}}x^j y^k.$$

The following recurrence relations for monic hypergeometric polynomials in one and two variables, necessary for proving the Theorem 1, are given. Based on Gauss relations [2, pp.3,8]

$$\text{(6)}\quad \begin{aligned} &\frac{\beta}{\gamma}tF(\alpha+1,\beta+1;\gamma+1;t) = F(\alpha+1,\beta;\gamma;t) - F(\alpha,\beta;\gamma;t),\\ &\alpha F(\alpha+1,\beta;\gamma;t) - \beta F(\alpha,\beta+1;\gamma;t) = (\alpha-\beta)F(\alpha,\beta;\gamma;t), \end{aligned}$$

and putting $\alpha = a$, $\beta = -m-1$, $\gamma = s$, $F(a,-m;s;t) = (-1)^m(a)_m/(s)_m\, G_m(a;s;t)$, by elimination of $F(a+1,-m-1;s;t)$, one yields the recurrence relation

$$\text{(7)}\quad G_{m+1}(a;s;t) = tG_m(a+1;s+1;t) - \frac{s+m}{a+m}G_m(a;s;t).$$

Starting from the relations for Appell hypergeometric function F_2 [2, pp.20–21]

$$\text{(8)}\quad \begin{aligned} &\alpha F_2(\alpha+1,\beta,\beta';\gamma,\gamma';x,y) - \beta F_2(\alpha,\beta+1,\beta';\gamma,\gamma';x,y)\\ &\quad - \beta' F_2(\alpha,\beta,\beta'+1;\gamma,\gamma';x,y) = (\alpha-\beta-\beta')F_2(\alpha,\beta,\beta';\gamma,\gamma';x,y),\\ &\qquad\qquad \frac{\beta x}{\gamma}F_2(\alpha+1,\beta+1,\beta';\gamma+1,\gamma';x,y)\\ &\qquad\qquad\quad + \frac{\beta' y}{\gamma'}F_2(\alpha+1,\beta,\beta'+1;\gamma,\gamma'+1;x,y)\\ &\qquad\qquad\quad = F_2(\alpha+1,\beta,\beta';\gamma,\gamma';x,y) - F_2(\alpha,\beta,\beta';\gamma,\gamma';x,y), \end{aligned}$$

by eliminating $F_2(\alpha+1,\beta,\beta';\gamma,\gamma';x,y)$, and by setting $\alpha = a$, $\beta = -m-1+i$, $\beta' = -i$, $\gamma = p$, $\gamma' = q$.

$$G_{m-i,i}(a;p,q;x,y) := (-1)^m\frac{(p)_{m-i}(q)_i}{(a)_m}F_2(a,-m+i,-i;p,q;x,y),$$

the recurrence relation for monic Appell hypergeometric polynomials $G_{m+1-i,i}(a;p,q;x,y)$ of degree $m+1$ as a sum of degrees of variables x and y

$$G_{m+1-i,i}(a;p,q;x,y)=\frac{m+1-i}{m+1}xG_{m-i,i}(a+1;p+1,q;x,y)\\ +\frac{iy}{m+1}G_{m+1-i,i-1}(a+1;p,q+1;x,y)\\ -\frac{(m+1-i)(p+m-i)}{(m+1)(a+m)}G_{m-i,i}(a;p,q;x,y)\\ -\frac{i(q+i-1)}{(m+1)(a+m)}G_{m+1-i,i-1}(a;p,q;x,y). \tag{9}$$

Now, by replacing (3) in (2), one obtains the following

Theorem 1. *Let a,p,q be real numbers $(a,p,q>0)$ and $m\in N_0$. Then*

$$\sum_{i=0}^{m}\binom{m}{i}G_{m-i,i}(a;p,q;x,y)=G_m(a;p+q;x+y). \tag{10}$$

Proof. We implement again the mathematical induction. Using (4b) and (5b) one can prove (10) by simple testing for $m=1,2,3$. Suppose (10) holds for $k=m$, Then we prove that (10) holds for $k=m+1$, i.e.,

$$\sum_{i=0}^{m+1}\binom{m+1}{i}G_{m+1-i,i}(a;p,q;x,y)=G_{m+1}(a;p+q;x+y). \tag{11}$$

Starting from the recurrence relation (9), the left-hand size of (11) becomes

$$\sum_{i=0}^{m+1}\tbinom{m+1}{i}\Big[\tfrac{m+1-i}{m+1}xG_{m-i,i}(a+1;p+1,q;x,y)\\ +\tfrac{i}{m+1}yG_{m+1-i,i-1}(a+1;p,q+1;x,y)-\tfrac{(m+1-i)(p+m-i)}{(m+1)(a+m)}\times\\ \times G_{m-i,i}(a;p,q;x,y)-\tfrac{i(q+i-1)}{(m+1)(a+m)}G_{m+1-i,i-1}(a;p,q;x,y)\Big] \tag{12}$$

$$=\sum_{i=0}^{m}\tbinom{m}{i}\Big[xG_{m-i,i}(a+1;p+1,q;x,y)-\tfrac{(p+m-i)}{(a+m)}G_{m-i,i}(a;p,q;x,y)\Big]\\ +\sum_{i=1}^{m+1}\tbinom{m}{i-1}\Big[yG_{m+1-i,i-1}(a+1;p,q+1;x,y)\\ -\tfrac{(q+i-1)}{(a+m)}G_{m+1-i,i-1}(a;p,q;x,y)\Big]$$

$$= \sum_{i=0}^{m} \binom{m}{i} \Big[xG_{m-i,i}(a+1;p+1,q;x,y)$$

$$+ yG_{m-i,i}(a+1;p,q+1;x,y) - \tfrac{p+q+m}{a+m} G_{m-i,i}(a;p,q;x,y)\Big]$$

$$= xG_m(a+1;p+q+1;x+y)$$

$$+ yG_m(a+1;p+q+1;x+y) - \tfrac{p+q+m}{a+m} G_m(a;p+q;x+y)$$

$$= (x+y)G_m(a+1;p+q+1;x+y) - \tfrac{p+q+m}{a+m} G_m(a;p+q;x+y)\,.$$

According to (7), the expression (12) is $G_{m+1}(a;p+q;x+y)$, i.e. the right-hand size of (11).

Remark 1. The equality (11) leads to the relation between the orthogonal polynomilas in two and one variable. At first we have the relations between orthogonal and hypergeometric polynomilas.

The monic Jacobi polynomial $\mathcal{P}_m^{(a,b-1)}(t)$ on $[0,1]$ and weight function $t^{b-1}(1-t)^a$ and monic Gauss polynomial $G_m(a+b+m,b,t)$ have a relation

$$\mathcal{P}_m^{(a,b-1)}(t) := \sum_{j=0}^{m}(-1)^{m+j}\binom{m}{j}\frac{(b+j)_{m-j}}{(a+b+m+j)_{m-j}}t^j = G_m(a+b+m;b;t).$$

The monic Appell hypergeometric polynomials $G_{m-i,i}(a+p+q+m;p,q;x,y)$ and basic Appell orthogonal polynomials $E_{m-i,i}(a,p,q;x,y)$ on the triangle $T_2 := \{(x,y) \mid x \geq 0,\ y \geq 0,\ x+y \leq 1\}$ and weight $x^{p-1}y^{q-1}(1-x-y)^a$ are connected by the relation

$$E_{m-i,i}(a,p,q;x,y) = G_{m-i,i}(a+p+q+m;p,q;x,y).$$

The basic orthogonal polynomials $V_{m-i,i}(a,p,q;x,y)$ on the circle $C_2 := \{(x,y) \mid x^2+y^2 \leq 1\}$ and weight function $|x|^p|y|^q(1-x^2-y^2)^a$ and basic Appell orthogonal polynomials $E_{m-i,i}(a,p,q;x;y)$ are connected by the following four equalities:

$$V_{2m-2i,2i}(a,p,q;x,y) = E_{m-i,i}(a,(p+1)/2,(q+1)/2;x^2,y^2)\,,$$
$$V_{2m-2i+1,2i}(a,p,q;x,y) = xE_{m-i,i}(a,(p+3)/2,(q+1)/2;x^2,y^2)\,,$$
$$V_{2m-2i,2i+1}(a,p,q;x,y) = yE_{m-i,i}(a,(p+1)/2,(q+3)/2;x^2,y^2)\,,$$
$$V_{2m-2i+1,2i+1}(a,p,q;x,y) = xyE_{m-i,i}(a,(p+3)/2,(q+3)/2;x^2,y^2)\,.$$

The corollary of Theorem 1 are the following equalities which connect the orthogonal polinomials in two and one variable:

$$\sum_{i=0}^{m}\binom{m}{i}E_{m-i,i}(a,p,q;x,y)=\mathcal{P}_m^{(a,p+q-1)}(x+y)\,,$$

$$\sum_{i=0}^{m}\binom{m}{i}V_{2m-2i,2i}(a,p,q;x,y)=\mathcal{P}_m^{(a,(p+q)/2)}(x^2+y^2)\,,$$

$$\sum_{i=0}^{m}\binom{m}{i}V_{2m-2i+1,2i}(a,p,q;x,y)=x\mathcal{P}_m^{(a,(p+q)/2+1)}(x^2+y^2)\,,$$

$$\sum_{i=0}^{m}\binom{m}{i}V_{2m-2i,2i+1}(a,p,q;x,y)=y\mathcal{P}_m^{(a,(p+q)/2+1)}(x^2+y^2)\,,$$

$$\sum_{i=0}^{m}\binom{m}{i}V_{2m-2i+1,2i+1}(a,p,q;x,y)=xy\mathcal{P}_m^{(a,(p+q)/2+2)}(x^2+y^2)\,.$$

For a family of polynomials $\{P_m(r,x)\}$, where r is real parameter and x is real variable, the particular case of (2) is a functional equation

$$\sum_{i=0}^{m}P_{m-i}(p,x)P_i(q,y)=P_m(p+q,x+y). \tag{13}$$

The polynomial family $\{P_m(r,x)=\dfrac{1}{m!}G_m(r;x)\}$, where

$$G_m(r;x):=(-1)^m(r)_mF(-m;r;t)=(-1)^m(r)_m\sum_{j=0}^{m}\frac{(-m)_j}{(r)_j}\frac{x^j}{j!}, \tag{14}$$

i.e.,

$$G_m(r;x):=\sum_{j=0}^{m}(-1)^{m+j}\binom{m}{j}(r+j)_{m-j}x^j,$$

satisfies (13).

The monic confluent hypergeometric polynomial in two variables, x,y, of degree m (as a sum of degees in variables x and y) is denoted as

$$G_{m-i,i}(p,q;x,y):=G_{m-i}(p;x)\cdot G_i(q;y), \tag{15}$$
$$(m=0,1,\dots,i=0,1,\dots,m).$$

Replacing the polynomials defined in (14) into the (13), we obtain the following

Theorem 2. *Let p, q be real numbers $(p, q > 0)$ and $m \in N_0$. Then holds*

$$\sum_{i=0}^{m} \binom{m}{i} G_{m-i,i}(p,q;x,y) = G_m(p+q;x+y). \tag{16}$$

Proof. Using the recurrence relation

$$G_{m+1}(s;t) = tG_m(s+2;t) - sG_m(s+1;t) \tag{17}$$

proof is identical of proof of Theorem 1.

Remark 2. The relation among the generalized Laguerre polynomial $L_l^{s-1}(t)$ on $[0,+\infty)$ and weight $t^{s-1}e^{-t}$, generalized Hermite polynomial $H_m(s,t)$ on $(-\infty,+\infty)$ with weight $|t|^p e^{-t^2}$, and confluent hypergeometric polynomial $G_l(s;t)$ is

$$H_m(s,t) = (-1)^m 2^m t^\delta L_l^{(s+2\delta+1)/2-1}(t^2) = 2^m t^\delta G_l((s+2\delta+1)/2;t^2), \tag{18}$$

where $l = [m/2], \delta = m - 2l$. The direct corollary of equality (16) and (18) are the following equalities, for the cases when degrees of both variables are even, then when degree of one variable is odd, and when degrees of both variables are odd, respectively:

$$\sum_{i=0}^{m} \binom{m}{i} H_{2m-2i}(p,x)H_{2i}(q,y) = (-1)^m 2^{2m} L_m^{(p+q)/2}(x^2+y^2), \tag{19}$$

$$\sum_{i=0}^{m} \binom{m}{i} H_{2m+1-2i}(p,x)H_{2i}(q,y) = (-1)^m 2^{2m+1} x L_m^{(p+q)/2+1}(x^2+y^2),$$

$$\sum_{i=0}^{m} \binom{m}{i} H_{2m-2i}(p,x)H_{2i+1}(q,y) = (-1)^m 2^{2m+1} y L_m^{(p+q)/2+1}(x^2+y^2),$$

$$\sum_{i=0}^{m} \binom{m}{i} H_{2m-2i+1}(p,x)H_{2i+1}(q,y) = (-1)^m 2^{2m+2} xy L_m^{(p+q)/2+2}(x^2+y^2).$$

The polynomial generalization

Like

$$\sum_{m_1+\cdots+m_n=m} P_{m_1}(x_1)\cdots P_{m_n}(x_n) = P_m(x_1+\cdots+x_n)$$

generalizes (1), one can define the n-dimensional ($n \geq 2$) generalizations of (2) and (4),

$$(20) \qquad \sum_{m_1+\cdots+m_n=m} P_{m_1,\ldots,m_n}(a, p_1, \ldots, p_n, x_1, \ldots, x_n) = P_m(a, p_1 + \cdots + p_n, x_1 + \cdots + x_n),$$

and

$$(21) \qquad \sum_{m_1+\cdots+m_n=m} P_{m_1}(p_1, x_1) \cdots P_{m_n}(p_n, x_n) = P_m(p_1+\cdots+p_n, x_1+\cdots+x_n),$$

respectively.

References

[1] D. E. Knuth, *Convolution Polynomials*, Mathematica Journal, **2** (1992,), no. 4, 67–78.

[2] P. Appell and J. Kampé de Fériet, *Fonctions Hypergéométriques et Hypersphériques*, Gauthier-Villars et Cie, Paris, 1926.

[3] L. Charlitz, *A combinatorial identity of L.N. Djordjević*, Univ. Beograd. Publ. Elektroteh. Fak. Ser. Mat. Fiz. **557–598** (1977), 86–88.

[4] L.N. Djordjević, *Proof of combinatorial identity using multiple integration*, Univ. Beograd. Publ. Elektroteh. Fak. Ser. Mat. Fiz. **557–598** (1977), 63–64.

[5] L.N. Djordjević, DJ. R. Djordjević, Z.Ilić, *Some symmetric cases of multivariate orthogonal polynomials – symbolic generation and reduction to one dimension* (1994), IX Workshop in Appl. Math., Budva.

Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia

Faculty of Civil Engineering, Beogradska 14, 18000 Niš, Yugoslavia

Faculty of Civil Engineering, Beogradska 14, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 647–656
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON AN EMBEDDING OF A CLASS OF SEMIGROUPS INTO RELATIONAL ALGEBRAS

Zoran D. Đorđević

ABSTRACT. In this paper we give a construction of relational algebras in which finite semigroups which are orthogonal sums of groups with zero adjoined could be embedded. It is proved that if two semigroups are isomorphic, then the corresponding relational algebras are also isomorphic.

1. Terminology, notations and basic definitions are taken from [1], [3] and [4].

Definition 1.1. [4] A relational algebra is an algebra

$$\mathcal{A} = (A, +, \bullet, ^{-}, 0, 1, \circ, 1', ^{-1})$$

of type $(2,2,1,0,0,2,0,1)$ which satisfies the following axioms:

(R1) $(A, +, \bullet, ^{-}, 0, 1)$ is a Boollean algebra;

(R2) $(A, \circ, 1')$ is a monoid;

(R3) Operation $^{-1}$ is an involution of the semigroup $(A, \circ)$, i.e. for all $x, y \in A$,

$$(x \circ y)^{-1} = y^{-1} \circ x^{-1}, \quad (x^{-1})^{-1} = x;$$

(R4) For all $x, y \in A$, $(x+y)^{-1} = x^{-1}+y^{-1}$, $x \circ (y+z) = (x \circ y)+(x \circ z)$;

(R5) For all $x, y \in A$, $(x^{-1} \circ (\overline{x \circ y})) \circ y = 0$.

We denote the class of relational algebras by RA. The Boolean part of relational algebra $\mathcal{A}$ we call the *Boolean reduct* of $\mathcal{A}$, and we denote it by $Rd_B(\mathcal{A})$,. The semigroup part will be denoted by $Rd_S(\mathcal{A})$. Therefore $Rd_B(\mathcal{A}) = (A, +, \bullet, ^{-}, 0, 1)$ and $Rd_S(\mathcal{A}) = (A, \circ)$. The set of all atoms of a relational algebra we will denote by $At(\mathcal{A})$. For a relational algebra $\mathcal{A}$ we say that it is atomic (complete), if the corresponding Boolean reduct is an atomic (complete) Boolean algebra.

Definition 1.2. [4] For a relational algebra $\mathcal{A}$ we say that it is:

1) *commutative*, if $x \circ y = y \circ x$, for all $x, y \in A$;
2) *symmetric*, if $x^{-1} = x$, for all $x \in A$;
3) *Boolean*, if $x \circ y = x \bullet y$, for all $x, y \in A$;
4) *integral*, if $x \circ y \neq 0$, for all $x, y \in A - \{0\}$.

An example of an relational algebra on the set $\mathcal{P}(A^2)$ of all binary relations of a set A is $(\mathcal{P}(A^2), \cup, \cap, ^-, \varnothing, A^2, \circ, \triangle_A, ^{-1})$. This relational algebra is called a *full relational algebra*.

2. Let $(G_\alpha, *_\alpha)$, where $\alpha \in I \neq \varnothing$, be groups for which $G_\alpha \cap G_\beta = \varnothing$, if $\alpha \neq \beta$. Define an operation $\bullet$ on the set:

$$(1) \qquad S = \bigcup_{\alpha \in I} G_\alpha^0, \qquad \text{where } G_\alpha^0 = G_\alpha \cup \{0\} \text{ and } 0 \notin \bigcup_{\alpha \in I} G_\alpha$$

in the following way:

$$(2) \qquad a \bullet b = \begin{cases} a *_\alpha b, & \text{if } a, b \in G_\alpha, \text{ for some } \alpha \in I \\ 0, & \text{otherwise.} \end{cases}$$

Then $(S, \bullet)$ is a semigroup. Following the terminology from [1] and [2] we will say that S is an orthogonal sum of semigroups G_α^0, i.e. an orthogonal sum of groups with zero adjoined. The class of all such semigroups will be denoted by KG^0, and by KG_n^0 it will be denoted the subclass of KG^0 consisiting of finite semigroups with exactly $n+1$ elements.

Since every group G is isomorphic to a permutation group of a set (G'), for groups $(G_\alpha, *_\alpha)$ from $S = \bigcup_{\alpha \in I} G_\alpha^0$ we have that for every $\alpha \in I$, $G_\alpha \cong G'_\alpha$, where an isomorphism is given by

$$a \mapsto f_a = \varrho_a = \{(x, a *_\alpha x) | x \in G_\alpha\}.$$

Therefore, to an element a from G_α $(\alpha \in I)$ of a semigroup S, there corresponds a function, i.e. the relation ρ_a in the group G'_α $(\alpha \in I)$. The image of 0 is $\varnothing$, i.e. $\rho_0 = \varnothing$. Now in

$$(3) \qquad S' = \bigcup_{\alpha \in I} (G_\alpha \cup \{\varnothing\})$$

for the operation $\circ$ (the composition of relations) we have that

$$(4) \qquad \rho_\alpha \circ \rho_\beta = \begin{cases} \rho_{a *_\alpha b}, & \text{if } a, b \in G_\alpha, \text{ for some } \alpha \in I, \\ \varnothing, & \text{otherwise.} \end{cases}$$

Hence, $(S', \circ)$ is a semigroup, belonging to the class KG^0.

Lemma 2.1. *Let S and S' be semigroups from the class KG^0, given by* (1) *and* (3). *Then they are isomorphic.*

In the sequel, we will consider finite semigroups from the class KG^0_n. Let $S \in KG^0_n$, i.e. let $S = \bigcup_{\alpha\in I} G^0_\alpha = \{0, a_1, a_2, a_3, \ldots, a_n\}$ be a finite semigroup.

Let $|I| = k$ $(1 \le k \le n)$. From $S \cong S'$ for semigroup $S \in KG^0_n$ it follows that

$$S' = \bigcup_{\alpha\in I} (G_\alpha \cup \{\varnothing\}) = \{\varnothing, \rho_{a_1}, \rho_{a_2}, \ldots, \rho_{a_n}\}.$$

Let us consider the set

$$A_S = \left\{\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_p}} \,\middle|\, \rho_{a_{i_\nu}} \in S', a_{i_\nu} \neq 0, 1 \le \nu \le p, 1 \le p \le n\right\} \cup \{0\}.$$

Elements of the set $S' - \{\varnothing\}$ are called atoms. Therefore, elements of the set A_S are unions of atoms and the empty set.

Theorem 2.1. *Let $S = \bigcup_{\alpha\in I} G^0_\alpha$ from KG^0_n and let $|I| = k$ $(k \le n)$. Then $(A_S, \circ)$, where $\circ$ is the composition of relations, is a semigroup with unit $1'$, where*

$$1' = \rho_{e_1} \cup \rho_{e_2} \cup \cdots \cup \rho_{e_k},$$

and e_i $(1 \le i \le k)$ are the units of groups G_{a_i} $(1 \le i \le k)$.

Proof. It is clear. □

For the unary operation $^{-1}$ in the semigroup $(S', \circ)$ we take the mapping

$$\rho_a \mapsto \rho_a^{-1}$$

where ρ_a^{-1} is the inverse element of ρ_a in the group G'_α, for some $\alpha \in I$. We take that $\varnothing^{-1} = \varnothing$.

Since the operaton $^{-1}$ is the usual inversion of relations, it can be extended to the semigroup $(A_S, \circ)$ where $(x^{-1})^{-1} = x$, $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$, for all $x, y \in A_S$. Therefore, according to Theorem 2.1, we have that $(A_S, \circ, 1', ^{-1})$ is an involutive semigroup with unit $1'$, where $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$ for all $x, y \in A_S$.

Since the elements of the semigroup A_S are relations, the operations $\cup, \cap, ^-$(union, intersection and complement) are defined in A_S. The complement is related to 1, where

$$1 = \rho_{a_1} \cup \rho_{a_2} \cup \cdots \cup \rho_{a_n} \quad \text{(union of all atoms).}$$

So for $x \in A_S$ where $x = \rho_{b_1} \cup \rho_{b_2} \cup \cdots \cup \rho_{b_s}$ $(s \le n)$, will be $\bar{x} = 1 - x = \rho_{c_1} \cup \rho_{c_2} \cup \cdots \cup \rho_{c_{n-s}}$, where $\rho_{c_\nu} \cap \rho_{b_\mu} = \varnothing$ for all ν $(1 \le \nu \le n - s)$ and all μ $(1 \leq \mu \leq s)$. Especially, $\bar{\varnothing} = 1$ and $\bar{1} = \varnothing$. So we get an algebra $(A_S, \cup, \cap, ^{-}, \varnothing, 1)$ of type $(2, 2, 1, 0, 0)$.

Let us notice Boolean algebra $(\mathcal{P}(At(A_S)), \cup, \cap, ^{-}, \varnothing, At(A_S))$ and the mapping $f : \mathcal{P}(At(A_S)) \mapsto A_S$ defined by

$$f : \{\rho_{a_{i_1}}, \rho_{a_{i_2}}, \dots, \rho_{a_{i_r}}\} \mapsto \rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \dots \cup \rho_{a_{i_r}} \text{ and } f(\varnothing) = \varnothing,$$

is an isomorphism. Then

$$(A_S, \cup, \cap, ^{-}, \varnothing, 1) \tag{5}$$

is also Boolean algebra.

Both Boolean algebras have same number of atoms (n-atoms). Therefore

$$\mathcal{P}(At(A_S)) \cong A_S.$$

Using the operations of the set A_S, we obtain:

Theorem 2.2. *The algebra* $\mathcal{A} = (A_S, \cup, \cap, ^{-}, 1, \circ, 1', ^{-1})$ *of type* $(2, 2, 1, 0, 0, 2, 0, 1)$ *is a relational algebra in which the semigroup* $(S, \bullet)$ *from the class* KG_n^0 *is embedded.*

Proof. Let $S = \cup_{\alpha \in I} G_\alpha \cup \{0\} = \{0, a_1, a_2, \dots, a_n\}$ and $|I| = k$ $(1 \leq k \leq n)$. The axioms $(R1), (R2), (R3)$ from Definition 1.1 follow by Theorem 2.1, and the axiom $(R4)$ is satisfied for elements from A_S since they are relations. The only thing left to prove is the axiom $(R5)$.

Let $x = \rho_{a_i}$, $y = \rho_{b_j}$ $(1 \leq i, j \leq n)$ be arbitrary atomc A_S. There are two cases:

Case 1: Let $a_i, a_j \in G_\alpha$ for some $\alpha \in I$. Then

$$x \circ y = \rho_{a_i} \circ \rho_{a_j} = \rho_{a_i *_\alpha a_j} = \rho_{a_l}$$

where $a_l \in G_\alpha$, and hence

$$\overline{x \circ y} = 1 - \rho_{a_1} = \rho_{a_1} \cup \rho_{a_2} \cup \dots \cup \rho_{a_{l-1}} \cup \rho_{a_{l+1}} \cup \dots \cup \rho_{a_n},$$

i.e.

$$x^{-1} \circ (\overline{x \circ y}) = \rho_{a_i^{-1}} \circ (\rho_{a_1} \cup \rho_{a_2} \cup \dots \cup \rho_{a_{l-1}1} \cup \rho_{a_{l+1}} \cup \dots \cup \rho_{a_n}).$$

It follows that

$$\begin{aligned} x^{-1} \circ (\overline{x \circ y}) = {} & \rho_{a_i^{-1}} \circ \rho_{a_1} \cup \rho_{a_i^{-1}} \circ \rho_{a_2} \cup \dots \\ & \dots \cup \rho_{a_i^{-1}} \circ \rho_{a_{l-1}} \cup \rho_{a_i^{-1}} \circ \rho_{a_{l+1}} \cup \dots \cup \rho_{a_i^{-1}} \circ \rho_{a_n}. \end{aligned} \tag{6}$$

Now we have two different subcases:

(a) $(1 - \rho_{a_l}) \cap G'_\alpha = \varnothing$.
(b) $(1 - \rho_{a_l}) \cap G_\alpha \neq \varnothing$, where $G'_\alpha \cong G_\alpha$.

In the case (a), $x^{-1} \circ (\overline{x \circ y}) = \varnothing$, since all the members of the union (6) are empty sets, and hence $(x^{-1} \circ (\overline{x \circ y})) \cap y = \varnothing \cap y = \varnothing$.

In the case (b), for every $\rho_u \in (1 - \rho_{a_l} \cap G_\alpha)$, $u \neq a_1$ is satisfied. Hence

$$\rho_{a_1^{-1}} \circ \rho_u = \rho_{a_i^{-1} *_\alpha u} = \rho_{a_{p_u}} \qquad (a_{p_u} \in G_\alpha).$$

Therefore $\rho_{a_{p_u}} \neq \rho_{a_j}$ for all u, because in the opposite case from $\rho_{a_{p_u}} = \rho_{a_j}$ i.e. $\rho_{a_i^{-1} *_\alpha u} = \rho_{a_j}$ it follows that $a_i^{-1} *_\alpha u = a_j$, i.e. $a_i *_\alpha a_j = a_l$, which gives a contradiction. Therefore

$$(x^{-1} \circ (\overline{x \circ y})) \cap y = \varnothing.$$

Case 2: Let $a_i \in G_\alpha, a_j \in G_\beta$ and $\alpha \neq \beta$. Then $x \circ y = \varnothing$, and hence

$$\begin{aligned}(x^{-1} \circ (\overline{x \circ y})) \cap y &= (\rho_{a_i^{-1}} \circ 1) \cap \rho_{a_j} \\ &= (\rho_{a_i^{-1}} \circ \rho_{a_1} \cup \rho_{a_i^{-1}} \circ \rho_{a_2} \cup \cdots \cup \rho_{a_i^{-1}} \circ \rho_{a_n}) \cap \rho_{a_j}.\end{aligned}$$

By (4), the members of the union $\rho_{a_i^{-1}} \circ 1$ are for all $a_m \in G_\alpha$ $(1 \leq m \leq n)$ from G'_α, otherwise they are empty sets. Since $\rho_{a_j} \in G'_\beta$ we have that $\rho_{a_j} \neq \rho_{a_i^{-1} *_\alpha a_m}$ for every m for which $a_m \in G_\alpha$. Therefore, from (7) it follows that $(x^{-1} \circ (\overline{x \circ y})) \cap y = \varnothing$. If at least one of the sets x, y is empty, then $(x^{-1} \circ (\overline{x \circ y})) \cap y = \varnothing$ is obviously satisfied. Hence, the theorem is proved. □

Theorem 2.3. *Let $S \in KG_n^0$ where $S = \bigcup_{\alpha \in I} G_\alpha^0$. Then in corresponding relational algebra $\mathcal{A}_S$ we have that $1'_s = 1_s$ if and only if $|G_\alpha| = 1$, for all $\alpha \in I$.*

Proof. If $S = \{0, a_1, a_2, \ldots, a_n\}$, then the atoms in the relational algebra $\mathcal{A}_S$ are $\rho_{a_1}, \rho_{a_2}, \ldots, \rho_{a_n}$ (by Theorem 2.2). The unit 1_S in $\mathcal{A}_S$ is the union of all the atoms, i.e. $1_S = \rho_{a_1} \cup \rho_{a_2} \cup \cdots \cup \rho_{a_n}$. The unit $1'_S$ of the semigroup reduct of $\mathcal{A}_S$ is $1'_S = \rho_{e_1} \cup \rho_{e_2} \cup \cdots \cup \rho_{e_k}$ $(1 \leq k \leq n)$ where $|I| = k$, and e_i $(1 \leq i \leq k)$ are the units of the groups G_{α_i} $(1 \leq i \leq k)$. Let $|G_{\alpha_i}| = n_i$ $(1 \leq i \leq k)$. Suppose that $1'_S = 1_S$. Then for every i $(1 \leq i \leq k)$ we have $\rho_{e_i} \circ 1'_S = \rho_{e_i} \circ 1_S$ i.e.

$$\rho_{e_i} = \rho_i \circ (\rho_{a_1} \cup \rho_{a_2} \cup \cdots \cup \rho_{a_n}) = \bigcup_{1 \leq i \leq n} (\rho_{e_i} \circ \rho_{a_i}) = \bigcup_{1 \leq \nu \leq n_i} (\rho_{e_i} \circ \rho_{a_{j_\nu}}).$$

Since $G_{\alpha_i} = \{a_{j_1}, a_{j_2}, \ldots, a_{j_{n_1}}\}$ and we also have that $\rho_{e_i} = \rho_{a_{j_1}} \cup \rho_{a_{j_2}} \cup \cdots \cup \rho_{a_{j_{n_i}}}$, whence $n_i = 1$, since ρ_{e_i} is the atom. Therefore $|G_\alpha| = 1$, for all $\alpha \in I$.

Convesely, let $|G_\alpha| = 1$, for all $\alpha \in I$. Then $|I| = n$, i.e. $n = k$, hence $1'_S = \rho_{e_1} \cup \rho_{e_2} \cup \cdots \cup \rho_{e_n}$, i.e. $1'_S$ is the union of all the atoms. Therefore $1'_S = 1_S$. Thus, the theorem is proved. . □

Theorem 2.4. *Let $S \in KG_n^0$ where $S = \bigcup_{\alpha\in I} G_\alpha^0$. Then the corresponding relational algebra $\mathcal{A}_S$ is integral if and only if $|I| = 1$.*

Proof. Suppose that $|I| > 1$ and that $\mathcal{A}_S$ is an integral relational algebra, i.e. $x \circ y \neq 0$ for all $x, y \in A_S$ for which $x \neq 0$ and $y \neq 0$ and $|I| = k > 1$. Then there are at least two different groups G_α and G_β $(\alpha \neq \beta)$ in s,which follows from the fact that for all $|G_{\alpha_i}| = n_i$ $(1 \leq i \leq k)$ we have that $n_1 + n_2 + \cdots + n_k = n$ and the assumption $k > 1$. Now for $a \in G_\alpha$ and $b \in G_\beta, \rho_a \neq \rho_b$, hence in $\mathcal{A}_S$ we have $\rho_a \circ \rho_b = \varnothing$ which contradicts the assumption that A_S is integral.

Conversely, let $|I| = 1$. Then the semigroup S contains only one group G_α, i.e. $S = G_\alpha \cup \{0\}$, and hence, all the elements $x, y \in A_S$ are the unions of the relations from the group G'_α different from $\varnothing$. Since $x \circ y$ is the union of elements of G'_α we have that $x \circ y \neq 0$, and relational algebra A_S is integral, which was to be proved. □

Theorem 2.5. *Let $S \in KG_n^0$ where $S = \cup_{\alpha\in I} S_\alpha \cup \{0\}$. Then the corresponding relational algebra $\mathcal{A}_S$ is Boolean, if and only if $|I| = n$.*

Proof. Since $(\forall\alpha \in I)(|G_\alpha| = 1)$ is equivalent to $|I| = n$, by the property 2.19.[1] and by Theorem 2.3 we have that the assertion holds. □

3. Let a semigroup S be from the class KG_n^0, and let $\mathcal{A}_S$ be the corresponding relational algebra in which the semigroup S is embedded. Denote the elements of $S = \bigcup_{\alpha\in I} G_\alpha^0$ by $0, a_1, a_2, \ldots, a_n$ and the corresponding atoms in the relational algebra A_S by $\rho_{a_1}, \rho_{a_2}, \ldots, \rho_{a_n}$. Before we give an example of an embedding of a semigroup into relational algebra we introduce the following notations. Denote by $0, 1, 2, \ldots, n$ the elements of a semigroup S and by $i_1 i_2 \ldots i_m$ the element $\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_m}}$ from A_S. With these notations we have

$$i_1 i_2 \ldots i_m \circ j_1 j_2 \ldots j_t = (i_1 \circ j_1)(i_1 \circ j_2)\cdots(i_1 \circ j_t)(i_2 \circ j_1)(i_2 \circ j_2)\cdots$$
$$\cdots(i_2 \circ j_t)\cdots(i_m \circ j_1)(i_m \circ j_2)\cdots(i_m \circ j_t)$$

where

$$i \circ j = \begin{cases} i *_\alpha j, & \text{if } i, j \in G_\alpha \\ 0, & \text{otherwise.} \end{cases}$$

Now, we give an example of an embedding of a semigroup $S \in KG_n^0$ into relational algebra A_S.

Example. Let a semigroup $S = \bigcup_{\alpha\in I} G_\alpha^0 = \{0, 1, 2, 3, 4\}$, where $G_{\alpha_1} = \{1\}$, $G_{\alpha_2} = \{2, 3\}$, $G_{\alpha_3} = \{4\}$ are groups, be given by Table 1 (The operations of groups G_{α_i} $(1 \leq i \leq 3)$ are given in the table, i.e. $x *_\alpha y = x \bullet y$).

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	0	0	0
2	0	0	2	3	0
3	0	0	3	2	0
4	0	0	0	0	4

TABLE 1

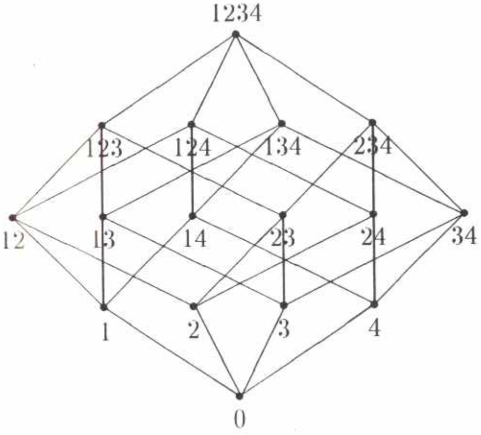


FIGURE 1

Then, for the relational algebra A_S in which the semigroup S is embedded, Boolean reduct is given by the lattice from Figure 1.

The semigroup reduct is presented by the following table

·	0	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
2	0	0	2	3	0	2	3	0	23	2	3	23	2	3	23	23
3	0	0	3	2	0	3	2	0	23	3	2	23	3	2	23	23
4	0	0	0	0	4	0	0	4	0	4	4	0	4	4	4	4
12	0	1	2	3	0	12	13	1	23	2	3	123	12	13	23	123
13	0	1	3	2	0	13	12	1	23	3	2	123	13	12	23	123
14	0	1	0	0	4	1	1	14	0	4	4	1	14	14	4	14
23	0	0	23	23	0	23	23	0	23	23	23	23	23	23	23	23
24	0	0	2	3	4	2	3	4	23	24	34	23	24	34	234	234
34	0	0	3	2	4	3	2	4	23	34	24	23	34	24	234	234
123	0	1	23	23	0	123	123	1	23	23	23	123	123	123	23	123
124	0	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
134	0	1	3	2	4	13	12	14	23	34	24	123	134	124	234	1234
234	0	0	23	23	4	23	23	4	23	234	234	23	234	234	234	234
1234	0	1	23	23	4	123	123	14	23	234	234	123	1234	1234	234	1234

In this algebra, we have that:

$$1'_S = 124, \quad 1_s = 1234.$$

The algebra $\mathcal{A}_S$ is symmetric, since $x^{-1} = x$, for all $x \in \mathcal{A}_S$.

Theorem 3.1. *Let S and S_1 be semigroups from the class KG_n^0. Then the semigroups S and S_1 are isomorphic if and only if the corresponding relational algebras $\mathcal{A}_S$ and $\mathcal{A}_{S_1}$ are isomorphic.*

Proof. Let $S \cong S_1$. Since $S \cong S'$ and $S_1 \cong S_1'$, where $S' \subset A_S$ and $S_1' \subset A_{S_1}$, it follows that $S' \cong S_1'$. Let $\varphi : S' \to S_1'$ be an isomorphism. Define a mapping $f : A_S \to A_{S_1}$ by

$$f(\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_m}}) = \varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \cdots \cup \varphi(\rho_{a_{i_m}}) \tag{8}$$

Since $S' \cong S_1'$, the relational algebras $\mathcal{A}_S$ and $\mathcal{A}_{s1}$ have the same number of atoms, and $Rd_B(\mathcal{A}_S) \cong Rd_B(\mathcal{A}_{S1})$.

Let $x = \rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_r}}$, $y = \rho_{b_{j_1}} \cup \rho_{b_{j_2}} \cup \cdots \cup \rho_{b_{j_s}}$, be arbitrary elements of A_S. Then, by (8), using the isomorphism φ, we have that:

$$\begin{aligned}
f(x \circ y) &= f((\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_r}}) \circ (\rho_{b_{j_1}} \cup \rho_{b_{j_2}} \cup \cdots \cup \rho_{b_{j_s}})) \\
&= f\Big(\bigcup_{\substack{1\le\mu\le r\\ 1\le\nu\le s}} (\rho_{a_{i_\mu}} \circ \rho_{b_{j_\nu}})\Big) = \bigcup_{\substack{1\le\mu\le r\\ 1\le\nu\le s}} \varphi(\rho_{a_{i_\mu}} \circ \rho_{b_{j_\nu}}) \\
&= \bigcup_{\substack{1\le\mu\le r\\ 1\le\nu\le s}} (\varphi(\rho_{a_{i_\mu}}) \circ \varphi(\rho_{b_{j_\nu}})) = \bigcup_{1\le\mu\le r} \varphi(\rho_{a_{i_\mu}}) \circ \bigcup_{1\le\nu\le s} \varphi(\rho_{b_{j_\nu}}) \\
&= f(\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_r}}) \circ f(\rho_{b_{j_1}} \cup \rho_{b_{j_2}} \cup \cdots \cup \rho_{b_{j_s}}) \\
&= f(x) \circ f(y).
\end{aligned}$$

It is clear that $f(\varnothing) = \varnothing$. To prove that f is an injection, assume that $x = \rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_r}}$, $y = \rho_{b_{j_1}} \cup \rho_{b_{j_2}} \cup \cdots \cup \rho_{b_{j_s}}$, $r \le s$, and $f(x) = f(y)$. Then

$$f(\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_r}}) = f(\rho_{b_{j_1}} \cup \rho_{b_{j_2}} \cup \cdots \cup \rho_{b_{j_s}}),$$

and hence

$$\varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \cdots \cup \varphi(\rho_{a_{i_r}}) = \varphi(\rho_{b_{j_1}}) \cup \varphi(\rho_{b_{j_2}}) \cup \cdots \cup \varphi(\rho_{b_{j_s}}) \tag{9}$$

By (9) for all $\mu (1 \le \mu \le r)$, we obtain that

$$\begin{aligned}
&\varphi(\rho_{a_{i_\mu}}) \cap (\varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \cdots \cup \varphi(\rho_{a_{i_r}})) \\
&\qquad = \varphi(\rho_{a_{i_\mu}}) \cap (\varphi(\rho_{b_{j_1}}) \cup \varphi(\rho_{b_{j_2}}) \cup \cdots \cup \varphi(\rho_{b_{j_s}}))
\end{aligned}$$

i.e.

$$(10)\quad \varphi(\rho_{a_{i_\mu}}) = (\varphi(\rho_{a_{i_\mu}}) \cap \varphi(\rho_{b_{j_1}})) \cup (\varphi(\rho_{a_{i_\mu}}) \cap \varphi(\rho_{b_{j_2}})) \cup \dots \\ \dots \cup (\varphi(\rho_{a_{i_\mu}}) \cap \varphi(\rho_{b_{j_s}})).$$

On the right hand side of relation (10) at least one member of the union is different from the empty set, because in the opposite case $\varphi(\rho_{a_{i_\mu}})$ is not an atom in A_{S1} (which gives a contradiction).

Let $\varphi(\rho_{a_{i_\mu}}) \cap \varphi(\rho_{b_{j_\mu}}) \neq 0$, hence,

$$(11)\qquad \varphi(\rho_{a_{i_\mu}}) = \varphi(\rho_{b_{j_\nu}}).$$

It is clear that for different μ $(1 \leq \mu \leq r)$, we obtain diferent ν $(1 \leq \nu \leq s)$, (11) is satisfied, since φ is an isomorphism. By (9) we have that

$$(12)\quad \varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \dots \cup \varphi(\rho_{a_{i_r}}) = \\ (\varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \dots \cup \varphi(\rho_{a_{i_r}})) \cup (\varphi(\rho_{b_{j_{r+1}}}) \cup \dots \cup \varphi(\rho_{b_{j_s}}))$$

Therefore, if we make intersection of the right and left hand sides of (12) with $\varphi(\rho_{b_{j_{r+1}}}) \cup \dots \cup \varphi(\rho_{b_{j_s}})$ and since the atoms are differnt, we obtain that

$$\varphi(\rho_{b_{j_{r+1}}}) \cup \dots \cup \varphi(\rho_{b_{j_s}}) = \varnothing.$$

Hence, $r = s$. By (11) we obtain that

$$\rho_{a_{i_\mu}} = \rho_{b_{j_\nu}}, \text{ for all } \mu, \nu \ (1 \leq \mu, \nu \leq r)$$

and hence $x = y$, which means that f is an injection. The mapping f is "onto". Indeed, for an arbitrary $z = \varphi(\rho_{a_1}) \cup \varphi(\rho_{a_2}) \cup \dots \cup \varphi(\rho_{a_m})$, $f(x) = z$ for $x = \rho_{a_1} \cup \rho_{a_2} \cup \dots \cup \rho_{a_m}$.

For $S' = \cup_{a \in I} G''_\alpha \cup \{\varnothing\}$ and $|I| = k$, $(1 \leq k \leq n)$ let $e_i \in G_{\alpha_i}$ $(1 \leq i \leq k)$ be units of groups G_{α_i}, then $1'_s = \rho_{e_1} \cup \rho_{e_2} \cup \dots \cup \rho_{e_k}$. Now

$$f(1') = f(\rho_{e_1} \cup \rho_{e_2} \cup \dots \cup \rho_{e_k}) = \varphi(\rho_{e_1}) \cup \varphi(\rho_{e_2}) \cup \dots \cup \varphi(\rho_{e_k}) = 1'_1.$$

$$\begin{aligned} f(x^{-1}) &= f((\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \dots \cup \rho_{a_{i_r}})^{-1}) = f({\rho_{a_{i_1}}}^{-1} \cup {\rho_{a_{i_2}}}^{-1} \cup \dots \cup {\rho_{a_{i_r}}}^{-1}) \\ &= \varphi({\rho_{a_{i_1}}}^{-1}) \cup \varphi({\rho_{a_{i_2}}}^{-1}) \cup \dots \cup \varphi({\rho_{a_{i_r}}}^{-1}) \\ &= \varphi(\rho_{a_{i_1}})^{-1} \cup \varphi(\rho_{a_{i_2}})^{-1} \cup \dots \cup \varphi(\rho_{a_{i_r}})^{-1} \\ &= (\varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \dots \cup \varphi(\rho_{a_{i_r}}))^{-1} \\ &= f(\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \dots \cup \rho_{a_{i_r}})^{-1} = f(x)^{-1}. \end{aligned}$$

$$f(1_S) = f(\rho_{a_{i_1}} \cup \rho_{a_{i_2}} \cup \cdots \cup \rho_{a_{i_n}}) = \varphi(\rho_{a_{i_1}}) \cup \varphi(\rho_{a_{i_2}}) \cup \cdots \cup \varphi(\rho_{a_{i_n}}) = 1_{S_1}.$$

Let us prove that f achieves (8).

It is obvius that for arbitrary $x, y \in \mathcal{A}_S$ it is $f(x \cup y) = f(x) \cup f(y)$. For $x = \rho_{a_{i_1}} \cup \cdots \cup \rho_{a_{i_r}}$ will be

$$f(\bar{x}) = f(\overline{\rho_{a_{i_1}}} \cup \cdots \cup \rho_{a_{i_s}}) = f(\rho_{a_{i_1}} \cup \cdots \cup \rho_{a_{i_{n-s}}})$$

where $\rho_{a_{i_\nu}} \cap \rho_{a_{i_{r\mu}}} = \varnothing$ for all $1 \leq \nu \leq s$ and $1 \leq \mu \leq n - s$. From there

$$f(\bar{x}) = \varphi(\rho_{e_{i_1}}) \cup \cdots \cup \varphi(\rho_{e_{i_{nus}}}) = \overline{\varphi(\rho_{a_{i_1}}) \cup \cdots \cup \varphi(\rho_{a_{i_s}})}$$

because of isomorphism φ will be $\varphi(\rho_{e_{i_\nu}}) \cap \varphi(\rho_{e_{i_{mus}}}) = \varnothing$, for every $1 \leq \nu \leq s$ and $1 \leq \mu \leq n - s$. So,

$$f(\bar{x}) = \overline{f(\rho_{a_{i_1}} \cup \cdots \cup \rho_{a_{i_s}})} = \overline{f(x)}.$$

Because of $x \cap y = \overline{\bar{x} \cup \bar{y}}$, f is also an isomorphism for $\cap$. Thus, $\mathcal{A}_S \cong \mathcal{A}_{S_1}$.

On the other hand if $f : A_S \to A_{S_1}$ is an isomorphism of algebras $\mathcal{A}_S$ and $\mathcal{A}_{S_1}$, then the restriction $f|_{S'} : S' \to S_1'$ is an isomorphism (since the isomrophism f maps atoms from $\mathcal{A}_S$ in atoms from $\mathcal{A}_{S_1}$), and hence $S \cong S_1$. Thus the theorem is proved. □

References

[1] S. Bogdanović and M. Ćirić, *Orthogonal sums of semigroups*, Israel J. Math. **90** (1995), 423–428.

[2] S. Bogdanović and M. Ćirić, *Polugrupe*, Prosveta, Niš, 1993.

[3] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, Vol 1, Amer. Math. Soc., 1961.

[4] R. Sz. Madarász and S. Crvenković, *Relacione algebre*, Matematički institut Beograd, Beograd, 1992.

Ul. 27. marta br 5/7, Beograd, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 657–674
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

VARIETIES OF POLYADIC GROUPS

Wieslaw A. Dudek

ABSTRACT. In this note the class of all n-ary groups is considered as the class of some universal algebras with different systems of fundamental operations. In any such case we give the minimal systems of identities defining this class.

1. Introduction

Wilhelm Dörnte, inspired by E. Noether, introduced in 1928 (see [1]) the notion of n-group (called also n-ary group or polyadic group), which is a natural generalization of the notion of group. The idea of such investigations seems to be going back to E. Kasner's lecture at the fifty-third annual meeting of the American Association for the Advancement of Science, reported (by L. G. Weld) in The Bulletin of the American Mathematical Society in 1904 (see [2]). The second paper which plays a very important role in the theory of n-ary groups is the large paper (143 pages) of E. L. Post [3].

We shall use the following abreviated notation: the sequence $x_i, x_{i+1}, ..., x_j$ will be denoted by x_i^j. For $j < i$ x_i^j is the empty symbol. In this convention $f(x_1^n)$ denotes $f(x_1, x_2, ..., x_n)$. The word

$$f(x_1, x_2, ..., x_k, x, ..., x, x_{k+t+1}, ..., x_n),$$

where x appears t times, will be denoted by $f(x_1^k, \overset{(t)}{x}, x_{k+t+1}^n)$. For $t \leq 0$ the symbol $\overset{(t)}{x}$ will be empty.

If $m = k(n-1)+1$, then the m-ary operation g given by

$$g(x_1^{k(n-1)+1}) = \underbrace{f(f(..., f(f}_{k}(x_1^n), x_{n+1}^{2n-1}), ...), x_{(k-1)(n-1)+2}^{k(n-1)+1})$$

will be denoted by $f_{(k)}$. In certain situations, when the arity of g does not play a crucial role, or when it will differ depending on additional assumptions, we write $f_{(.)}$, to mean $f_{(k)}$ for some $k = 1, 2, \ldots$.

A non-empty set G with an n-ary operation $f : G^n \longrightarrow G$ will be called *an n-groupoid* or *an n-ary groupoid* and will be denoted by $(G; f)$. An n-groupoid $(G; f)$ will be called *an n-group* or *an n-ary group* if and only if

1^0 for all $x_1, x_2, \ldots, x_{2n-1} \in G$ *the* (i, j)*-associative law*

$$f(x_1^{i-1}, f(x_i^{n+i-1}), x_{n+i}^{2n-1}) = f(x_1^{j-1}, f(x_j^{n+j-1}), x_{n+j}^{2n-1}) \tag{1}$$

hold for every $i, j \in \{1, 2, \ldots, n\}$,

2^0 for all $x_0, x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n \in G$ $(k = 1, 2, \ldots, n)$ there exist a unique $z \in G$ such that

$$f(x_1^{k-1}, z, x_{k+1}^n) = x_0. \tag{2}$$

Condition 1^0 is called *associativity*, and algebras (G, f) fulfilling 1^0 are called *n-semigroups*. Algebras fulfilling only 2^0 are called *n-quasigroups*.

The above definition is a generalization of H. Weber's formulation of axioms of a group (from 1896). Similar generalization of L. E. Dickson's (with the neutral element) one leads to some narrower class of n-groups derived from 2-groups (i.e. classical groups).

It is interesting that there exists no nontrivial (on a non one-element set) theory of infinitary groups, i.e. ω-groups for countable infinite ordinal ω, but there exist infinitary quasigroups of any (finite and infinite) order [19]. Therefore we shall consider n-ary groups (n-ary groupoids) only in the case when $n \geq 2$ is a fixed (but arbitrary) natural number.

It is worthwhile to note that, under the assumption 1^0, it suffices only to postulate the existence of a solution of (2) at the places $k = 1$ and $k = n$ or at one place k other than 1 and n. Then one can prove uniqueness of the solution of (2) for all $k = 1, \ldots, n$ (see [3], p. 213^{17}). Also the following Proposition is true (see [4]).

Proposition 1.1. *An n-groupoid $(G; f)$ is an n-group if and only if (at least) one of the following conditions is satisfied:*

(a) *the $(1, 2)$-associative law holds and the equation (2) is solvable for $k = n$ and uniquely solvable for $k = 1$,*

(b) *the $(n-1, n)$-associative law holds and the equation (2) is solvable for $k = 1$ and uniquely solvable for $k = n$,*

(c) *the $(i, i+1)$-associative law holds for some $i \in \{2, \ldots, n-2\}$ and the equation (2) is uniquely solvable for i and some $k > i$.*

2. Varieties of n-ary groups

In an n-quasigroup $(G; f)$ for every $s \in \{1, 2, ..., n\}$ one can define *the s-th inverse n-ary operation* $f^{(s)}$ putting

$$f^{(s)}(x_1^n) = y \quad \textit{if and only if} \quad f(x_1^{s-1}, y, x_{s+1}^n) = x_s \, .$$

Obviously, the operation $f^{(s)}$ is the s-th inverse operation for f if and only if

$$f^{(s)}(x_1^{s-1}, f(x_1^n), x_{s+1}^n) = x_s \, , \tag{3}$$

for all $x_1, ..., x_n \in G$. Therefore (as in the binary case) the class of all n-quasigroups (and in the consequence the class of all n-groups) may be treated as the variety of equationally definable algebras with $n+1$ fundamental n-ary operations $f, f^{(1)}, f^{(2)}, ..., f^{(n)}$. Such variety is defined by (1) and (3). Obviously (1) and (3) must hold for all $i, j, s \in \{1, 2, ..., n\}$.

An n-group ($n > 2$) may be considered also as an algebra with three n-ary operations. Namely, as a consequence of Proposition 1.1 we obtain the following characterization.

Corollary 2.1. *Every n-ary group* ($n > 2$) *may be considered as an algebra* $(G; f, f^{(j)}, f^{(k)})$ *of the type* (n, n, n) *with the* $(i, i+1)$*-associative operation* f *where*

(a) $i = j = 1$ *and* $k = n$, *or*
(b) $i = n-1$, $j = 1$ *and* $k = n$, *or*
(c) $i \in \{2, ..., n-2\}$ *is fixed and* $k > j = i$.

Corollary 2.2. *The class of algebras with three n-ary* ($n > 2$) *operations* f, g, h *is the variety of all n-ary groups* $(G; f)$ *if and only if (at least) one of the following axiom systems is satisfied:*

$$(a) \quad \begin{cases} f(f(x_1^n), x_{n+1}^{2n-1}) = f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1}) \, , \\ g(f(y, x_2^n), x_2^n) = y \, , \\ h(x_1^{n-1}, f(x_1^{n-1}, y)) = y \, , \end{cases}$$

$$(b) \quad \begin{cases} f(x_1^{n-2}, f(x_{n-1}^{2n-2}), x_{2n-1}) = f(x_1^{n-1}, f(x_n^{2n-1})) \, , \\ g(f(y, x_2^n), x_2^n) = y \, , \\ h(x_1^{n-1}, f(x_1^{n-1}, y)) = y \, , \end{cases}$$

$$(c) \quad \begin{cases} f(x_1^{i-1}, f(x_i^{n+i-1}), x_{n+i}^{2n-1}) = f(x_1^i, f(x_{i+1}^{n+i}), x_{n+i+1}^{2n-1}) \, , \\ g(x_1^{i-1}, f(x_1^{i-1}, y, x_{i+1}^n), x_{i+1}^n) = y, \\ h(x_1^{s-1}, f(x_1^{s-1}, y, x_{s+1}^n), x_{s+1}^n) = y, \ \textit{where } 1 < i < s < n \textit{ are fixed.} \end{cases}$$

Note that axiom systems given by (a) and (b) (also in Corollary 2.1) are valid for $n = 2$, too. But the greater part of characterizations of n-ary groups obtained by several authors are valid only for $n > 2$. Characterizations which are valid also for $n = 2$ are given for example in [7], [9] and [8]. Since in all these characterizations f is an associative operation, then founded systems of defining identities are not minimal.

We give such minimal system basing on result obtained in [8].

Corollary 2.3. *The class of all n-ary groups* $(n \geq 2)$ *may be considered as the variety of algebras with one* $(1,2)$*-associative (or* $(n-1,n)$*-associative) n-ary operation* f *and one* $(n-1)$*-ary operation* h *satisfying the following two axioms:*

$$\text{(a)}\quad f(h(x_1^{n-2}, z), x_1^{n-2}, f(z, x_1^{n-2}, y)) = y\,,$$

$$\text{(b)}\quad f(f(y, x_1^{n-2}, z), x_1^{n-2}, h(x_1^{n-2}, z)) = y\,.$$

Proof. If an algebra $(G; f, h)$ satisfies the above conditions, then as in [8] one can prove that (2) has a unique solution at the place $k = 1$ and $k = n$, which together with our Proposition 1.1 proves that $(G; f)$ is an n-group.

Conversely, if $(G; f)$ is an n-group then for every $x_1, ..., x_{n-2} \in G$ there exists a unique element $v \in G$ such that

$$y = f(y, x_1^{n-2}, v) = f(x_1^{n-2}, v, y) = f(y, v, x_1^{n-2}) = f(v, x_1^{n-2}, y)$$

for each $y \in G$ (see [3], 214-215). Hence for every $z, x_1, ..., x_{n-2} \in G$ there exists only one $u \in G$ such that

$$f(u, x_1^{n-3}, f(x_{n-2}, z, x_1^{n-2}), y) = y$$

holds for each $y \in G$. Since u depends on $n-1$ elements $z, x_1, ..., x_{n-2}$, it may be treated as the value of an $(n-1)$-ary operation h. Obviously h satisfies (a) and (b). This completes the proof.

As it is well known in an n-group $(G; f)$ the equation

$$(4)\qquad f(\overset{(n-1)}{x}, z) = x$$

has a unique solution $z \in G$, which is called *the skew element* to x and is denoted by $\overline{x}$. Since for every $x \in G$ there exists only one skew element, then the solution of (4) induces on G a new unary operation $x \to \overline{x}$. Thus an n-group $(G; f)$ may be considered as an algebra $(G, f; ^{-})$ with two fundamental opertations: an n-ary one f and an a unary one $x \to \overline{x}$. The variety of such n-groups is defined (see [6]) by three identities: one of the type (1) and two so-called *Dörnte's identities*

$$(5)\qquad f(\overset{(i-1)}{x}, \overline{x}, \overset{(n-i-1)}{x}, y) = y\,,$$

$$(6) \qquad f(y,\overset{(n-j-1)}{x},\overline{x},\overset{(j-1)}{x}) = y\,.$$

In an n-group the last two identities hold for all $i,j \in \{1,2,...,n-1\}$, but one can prove (see for example [6],[4]) that (5) and (6) determine (together with (1)) an n-group if it hold for some fixed i,j. The minimal base of such variety is given by the following theorem (proved in [7]).

Theorem 2.4. *Let $(G;f,^{-})$ be an n-ary groupoid $(n>2)$ with a unary operation $x \to \overline{x}$. Then $(G;f,^{-})$ is an n-group if and only if f is $(1,2)$ or $(n-1,n)$-associative and Dörnte's identities hold for some fixed $i,j \in \{1,2,...,n-1\}$.*

As a consequence we obtain

Corollary 2.5. *The class of all n-ary groups $(n \geq 2)$ may be considered as the variety of algebras with one $(1,2)$-associative (or $(n-1,n)$-associative) n-ary operation f and one unary operation $x \to \hat{x}$ satisfying the following two axioms:*

$$\text{(a)} \quad f(\hat{x},\overset{(n-2)}{x},f(\overset{(n-1)}{x},y)) = y\,,$$

$$\text{(b)} \quad f(f(y,\overset{(n-1)}{x}),\overset{(n-2)}{x},\hat{x}) = y\,.$$

Theorem 2.6. *The class of algebras $(G;f,g,h)$ with one $(1,2)$-associative (or $(n-1,n)$-associative) n-ary $(n>2)$ operation f and two $(n-2)$-ary operations g and h is the variety of n-ary groups if and only if the following two identities*

$$(7) \qquad f(x_1^{i-1},g(x_1^{n-2}),x_i^{n-2},y) = y\,,$$

$$(8) \qquad f(y,x_1^{j-1},h(x_1^{n-2}),x_j^{n-2}) = y$$

hold for some fixed $i,j \in \{1,2,...,n-1\}$.

Proof. From [3] (p.215) follows that in every n-group $(G;f)$ there exists an $(n-2)$-ary operation g satisfying (7). Similarly there exists an $(n-2)$-ary operation h satisfying (8). Thus (7) and (8) hold in every n-group.

To prove the converse observe first that putting in (7) $x = x_1 = ... = x_{n-2}$ and $g(\overset{(n-2)}{x}) = \overline{x}$ we obtain the identity

$$(9) \qquad f(\overset{(i-1)}{x},\overline{x},\overset{(n-i-1)}{x},y) = y\,.$$

Similarly, for $h(\overset{(n-2)}{x}) = \hat{x}$ from (8) follows

$$(10) \qquad f(y,\overset{(j-1)}{x},\hat{x},\overset{(n-j-1)}{x}) = y\,.$$

If f is $(1,2)$-associative, then (10) implies

$$\begin{aligned}
&f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1}) = f(f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1}), \overset{(j-1)}{x}, \hat{x}, \overset{(n-j-1)}{x}) = \\
&f(x_1, f(f(x_2^{n+1}), x_{n+2}^{2n-1}, x), \overset{(j-2)}{x}, \hat{x}, \overset{(n-j-1)}{x}) = \\
&f(x_1, f(x_2, f(x_3^{n+2}), x_{n+3}^{2n-1}, x), \overset{(j-2)}{x}, \hat{x}, \overset{(n-j-1)}{x}) = \\
&f(f(x_1^2, f(x_3^{n+2}), x_{n+3}^{2n-1}), \overset{(j-1)}{x}, \hat{x}, \overset{(n-j-1)}{x}) = f(x_1^2, f(x_3^{n+2}), x_{n+3}^{2n-1}).
\end{aligned}$$

This proves $(1,3)$-associativity of f. Now, using $(1,2)$ and $(1,3)$-associativity we prove $(1,4)$-associativity. Similarly we can prove $(1,k)$-associativity for all $k = 5, 6, \ldots, n$. Thus $(G; f)$ is an n-semigroup.

In the case of $(n-1, n)$-associativity the proof is analogous.

To prove that $(G; f)$ is an n-group it is sufficient to solve (2) for $k = 1$ and $k = n$. In the same manner as in the proof of Theorem 2 in [4] one can verify that if (9) holds for $2 \leq i \leq n-1$ then the element

$$z = f_{(.)}(\overset{(i-2)}{x_{n-1}}, \overline{x}_{n-1}, \overset{(n-i-1)}{x_{n-1}}, \overset{(i-2)}{x_{n-2}}, \overline{x}_{n-2}, \overset{(n-i-1)}{x_{n-2}}, \ldots, \overset{(i-2)}{x_1}, \overline{x}_1, \overset{(n-i-1)}{x_1}, x_0)$$

is a solution of the equation $f(x_1^{n-1}, z) = x_0$.

Similarly, under the assumption $1 \leq j \leq n-2$ in (10), the element

$$z = f_{(.)}(x_0, \overset{(j-1)}{x_n}, \hat{x}_n, \overset{(n-j-2)}{x_n}, \overset{(j-1)}{x_{n-1}}, \hat{x}_{n-1}, \overset{(n-j-2)}{x_{n-1}}, \ldots, \overset{(j-1)}{x_2}, \hat{x}_2, \overset{(n-j-2)}{x_2})$$

is a solution of the equation $f(z, x_2^n) = x_0$.

Thus $(G; f)$ is an n-group if (9) and (10) hold with the restriction:

(11) $\quad 2 \leq i \leq n-1 \quad$ *and* $\quad 1 \leq j \leq n-2.$

We have still to consider the following cases:

(12) $\quad i = 1, \qquad j = n-1,$

(13) $\quad i = 1, \qquad 2 \leq j \leq n-2,$

(14) $\quad j = n-1, \quad 2 \leq i \leq n-2,$

(15) $\quad i = n-1, \quad j = n-1,$

(16) $\quad i = 1, \qquad j = 1.$

Let (9) and (10) hold for $i = 1$ and $j = n-1$. Then

$$f(\overline{x}, \overset{(n-2)}{x}, y) = f(y, \overset{(n-2)}{x}, \hat{x}) = y,$$

which gives

$$f(\overline{x}, \overset{(n-1)}{x}) = f(\overset{(n-2)}{x}, \hat{x}) = x \quad \textit{and} \quad \overline{x} = f(\overline{x}, \overset{(n-2)}{x}, \hat{x}) = \hat{x}\,.$$

As a consequence we obtain

$$y = f(\overline{x}, \overset{(n-2)}{x}, y) = f(\overline{x}, f(\overset{(n-1)}{x}, \overline{x}), \overset{(n-3)}{x}, y) =$$

$$f(f(\overline{x}, \overset{(n-1)}{x}), \overline{x}, \overset{(n-3)}{x}, y) = f(x, \overline{x}, \overset{(n-3)}{x}) .$$

By a similar calculation we get

$$y = f(y, \overset{(n-3)}{x}, \overline{x}, x) = f(y, \overset{(n-3)}{x}, \widehat{x}, x) .$$

Thus the case (12) is reduced to (11) and $(G; f)$ is an n-group.

If (9) and (10) hold with the restriction (13), then

$$x = f(\overline{x}, \overset{(n-1)}{x}) = f(\overset{(j)}{x}, \widehat{x}, \overset{(n-j-1)}{x}),$$

which implies

$$y = f(\overline{x}, \overset{(n-2)}{x}, y) = f(\overline{x}, \overset{(n-j-1)}{x}, f(\overset{(j)}{x}, \widehat{x}, \overset{(n-j-1)}{x}), \overset{(j-2)}{x}, y) =$$

$$f(f(\overline{x}, \overset{(n-1)}{x}), \widehat{x}, \overset{(n-3)}{x}, y) = f(x, \widehat{x}, \overset{(n-3)}{x}, y).$$

Hence

$$f(x, \overline{x}, \overset{(n-3)}{x}, y) = f(x, \overline{x}, \overset{(n-3)}{x}, f(x, \widehat{x}, \overset{(n-3)}{x}, y)) =$$

$$f(x, f(\overline{x}, \overset{(n-2)}{x}, \widehat{x}), \overset{(n-3)}{x}, y) = f(x, \widehat{x}, \overset{(n-3)}{x}, y) = y.$$

This proves that (9) holds also for $i = 2$. Therefore (13) may be reduced to (11) and $(G; f)$ is an n-group. By a similar argumentation the case (14) may be reduced to (11).

Now we consider the case (15). In this case the identity (9) has the form $f(\overset{(n-2)}{x}, \overline{x}, y) = y$. which in particular implies

$$f(\overset{(n-2)}{x}, \overline{x}, x) = f(\overset{(n-2)}{\overline{x}}, \overline{\overline{x}}, x) = x ,$$

where $\overline{\overline{x}} = g(\overset{(n-2)}{\overline{x}})$. Using these identities it is not difficult to verify that the solution z of the equation

$$f(\overset{(n-3)}{x}, \overline{x}, x, z) = y$$

has the form

$$z = f_{(n-2)}(\overset{(n-3)}{x}, \overline{x}, \overset{(n-3)}{\overline{x}}, \overline{\overline{x}}, \underbrace{\underbrace{\overset{(n-3)}{x}, \overline{x}}, \dots, \underbrace{\overset{(n-3)}{x}, \overline{x}}}_{(n-3)\ times}, y) =$$

$$f_{(n-3)}(\overset{(n-3)}{x}, f(\overset{(n-2)}{\overline{x}}, \overline{\overline{x}}, x), \overset{(n-4)}{x}, \overline{x}, \overset{(n-3)}{x}, \overline{x}, \dots, \overset{(n-3)}{x}, \overline{x}, y) =$$

$$f_{(n-3)}(\overset{(n-3)}{x}, \overset{(n-3)}{x}, \overline{x}, \dots, \overset{(n-3)}{x}, \overline{x}, y) =$$

$$f_{(n-4)}(\overset{(n-4)}{x}, f(\overset{(n-2)}{x}, \overline{x}, x), \overset{(n-4)}{x}, \overline{x}, \dots, \overset{(n-3)}{x}, \overline{x}, y) =$$

$$f_{(n-4)}(\overset{(n-4)}{x}, \overset{(n-3)}{x}, \overline{x}, \dots, \overset{(n-3)}{x}, \overline{x}, y) = \dots =$$

$$f(x, f(\overset{(n-2)}{x}, \overline{x}, x), \overset{(n-4)}{x}, \overline{x}, y) = f(\overset{(n-2)}{x}, \overline{x}, y) = y.$$

Hence in this case holds also $f(\overset{(n-3)}{x}, \overline{x}, x, y) = y$, which reduces (15) to (14). Analogously (16) may be reduced to (13). This completes our proof.

Note that in general $g(x_1^{n-2}) \neq h(x_1^{n-2})$, but as it easy to show $g(x, \dots, x) = h(x, \dots, x)$ for all $x \in G$. Moreover, using the Post's Coset Theorem (see [3]), one can prove that in the case $i = j$ we have $g(x_1^{n-2}) = h(x_1^{n-2})$. Hence as a simple consequence of Theorem 2.6 we obtain

Corollary 2.7. *The class of algebras $(G; f, g)$ with one $(1,2)$-associative (or $(n-1, n)$-associative) n-ary $(n > 2)$ operation f and one $(n-2)$-ary operation g is the variety of n-ary groups if and only if the following two identities*

(a) $f(x_1^{i-1}, g(x_1^{n-2}), x_i^{n-2}, y) = y$,
(b) $f(y, x_1^{i-1}, g(x_1^{n-2}), x_i^{n-2}) = y$

hold for some fixed $i = 1, 2, \dots, n-1$.

Corollary 2.8. *The variety of n-ary groups $(n > 2)$ is the class of algebras $(G; f, g, h)$ with one associative n-ary operation f and two $(n-2)$-ary operations g and h satisfying for some fixed $i, j \in \{1, 2, \dots, n-1\}$ the identity*

$$f_{(2)}(x_1^{i-1}, g(x_1^{n-2}), x_i^{n-2}, y, x_1^{j-1}, h(x_1^{n-2}), x_j^{n-2}) = y\,. \tag{17}$$

Proof. In every n-group $(G; f)$ there exist (by Theorem 2.6) two $(n-2)$-ary operations g and h satisfying (7) and (8). Hence (17) is satisfied, too.

Conversely, if the identity (17) holds in an n-semigroup $(G; f)$, then putting $x = x_1 = \dots = x_{n-2}$, $g(\overset{(n-2)}{x}) = \tilde{x}$ and $h(\overset{(n-2)}{x}) = \hat{x}$ in (17) we obtain

$$f_{(2)}(\overset{(i-1)}{x}, \tilde{x}, \overset{(n-i-1)}{x}, y, \overset{(j-1)}{x}, \hat{x}, \overset{(n-j-1)}{x}) = y\,.$$

Using the same method as in the proof of Theorem 4 from [7] one can prove that $(G;f)$ is an n-group, which completes the proof.

Analogously as in Theorem 2.6, using the Post's Coset Theorem, one can prove that $\tilde{x} = \hat{x}$ for every $x \in G$. Thus as a simple consequence we obtain

Corollary 2.9. *The variety of n-ary groups $(n > 2)$ may be considered as the class of n-ary semigroups $(G;f)$ with one unary operation $x \to \hat{x}$ satisfying for some fixed $i, j \in \{2, 3, ..., n\}$ the identity*

$$f_{(2)}(\overset{(i-2)}{x}, \hat{x}, \overset{(n-i)}{x}, y, \overset{(n-j)}{x}, \hat{x}, \overset{(j-2)}{x}) = y\,.$$

Observe that from Corollary 2.1 (a) follows that the class of n-ary groups $(n \geq 2)$ may be considered as the subvariety of the variety of n-ary quasigroups. For $n \geq 3$ this class may be considered also as the subvariety of the class of inversive n-ary semigroups described in [20] and may be defined by a system of identities containing some identities which are characteristic for inversive n-semigroups.

Proposition 2.10. *The class of all n-ary groups $(n > 2)$ may be considered as the variety of algebras $(G; f, g, h)$ of the type $(n, n-2, 3)$ defined by*

(a) $f(x_1^n) = h(x_1, g(x_2^{n-1}), x_n)$,

(b) $h(y, x, x) = h(x, x, y) = y$,

(c) $h(h(x_1^3), x_4^5) = h(x_1, h(x_4, x_3, x_2), x_5) = h(x_1^2, h(x_3^5))$,

(d) $g(x_1^{n-3}, g(x_1^{n-2})) = x_{n-2}$,

where the operation f is $(1,2)$ or $(n-1, n)$-associative.

Proof. Any n-group $(n \geq 3)$ is an inversive n-semigroup in which there exist two operations g and h satisfying the above identities (see [20]).

Conversely, if an algebra $(G; f, g, h)$ satisfies the above conditions, then for all $x, y \in G$ and $\hat{x} = g(\overset{(n-2)}{x})$, we obtain

$$f(y, \overset{(n-2)}{x}, \hat{x}) = h(y, g(\overset{(n-2)}{x}), \hat{x}) = h(y, \hat{x}, \hat{x}) = \cdot y$$

and

$$f(\hat{x}, \overset{(n-2)}{x}, y) = h(\hat{x}, \hat{x}, y) = y\,,$$

which together with the $(1,2)$-associativity of f implies the $(1,3)$-associativity. Indeed,

$$f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1}) = f(f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1}), \overset{(n-2)}{x}, \widehat{x}) =$$

$$f(x_1, f(f(x_2^{n+1}), x_{n+2}^{2n-1}, x), \overset{(n-3)}{x}, \widehat{x}) = f(x_1, f(x_2, f(x_3^{n+2}), x_{n+3}^{2n-1}, x), \overset{(n-3)}{x}, \widehat{x}) =$$

$$f(f(x_1^2, f(x_3^{n+2}), x_{n+3}^{2n-1}), \overset{(n-2)}{x}, \widehat{x}) = f(x_1^2, f(x_3^{n+2}), x_{n+3}^{2n-1}) .$$

Now, using $(1,2)$ and $(1,3)$-associativity we prove the $(1,4)$-associativity. Similarly we can prove $(1,j)$-associativity for $j = 5, 6, ..., n$. Thus $(G; f)$ is an n-semigroup. By Theorem 13 from [20] it is an n-group.

In the case of the $(n-1, n)$-associativity the proof is analogous.

Moreover, the above proof suggest the following characterization of n-groups.

Corollary 2.11. *The class of all n-ary groups* $(n > 2)$ *may be considered as the variety of algebras* $(G; f, g, h)$ *of the type* $(n, n-2, 3)$ *defined by*

(a) $f(x_1^n) = h(x_1, g(x_2^{n-1}), x_n)$,

(b) $h(y, x, x) = h(x, x, y) = y$,

(c) $f(f(x_1^n), x_{n+1}^{2n-1}) = f(x_1, f(x_2^{n+1}), x_{n+2}^{2n-1})$.

Proof. As in the previous proof we can prove that $(G; f)$ is an n-semigroup with a unary operation $x \to \overline{x} = g(\overset{(n-2)}{x})$ and satisfies the assumption of Theorem 2.4. Hence it is an n-group.

Conversely, if $(G; f)$ is an n-group, then by Post's Coset Theorem (see [3]) there exists a binary group $(G^*, \cdot)$ such that $f(x_1^n) = x_1 \cdot x_2 \cdot ... \cdot x_n$ for all $x_1, x_2, ..., x_n \in G$. Hence $g(x_1^{n-2}) = (x_1 \cdot x_2 \cdot ... \cdot x_{n-2})^{-1}$ and $h(x, y, z) = x \cdot y^{-1} \cdot z$ are operations fulfilling (a) and (b), which completes the proof.

Remark that in general the operations g from Proposition 2.10 and Corollary 2.11 are not identical because the second not satisfies (d), in general.

It is worth remaining that the operation h satisfying (b) is so-called *Mal'cev operation.* The existence of such operation in the set of all polynomials of some variety of general algebras is equivalent to the commutativity of congruence on each algebra from this variety. Moreover, the lattice of all congruences of a fixed algebra from such variety is modular (see for example [21]). Thus for every fixed $n \geq 2$ the class of all n-groups is a Mal'cev variety and the lattice of all congruences of a fixed n-groups is modular. For $n = 2$ this fact is known, for $n \geq 2$ it was proved in [22].

Theorem 2.12. *The class of algebras $(G; f, g, h)$ with one associative n-ary $(n \geq 2)$ operation f and two binary operations g and h is the variety of all n-ary groups if and only if for some fixed $i, j \in \{1, 2, ..., n-1\}$ the following two identities hold:*

$$(18) \qquad f(\overset{(i)}{x}, \overset{(n-1-i)}{y}, g(x, y)) = y,$$

$$(19) \qquad f(h(x, y), \overset{(n-1-j)}{y}, \overset{(j)}{x}) = y.$$

Proof. It is well known that in every n-group $(n \geq 2)$ the solution z of the equation $f(\overset{(i)}{x}, \overset{(n-1-i)}{y}, z) = y$ there exists and depends only on x and y. Thus z may be treated as the value of a binary operation g satisfying (18). The similar argumentation shows that there exists a binary operation h satisfying (19). (In general $g(x, y) \neq h(x, y)$, but $g(x, x) = h(x, x)$ for all $x \in G$.)

Conversely, let $(G; f)$ be an n-semigroup with two binary operation satisfying (18) and (19). Then in a similar way as in the proof of Theorem 2 in [4] one can verify that for $2 \leq i \leq n-1$ the element

$$z = f_{(\cdot)}(\overset{(i-1)}{x_{n-1}}, \overset{(n-1-i)}{x_{n-2}}, g(x_{n-1}, x_{n-2}), \overset{(i-2)}{x_{n-2}}, \overset{(n-1-i)}{x_{n-3}}, g(x_{n-2}, x_{n-3}), \dots$$

$$\dots, \overset{(i-2)}{x_2}, \overset{(n-1-i)}{x_1}, g(x_2, x_1), \overset{(i-2)}{x_1}, \overset{(n-1-i)}{x_0}, g(x_1, x_0))$$

is a solution of the equation $f(x_1^{n-1}, z) = x_0$.

For $i = 1$ this solution has the form

$$z = f_{(\cdot)}(\overset{(n-2)}{x_0}, g(x_{n-1}, x_0), \overset{(n-3)}{x_0}, g(x_2, x_0), \dots, \overset{(n-3)}{x_0}, g(x_1, x_0)).$$

Similarly, the solution of $f(z, x_2^n) = x_0$ has the form

$$z = f(h(x_n, x_0), \overset{(n-1-j)}{x_0}, \overset{(j-2)}{x_n}, h(x_n, x_{n-1}), \overset{(n-1-j)}{x_n}, \overset{(j-1)}{x_{n-1}}, \dots$$

$$\dots, h(x_4, x_3), \overset{(n-1-j)}{x_4}, \overset{(j-2)}{x_3}, h(x_3, x_2), \overset{(n-1-j)}{x_3}, \overset{(j-1)}{x_2})$$

for $j \in \{2, ..., n-1\}$, and

$$z = f(h(x_n, x_0), \overset{(n-3)}{x_0}, \dots, h(x_3, x_2), \overset{(n-3)}{x_0}, h(x_2, x_0), \overset{(n-2)}{x_0})$$

for $j = 1$.

This proves (by Proposition 1.1) that $(G; f)$ is an n-group.

As a simple consequence of Theorem 2.12 we obtain

Corollary 2.13. *The class of algebras $(G; f, g, h)$ with one associative n-ary $(n \geq 2)$ operation f and two binary operations g and h is the variety of all n-ary groups if and only if the following two identities hold:*

(i) $f(\overset{(n-1)}{x}, g(x, y)) = y$,

(ii) $f(h(x, y), \overset{(n-1)}{x}) = y$.

Corollary 2.14. *An n-semigroup $(G; f)$ is an n-group $(n \geq 2)$ if and only if for every $x, y \in G$ and some fixed $i, j \in \{1, 2, ..., n-1\}$ there exists $z \in G$ such that*

(i) $f(\overset{(i)}{x}, \overset{(n-1-i)}{y}, z) = y$,

(ii) $f(z, \overset{(n-1-j)}{y}, \overset{(j)}{x}) = y$.

In particular, for $i = j = n - 1$ we obtain the following result proved in [10].

Corollary 2.15. *An n-semigroup $(G; f)$ is an n-group $(n \geq 2)$ if and only if for every $x, y \in G$ there exists $z \in G$ such that*

(i) $f(\overset{(n-1)}{x}, z) = y$,

(ii) $f(z, \overset{(n-1)}{x}) = y$.

3. Subvarieties

In this part basing on the results of previous section we describe some subvarieties of the variety of all n-groups.

In the first place we consider the class of idempotent n-groups. This class is the variety selected from the variety of n-groups by the identity $f(x, ..., x) = x$. Since in idempotent n-groups $(G; f)$ the operation $x \longrightarrow \overline{x}$ is the identity mapping, i.e. $x = \overline{x}$ for all $x \in G$, then by Theorem 2.4 this class has the following description, which for $n = 2$ trivially yields one-element groups.

Proposition 3.1. *The class of all idempotent n-ary groups $(n \geq 2)$ is the variety of algebras $(G; f)$ with one $(1, 2)$ or $(n-1, n)$-associative n-ary operation f such that the equalities*

$$f(\overset{(n-1)}{x}, y) = f(y, \overset{(n-1)}{x}) = y$$

holds for every $x, y \in G$.

As a consequence of Theorem 2.6 we obtain

Corollary 3.2. *The class of algebras* $(G: f, g)$ *with one* $(1,2)$*-associative (or* $(n-1,n)$*-associative)* n*-ary* $(n > 2)$ *operation* f *and two idempotent* $(n-2)$*-ary operations* g *and* h *is the variety of idempotent n-ary groups if and only if for some fixed* $i, j \in \{1, 2, ..., n-1\}$ *the following two identities hold*

(a) $f(x_1^{i-1}, g(x_1^{n-2}), x_i^{n-2}, y) = y\,,$

(b) $f(y, x_1^{j-1}, h(x_1^{n-2}), x_j^{n-2}) = y\,.$

In a similar way as Theorem 2.12 we can prove

Proposition 3.3. *The class of algebras* $(G; f, g, h)$ *with one associative n-ary* $(n \geq 2)$ *operation* f *and two idempotent binary operations* g *and* h *is the variety of all idempotent n-ary groups if and only if for some fixed* $i, j \in \{1, 2, ..., n-1\}$ *the following two identities hold:*

(i) $f(\overset{(i)}{x}, \overset{(n-1-i)}{y}, g(x, y)) = y\,,$

(ii) $f(h(x, y), \overset{(n-1-j)}{y}, \overset{(j)}{x}) = y\,.$

Corollary 3.4. *The class of algebras* $(G; f, g, h)$ *with one associative n-ary* $(n \geq 2)$ *operation* f *and two idempotent binary operations* g *and* h *is the variety of all idempotent n-ary groups if and only if the following identities hold:*

(i) $f(\overset{(n-1)}{x}, g(x, y)) = y\,,$

(ii) $f(h(x, y), \overset{(n-1)}{x}) = y\,.$

The variety of idempotent n-ary groups may be considered also as the variety of n-groups in which all inverse operations are idempotent. The minimal system of identities defining such variety is given (for example) by Corollary 2.1 and Corollary 2.2, where all operations $f^{(i)}, g, h$ are idempotent.

On the other hand, it is easy to see that if in Corollary 2.3 an operation f is idempotent, then also g and h are idempotent. The converse is not true. For example, in an algebra $(Z_4; f, g, h)$, where $f(x, y, z) = (x+y+z)(mod\,4)$ and $g(x, y) = h(x, y) = (2x + 3y)(mod\,4)$, the conditions (a) and (b) are satisfied. Moreover, g and h are idempotent, but a 3-group (Z_4, f) is not idempotent.

We say that an n-group $(G;f)$ is *σ-commutative* if $f(x_{\sigma(1)},...,x_{\sigma(n)})$ is invariant under a permutation $\sigma \in S_n$. An n-group which is σ-commutative for every $\sigma \in S_n$ is called *commutative*. It is not difficult to prove (see [7]) that an n-group is commutative iff it is σ-commutative for some fixed $\sigma = (i, i+1)$. Moreover, this fact together with Hosszú Theorem [11] gives

Lemma 3.5. *An n-group $(G;f)$ is commutative if and only if there exists an element $a \in G$ such that for all $x, y \in G$ and some $2 \le i \le n$ holds*

$$f(\overset{(i-2)}{a}, x, y, \overset{(n-i)}{a}) = f(\overset{(i-2)}{a}, y, x, \overset{(n-i)}{a}).$$

Theorem 3.6. *The class of all n-ary commutative groups $(n > 2)$ may be considered as the variety of algebras with one $(1,2)$-associative n-ary operation f and one unary operation $x \to \hat{x}$ satisfying for some fixed $2 \le i \le n$ and $3 \le j \le n$ the following two identities:*

(a) $f(y, \overset{(i-2)}{x}, \hat{x}, \overset{(n-i)}{x}) = y$,

(b) $f(x, y, \overset{(j-3)}{x}, \hat{x}, \overset{(n-j)}{x}) = y$.

Proof. Since every commutative n-group satisfies these conditions we prove the converse. Let $(G;f)$ be an $(1,2)$-associative n-groupoid satisfying (a) and (b). Since (a) implies $f(\overset{(i-1)}{x}, \hat{x}, \overset{(n-i)}{x}) = x$, then (b) together with the $(1,2)$-associativity gives

$$y = f(x, y, \overset{(j-3)}{x}, \hat{x}, \overset{(n-j)}{x}) = f(f(\overset{(i-1)}{x}, \hat{x}, \overset{(n-i)}{x}), y, \overset{(j-3)}{x}, \hat{x}, \overset{(n-j)}{x}) =$$

$$f(x, f(\overset{(i-2)}{x}, \hat{x}, \overset{(n-i)}{x}, y), \overset{(j-3)}{x}, \hat{x}, \overset{(n-j)}{x}) = f(\overset{(i-2)}{x}, \hat{x}, \overset{(n-i)}{x}, y).$$

Thus by Theorem 2.4 an algebra $(G; f, \hat{\ })$ is an n-group and $\hat{x}$ is the skew element. Therefore (a) and (b) are valid for all $2 \le i \le n$ and $3 \le j \le n$. Moreover,

$$f(x, y, \overset{(n-2)}{a}) = f(f(y, x, \overset{(j-3)}{y}, \hat{y}, \overset{(n-j)}{y}), y, \overset{(n-2)}{a}) =$$

$$f(y, f(x, \overset{(j-3)}{y}, \hat{y}, \overset{(n-j+1)}{y}), \overset{(n-2)}{a}) = f(y, x, \overset{(n-2)}{a})$$

for all $a, x, y \in G$, which by Lemma 3.5 completes the proof.

As a consequence of the above Theorem and Theorem 2.6 we obtain the following characterization of commutative n-groups.

Corollary 3.7. *The class of algebras $(G; f, g)$ with one $(1,2)$-associative n-ary $(n > 2)$ operation f and one $(n-2)$-ary operation g is the variety of commutative n-ary groups if and only if for some fixed $j \in \{1, 2, ..., n-1\}$ the following two identities hold*

(a) $f(y, x_1^{j-1}, g(x_1^{n-2}), x_j^{n-2}) = y$,

(b) $f(, x_1, y, x_2^{j-1}, g(x_1^{n-2}), x_j^{n-2}) = y$.

In the theory of n-semigroups the following identities

$$f(x_1, x_2^{n-1}, x_n) = f(x_n, x_2^{n-1}, x_1)$$

and

$$f(f(x_{11}^{1n}), f(x_{21}^{2n}), ..., f(x_{n1}^{nn})) = f(f(x_{11}^{n1}), f(x_{12}^{n2}), ..., f(x_{1n}^{nn}))$$

play a very important role.

The first of them is called *semi-commutativity* (an n-group with this identity is called, by Dörnte [1], *semiabelian.*) The second of them is a natural generalization of the medial (entropic) law for groupoids. An n-semigroup satisfying this identity is called *medial* or *Abelian* (see [12]) since an n-semigroup $(G; f)$ treated as an algebra $(G; f, f)$ of the type (n, n) is Abelian in the sense of [13] (p. 87).

Each semi-commutative n-semigroup is medial [12], but for every $n \geq 2$ there exist medial n-semigroups which are not semi-commutative [5]. An n-ary group is medial iff it is semi-commutative [12], or equivalently (see [5] and [14]), iff for some fixed $a \in G$ the identity $f(x, \overset{(n-2)}{a}, y) = f(y, \overset{(n-2)}{a}, x)$ is true. Hence the class of all medial n-groups is the variety defined by the last identity, the $(1,2)$-associativity and (6) (or by the $(n-1, n)$-associativity and (5)).

4. Open problems

From the proof of Theorem 3 in [12] follows that any medial n-group satisfies the identity

(20) $\overline{f(x_1^n)} = f(\overline{x}_1, \overline{x}_2, ..., \overline{x}_n)$.

Hence an n-group $(G; f)$ is Abelian as an algebra $(G; f, ^{-})$. Note that (20) holds also in some non-medial n-groups. It holds for example in all idempotent n-groups. Therefore the following problems (announced in [5]) seems to be interesting:

Problem 1. *Describe the variety of all n-groups satisfying* (20).

Let $(G; f, ^{-})$ be an n-group and let $\overline{x}$ be the skew element to x. Moreover, let $\overline{x}^{(0)} = x$ and let $\overline{x}^{(s+1)}$ be the skew element to $\overline{x}^{(s)}$ for $s \geq 0$. In the other words: $\overline{x}^{(1)} = \overline{x}$, $\overline{x}^{(2)} = \overline{\overline{x}}$, $\overline{x}^{(3)} = \overline{\overline{\overline{x}}}$, etc.

Problem 2. *Describe the class of n-groups in which there exists s such that $\overline{x}^{(s)} = \overline{x}^{(t)}$ for all elements and all $t \geq s$.*

Some results connected with this problem are obtained in [15] and [16].

Problem 3. *Describe the class of n-groups in which $\overline{x}^{(s)} \neq \overline{x}^{(t)}$ for all $s \neq t$ and $x \in G$.*

Problem 4. *Describe the variety $\mathbf{V}_s$ of n-groups in which $\overline{x}^{(s)} = x$ for all $x \in G$.*

The class $\mathbf{V}_1$ is the variety of idempotent n-groups. Obviously $\mathbf{V}_1 \subset V_s$ for every natural s. Moreover, $\mathbf{V}_s \cap V_{s+1} = V_1$ and $\mathbf{V}_s \subset V_{st}$ for any natural s, t. Any $\mathbf{V}_s$ contains the variety of medial n-groups (and in the consequence - the varierty of commutative n-groups). Since $\overline{\overline{x}} = x$ for all 3-groups [1], the variety $\mathbf{V}_{2s}$ contains the variety of all 3-groups.

As it is known (see [18]) in some n-ary algebras there exist so-called *splitting automorphisms*, i.e. automorphism ψ satisfying for every $i = 1, 2, ...n$ the condition $\psi(f(x_1^n)) = f(x_1^{i-1}, \psi(x_i), x_{i+1}^n)$. Such automorphisms there exist also in some n-ary groups $(n > 2)$. For example, it is easy to see that $\psi_a(x) = (x+a)(mod\, n)$ is a splitting automorphism of an (n+1)-group $(Z_n; f)$ defined by $f(x_1^{n+1}) = (x_1 + ... + x_{n+1} + b)(mod\, n)$. Moreover, in some n-groups the unary operation $x \longrightarrow \overline{x}$ is a splitting automorphism. Such n-groups are called *distributive*. The class of distributive n-groups forms a variety selected from the variety of all n-groups by the identity

$$(21) \qquad \overline{f(x_1^n)} = f(x_1^{i-1}, \overline{x}_i, x_{i+1}^n),$$

where $i = 1, 2, ..., n$.

Every distributive n-group satisfies (20) and it is a set-theoretic union of disjoint and isomorphic subgroups of the form $\{x, \overline{x}, ..., \overline{x}^{(t-1)}\}$, where t is fixed. Hence a distributive n-group is idempotent or has no any idempotents [17]. Moreover, $\{\phi, \phi^2, \phi^3, ..., \phi^t\}$, where $\phi(x) = \overline{x}$ is an invariant subgroup of the group of all splitting automorphisms.

In every medial distributive n-group $(G; f)$ an operation f is distributive with respect to itself, i.e. the identity

$$f(x_1^{i-1}, f(y_1^n), x_{i+1}^n) = f(f(x_1^{i-1}, y_1, x_{i+1}^n), ..., f(x_1^{i-1}, y_n, x_{i+1}^n)),$$

holds for all $i = 1, 2, ..., n$. Such n-groups, called *autodistributive*, are described in [16] and [5]. The class of autodistributive n-groups $(n > 3)$ is a proper subvariety of the variety of distributive n-groups. For $n = 3$ these varieties are equal; for $n = 2$ are trivial.

Problem 5. *Describe the variety of all n-groups satisfying* (20) *and* (21).

Problem 6. *Describe the class of all n-groups in which there exists at least one non-trivial splitting automorphism.*

References

[1] Dörnte W., *Untersuchungen über einen verallgemeinerten Gruppenbegriff*, Math. Z. **29**(1928), 1-19.

[2] Kasner E., *An extension of the group concept* (reported by L. G. Weld), Bull. Amer. Math. Soc. **10**(1904), 290-291.

[3] Post E. L., *Polyadic groups*, Trans. Amer. Math. Soc. **48**(1940), 208-150.

[4] Dudek W. A., Glazek K., Gleichgewicht B., *A note on the axioms of n-groups*, Coll. Math. Soc. J. Bolyai. 29. Universal Algebra, Esztergom (Hungary), 1977, 195-202.

[5] Dudek W. A., *Medial n-groups and skew elements*, Proceedings of the V Universal Algebra Symposium "Universal and Applied Algebra", Turawa 1988, World Scientific, Singapore 1989, 55-80.

[6] Gleichgewicht B., Glazek K., *Remarks on n-groups as abstract algebras*, Colloq. Math. **17**(1967), 209-219.

[7] Dudek W. A., *Remarks on n-groups*, Demonstratio Math. **13**(1980), 165-181.

[8] Ušan J., *A comment on n-groups*, Zb. Rad. Prirod.-Mat. Fak. Univ. u Novom Sadu, ser. Mat. **24.1**(1994), 281-288.

[9] Rusakov S. A., *On the definition of n-ary groups*, (in Russian), Doklady Akad. Nauk BSSR **23**(1979), 965-967.

[10] Tyutin V.I., *About the axiomatics of n-ary groups*, (in Russian), Doklady Akad. Nauk BSSR **29**(1985), 691-693.

[11] Hosszú M., *On the explicit form of n-group operations*, Publ. Math., Debrecen, **10**(1963), 88-92.

[12] Glazek K., Gleichgewicht B., *Abelian n-groups*, Coll. Math. Soc. J. Bolyai, 29. Universal Algebra, Esztergom (Hungary), 1977, 321-329.

[13] Kuroš A. G., *General algebra*, Lecture notes, 1969-1970, Nauka, Moscow, 1974 (in Russian).

[14] Kolesnikov O. V., *Decomposition of n-groups*, (in Russian), Mat. Issled. **51**(1979), 88-92.

[15] Dudek I. M., Dudek W. A., *On skew elements in n-groups*, Demonstratio Math. **14**(1981), 827-833.

[16] Dudek W. A., *Autodistributive n-groups*, Annales Soc. Math. Polonae, Comm. Math. **23**(1983), 1-11.

[17] Dudek W. A., *On distributive n-ary groups*, Quasigroups and Related Systems (Kishiniev), (in print).

[18] Płonka J., *On splitting automorphisms of algebras*, Bull. Soc. Royale Sci. Liège **42**(1973), 302-306.

[19] Belousov V. D., Stojaković Z., *On infinitary quasigroups*, Publ. l'Inst. Math. (Beograd) **16(30)**(1973), 31-42.

[20] Kolesnikov O. V., *Inverse n-semigroups*, (in Russian), Comment. Math. **21** (1980), 101-108.

[21] Smith J. D. H., *Mal'cev varieties*, Lecture notes in Mathematics, 554, Springer, Berlin, 1976.

[22] Monk J. D., Sioson F. M., *On the general theory of m-groups*, Fundamenta Math. **72** (1971), 233-244.

Wiesław A. Dudek, Institute of Mathematics, Technical University, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

E-mail address: dudek@graf.im.pwr.wroc.pl

FILOMAT (Niš) 9:3 (1995), 675-687
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SEMIGROUPS OF INTEGRAL FUNCTIONS IN VALUED FIELDS

Ghiocel Groza

ABSTRACT. Let K be a valued field and IK[[X]] the commutative algebra of integral functions over K. This paper is devoted to study some semigroups S of $(IK[[X]], \circ)$, where $f \circ g$ is the composite function of $f, g \in IK[[X]]$. In the first section we define a topology $Inv_K S$ on K and we extend to integral functions some notions used for polynomials (see [5] and [6]). Here we study some connections between the subsemigroups $(S, \circ)$ of $(IK[[X]], \circ)$ and the topologies $Inv_K S$ on K. In the second section we study when a particular subset of K is an open set in the topology defined on K by some semigroup of integral functions.

1. Semigroups and topologies

Let K be a field admitting a rank 1 nontrivial valuation $| \, |$(see [2] or [3]), this is a mapping from K into $\in R$ such that for all $x, y \in K$

i) $| x | \geq 0$ and $| x | = 0$ iff $x = 0$;

ii) $| xy | = | x || y |$;

iii) $| x + y | \leq | x | + | y |$;

iv) there exists an element $z \in K \setminus \{0\}$ such that $| z | \neq 1$.

For $x, y \in K$, define $d(x, y) = | x - y |$. Thus (K, d) is a metric space and we can, therefore, introduce the customary topological concepts into such a space in terms of the metric.

A formal power series

$$f(X) = \sum_{k=0}^{\infty} a_k X^k \in K[[X]] \tag{1}$$

is called an integral function over K if for every $x \in K$ the sequence

$$S_n(X) = \sum_{k=0}^{n} a_k X^k \tag{2}$$

is a Cauchy sequence. We denote by $IK[[X]]$ the commutative algebra of integral functions over K. If $f, g \in IK[[X]]$ we consider $f \circ g \in I\hat{K}[[X]]$ the composite function of f and g, where $\hat{K}$ is a completion of K. We consider $(S, \circ)$ a semigroup of integral functions over K and we denote by

$$Inv_K S = \{D \subset K;\ f(D) \subset D, \forall f \in S\}.$$

Obviously, if K is a complete field, then $(IK[[X]], \circ)$ is a semigroup and for every subsemigroup $S \subset IK[[X\}], K \in Inv_K S$.

Proposition 1. *LetKbe a valued field and let $(S, \circ)$ be a semigroup of integral functions over K. If $K \in Inv_K S$, then $Inv_K S$ defines a topology on K such that K is a locally quasi-compact and locally connected topological space. Furthermore for every $a \in K$ there exists $D_a \in Inv_K S$ such that D_a is the smallest open set from $Inv_K S$ which contains a.*

Proof. Suppose that $\{D_i\}, i \in I$ is a family of sets from $Inv_K S$. It is easily to see that

$$\bigcup_{i \in I} D_i \in Inv_K S \text{ and } \bigcap_{i \in I} D_i \in Inv_K S.$$

Thus $Inv_K S$ is a topology on K. If $a \in K$ we consider

$$D_a = \bigcup_{f \in S} \{f(a)\} \bigcup \{a\}.$$

Then $D_a \in Inv_K S$ and D_a is the smallest open set from $Inv_K S$ which contains a. Since D_a is a quasi-compact and connected subspace of K (see [4]) it follows that $(K, Inv_K S)$ is a locally quasi-compact and locally connected topological space. □

Remark 1. If

$$(S_1, \circ), (S_2, \circ)$$

are two semigroups of integral functions over K, then $Inv_K S_1$ is not necessarily different from $Inv_K S_2$. For example we consider

$$K = \in C,\ S_1 = I\in C[[X]] \text{ and } S_2 = \in C[X].$$

Then

$$Inv_K S_1 = Inv_K S_2$$

is the coarsest topology on $\in C$. However for cyclic semigroups we have:

Proposition 2. *Let K be a valued field of characteristic zero and let $(S_1, \circ)$, $(S_2, \circ)$ be two cyclic semigroups of integral functions over K. If $S_1 \neq S_2$, then*

$$Inv_K S_1 \neq Inv_K S_2.$$

Proof. Let f_i be a generator of S_i, $i = 1, 2$. Since the set of zeros from K of an integral function over K is countable (see [1], p. 144 for a non-archimedean valuation), we consider the countable set M of zeros of the integral functions

$$f_1^j(X) - f_2^k(X), j, k \in \in N, j^2 + k^2 \neq 0.$$

There exists then $a \in K \setminus M$ and we denote

$$D_a^i = \bigcup_{k \in \in N} \{f_i^k(a)\} \bigcup \{a\},\ i = 1, 2.$$

Hence it follows that

$$D_a^1 \neq D_a^2 \text{and} Inv_K S_1 \neq Inv_K S_2. \quad \square$$

We now raise the question as to when the topological space $(K, Inv_K S)$ is separable. Since $h(a) \in D_a$, for every $h \in S$, from Proposition 1 it follows immediately:

Proposition 3. *Let K be a valued field and let $(S, \circ)$ be a semigroup of integral functions over K. If $K \in Inv_K S$, the following conditions are equivalent:*

a) *$(K, Inv_K S)$ is a Hausdorff space.*
b) *$S = \{X\}$.*
c) *$Inv_K S$ is the discrete topology on K.*

We recall that the assertion that for every two distinct points at least one of them has a neighbourhood that does not contain the other is called axiom T_0.

Proposition 4. *Let be K a valued field and let $(S, \circ)$ be a semigroup of integral functions over K. If $K \in Inv_K S$, then $(K, Inv_K S)$ is a T_0- (Kolmogoroff) space if and only if for every $a \in K$, either a is a fixed point of S, that is $h(a) = a$, for all $h \in S$, or, if there exists $h_1 \in S$ such that $h_1(a) \neq a$, then $h_2 h_1(a) \neq a$, for all $h_2 \in S$.*

Proof. If $(K, Inv_K S)$ is a T_0-space, then we consider $a \in K$ such that there exists $h_1 \in S$ for which $h_1(a) \neq a$. Suppose there exists $h_2 \in S$, such that

$h_2h_1(a) = a$. By Proposition 1, either $h_1(a) \notin D_a$, or $a \notin D_{h_1(a)}$, which is absurd since $h_1(a) \in D_a$ and $a = h_2h_1(a) \in D_{h_1(a)}$.

Conversly, let $a, a' \in K, a \neq a'$. If a, for example, is a fixed point of S, then $a' \notin D_a = \{a\}$, otherwise suppose $a' \in D_a$ and $a \in D_{a'}$. Hence there exist $h_1, h_2 \in S$ such that $a' = h_1(a)$ and $a = h_2(a')$. Since $h_1(a) \neq a$ it follows that $h_2h_1(a) \neq a$, which is absurd since $a = h_2(a') = h_2h_1(a)$. This shows that $(K, Inv_K S)$ is a Kolmogoroff space. □

Corollary. *Let K be a valued field, $f(X) \in IK[[X]]$ and let $S = (f)$ a cyclic semigroup of integral functions over K. If $K \in Inv_K S$, then $(K, Inv_K S)$ is a Kolmogoroff space if and only if for all $a \in K$ either a is a fixed point of $f(X)$ or for all $k \in \in N, k \geq 2$, there exists an integral functions $g_k(X)$ over K such that $g_k(a) \neq 0$ and $f^k(X) = g_k(X) + X$.*

The proof follows directly from Proposition 4.

Example 1. Suppose that $K_1 = \in R, K_2 = \in C$ and $||$ is the usual archimedean valuation. Let

$$f(X) = e^X + X.$$

If $S = (f)$, then by Corollary it follows that $(\in R, Inv_{\in R}S)$ is a Kolmogoroff space and $(\in C, Inv_{\in C}S)$ is not a Kolmogoroff space.

Remark 2. If $(K, Inv_K S)$ is a Kolmogoroff space, we define a partial ordering $\leq$ on K such that $a \leq a'$ if and only if a belongs to the closure of $\{a'\}$ in $Inv_K S$ (see [4], Ch. 1). Then the open intervals of $(K, \leq)$ form a basis for the topology $Inv_K S$. The assertion follows by Proposition 1 and by [4], Ch. 1.

2. Invariant sets and semigroups

In this section we study the connection between particular subsets of K and particular semigroups of integral functions. We shall use the terminology and notation introduced in Section 1. We shall need the following result from [7].

Theorem 1. *Let K be a complete valued field, $\{x_n\}_{n\geq 1}$ an infinite sequence of distinct elements in K such that*

$$\lim_{n\to\infty} | x_n |= \infty \tag{3}$$

and $\{y_n\}_{n\geq 1}$ an arbitrary infinite sequence of elements in K. Then there exists a function $f(X) \in IK[[X]]$ such that

$$f(x_j) = y_j, \ \forall j \geq 1. \tag{4}$$

Theorem 2. *Let K be a complete valued field and let $M = \{x_n\}_{n \geq 1}$ be a countable subset of K which satisfies (3). Then there exists an infinite cyclic semigroup S of integral functions over K such that $M \in Inv_K S$.*

The proof follows immediately from Theorem 1.

We shall now study some particular cases when K is not necessarily a complete field. We begin with a lemma on a determinant which is a generalization of the Vandermonde determinant.

Lemma 1. *Let K be a field of characteristic zero, $m, n, k \in \in N$ and $m \geq k$. We consider the polynomial $D_{m,n,k}(X_0, X_1, \dots, X_n) \in K[X_0, X_1, \dots, X_n]$ defined by the determinant which has the order $(k+1)(n+1)$, its j-th row, $j = 1, \dots, n+1$ has the form*

$$(X_{j-1}^{m}, X_{j-1}^{m+1}, \dots, X_{j-1}^{m-1+(k+1)(n+1)})$$

and the following rows are their derivatives up to order k inclusive. Then there exists $C \in K \setminus \{0\}$ such that

$$D_{m,n,k}(X_0, X_1, \dots, X_n) = C \prod_{i=0}^{n} X_i^{m(k+1)} \prod_{0 \leq i < j \leq n} (X_j - X_i)^{(k+1)^2} \tag{5}$$

Proof. By induction on k, using Laplace's theorem, it is easily verified that the (total) degree of $D_{m,n,k}$ is

$$\deg D_{m,n,k}(X_0, X_1, \dots, X_n) = \frac{(n+1)(k+1)}{2}(2m + n + kn) \tag{6}$$

We shall denote $D_{m,n,k}$ by D, for simplicity. Let $(X_1 - X_0)^q$ be the highest power of $X_1 - X_0$ which divides D in $K[X_0, X_1, \dots, X_n]$. Then

$$\frac{\partial^j D}{\partial X_0^j}(X_1, X_1, X_2, \dots, X_n) \equiv 0, \ j = 0, 1, 2, \dots, q-1 \tag{7}$$

and

$$\frac{\partial^q D}{\partial X_0^q}(X_1, X_1, X_2, \dots, X_n) \tag{8}$$

is not identically equal to zero. Since the derivative of a determinant Δ of order N is the sum of N determinants Δ_s in which all rows (except the s-th) are the same as in Δ and the s-th row in Δ_s is the derivative of s-th row in Δ, it follows that

$$\frac{\partial D}{\partial X_0}(X_0, X_1, \dots, X_n)$$

is a sum of such determinants in which all rows (excepts the i-th rows, $i = 1, n+2, 2n+3, \dots, kn+k+1$) are the same as in D and i-th rows are the i-th rows in D or a derivative of the i-th rows in D.

On the other hand, by using suitable derivative of D, it follows that D is not identically equal to zero. To obtain (7) it is enough to prove that

$$\frac{\partial^j D}{\partial X_0^j}(X_0, X_1, \dots, X_n) \tag{9}$$

is a sum of such determinants in which there exists a row which is equal to the first row of D or is equal to a derivative up to order k inclusive of the first row of D. If q_1 is the smallest value of j such (9) has not this property, it follows that

$$q \geq q_1 \geq (k+1)^2 \tag{10}$$

Since D is a homogeneous polynomial and it remains unchanged, to within sign, under any transposition of two unknows, it follows that the degree of the product of all the factors $(X_j - X_i)^q$, $j > i$, where $(X_j - X_i)^q$ is the highest power of $X_j - X_i$ which divides D is equal to

$$C_{n+1}^2 q \geq \frac{n(n+1)}{2}(k+1)^2 \tag{11}$$

Similarly, if we denote by X_i^p the highest power of X_i which divides D, it follows that the degree of the product of all the factors

$$X_i^p,\ i = 0, 1, \dots, n, \quad \text{is equal to} \quad (n+1)p \geq (n+1)(k+1)m \tag{12}$$

Since

$$\frac{n(n+1)}{2}(k+1)^2 + (n+1)(k+1)m = \frac{(n+1)(k+1)}{2}(2m+n+kn) = \deg D,$$

by (11) and (12) it follows that $q = (k+1)^2$ and $p = m(k+1)$, which gives the assertion. □

Theorem 3. *Let K be a field of characteristic zero and $|\,|$ a rank 1 nontrivial valuation of K. We denote by $\hat{K}$ a completion of K for its topology defined by $|\,|$. We consider K_1 a countable subset of $\hat{K} \setminus \{0\}$ and K_2 a dense subset of $\hat{K}$. If $\{L_n\}_{n\in\in N}$ is a family of dense subsets of $\hat{K}$, then there exists a function*

$$f(X) = \sum_{n=0}^{\infty} a_n X^n \in I\hat{K}[[X]] \setminus \hat{K}[[X]] \tag{13}$$

such that

a) $a_n \in L_n$ *for all* $n \in \in N$;
b) $f^{(k)}(x) \in K_2$ *for all* $x \in K_1$ *and* $k \in \in N$.

Proof. Let $\{x_i\}_{i\in\in N}$ be the elements of K_1 and we denote by

$$S_n(X) = \sum_{k=0}^{n} a_k X^k \tag{14}$$

We consider the sequences $u_n = 1\cdot 2 + 2\cdot 3 + \ldots + n(n+1) - 1$,

$$v_n = u_n^{-u_n}.$$

Because K_2 is a dense subset of $\hat{K}$ and the polynomials are continuous functions we can find $y_{0,0}, y_{1,0} \in K_2$ such that the system

$$\begin{cases} b_0 + b_1 x_0 = y_{0,0} \\ b_0 + b_1 x_1 = y_{1,0} \end{cases} : \tag{15}$$

has the solutions $b_0, b_1 \in \hat{K}$ with the following property

$$| b_i | < v_2, \; i = 0, 1. \tag{16}$$

With the notations of Lemma 1 we have

$$D_{2,2,1}(x_0, x_1, x_2) \neq 0.$$

Let F_2 be the finite set of the cofactors of the elements in $D_{2,2,1}(x_0, x_1, x_2)$. Since L_0, L_1 are dense subsets in $\hat{K}$, there exist $a_i \in L_i$, $i = 0, 1$, such that

$$\begin{cases} | a_i | < v_2, \\ | S_1(x_j) - y_{j,0} | < v_2 \\ \left| \dfrac{c}{D_{2,2,1}} (S_1(x_j) - y_{j,0}) \right| < \dfrac{1}{2\cdot 3} v_3, \; \forall c \in F_2, \; j = 0, 1. \end{cases} \tag{17}$$

Because K_2 is a dense subset of $\hat{K}$ there exist the elements $y_{2,0}, y_{0,1}, y_{1,1}, y_{2,1} \in K_2$ such that

$$(18) \qquad \begin{cases} \left| \dfrac{c}{D_{2,2,1}}(S_1(x_2) - y_{2,0}) \right| < \dfrac{1}{2 \cdot 3} v_3, \\ \left| \dfrac{c}{D_{2,2,1}}(a_1 - y_{j,1}) \right| < \dfrac{1}{2 \cdot 3} v_3, \ \forall c \in F_2, \ \ j = 0, 1, 2. \end{cases}$$

Applying Cramer's rule it follows that the system

$$(19) \qquad \begin{cases} a_0 + a_1 x_0 + b_2 x_0^2 + b_3 x_0^3 + b_4 x_0^4 + b_5 x_0^5 + b_6 x_0^6 + b_7 x_0^7 = y_{0,0} \\ a_0 + a_1 x_1 + b_2 x_1^2 + b_3 x_1^3 + b_4 x_1^4 + b_5 x_1^5 + b_6 x_1^6 + b_7 x_1^7 = y_{1,0} \\ a_0 + a_1 x_2 + b_2 x_2^2 + b_3 x_2^3 + b_4 x_2^4 + b_5 x_2^5 + b_6 x_2^6 + b_7 x_2^7 = y_{2,0} \\ a_1 + 2b_2 x_0 + 3b_3 x_0^2 + 4b_4 x_0^3 + 5b_5 x_0^4 + 6b_6 x_0^5 + 7b_7 x_0^6 = y_{0,1} \\ a_1 + 2b_2 x_1 + 3b_3 x_1^2 + 4b_4 x_1^3 + 5b_5 x_1^4 + 6b_6 x_1^5 + 7b_7 x_1^6 = y_{1,1} \\ a_1 + 2b_2 x_2 + 3b_3 x_2^2 + 4b_4 x_2^3 + 5b_5 x_2^4 + 6b_6 x_2^5 + 7b_7 x_2^6 = y_{2,1} \end{cases}$$

in the unknowns b_i, has solutions with the following property

$$(20) \qquad | b_i | < v_3, \ i = u_1 + 1, \dots , u_2.$$

We now consider

$$D_{8,3,2}(x_0, x_1, x_2, x_3) \neq 0$$

and we denote by F_3 the set of the cofactors of the elements in $D_{8,3,2}$. Since L_i, $i = u_1 + 1, \dots , u_2$, are dense subsets in $\hat{K}$, by (19) and (20) it follows that there exist $a_i \in L_i$, $i = u_1 + 1, \dots , u_2$, such that

$$(21) \qquad \begin{cases} | a_i | < v_3, \ i = u_1 + 1, \dots , u_2, \\ | S_{u_2}(x_j) - y_{j,0} | < v_3, \\ | S'_{u_2}(x_j) - y_{j,1} | < v_3, \\ \left| \dfrac{c}{D_{8,3,2}}(S_{u_2}(x_j) - y_{j,0}) \right| < \dfrac{1}{3 \cdot 4} v_4, \\ \left| \dfrac{c}{D_{8,3,2}}(S'_{u_2}(x_j) - y_{j,1}) \right| < \dfrac{1}{3 \cdot 4} v_4, \ \forall c \in F_3, \ j = 0, 1, 2. \end{cases}$$

Now by induction on r, we consider

$$D_{u_r+1, r+1, r}(x_0, x_1, \dots , x_{r+1}) \neq 0$$

and we denote by F_{r+1} the set of the cofactors of its elements.

We suppose that we have found $y_{j,k}\in K_2$, $j=0,1,\dots,r, k=0,1,\dots,r-$ 1 and $a_i\in L_i$, $i=0,1,\dots,u_r$, such that

$$\text{(22)}\qquad |\, a_i \,| < v_{t+1},\ i=u_{t-1}+1,\dots,u_t,\ \forall t=1,\dots,r$$

$$\text{(23)}\qquad |\, S^{(k)}_{u_r}(x_j)-y_{j,k} \,| < v_{r+1},\ \forall j=0,\dots,r,\ k=0,\dots,r-1$$

$$\text{(24)}\qquad \left|\frac{c}{D_{u_r+1,r+1,r}}(S^{(k)}_{u_r}(x_j)-y_{j,k})\right| < \frac{1}{(r+1)(r+2)}v_{r+2},$$
$$\forall c\in F_{r+1}, j=0,..,r,\ k=0,\dots,r-1.$$

Since K_2 is a dense subset in $\hat{K}$ there exist the elements $y_{r+1,0}, y_{r+1,1},\dots,y_{r+1,r}, y_{0,r},\dots,y_{r+1,r}\in K_2$ such that the condition (24) hold true for all $j=0.1,\dots,r+1$ and $k=0,1,\dots,r$. Then the system

$$\text{(25)}\qquad S^{(k)}_{u_r}(x_j)+(b_{u_r+1}X^{u_r+1}+\dots+b_{u_{r+1}}X^{u_{r+1}})^{(k)}_{X=x_j}=y_{j,k},$$
$$0\le j\le r+1,\ 0\le k\le r$$

in the unknowns b_i, which for $r=1$ coincides with the system (19), has the solutions b_i with the following property

$$\text{(26)}\qquad |\, b_i \,| < v_{r+2},\ i=u_r+1,\dots,u_{r+1}$$

Since $L_i, i=u_r+1,\dots,u_{r+1}$ are dense subsets in $\hat{K}$, by (25) and (26) it follows that there exist $a_i\in L_i$ such that the conditions (22) - (24) are satisfied for $r+1$. This proves (22) - (24) for every r.

We consider now $n\in\in N$. Then there exists $r\in N$ such that

$$u_r < n \le u_{r+1}$$

and by (22) it follows that

$$|\, a_n \,|^{\frac{1}{n}} < v_{r+2}^{\frac{1}{n}} < u_{r+2}^{-u_{r+2}/n} < \frac{1}{u_{r+2}}.$$

Hence

$$\lim_{n\to\infty} |\, a_n \,|^{\frac{1}{n}} = 0$$

and

$$f(X)=\sum_{n=0}^{\infty} a_n X^n \in I\hat{K}[[X]].$$

We remark that we can find $a_n\neq 0$. To prove b) we consider $k,j\in\in N$ and we chose $r>k$ and $r>j$. Then by (23) it follows that

$$f^{(k)}(x_j)=y_{j,k}\in K_2$$

and this establishes the theorem. □

Corollary. *Let K be a countable field of characteristic zero and $||$ a rank 1 nontrivial valuation of K. We denote by*

$$S_K = \{f(X) \in IK[[X]], f^{(k)}(x) \in K, forallx \in K andk \in \in N\}. \tag{27}$$

Then S_K is a semigroup which contains some integral functions which are not polynomials.

The assertion follows from Theorem 3 by taking $K_2 = L_n = K$ for all $n \in \in N$ and $K_1 = K \setminus \{0\}$.

Let

$$S_\infty = \{S_1; \exists f(X) \in IK[[X]] \setminus K[X],\ f \in S_1\},$$

where $(S_1, \circ)$ is a subsemigroup of $(S_K, \circ)$. In the last part of this paper we shall prove that we can find an infinite subset D of $\in Q$ such that, for all $S_1 \in S_\infty$, the topology $Inv_{\in Q}S_1$ does not contain the set D. More precise we have the following assertion:

Theorem 4. *Suppose $K = \in Q$ and $||$ is the usual absolute value function. Let $D = \{1/n\}_{n \in \in N^*}$ and let $f(X) \in S_{\in Q}$ such that*

$$f(D) \subset D \tag{28}$$

then $f(X)$ is a polynomial which is of the form

$$f(X) = \frac{1}{r} X^s,\ r \in \in N^*,\ s \in \in N \tag{29}$$

Proof. If

$$f(X) = \sum_{j=0}^{\infty} a_j X^j,\ a_j = \frac{\alpha_j}{\beta_j},\ \ \alpha_j, \beta_j \in \in Z,\ \beta_j \neq 0 \tag{30}$$

we may assume that $\beta_j > 0$ and $\beta_j \mid \beta_{j+1}$ for all $j \in \in N$. We denote

$$f(\frac{1}{n}) = \frac{1}{k_n},\ k_n \in \in N^*. \tag{31}$$

Since $f(X)$ is a continuous function it follows that

$$\lim_{n\to\infty} f(\frac{1}{n}) = a_0 = \lim_{n\to\infty} \frac{1}{k_n}. \tag{32}$$

We may assume that $f(X) \notin K$ and because the zeros of an integral function which does not vanish identically are isolated, it follows that

$$\lim_{n\to\infty} \frac{1}{k_n} = 0 = a_0.$$

Let a_i be the first coefficient which is not equal to zero. Since $f(X)$ is an integral function we have

$$\lim_{m\to\infty} \mid a_m \mid^{\frac{1}{m}} = 0$$

and then there exists $m_0 \in \in N, m_0 \geq i$ such that for all $m \geq m_0$

$$\mid f(x) - \sum_{j=1}^{m} a_j x^j \mid \leq x^{m+1}, \forall x \in [0,1]. \tag{33}$$

By (31) and (33) it follows that, for all $m \geq m_0$ and $n \in \in N^*$,

$$\left| \frac{1}{k_n} - \left(\frac{\alpha_i}{\beta_i} \cdot \frac{1}{n^i} + \ldots + \frac{\alpha_m}{\beta_m} \cdot \frac{1}{n^m}\right) \right| \leq \frac{1}{n^{m+1}}. \tag{34}$$

Hence

$$\lim_{n\to\infty} \frac{n^i}{k_n} = \frac{\alpha_i}{\beta_i} \tag{35}$$

and for all $n \in \in N^*$ and $m \geq m_0$

$$\left| \beta_m n^m - k_n(\alpha_i \beta_m \beta_i^{-1} n^{m-i} + \ldots + \alpha_m) \right| \leq \beta_m \frac{k_n}{n}. \tag{36}$$

Suppose that there exists a fixed

$$m \geq m_0,\ m \geq 2i\, suchthatalpha_m \neq 0. \tag{37}$$

Then by (35) and (36) there exists $r_m \in \in Z$ such that for all $n \in \in N^*$

$$\beta_m n^m - k_n(\alpha_i \beta_m \beta_i^{-1} n^{m-i} + \ldots + \alpha_m) - r_m = 0, \tag{38}$$

where $r_m = O(n^{i-1})$. We consider the polynomials

$$P_1(X) = \beta_m X^m,$$
$$P_2(X) = \alpha_i \beta_m \beta_i^{-1} X^{m-i} + \ldots + \alpha_m.$$

Then there exist $R_1(X), Q_1(X) \in \in Q[X]$ such that

$$P_1(X) = Q_1(X)P_2(X) + R_1(X), \tag{39}$$

where $degR_1(X) < m - i$ and $degQ_1(X) = i$. By (38) and (39) it follows that

$$k_n = Q_1(n) + \frac{R_1(n) - r_m}{P_2(n)} = 0. \tag{40}$$

Since

$$\lim_{n\to\infty} \frac{R_1(n) - r_m}{P_2(n)} = 0$$

there exists $n_0 \in \in N^*$ such that

$$\left| \frac{R_1(n) - r_m}{P_2(n)} \right| < \frac{1}{d+1}, \ \forall n \geq n_0, \tag{41}$$

where d is the least common multiple of the denominators of the coefficients of $Q_1(X)$. Because $k_n \in \in N$, by (40), it follows that there exists $n_1 \in \in N^*$ such that

$$k_n = Q_1(n), \forall n \geq n_1.$$

Hence

$$f(\frac{1}{n}) = \frac{1}{Q_1(n)} = \frac{n^{-i}}{Q_2(n^{-1})}, \tag{42}$$

where

$$Q_2(X) = X^i Q_1(\frac{1}{X}).$$

Since D has a limit point, by (42), it follows that

$$f(X) = \frac{X^i}{Q_2(X)}.$$

Since also $f(X)$ is an integral function we must have $Q_2(X) \in \in Q$ and $i = 0$. Then there exists $m \in \in N$ such that $\alpha_m \neq 0$ and for all $m_1 > m$, $\alpha_{m_1} = 0$. Thus $f(X)$ is a polynomial and by (34)

$$\mid \beta_m n^m - k_n(\alpha_i \beta_m \beta_i^{-1} n^{m-i} + \ldots + \alpha_m) \mid < \frac{\beta_m}{n}.$$

Hence there exists $n_2 \in \in N^*$ such that for all $n \geq n_2$

$$\beta_m n^m = k_n(\alpha_i \beta_m \beta_i^{-1} n^{m-i} + \ldots + \alpha_m). \tag{43}$$

We denote $(n, \alpha_m) = d_m$ and $n = d_n v_n$. Then $\lim v_n = \infty$ and by (43) $v_n^m \mid k_n$. Hence, if $m > i$, then

$$\lim_{n \to \infty} \frac{k_n}{n^i} = \infty,$$

which is absurd. Then $m = i$ and

$$f(X) = \frac{\alpha_i}{\beta_i} X^i.$$

Hence by (28) it follows (29). □

References

[1] Y.Amice, *Les nombres p-adiques*, Presses Universitaires de France, 1975.

[2] E. Artin, *Algebraic numbers and algebraic functions*, Gordon and Breach, 1967.

[3] G. Bachman, *Introduction to p-adic number and valuation theory*, Academic Press, New-York-London, 1964.

[4] N. Bourbaki, *Topologie générale*, Chap. 1 et 2, 3rd ed., Hermann, Paris, 1961.

[5] P. J. Cahen, *Integer-valued polynomials on a subset*, Proc. Amer. Math. Soc. **117** (1993), 919–929.

[6] J. L. Chabert, *Sur le théorème de Stone-Weierstrass en algèbre commutative*, Rend. Circ. Mat. Palermo, serie II XVII (1994), 51–70.

[7] G. Groza, *Construction of integral functions in valued fields* (to appear).

Department of Mathematics, Technical University of Civil Engineering,
Lacul Tei 124, Sec. 2, R-72302, Bucharest, Romania

FILOMAT (Niš) 9:3 (1995), 689–697
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ТРИДЦАТЬ ЛЕТ МЕТОДА РЕЗОЛЮЦИИ

Петар Хотомски

РЕЗЮМЕ. Статья посвещена отмечению 30–ти летия метода резолюции, обоснованного в статье Робинсона 1965. года. В суммарном очерке приведены ключевые результаты (правило резолю ции, теорема об резолюцие, процедура опровержения) с указанием интенсивного развития рестриктивных форм, техник и стратегий , а также и расширений этого метода, включая логическую систему языка Пролог. Отдельно приведенны наши усилия по развитию метода и его приложений . Кроме описания результатов в рамках развития системы "GRAPH", приведенны сведенья об системе "А-ДТ с вариябильными стратегиями поиска" и об системе "DEDUC" для автоматического порождения комбинаторных расположений . Приведен обзор опубликованных статьей и других наших трудов связанных с методом резолюции. В заключении отмечено что системы с резолюцией могут оказатся полезным в качестве "интеллигентного мотора" для решения задач в различных проблемных средах, особенно тех которые подаются последовательности несложных дедукций. Таким образом возможно построить и различные непроцедурные языки программирования и преодолеть известные помехи присущие в Прологе.

1. Введение

В январе 1995. года совершилось 30 лет со дня опубликования статьи Робинсона [10]. В цей обоснованна новая логическая система значительная не только для теоретических расуждений в рамках математики, но и для практических приложений в других областях. В течении 30–ти лет метод резолюции добился не только к математикам, инженерам, студентам, но даже и к другим пользователям, превасходно посредством языка Пролог.

Поэтому в настоящей статье мы не будем заниматся хронологическим докладом результатов и исторических данных. Мы ограничимся суммарным очерком ключевых результатов с указанием наших усилий по развитию этого метода и его приложений .

2. Суммарный очерк развития метода резолюции и его приложений

Теоретическим окружением в котором развился метод резолюции является автоматическое доказательство теорем на языке исчисления предикатов первого порядка. По отношению к состоянию этой проблематики,ќоторое описанно в фундаментальной статье Ван Хао [4] со 1960. года и статье [5] со 1965. года, существенный прогресс осуществлен в статье Робинсона [10] путем создания логической системы первого порядка не потребующей логических аксиом и обоснованной на только одном правиле вывода (принциппе резолюции). Логическая корректность и полнота этой системы обоснованны на существующих результатах Эрбрана.

Ключевыми результатами являются : правило резолюции, теорема об резолюцие и процедура опровержения . Их можно сформулировать следующим образом.

Правило резолюции:

Из дизьюнктов D_1 и D_2 не имеющих общих переменных, (это можно всегда получить переименованием переменных), при условии существования наиболее общего унификатора Θ для непустых множеств литерь $L_1 \subseteq D_1$ и $L_2 \subseteq D_2$, причем $L_1\Theta$ и $L_2\Theta$ являются дополнительными по отношению к отрицанию , выводим дизьюнкт — резольвента:

$$(D_1 \backslash L_1)\Theta \cup (D_2 \backslash L_2)\Theta \ .$$

Специально: из дизьюнктов: $\neg A \vee C_1$, $B \vee C_2$ выводится резольвента: $C_1\Theta \vee C_2\Theta$, причем $A\Theta$ совпадает с $B\Theta$.

Таким образом правило резолюции обьединяет процесс подстановки переменных с процессом логического вывода в исчислении высказываний. Наиболее общий унификатор находится при помощи алгоритма унификации.

Теорема об резолюцие:

Пусть $R(S)$ обозначает обьединение множества S с множеством всех резольвент выводимых из дизьюнктов множества S, $R_0(S) = S$ и $R_n(S) = R(R_{n-1}(S))$, $n > 0$. Тогда справедлива следующая теорема:

Конечное множество дизьюнктов S невыполнимо тогда и только тогда, когда $R_n(S)$ содержит для некоторого $n > 0$ пустой дизьюнкт.

Опровержением исходного множества дизьюнктов S называется последовательность дизьюнктов $B_1, B_2, ..., B_k$, такая что для каждого члена B_i , $1 \leq i \leq k$, справедливо:

$B_i \in S$, либо B_i является резольвентой некоторых предшествующих членов, и B_k пустой дизьюнкт.

Из теоремы об резолюцие следует:

Конечное множество дизьюнктов S невыполнимо тогда и только тогда, когда существует опровержение множества S.

Поэтому теорема об резолюцие установляет полноту этой логической системы. С целью повысшения эффективности процедуры опровержения в [10] предложенны первые стратегии поиска (search principles).

Логическая простота и открытые перспективы привлекли большое число исследователей по фундаментальным и аппликативным вопросам. Выработанны различные рестриктивные формы резолюции и оссобые техники для повысшения эффективности процедур поиска, такие как: семантическая резолюция , гиперрезолюция , упорядоченая резолюция , линейная резолюция , OL-резолюция с маркированными литерами и другие. Кроме синтаксических стратегий поиска (в ширину либо в глубину) развиты различные эвристические стратегии. С целью повысшения их эффективности добавленыны оссобые правила которые учитывают оссобенности отдельных проблемных областей (комутативность , асоциативность и т.п.)

Кроме рестриктивных форм исследованны и расширения метода резолюции на случай когда исходное множество не является конечным и содержит схэмы аксиом, такие как подстановочность равенства либо математической индукции. Метод резолюции дополнен правилом парамодуляции для теорий с равенством и правилом индукции для теорий с математической индукцией . Эти правила позволяют элиминировать схэмы аксиом из исходного множества дизьюнктов. Сведенья об этом этапе развития можно найти в [18], где приведен список превысшающий 300 библиографических единиц.

В рамках эвристического программирования осуществленны различные программные системы которые ориентированны на различные теоретические или практические приложения . Из сферы чистой математики, т.е. доказательства теорем в различных математических теориях, открыты приложения этого метода в других областьях, вне математики.

Оказалось что метод резолюции применим к широкому спектру проблем которые выразимы на языке исчисления предикатов, такие как: вопросно-ответные системы, ситуационное управление и принятие решений , порождение комбинаторных расположений , автоматическое генерырование программ и определение их корректности, логическое программирование и непроцедурные языки, базы данных и т.д. Одна из модификаций резолюционной процедуры опровержения встроенна в темель языка программирования Пролог. Этим, а также и другими полезными практическими приложениями, уменьшенно разочарение которое произошло в последствии сверхмерного начального оптимизма с начала 70-тых, с одной стороны и скромного эффекта построенных программных систем, с другой стороны.

Полученные теоретические результаты и практические эксперименты создали возможность с больше реализма оценивать положение и роль метода резолюции в усилиях автоматизации логического умозаключения и расуждения , оссобенно в областьях математической деятельности. Ожиданная автономность и мощность программных систем с резолюцией уступает место новой роли консультанта-асистента человеку исследователю в процессе решения сложных проблем. При этом резолюционный подход вкладывается в другие нерезолюционные, интерактивные системы вывода.

3. Обзор наших результатов

Обзор наших усилий начнем с очерком об первой более комплексной системе доказательства теорем которая построенна в рамках экспертной системы GRAPH, [30] на Электротехническом факултете в Белграде 1980–1985. года.

Модуль THEOR в составе системы GRAPH содержит эвристический доказатель теорем с естественным выводом [29] и доказатель с резолюцией и индукцией [14]. Первый из их включает интерактивный режим работы, а другой является вполне самостоятельной автоматической процедурой поиска доказательства без вмешательства человека. Сопряжение этих доказателей осуществленно таким образом что цель малой комплексности отсылается из интерактивной части на доказательство в часть с резолюцией . Система с резолюцией и индукцией обстоит на OL-резолюцие с маркированными литерами и на оригинальных результатах в связи расширения метода резолюции на теории первого порядка с математической индукцией , которые описанны в [11], [12], [13].

С 1986. года на Техническом факултете в Зренянине продолженно совершенствование этой системы. Разработанна и встроенна в систему новая эвристическая техника автоматического упорядочения дизьюнктов в исходном множестве, при помощи вычисления их весов, [7], [8].

Система с резолюцией и индукцией в рамках системы GRAPH обоснованна на стратегии поиска в ширину. Повысшение эффективности возможно при помощи использования различных стратегий . На этой почве построенна система АДТ которая описанна в [2], [24], [26]. Первые эксперименты на этой системе описанны в [3]. Система АДТ с вариабильными стратегиями поиска является частью большей системы АДТ которая на факультете в Зренянине разработанна с 1990. года. В самом деле, система доказательства дополненна экспертно-обучающей системой , которая развивалась в сотрудничестве с институтом кибернетики из Киева в эдукативном направлении, [19], [20]. Обе части являются независимыми и имеют самостоятельное значение в научно-исследовательском, обучающем и практическом смысле.

Одно из практических приложений системы АДТ осуществленно в области порождения комбинаторных расположений , [21], [22]. Разработанна и построенна на PC компьютере система DEDUC для автоматического порождения распоряжения уроков, которая использует систему АДТ в качестве дедуктивного механизма [9], [25], [26]. Достоинство системы DEDUC состоит в возможности менять исходные условия без неопходимости перемен в процедурной части системы.

Дальнейшие исследования , которые сейчас проводятся, ориентированны на развитие непроцедурных языков программирования , при чем используется АДТ система.

Обзор наших активностей дополним следующими статьями, которые не упомянуты высше, а которые связанны с методом резолюции. Некоторые детали об системе с резолюцией и индукцией обсужденны в трудах [15], [6]. В первом из их описанна техника выделения только релевантных шагов опровержения из множества всех шагов которые осуществленны в процессе поиска доказательства. В другом описанны эксперименты показывающие возможность получить различные доказательства одной и той же теоремы.

В [17] представлен короткий обзор развития автоматического доказательства теорем, приведенны сведенья об системе GRAPH, примери доказательств и возможности дальнейшего совершенствования системы с резолюцией . Это подстрекнуло дальнейшие исследования, опубликованные в статье [1]. Осуществленно элиминирование идентичных резольвент которые могут быть порожденны на соседних уровнях поиска, а также внесенны некоторые перемены в связи с применением правила симметрии.

В [16] описанна возможность использования системы с резолюцией для решения задач стабилизирования и трансформирования технологических процессов.

В периоде 1983–1988. года проведенно несколько обзорных докладов и семинаров посвященных методу резолюции, которые припомогли увеличению числа исследователей .

Теоретические знания и практический опыт отразились и на обучение студентов. Метод фундаментальной резолюции вошел в состав курса "Математическая логика и принципы программирования " (как один из методов исследования тавтологий) на 1. курсе обучения профессоров информатики, [23]. Дополнительное знание об методе резолюции эти студенты получают на 3. курсе в рамках предмета "Системы искусственного интеллекта", [28]. На упражнениях используются системы АДТ и DEDUC на PC компьютерах (раньше использованна система GRAPH на VAX компьютере).

На аспирантном уровне "Информатика в образовании" выработанно несколько семинарных работ посвященных приложениям метода резолюции.

Наконец, приведены и некоторые числовые указатели наших усилий по расширению метода резолюции:

- опубликованные труды: 24
- обзорные доклады: 16
- монографии и учебники: 2 + 2
- кандидатские тезисы (магистратура): 3
- дисертации (доктораты): 1 + 1 (в работе).

Почти все из приведенных исследований были осуществленны в рамках проектов, либо отдельных тем (4 проекта, 6 тем) с фонансовой поддержкой научных учреждений Республики Сербии или Воеводины.

4. Заключение

Логически полная система, обоснованная на принципе резолюции, интенсивно развивалась в рамках теории и практики автоматического доказательства теорем и приобрела форму метода резолюции — одного из возможных методов автоматического воспроизведения логического вывода на компьютере.

Недостатки этого подхода, такие как:

- непрыродное представление в форме дизьюнктов,
- экспоненциальный взрыв пространства поиска,
- невозможность совершить доказательства сложных математических теорем в реальных ресурсах,

в определеной мере обескуражили использовать этот метод в различных проблемных областях. Ослабление этих недостатков осуществленно дальнейшими модификациями исходного метода. Так построенна дедуктивная система Пролог-языка которая использует импликативную форму представления и содержит развитые техники сужения пространства поиска, но и некоторые недостатки. Помимо известных недостатков, можно утверждать что именно программирование в Прологе внесло большой вклад в расширение приложений метода резолюции к различным областям.

Наши анализы и опыт использования систем с резолюцией позволяют сделать следующие выводы:

- В тех случаях, когда внутренее представление скрыто от пользователя , нет неопходимости устранять форму дизьюнктов.
- Существует значительное число задач, для решения которых ”глубокие” расуждения можно привести к последовательности простых дедукций , с которыми система АДТ справляется совсем эффективно и практически удоволетворимо.
- Систему АДТ с резолюцией можно использовать в качестве ”интеллигентного мотора” в различных проблемных средах с целью решения определенных задач.
- Используя систему АДТ с резолюцией как дедуктивный базис, возможно построить различные непроцедурные языки программирования и преодолеть некоторые недостатки Пролога.

Литература

[1] Берковић И., Хотомски П., Радуловић Б., *Имплементација неких побољшања резолуцијског доказивача теорема*, Зборник симпозијума о информационим технологијама ”Сарајево-Јахорина”, ЕТФ Сарајево (1991), 227.1 -227.8.

[2] Берковић И., *Варијабилне стратегије претраживања у наставно оријентисаном систему за аутоматско доказивање теорема*, Магистарски рад, Технички факултет ”М. Пупин”, Зрењанин (1994).

[3] Berkovič I., *Some experiments with the system for automatic theorem proving*, Workshop on information technologies. Novi Sad (1994), 1-8 (to appear).

[4] Wang H., *Toward Mechanical Mathematics*, IBM J. Res.Devel. **1** (1960), 2-22.

[5] Wang H., *Formalization and Axiomatic Theorem Proving*, Proc. IFIP Congr. 1965 **1** (1965), Spartan Books, Washington.

[6] Kujačić M., Hotomski P., *Different computer proofs of the same theorem*, Zbornik XI međunarodnog simpozija "Kompjuter na sveučilištu", Cavtat 1989, SRCE Zagreb (1989), 8.14.1-8.14.6.

[7] Кујачић М., *Могућности за побољшање доказивача теорема у систему "Граф"*, Зборник симпозијума "Остварења и примене вештачке интелигенције", Дубровник 1989, Технички факултет "М. Пупин" Зрењанин (1990), 67-76.

[8] Кујачић М., *Усавршавање технике резолуцијског доказивања теорема са концепцијом примене у образовању*, Магистарски рад, Технички факултет "М. Пупин" Зрењанин (1991).

[9] Прохаска Д., *Дедуктивна метода генерисања комбинаторних распореда и њена програмска алгоритмизација*, Магистарски рад, Технички факултет "М. Пупин", Зрењанин (1994).

[10] Robinson J. A., *A Machine-Oriented Logic Based on the Resolution principle*, J. ACM **12** (1965), no. 1, 23-41.

[11] Хотомски П., *Правило индукции в доказательствах опровержением с применением к автоматическому доказательству теорем*, Publication de l'Institut mathematique, Belgrade **31(45)** (1982), 51-63.

[12] Хотомски П., *Способ встроения правила индукции в процедуры автоматического доказательства теорем с резолюцией*, Publication de l'Institut mathematique, Belgrade **33(47)** (1983), 89-95.

[13] Hotomski P.Z., *An induction law in proofs by contradiction with an application to automatic theorem proving (autoreferat)*, Zentralblatt für Mathematik **521.03006** (1984), 19-20.

[14] Хотомски П., *Система автоматического доказательства теорем с резолюцией, индукцией и симметрией*, Proc. of the conf. algebra and logic, Institute of mathematics, Novi Sad (1985), 55-61.

[15] Hotomski P., *Determination of the proof tree in an automated theorem proving system*, Zbornik X međ unarodnog simpozija "Kompjuter na sveučilištu", Cavtat 1988, SRCE Zagreb (1988), 8.10.1-8.10.8.

[16] Хотомски П., Кујачић М., *Логичке методе у аутоматском управљању*, Зборник II симпозијума ИЕС у процесној индустрији, САУМ-ЗЗЕЕ, Београд (1988), 3.1-3.11.

[17] Хотомски П., *Аутоматско доказивање теорема — резултати и перспективе*, Зборник радова техничког факултета "М. Пупин", Зрењанин (1990), 66-73.

[18] Хотомски П., Певац И., *Математички и програмски проблеми вештачке интелигенције у области аутоматског доказивања теорема*, Научна књига, Београд, 1988 (друго издање 1991).

[19] Хотомски П., Кујачић М., Кудрявцева С., *Експертно—обучавајући систем АДТ за учење о аутоматском доказивању теорема*, Зборник "Информатика у образовању и нове технологије", Нови Сад 1991., Технички факултет "М. Пупин", Зрењанин (1991), 39.1- 39.8.

[20] Хотомски П., *Учење методе резолуције помоћу аутоматизованог курса*, Зборник "Информатика у образовању", Технички факултет "М. Пупин", Зрењанин (1992), 71.1-71.8.

[21] Хотомски П., *Дедуктивный подход к автоматическому порождению комбинаторных расположений*, Proc. of the VI Conf. on Logic and Comp. Science LIRA 92, Novi Sad (1992), 35-42.

[22] Хотомски П., *Креирање комбинаторних распореда методом аутоматске дедукције*, Зборник радова Техничког факултета "М. Пупин", Зрењанин, **3/92** (1993), 1.1-1.11.

[23] Хотомски П., Кујачић М., *Математичка логика и принципи програмирања*, Технички факултет "М. Пупин", Зрењанин, 1992.

[24] Хотомски П., Берковић И., *Имплементација система АДТ за аутоматско доказивање теорема са варијабилним стратегијама*, Зборник "Информатика у образовању". Технички факултет "М. Пупин" , Зрењанин (1994), 101-109.

[25] Хотомски П., Прохаска Д., *Имплементација система* DEDUC *за аутоматско генерисање комбинаторних распореда*, Зборник "Информатика у образовању", Технички факултет "М. Пупин" , Зрењанин (1994), 110-118.

[26] Хотомски П., *Интелигентни програмски системи у области аутоматског резоновања*, "Информатика и друштвене промене", Технички факултет "М. Пупин", Зрењанин (1994), 1-17 (to appear).

[27] Hotomski P., *Automated reasoning in Resolution based systems*, Workshop on information technologies. University of Novi Sad and Aristotle University of Thessaloniki (1994), 1-8 (to appear).

[28] Хотомски П., *Системи вештачке интелигенције*, Технички факултет "М. Пупин", Зрењанин, 1995.

[29] Cvetkovič D., Pevac I., *Man-machine theorem proving in Graph theory*, Artificial intelligence vol 35 (1988).

[30] Цветковић Д. и други, *Десет година развоја и примене експертног система "Граф"*, Зборник симпозијума "Остварења и примене вештачке интелигенције", Дубровник 1989 (1990), Технички факултет "М. Пупин", Зрењанин, 25-46.

Технички факултет "М. Пупин", Зрењанин, Ђ. Ђаковића б.б. Југославија

FILOMAT (Niš) 9:3 (1995), 699–710
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ЕЩЕ О ПРОБЛЕМЕ ВКЛЮЧЕНИЯ РЕГУЛЯРНЫХ И ЛИНЕЙНЫХ ЯЗЫКОВ В ГРУППОВЫЕ ЯЗЫКИ

Красимир Янков Йорджев

1. Введение

Пусть G - группа с множеством образующих

$$X = \Sigma \cup \Sigma^{-1} = \{x_1, x_2, ..., x_n, x_1^{-1}, x_2^{-1}, ..., x_n^{-1}\},$$

определяющими соотношениями θ, единицей e и с разрешимой проблемой равенства слов. Тогда множество слов $\mathcal{M} = \{\omega = x_{i_1}^{k_1} x_{i_2}^{k_2} \cdots x_{i_t}^{k_t} | \omega = e\} \subseteq \Sigma^*$ называется групповым языком, задающим группу G. Σ^* - это свободный моноид над Σ. Группа G задается контекстно-свободным языком, если соответствующий групповой язык $\mathcal{M}$ является контекстно-свободным. Группа G в этом случае называется контекстно-свободной группой. Ряд свойств групповых языков рассмотрены в [1, 2, 3, 4].

В [12] показано, что проблема включения и эквивалентности конечно-автоматных отображений алгоритмически разрешима. Этот результат стимулирует попытки найти полиноминальный алгоритм проверяющий однозначности конечно-автоматных отображений. Такие алгоритмы найдены в [10, 12, 13]. Раньше до выхода этих результатов А. В. Анисимов [3], решает более общую задачу. Он доказывает, что проблема определения однозначности конечно-автоматных отображений являются частным случаем проблемы включения контекстно-свободных языков в групповые языки. В работе [3] А. В. Анисимов предлагает алгоритм для проверки включения произвольного контекстно-свободного языка L в групповой язык $\mathcal{M}$ группы с разрешимой проблемой равенства слов. В предлагаемом алгоритме находится конечное множество W_1 тако е, что $W_1 \subseteq \mathcal{M}$ тогда и т олько тогда, когда $L \subseteq \mathcal{M}$. Более конкретно доказывается следующая теорема.

Теорема 1. *Пусть $\Gamma = (N, \Sigma, \Pi)$ контекстно-свободная грамматика, порождающая контекстно-свободный язык L, а $\mathcal{M}$ - групповой язык группы G с разрешимой проблемой равенства слов. Пусть Ω_1 - множество всех слов* *из* L *длиной меньше или равной* p , *а* $\Omega_2 = \{uwvw^{-1}| \quad |uwv| \leq q, \quad uv \neq \varepsilon, \quad \exists S \in N : S \Rightarrow uSv, S \Rightarrow w\}$, *где* p *и* q *суть константы из известной* $xuwvy$ *- теоремы (см. напр.* [5]). *Тогда* $L \subseteq \mathcal{M}$ *тогда и только тогда, когда* $W_1 = \Omega_1 \cup \Omega_2 \subseteq \mathcal{M}$.

Все необходимые сведения и обозначения из теории контекстно-свободных языков можно найти например в [5] или [7].

Цель настоящей работы - модифицировать алгоритм А. В. Анисимова так, чтобы он работал за полиномиальное время. Это мы сделаем для регулярных языков и для линейных языков, используя специфические свойства этих языков. Эта работа продолжение и дополнение работы [8]. Здесь мы опишем новые конечные множества, с помощью которых проверяется включение $L \subseteq \mathcal{M}$.

Диаграммой переходов будем называть четверку $H = (V, R, S, l)$, где (V, R) - конечный ориентированный граф с множеством вершин V и множеством ребер R, S - полугруппа, элементы которой будем называть метками, а l - функция из R в S, называемая функция разметок. Другими словами, каждое ребро графа помечено некоторым элементом полугруппы S. Если π - путь в диаграмме переходов H , то метка пути $l(\pi)$ - это произведение меток ребер, составляющих этот путь, причем метки берутся в порядке прохождения ребер. Если P - множество путей в H , то $l(P)$ будет множество $\{\omega \,|\, \exists \pi \in P : l(\pi) = \omega\}$

2. Включение регулярных языков в групповые языки

В теореме 1 утверждается, что для контекстно-свободного языка L можно конструктивным путем найти конечное множество $W_1 = \Omega_1 \cup \Omega_2$, такое, что L включается в данный групповой язык $\mathcal{M}$ группы G с разрешимой проблемой равенства слов, тогда и только тогда, когда W_1 включается в $\mathcal{M}$. В этом параграфе решим более конкретную задачу для регулярных языков, которые являются важным частным случаем контекстно-свободных языков и найдем еще два множества, обладающих этим свойством. Эти мно -жества будут различаться W_1 и между собой.

Пусть L - регулярный язык. Тогда существует конечный автомат - распознаватель (детерминированный или недетерминированный)

$$A = (Q, \Sigma, \delta, q_1, Z)$$

такой, что $L = T(A)$, где $Q = \{q_1, q_2, ..., q_n\}$- множество состояний;$\Sigma = \{x_1, x_2, ..., x_m\}$- входной алфавит; δ- функция переходов; q_1- начальное состояние;Z- множество заключительных состояний; $T(A)$- множество слов распознаваемых A. Построим диаграмму переходов $H_A = (Q, R, \Sigma^*, l_A)$, где множество вершин Q совпадает с множеством состояний автомата A; множество ребер R образовано следующим образом: $\rho \in R$, где $\rho =$

(q_i, q_j), тогда и только тогда, когда существует $x \in \Sigma \cup \{\varepsilon\}$ такое.что $q_j \in \delta(q_i, x)$ и при этом $l_A(\rho) = x$. ε - это пустое слово. Тогда слово $\omega \in L = T(A)$ тогда и только тогда, когда существует путь π в H_A с началом q_1, концом - элемент из Z и с метко цхар26 пути ω. Пусть G - группа с множеством образующих $X = \{x_1, x_2, ..., x_m, x_1^{-1}, x_2^{-1}, ..., x_m^{-1}\}$ и пусть $\mathcal{M}$ - соответствующий групповой язык. Рассмотрим диаграмму переходов $H_G = (Q, R, G, l_G)$ с тем же множеством вершин и ребер как и в H_A, только метки ребер считаем как элементы группы G.

Пусть $F_G = (2^G, \cup, \cdot, \phi. \{e\})$ - замкнутое полукольцо с элементами всех подмножеств группы G, включая и пустое множество. Операции в F_G будут соответственно обьединение и произведение множеств, единичный элемент - это множество $\{e\}$, содержащее только единицу e группы G, а нулевой элемент - пустое множество ϕ. Замкнутые полукольца и их приложения хорошо изучены в [11]. Это понятие дефинированно и используется и в [6, 9].

В замкнутом полукольце F_G определяем бинарную операцию $[x, y]$ следующим образом: если $a, b \in G$, то $[\{a\}, \{b\}] = \{aba^{-1}\}$ и для $x, y, z \in F_G$ выполнено $[x \cup y, z] = [x, z] \cup [y, z]$ и $[x, y \cup z] = [x, y] \cup [x, z]$. Очевидно, это корректно введенная операция в силу дистрибутивного закона в F_G.

Пусть P - множество всех путей в H_G с началом q_1 и концом - элемент из Z. Тогда очевидно $L \subseteq \mathcal{M}$ тогда и только тогда, когда $l_G(P) = \{e\}$.

Пусть P_1 - множество всех путей из P не содержащих циклов и пусть $\Omega_3 = l_G(P_1)$.

Элементарным циклом назовем цикл в H_G без кратных вершин, т.е. цикл типа $(q_{i_1}, q_{i_2})(q_{i_2}, q_{i_3}) \cdots (q_{i_{k-1}}, q_{i_k})(q_{i_k}, q_{i_1})$, где $q_{i_s} \neq q_{i_t}$ для $s \neq t$ и пусть C - множество элементарных циклов в H_G. Тогда, если π - путь из P , то, очевидно, π принадлежит P_1 или π можно представит, в виде $\pi = \pi_1\pi_2\pi_3$, где π_1 не содержит циклов $\pi_2 \in C$, а π_3 начинает с конца π_1 и кончает в элементе из Z. Но тогда и пути $\pi_1\pi_2^k\pi_3$, $k = 0, 1, 2, ...$ тоже принадлежат P.

Рассмотрим множество Ω_4 , состоящее из всех элементов группы G вида $\alpha = uvu^{-1}$, где $u = l_G(\pi_1)$ для некоторого пути $\pi_1 \in H_G$ с началом q_1 и не имеющих циклов, $v = l_G(\pi_2)$ для некоторого пути $\pi_2 \in C$, переходящих через конца π_1 и существует путь $\pi_3 \in H_G$ с началом конец π_1 и концом - некоторый элемент из Z. Ω_3 и Ω_4 являются элементами замкнутого полукольца F_G. Полагаем $W_2 = \Omega_3 \cup \Omega_4$.

Обазуем множества путей в H_G, C_{ij}^k, где $i, j, k = 1, 2, ..., n$; $n = |Q|$ следующим образом:

$$C_{ij}^{0} = \{\rho | \rho \text{ - ребро из } q_i \text{ в } q_j \}$$

$$C_{ij}^k = C_{ij}^{k-1} \cup C_{ik}^{k-1} C_{kj}^{k-1}, \quad k = 1, 2, ..., n$$

Нетрудно заметить, что C'^{k}_{ij} состоит только из путей длиной меньше или равной $k+1$ с началом q_i, концом q_j и все узлы, которых кроме быть может начала или конца принадлежат множеству $\{q_1, q_2, ..., q_k\}$.

Рассмотрим следующие элементы замкнутого полукольца F_G:

$\Omega_5 = \{l_G(C'^{n}_{1j})\}$, где j такое, что $q_j \in Z$;

$\Omega_6 = \{[l_G(C'^{n}_{1j}), l_G(C'^{n}_{jj})]\}$, где j такое, что существует путь с началом q_j и концом q_t для некоторого $q_t \in Z$, т.е. $C'^{n}_{jt} \neq \phi$.

Лемма 1. *В C'^{k}_{ij}, $k = 2, 3, ..., n$, возможно существование пути содержащего цикла.*

Доказательство. Очевидно для всех $s, t = 1, 2, ..., n$ выполнено $C'^{0}_{st} \subseteq C'^{1}_{st} \subseteq \cdots \subseteq C'^{n}_{st}$. Кроме этого для всех $r, s, t = 1, 2, ..., n$ докажем индукцией по r , что в C'^{r}_{st} возможно существование пути $(q_s, q_1)(q_1, q_2)\cdots(q_r, q_t)$. В самом деле для $r = 1$ утверждение очевидно. Пусть для $r \leq r_0$ в C'^{r}_{st} возможно существование пути $(q_s, q_1)(q_1, q_2)\cdots(q_r, q_t)$ и рассмотрим $C'^{r+1}_{st} \subseteq C'^{r}_{s\ r+1}C'^{r}_{r+1\ t}$. Из индукционного предположения имеем, что в $C'^{r}_{s\ r+1}$ возможен путь $(q_s, q_1)(q_1, q_2)\cdots(q_r, q_{r+1})$, и так как $C'^{0}_{r+1\ t} \subseteq C'^{r}_{r+1\ t}$ и в $C'^{0}_{r+1\ t}$ возможно существование реб ра (q_{r+1}, q_t) , то в C'^{r+1}_{st} возможно существование пути $(q_s, q_1)(q_1, q_2)\cdots(q_r, q_{r+1})(q_{r+1}, q_t)$.

Следовательно, в C'^{k-1}_{ik} возможен путь $(q_i, q_1)(q_1, q_2)\cdots(q_{k-1}, q_k)$, а в C'^{k-1}_{kj} возможен путь $(q_k, q_1)(q_1, q_2)\cdots(q_{k-1}, q_j)$. Тогда в $C'^{k}_{ij} \subseteq C'^{k-1}_{ik}C'^{k-1}_{kj}$ возможен путь $(q_i, q_1)(q_1, q_2)\cdots(q_{k-1}, q_k)(q_k, q_1)(q_1, q_2)\cdots(q_{k-1}, q_j)$ в котором содержится цикл $(q_1, q_2)(q_2, q_3)\cdots(q_{k-1}, q_k)(q_k, q_1)$. Лемма доказана. □

Следствие 1. *Для определенных выше множеств $\Omega_3, \Omega_4, \Omega_5$ и Ω_6 из элементов группы G имеем $\Omega_3 \subseteq \Omega_5$ и $\Omega_4 \subseteq \Omega_6$ при том в общем случае $\Omega_3 \neq \Omega_5$ и $\Omega_4 \neq \Omega_6$.*

Также нетрудно заметить, что в общем случае множества Ω_1 и Ω_2 (если ихние элементы рассматриваем как элементы группы G) из теоремы 1 , введенные А. В. Анисимовым различаются множествами введенными нами в этом параграфе. В силу следующая теорема:

Теорема 2. *Для введенных выше обозначений, следующие условия эквивалентны:*

(i) $L \subseteq \mathcal{M}$;
(ii) $W_1 = \Omega_1 \cup \Omega_2 = \{e\}$;
(iii) $W_2 = \Omega_3 \cup \Omega_4 = \{e\}$;
(iv) $W_3 = \Omega_5 \cup \Omega_6 = \{e\}$.

Доказательство. Так как регулярные языки являются частными случаями контекстно-свободных языков, то эквивалентность условий (i) и (ii) установлена А. В. Анисимовым в [3]. Кроме этого $W_2 \subseteq W_3$ (Следствие 1), т.е. из $W_3 = \{\epsilon\}$ следует $W_2 = \{\epsilon\}$, или доказали, что из (ив) следует (iii). Чтобы доказать теорему нам достаточно доказать, что из (iii) следует (i) и из (i) следует (iv).

(iii)→(i). Пусть $W_2 = \Omega_3 \cup \Omega_4 = \{\epsilon\}$ и пусть $\omega \in L$. Тогда существует путь π в H_A с началом q_1 и концом элемент из Z, такой, что $l_A(\pi) = \omega$. Если π не содержит циклов, то $l_G(\pi) \in l_G(P_1) = \Omega_3 = \{\epsilon\}$ и следовательно $\omega \in \mathcal{M}$.

Пусть π содержит цикл. Тогда π содержит элементарный цикл. Другими словами π можно представит, в виде $\pi = \pi_1\pi_2\pi_3$, где

$$l_G(\pi_1)l_G(\pi_2)(l_G(\pi_1))^{-1} \in \Omega_4 = \{\epsilon\}.$$

Отсюда следует, что $l_G(\pi_1)l_G(\pi_2) = l_G(\pi_1)$ или все равно $l_G(\pi_1\pi_2\pi_3) = l_G(\pi_1\pi_3)$. Так как $\pi_2 \in C$, то длина π_2 больше 1. Следовательно, существует путь в H_G с меньшей длиной, чем длина π, метка которого равна ω в группе G. Этот процесс сокращения можно продолжать конечное число раз, так как длина ω - конечная. В конце этого процесса получим путь без циклов с меткой равной ω в группе G. Но $l_G(P_1) = \Omega_3 = \{\epsilon\}$. Следовательно, $\omega = \epsilon$ в группе G т.е. $L \subseteq \mathcal{M}$.

(i)→(iv). Пусть $L \subseteq \mathcal{M}$, т.е. $l_G(P) = \{\epsilon\}$ и так как $\Omega_5 \subseteq l_G(P)$, то $\Omega_5 = \{\epsilon\}$.

Пусть $z \in \Omega_6$. Тогда z можно представит, в виде $z = uvu^{-1}$, где $u \in l_G(C^n_{1j})$, $v \in l_G(C^n_{jj})$ для некоторого j, такого, что существует путь π_3 с началом q_j, концом элемент из Z и пусть $l_G(\pi_3) = w$. Кроме этого очевидно u -это метка некоторого пути π_1 с началом q_1 и концом q_j, а v - это метка некоторого цикла, проходящего через q_j. Следовательно пути $\pi' = \pi_1\pi_2\pi_3$ и $\pi'' = \pi_1\pi_3$ принадлежат P и так как $L \subseteq \mathcal{M}$, то $l_G(\pi') = l_G(\pi'') = \epsilon$ и следовательно $uvw = uw$, т.е. $uvu^{-1} = \epsilon$. Следовательно $z = \epsilon$ и так как z произвольное, то $\Omega_6 = \{\epsilon\}$. Теорема доказана. □

Следующий алгоритм базируется на эквивалентности (i) и (iv) из теоремы 2. Для удобства $i \in Z$ будет означать $q_i \in Z$, а g^k_{ij} - это $l_G(C^k_{ij})$. Здесь k в g^k_{ij} индекс и не означает степень.

Алгоритм 1. *Проверяет включение $L \subseteq \mathcal{M}$ для регулярного языка L и группового языка $\mathcal{M}$ группы G с разрешимой проблемой равенства слов.*

Вход: $g^0_{ij} = l_G(C^0_{ij})$, $i, j = 1, 2, ..., n$

Выход: Логическая переменная T , получающая стоимость **Истина**, если $L \subseteq \mathcal{M}$ и стоимость **Ложь**, в противном случае. Алгоритм останавливает сразу после получения стоимости T :=**Ложь**.

```
Начало
1.  $T$ :=Истина;
2.  Для $1 \leq k \leq n$ Делать
3.          Для $1 \leq i, j \leq n$ Делать
4.          $g_{ij}^{k} := g_{ij}^{k-1} \cup g_{ik}^{k-1} g_{kj}^{k-1}$
5.                                  Конец Делать
6.                              Конец Делать
7.  Для $j \in Z$ Делать
8.  Если $g_{1j}^{n} \neq \phi$ и $g_{1j}^{n} \neq \{e\}$ То
9.                                  Начало $T$ :=Ложь ; Останов Конец ;
10.                             Конец Делать
11. Для $1 \leq j \leq n$ Делать
12.         Для $t \in Z$ Делать
13.         Если $g_{jt}^{n} \neq \phi$ и $g_{1j}^{n} \neq \phi$ и $g_{jj}^{n} \neq \phi$ То
14.                 Если $[g_{1j}^{n}, g_{jj}^{n}] \neq \{e\}$ То
15.                         Начало $T$ :=Ложь ; Останов Конец
16.                             Конец Делать
17.                         Конец Делать
Конец.
```

Теорема 3. *Алгоритм 1 выполняется не более $O(n^3)$ операции сложения, произведения и $[x, y]$ элементов из замкнутого полукольца F_G и проверяет включение $L \subseteq \mathcal{M}$, где L - регулярный язык, распознаваемый конечным автоматом A, $\mathcal{M}$ - групповой язык группы G с разрешимой проблемой равенства слов, n - число состояний автомата.*

Доказательство. Согласно теоремы 3 и учитывая аксиомы замкнутого полукольца F_G, то в рядах 9 и 15 алгоритма 1 логическая перементая T принимает стоимость **Ложь** тогда и только тогда, когда L не включается в $\mathcal{M}$. В противном случае T принимает стоимость **Истина**. Следовательно, алгоритм 1 коректно проверяет выполняется ли включение $L \subseteq \mathcal{M}$.

Легко проверить, что строка 4 выполняется не более n^3 раз, причем каждый раз выполняются две операции в замкнутом полукольце F_G. Строки 13 и 14 выполняется не более n^2 раз каждая. Следовательно, алгоритм 1 выполняет не более $O(n^3)$ операции сложения, умножения и $[x, y]$ в замкнутом полукольце F_G. Теорема доказана. □

Следствие 2. *Если операции сложение, умножение и $[x, y]$ в замкнутом полукольце F_G можно выполнить за полиномиальное время, то алгоритм 1 является полиномиальным.*

3. Включение линейных языков в групповые языки

В этом параграфе будем продолжать использовать идею А. В. Анисимова для нахождения конечных множеств, с помощью которых можно определить включается ли данный контекстно-свободный язык в групповой язык группы с разрешимой проблемой равенства слов. Это сделаем для линейных языков. Как известно, класс линейных языков включается в класс контекстно-свободных языков. Анализируя доказательство А. В. Анисимова[3], можно заключить, что задача о проверке включения линейного яз ыка в групповой язык группы с разрешимой проблемой равенства слов решает и задачу о проверке однозначности конечно-автоматного отображения.

В этом параграфе будем рассматривать линейную грамматику $\Gamma = (N, \Sigma, \Pi)$ где $N = \{A_1, A_2, ..., A_n\}$ - множество нетерминалов,

$$\Sigma = \{x_1, x_2, ..., x_m\}$$

- множество терминалов, Π - множество правил, а L будет означать линейный язык $L = L(\Gamma, A_1)$. Контекстно-свободная грамматика $\Gamma = (N, \Sigma, \Pi)$ называется линейной, если все правила в Π имеют вид $A_i \rightarrow \alpha A_j \beta$ или $A_i \rightarrow \alpha$, где $A, B \in N$ - нетерминалы, α, β - терминалные слова из свободного моноида Σ^*. Язык L называется линейным, если существует линейная грамма тика Γ и нетерминальный символ $A_i \in N$ такие, что $L = L(\Gamma, A_i)$.

Пусть S - моноид с единицей 1. Рассмотрим множество $U_S = S \times S = \{(x, y) | x, y \in S\}$. В U_S вводим операция "$\circ$" следующим образом: если (x, y), $(z, t) \in U_S$ то $(x, y) \circ (z, t) = (xz, ty)$. Нетрудно видеть, что U_S с этой операцией моноид с единицей (1,1). Если S - группа, то и U_S будет группа, при этом если $a = (x, y) \in U_S$, то обратным элементом a будет $a^{-1} = (x^{-1}, y^{-1})$. Образуем отображения f_l, f_r, f_d из U_S в S следующим образом:

$$f_l(x, y) = x$$
$$f_r(x, y) = y$$
$$f_d(x, y) = xy$$

Очевидно $f_d(x, y) = f_l(x, y) f_r(x, y)$

Рассмотрим диаграмму переходов $H_\Gamma = (V, R, U_{\Sigma^*}, l_\Gamma)$ с множеством вершин $V = N \cup \{A_{n+1}\}$, где $A_{n+1} \notin N$, U_{Σ^*} - рассматриваемый выше моноид с множеством элементов $\{(\alpha, \beta) | \alpha, \beta \in \Sigma^*\}$ и операцией "$\circ$"; множество ребер R образовано следующим образом:

а) если в Π существует правило $A_i \rightarrow \alpha A_j \beta$, $\quad A_j, A_j \in N$, то в R существует ребро с началом A_i, концом A_j и меткой (α, β);

b) если в Π существует правило $A_i \to \alpha$, где $A_i \in N$, $\alpha \in \Sigma^*$, то в R существует ребро с началом A_i, концом A_{n+1} и меткой (α, ε), ε -пустое слово.

v) не существуют другие ребра в R, кроме описанных в пунктах а) и б).

Пусть G - группа с множеством образующих

$$X = \Sigma \cup \Sigma^{-1} = \{x_1, x_2, ..., x_m, x_1^{-1}, ..., x_m^{-1}\},$$

множеством определяющих соотношений θ, единицей e и с разрешимой проблемой равенства слов. Пусть $\mathcal{M}$ - соответствующий групповой язык, а U_G - группа, полученная описанным выше способом. Рассмотрим диаграмму переходов $H_U = (V, R, U_G, l_U)$, где множества вершин V и ребер совпадают с соответствующими множествами в диаграмме переходов H_Γ, а метки р ебер считаем как элементы группы U_G. Как и в параграфе 2 можно рассмотреть замкнутые полукольца $F_U = (U_G, \cup, \circ, \phi, \{(e, e)\})$ и $F_G = (G, \cup, \cdot, \phi, \{e\})$. Тогда отображения f_l, f_r и f_d естественным способом можно продолжить до отображения из F_U в F_G.

В F_G вводим операцию $\langle x, y, z\rangle$ следующим образом: если $a, b, c \in G$, то $\langle \{a\}, \{b\}, \{c\}\rangle = \{abcb^{-1}\}$ и для $x, y, z, t \in F_G$ выполнено:

$$\langle x \cup y, z, t\rangle = \langle x, z, t\rangle \cup \langle y, z, t\rangle$$
$$\langle x, y \cup z, t\rangle = \langle x, y, t\rangle \cup \langle x, z, t\rangle$$
$$\langle x, y, z \cup t\rangle = \langle x, y, z\rangle \cup \langle x, y, t\rangle$$

В силу дистрибутивного закона в F_G, $\langle x, y, z\rangle$ - коректно введенная операция.

Пусть P_Γ - множество всех путей в H_Γ с началом A_1 и концом A_{n+1}, а P_U - множество всех путей в H_U с началом A_1 и концом A_{n+1}.

Лемма 2. *Для введеных выше обозначений исполнено* $L = f_d(l_\Gamma(P_\Gamma))$.

Доказательство. Пусть $\omega \in L$. Тогда существует вывод в Γ:

$$A_1 \to \alpha_1 A_{i_1}\beta_1 \to \alpha_1\alpha_2 A_{i_2}\beta_2\beta_1 \to \cdots \to \alpha_1\alpha_2\cdots\alpha_k A_{i_k}\beta_k\beta_{k-1}\cdots\beta_1$$
$$\to \alpha_1\alpha_2\cdots\alpha_k\gamma\beta_k\beta_{k-1}\cdots\beta_1,$$

где $A_{i_j} \in N$, $\alpha_j, \beta_j, \gamma \in \Sigma^*$. Но тогда существует путь в H_Γ $\pi = \rho_1\rho_2\cdots\rho_k\tau$, где ρ_1 - ребро из A_1 в A_{i_1}, ρ_j ребра из $A_{i_{j-1}}$ в A_{i_j}, τ - ребро из A_{i_k} в A_{n+1}; $l_\Gamma(\rho_j) = (\alpha_j, \beta_j)$, $l_\Gamma(\tau) = (\gamma, \varepsilon)$. Но тогда $l_\Gamma(\pi) = l_\Gamma(\rho_1) \circ l_\Gamma(\rho_2) \circ \cdots \circ l_\Gamma(\rho_k) \circ l_\Gamma(\tau) = (\alpha_1, \beta_1) \circ (\alpha_2, \beta_2) \circ \cdots \circ (\alpha_k, \beta_k) \circ (\gamma, \varepsilon) = (\alpha_1\alpha_2\cdots\alpha_k\gamma, \beta_k\beta_{k-1}\cdots\beta_1)$. Отсюда следует, что $f_d(l_\Gamma(\pi)) = \omega$. Следовательно $L \subseteq f_d(l_\Gamma(P_\Gamma))$.

Наоборот, если $\omega \in f_d(l_\Gamma(P_\Gamma))$, то ω можно представить в виде $\omega = \alpha\beta$, где (α, β) - метка некоторого пути π из P_Γ и пусть $\pi = \rho_1\rho_2\cdots\rho_k$, где ρ_1

- ребро из A_1 в A_{i_1}, ρ_j, $j = 2, 3, ..., k-1$ ребра из $A_{i_{j-1}}$ в A_{i_j}, ρ_k - ребро из A_{i_k} в A_{n+1} и пусть $l_\Gamma(\rho_j) = (\alpha_j, \beta_j)$ $l_\Gamma(\rho_k) = (\gamma, \varepsilon)$ для некоторых $\alpha_j, \beta_j, \gamma \in \Sigma^*$, таких что $A_1 \to \alpha_1 A_{i_1} \beta_1$, $A_{i_j} \to \alpha_{j+1} A_{i_{j+1}} \beta_{j+1}$, $A_{k-1} \to \gamma$ суть правила в Γ и $\alpha_1\alpha_2 \cdots \alpha_{k-1}\gamma = \alpha$, $\beta_{k-1}\beta_{k-2}\cdots\beta_1 = \beta$. Но тогда существует вывод в Γ:

$$A_1 \to \alpha_1 A_{i_1} \beta_1 - \cdots - \alpha_1\alpha_2\cdots\alpha_{k-1} A_{i_{k-1}} \beta_{k-1}\beta_{k-2}\cdots\beta_1$$
$$- \alpha_1\alpha_2\cdots\alpha_{k-1}\gamma\beta_{k-1}\beta_{k-2}\cdots\beta_1 = \alpha\beta = \omega.$$

Следовательно $f_d(l_\Gamma(P_\Gamma)) \subseteq L$. Лемма доказана. □

Следствие 3. *Для введенных выше обозначений* $L \subseteq \mathcal{M}$ *тогда и только тогда, когда* $f_d(l_U(P_U)) = \{\epsilon\}$

Пусть $P_1 \subseteq P_U$ - множество всех путей с началом A_1, концом A_{n+1} и не содержащих циклов и пусть

$\Omega_7 = f_d(l_U(P_1))$

Пусть C - множество элементарных циклов в H_U. Пусть

P' - множество всех путей из H_U, начинающихся в A_1

P'' - множество всех путей из H_U не содержащих циклов и кончающихся в A_{n+1}.

Очевидно $P_U \subseteq P'$, $P_1 \subseteq P''$ и в общем случае $P_U \neq P'$ и $P_1 \neq P''$. Рассмотрим множество путей $P_2 = \{\pi = \pi_1\pi_2\pi_3 | \pi_1 \in P', \pi_2 \in C, \pi_3 \in P''$. Для всех $\pi = \pi_1\pi_2\pi_3 \in P_2$ рассмотрим множество

$\Omega_8 = \{\langle f_l(l_u(\pi_2)), f_d(l_U(\pi_3)), f_r(l_U(\pi_2))\rangle |$ существует путь $\pi = \pi_1\pi_2\pi_3 \in P_2\}$.

Если проанализировать способ образования множества $\Omega_1, \Omega_2\Omega_7$ и Ω_8, где Ω_1 и Ω_2 суть множества, введенные А. В. Анисимовым (теорема 1), элементы которых рассматриваем как элементы группы G, то нетрудно заметить, что $\Omega_7 \subseteq \Omega_1$, $\Omega_8 \subseteq \Omega_2$ и в общем случае $\Omega_7 \neq \Omega_1$ и $\Omega_8 \neq \Omega_2$.

Как и в параграфе 2. образуем множества путей C'^k_{ij} (длиной меньше или равной $k+1$) из P_U $(1 \leq i, j, k \leq n+1)$.

Пусть $g^k_{ij} = l_U(C'^k_{ij}) \in F_U$. Здесь k индекс и не означает степень. Рассмотрим элементы замкнутого полукольца F_G:

$$\Omega_9 = f_d(g^n_{1\ n+1})$$
$$\Omega_{10} = \{\langle f_l(g^n_{ii}), f_d(g^n_{i\ n+1}), f_r(g^n_{ii})\rangle$$

Аналогично доказательства леммы 1 можно доказать, что в C'^k_{ij} возможно существование путей, содержащих циклы и, имея ввиду способ образования Ω_7, Ω_8, Ω_9 и Ω_{10}, получаем следующее утверждение:

Лемма 3. *Для введенных выше обозначений исполнено* $\Omega_7 \subseteq \Omega_9$ *и* $\Omega_8 \subseteq \Omega_{10}$, *при этом в общем случае* $\Omega_7 \neq \Omega_9$ *и* $\Omega_8 \neq \Omega_{10}$.

Теорема 4. *Для введенных выше обозначений следующие условия эквивалентны:*

(i) $L \subseteq \mathcal{M}$;
(ii) $W_1 = \Omega_1 \cup \Omega_2 = \{e\}$;
(iii) $W_4 = \Omega_7 \cup \Omega_8 = \{e\}$;
(iv) $W_5 = \Omega_9 \cup \Omega_{10} = \{e\}$.

Доказательство. Эквивалентность условий (i) и (ii) доказана А. В. Анисимовым в [3] (см. Теорему 1). Как заметили выше $W_4 \subseteq W_1$ и следовательно из $W_1 = \{e\}$ следует $W_4 = \{e\}$. Из Леммы 3 следует, что если выполнено $W_5 = \{e\}$, то выполнено $W_4 = \{e\}$. Чтобы доказать теорему осталось доказать, что (i) влечет (iv) и (iii) влечет (i).

(i)→(iv). Пусть $L \subseteq \mathcal{M}$. Тогда (Следствие 3) $f_d(l_U(P_U)) = \{e\}$. Но очевидно $C^n_{1\ n+1} \subseteq P_U$ и следовательно $\Omega_9 = \{e\}$.

Пусть $z \in \Omega_{10}$. Тогда $z = uvwv^{-1}$, где $u \in f_l(g^n_{ii})$, $v = f_d(g^n_{i\ n+1})$, $w = f_r(g^n_{ii}$ для некоторого i такое, что существует путь π_1 в H_U из A_1 в A_i и пусть $l_U(\pi_1) = (x, y)$. Тогда очевидно (u, w) метка некоторого цикла π_2, проходящего через A_i, а v можно представить в виде $v = v_1 v_2$, где (v_1, v_2) - метка некоторого пути π_3 из A_i в A_{n+1}. Рассмотрим пути $\pi' = \pi_1\pi_2\pi_3$ и $\pi'' = \pi_1\pi_3$. Очевидно они начинаются в A_1 и заканчиваются в A_{n+1}. Имеем:

$$l_U(\pi_1\pi_2\pi_3) = (x, y) \circ (u, w) \circ (v_1, v_2) = (xuv_1, v_2wy)$$
$$l_U(\pi_1\pi_3) = (x, y) \circ (v_1, v_2) = (xv_1, v_2y)$$

Согласно Следствию 3, $xuvwy = xvy = e$, откуда следует что $uvwv^{-1} = e$, т.е. $z = e$. Так, как z - произвольное из Ω_{10}, то $\Omega_{10} = \{e\}$.

(iii)→(i). Пусть $W_4 = \Omega_7 \cup \Omega_8 = \{e\}$ и пусть $\omega \in L$. Согласно леммы 2 $\omega \in f_d(l_\Gamma(P_\Gamma))$, т.е. ω можно представить в виде $\omega = \omega_1\omega_2$, где (ω_1, ω_2) - метка некоторого пути в H_Γ с началом A_1 и концом A_{n+1} и пусть π - соответствующий ему путь в H_U. Если $\pi \in P_1$ (т.е. если π не содержит цикла), то $f_d(l_U(\pi)) \in f_d(l_U(P_1)) = \Omega_7 = \{e\}$ и следовательно $\omega \in \mathcal{M}$.

Пусть π содержит цикл. Тогда π можно представить в виде $\pi = \pi_1\pi_2\pi_3$, где $\pi_1 \in P'$, $\pi_2 \in C$, $\pi_3 \in P''$ и пусть $l_U(\pi_1) = (a_1, b_1)$, $l_U(\pi_2) = (a_2, b_2)$, $l_U(\pi_3) = (a_3, b_3)$. Тогда

$$f_d(l_U(\pi)) = f_d((a_1, b_1) \circ (a_2, b_2) \circ (a_3, b_3)) = f_d(a_1a_2a_3, b_3b_2b_1) = a_1a_2a_3b_3b_2b_1.$$

Но $a_2a_3b_3b_2(a_3b_3)^{-1} \in \Omega_8$, т.е. $a_2a_3b_3b_2(a_3b_3)^{-1} = e$, или $a_1a_2a_3b_3b_2b_1 = a_1a_3b_3b_1$. Нетрудно заметить, что (a_1a_3, b_3b_1) - это метка пути $\pi_1\pi_3$, которая получается из π, опуская цикл π_2. Продолжая таким образом опускать циклы в π, то так как слово ω - конечное, через конечное число шагов получим, что $f_d(l_u(\pi)) = f_d(l_U(\pi'))$, где π' путь из A_1 в A_{n+1} не имеющий циклов. Но $f_d(l_U(\pi')) \in \Omega_7 = \{e\}$. Следовательно $f_d(l_U(P_2)) = \{e\}$ и согласно следствию 3 получаем, что $L \subseteq \mathcal{M}$. Теорема доказана. □

На базе Теоремы 4 ((i)↔(iv) получаем следующий алгоритм, проверяющий включения $L \subseteq \mathcal{M}$.

Алгоритм 2. *Проверяст включения $L \subseteq \mathcal{M}$, для линейного языка L и группового языка $\mathcal{M}$ группы G с разрешимой проблемой равенства слов.*

Вход: $g^0_{ij} = l_U(C^0_{ij}), \quad i,j = 1,2,...,n+1$

Выход: Логическая переменная T , получающая стоимость **Истина**, если $L \subseteq \mathcal{M}$ и стоимость **Ложь**, в противном случае.

```
Начало
T := Истина ;
Для $1 \le k \le n$ Делать
    Для $1 \le i,j \le n+1$ Делать
        $g^k_{ij} := g^{k-1}_{ij} \cup g^{k-1}_{ik} \circ g^{k-1}_{kj}$
    Конец Делать
Конец Делать;
Если $g^n_{1\ n+1} \ne \phi$ и $f_d(g^n_{1\ n+1}) \ne \{e\}$ То
    Начало $T$ := Ложь; Останов Конец;
Для $1 \le i \le n+1$ Делать
    Если $g^n_{1i} \ne \phi$ и $g^n_{ii} \ne \phi$ и $g^n_{i\ n+1} \ne \phi$ То
        Если $\langle f_l(g^n_{ii}), f_d(g^n_{i\ n+1}), f_r(g^n_{ii})\rangle \ne \{e\}$ То
            Начало $T$ := Ложь ; Останов Конец
    Конец Делать
Конец.
```

Теорема 5. *Алгоритм 2 выполняет не более $O(n^3)$ операции сложения и произведения в замкнутом полукольце F_U, не более $O(n^2)$ операции $\langle x,y,z\rangle$ в замкнутом полукольце F_G и проверяет включение $L \subseteq \mathcal{M}$, где L - линейный язык, порождаемый линейной грамматикой с n нетерминалов, $\mathcal{M}$ - групповой язык группы G с разрешимой проблемой равенства слов.*

Доказательство теоремы повторяет доказательство Теоремы 3.

Следствие 4. *Если существуют алгоритмы, выполняющие операции умножения и сложения в F_U и $\langle x,y,z\rangle$ в F_G за полиномиальное время, то алгоритм 2 является полиномиальным.*

Литература

[1] Анисимов А. В. *О групповых языках*, Кибернетика **4** (1971), 18–24.

[2] Анисимов А. В, *О некоторых алгоритмических вопросах для групп и контекстно-свободных языков*. Кибернетика **2** (1972), 4–11.

[3] Анисимов А. В, *Полугрупповые конечно-автоматные отображения*, Кибернетика **5** (1981), 1–7.

[4] Анисимов А. В. Лисовик Л. П, *Проблемы эквивалентности конечно-автоматных отображений в свободную и коммутативную полугруппу*, Кибернетика **3** (1978), 1–8.

[5] Гинсбург С, *Математическая теория контекстно-свободных языков*, М., Мир, 1970.

[6] Ахо А., Хопкрофт Дж., Ульман Дж, *Построение и анализ вычислителных алгоритмов*, М., Мир, 1978.

[7] Ахо А., Ульман Дж, *Теория синтактического анализа,перевода и компиляции*, М., Мир, 1978.

[8] Йорджев К. Я, *О включении контекстно-свободных языков в групповые языки*, Модели и системы обработки информации, вып. **10** (1991), 21–27.

[9] Кук Д., Бейс Г. *Компьютерная математика*, М., Наука, 1990.

[10] Kuich W. *Unambiguous automata* bull. EATCS (1989), 62–67.

[11] Kuich W. Salomaa A. *Semigroups, Automata, Languages* Springer, 1986.

[12] Stearns R. Hunt H, *On the equivalens and containment problēms for unambiguous regular expressions. regular grammars and finite automata* 22nd FOCS (1981), 74–81.

[13] Weber A. Seidl H. *On the degree ambiguity of finite automata*, Lecture Notes in Computer Science. **233** (1986), 620–629.

Пед. Институт. Ямбол. Болгария

FILOMAT (Niš) 9:3 (1995), 711–721
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

THEORY OF MULTIPLE ANTISYMMETRY

Slavik V. Jablan

ABSTRACT. Survey of problems in theory of multiple antisymmetry, which can be solved using antisymmetric characteristic method, is given.

0. Introduction and definitions

Originated from Speiser (1927) and realized by Weber (1929), the idea of representing symmetry groups of bands by black-white plane diagrams was the starting point for introducing the antisymmetry (Heesch, 1929). The color change white-black used as the possibility for the dimensional transition from the symmetry groups of friezes G_{21} to the symmetry groups of bands G_{321}, or from the plane groups G_2 to the layer groups G_{32}, applied on Fedorov space groups G_3 in order to derive the hyperlayer symmetry groups G_{43} (Heesch, 1930) was the beginning of the theory of antisymmetry. The further development of the theory of antisymmetry can be followed through the works by Shubnikov, Belov and Zamorzaev [1].

Its natural generalization, the multiple antisymmetry is suggested by Shubnikov (1945) and introduced by Zamorzaev (1957). Three months later, the different concept of the multiple antisymmetry is proposed by Mackay. During the next 30 years, mostly by the contribution of Kishinev school (Zamorzaev, Palistrant, Galyarskij...) the theory of multiple antisymmetry has become an integral part of mathematical crystallography and acquired the status of a complete theory extended to all categories of isometric symmetry groups of the space E^n $(n \leq 3)$, different kinds of non-isometric symmetry groups (of similarity symmetry, conformal symmetry...) and P-symmetry groups [1,2,3,4]. On the other hand, investigation of Mackay approach to the multiple antisymmetry was not continued.

Supported by Grant 0401 of FNS through Math. Inst. SANU

Let the discrete symmetry group G with a set of generators $\{S_1, \ldots, S_r\}$ be given by presentation [5]

$$g_n(S_1, \ldots, S_r) = E, \quad n = \overline{1,s}$$

and let $e_1, \ldots, e_l$ be antiidentities of the first,...,lth kind, satisfying the relations

$$e_ie_j = e_je_i \quad e_i^2 = E \quad e_iS_q = S_qe_i, \quad i,j = \overline{1,l}, \;\; q = \overline{1,r} \qquad (1).$$

The group consisting of transformations $S' = e'S$, where e' is the identity, antiidentity, or some product of antiidentities, is caled the (multiple) antisymmetry group. In particular, for $l = i = j = 1$ we have the simple antisymmetry. From the point of view of the mathematical logic or discrete mathematics the system of antiidentities can be considered as l-dimensional Boolean space.

The groups of simple and multiple antisymmetry can be derived by Shubnikov-Zamorzaev method: by replacing the generators of G by antigenerators of one or several independent patterns of antisymmetry. Having in mind the theorem on dividing all groups of simple and multiple antisymmetry into groups of C^k $(1 \leq k \leq l)$, C^kM^m $(1 \leq k, m;\, k+m \leq l)$ and M^m $(1 \leq m \leq l)$ types, and the derivation of the groups of C^k and C^kM^m types directly from the generating group G and from the groups of M^m-type respectively, the only non-trivial problem is the derivation of the M^m-type groups [1].

In this paper we will consider only the junior multiple antisymmetry groups of the M^m-type, i.e. the multiple antisymmetry groups isomorphic with their generating symmetry group, that possess the independent system of antisymmetries.

Every junior multiple antisymmetry group G' of the M^m-type can be (uniquely) defined by the extended group/subgroup symbol

$$G/(H_1, \ldots, H_m)/H,$$

where G is the generating group, H_i its subgroups of the index 2 satisfying the relationships $G/H_i \simeq C_2 = \{e_i\}$ $(1 \leq i \leq m)$, and H the subgroup of G of the index 2^m, the symmetry subgroup of G' $(G/H \simeq C_2^m = \{e_1\} \times \ldots \times \{e_m\})$.

For the equality of multiple antisymmetry groups can be used three different criteria:

(1) "strong" equality criterion according to which the antiidentities e_i are noneq uivalent. Consequently, in the symbol $G/(H_1, \ldots, H_m)/H$ the order of the subgroups $H_1, \ldots, H_m$ is important. In the sense of interpretation,

this means that the bivalent changes e_i are physically different (nonequivalent) (e.g. (white black), (+ -), $(S\ N)$, (0 1)...);

(2) "middle" equality criterion, where all e_i are treated as the equivalent ones (i.e. permutable), so the order of the subgroups mentioned it is not important; (3) "weak" equality criterion G/H.

Using the "strong" equality criterion, as the result we have Zamorzaev groups (Z-groups), and using the "middle" Mackay (or compound) multiple antisymmetry groups (M-groups) [6]. In this paper the consideration is restricted on Z-groups.

Theorem 1. (THE EXISTENTIAL CRITERION FOR M^m-TYPE GROUPS) *A Z-group G' will be of the M^m-type*

(a) *if all the relations (1) remain satisfied after replacing the generators by antigenerators; and*

(b) *if G' exausts all the antisymmetry patterns, for fixed m.*

For the derivation of Z-groups very efficiently used is the antisymmetric characteristic method [7,8,9].

Definition 1. Let all products of the generators of G, within which every generator participates once at the most, be formed and then subsets of transformations that are equivalent in the sense of symmetry with regard to the symmetry group G be separated. The resulting system is called the antisymmetric characteristic of group G ($AC(G)$).

The most of AC permit the reduction, i.e. a transformation into the simplest form; e.g., the AC of the plane symmetry group **pm** given by the presentation [5]

$$\{X,Y,R\} \quad XY = YX \quad R^2 = (RX)^2 = E \quad RY = YR$$

is $\{R, RX\}\{Y\}\{RY, RXY\}\{X\}\{XY\}$ and its reduced AC is $\{R, RX\}\{Y\}$.

Definition 2. Two or more Z-groups belong to a family iff they are derived from the same symmetry group G.

Theorem 2. *Two Z-groups G_1' and G_2' of the M^m-type for m fixed, with common generating group G, are equal iff they possess equal AC.*

Every $AC(G)$ completely defines the series $N_m(G)$, where by $N_m(G)$ is respectively denoted the number of Z-groups of the M^m-type derived from G, for m fixed $(1 \leq m \leq l)$. For example, $N_1(\mathbf{pm})$= 5, $N_2(\mathbf{pm})$= 24, $N_3(\mathbf{pm})$= 84.

Theorem 3. *Symmetry groups that possess isomorphic AC generate the same number of Z-groups of the M^m-type for every fixed m $(1 \leq m \leq l)$, which correspond to each other with regard to structure.*

Corolary. *The derivation of all Z-groups of the M^m-type can be completely reduced to the construction of all non-isomorphic AC and the derivation of the corresponding groups of the M^m-type from these AC.*

According to Theorem 3, it is possible to identify every AC with the corresponding isomorphic algebraic term, a representative of the equivalency class which consists of all isomorphic AC. For example, it is possible to identify $AC(\mathbf{pm}) = \{R, RX\}\{Y\}$ with the term $\{A, B\}\{C\}$.

1. The derivation of (P, l)-symmetry groups from P-symmetry groups using AC

Let G^P be a junior group of P-symmetry derived from G [3]. By replacing in Definition 1 the term "transformations that are equivalent with respect to symmetry" with a more general notion "transformations that are equivalent with respect to P-symmetry", the transition from G to G^P induces the transition from $AC(G)$ to $AC(GP)$, which makes possible the derivation of groups of (P, l)-symmetry of the M^m-type using the metod of AC.

The said can be illustrated by the example of derivation of groups $G_2^{l,4}$ from groups G_2^4: $\{a, b^{(4)}\}(m)$ and $\{a^{(2)}, b^{(4)}\}(m)$.

In the first case, in the transition from $G = \mathbf{pm}$ to $G^4 = \{a, b^{(4)}\}(m)$ AC remains unchanged. In the second case, in the transition from $G = \mathbf{pm}$ to $G^4 = \{a^{(2)}, b^{(4)}\}(m)$, the equivalency of symmetry transformations is disturbed and the term $\{m, ma^{(2)}\}\{b^{(4)}\}$ is transformed into a new AC: $\{m\}\{ma\}\{b\}$. In accordance with the facts already mentioned, we have

$$\{a, b^{(4)}\}(m) \quad AC: \{m, ma\}\{b\} \simeq \{A, B\}\{C\} \quad N_1 = 5 \;\; N_2 = 24 \;\; N_3 = 84$$

$$\{a^{(2)}, b^{(4)}\}(m) \quad AC: \{m\}\{ma\}\{b\} \simeq \{A\}\{B\}\{C\} \quad N_1 = 7 \;\; N_2 = 42 \;\; N_3 = 168.$$

The given numbers N_m denote the number of groups of the M^m-type of the uncomplete $(4, l)$-symmetry. In a general case, besides the numbers N_m for p-even, we can discuss also the numbers (N_{m-1}) $(1 \leq m \leq l)$, where by (N_{m-1}) is denoted the number of groups of the complete (p, l)-symmetry of the M^m-type. For p-odd, the relationsip $N_m = (N_m)$ holds, and for p-even

$$(N_m) = N_m - (2^m - 1)(N_m - 1), \quad (N_0) = 1, \quad 1 \leq m \leq l.$$

One of the most important results obtained using the mentioned method, is the derivation of the groups $G_3^{l,p}$ from the groups G_3^p $(p = 3, 4, 6, P \simeq C_p)$ [10] and calculation of the numbers N_m and (N_{m-1}):

$$N_1 = 4840 \;\; N_2 = 40996 \;\; N_3 = 453881 \;\; N_4 = 5706960 \;\; N_5 = 59996160$$

$$(N_1) = 4134 \quad (N_2) = 29731 \quad (N_3) = 260114 \quad (N_4) = 2048760 \quad (N_5) = 1249920.$$

By the same method, the crystallographic $(p2, l)$- and (p', l)-symmetry groups are derived from the P-symmetry groups G_3^{p2} and $G_3^{p'}$ $(p = 3, 4, 6, P \simeq D_n, D_{n(2n)})$ [11,12].

The derivation of (P, l)-symmetry groups of the M^m-type from P-symmetry groups using the AC-method can be reduced to a series of successive transitions

$$G \mapsto G^P \mapsto G^{P,1} \mapsto \ldots \mapsto G^{P,l}$$

and induced transitions

$$AC(G) \mapsto AC(G^P) \mapsto AC(G^{P,1}) \mapsto \ldots \mapsto AC(G^{P,l}).$$

Every induced AC consists of the same number of generators. Since every transition $G^{P,k-1} \mapsto G^{P,k}$, $(1 \leq k \leq l)$, is a derivation of simple antisymmetry groups using $AC(G^{P,k-1})$, for derivation of all multiple antisymmetry groups, the catalogue of all non-isomorphic AC formed by l generators and simple antisymmetry groups derived by these AC, is completely sufficient.

3. Reduction of multiple antisymmetry simple antisymmetry

The basis of this reduction is the idea already mentioned about the transition $G \mapsto G^P$ and induced transition $AC(G) \mapsto AC(G^P)$, where $AC(G)$ and $AC(G^P)$ consist of the same number of generators. This means that every step in the derivation of multiple antisymmetry groups

$$G \mapsto G^1 \mapsto G^2 \mapsto \ldots \mapsto G^{k-1} \mapsto G^k \mapsto \ldots \mapsto G^l,$$

i.e. the transition $G^{k-1} \mapsto G^k$, $(1 \leq k \leq l)$, is a derivation of simple antisymmetry groups using $AC(G^{k-1})$, followed by the induced transition $AC(G^{k-1}) \mapsto AC(G^k)$, $(1 \leq k \leq l-1)$. All the AC of induced series consist of the same number of generators.

The said can be illustrated by the example of derivation of multiple antisymmetry groups from the plane symmetry group **pm**:

pm $\{a, b\}(m)$ $AC : \{m, ma\}\{b\} \simeq \{A, B\}\{C\}$.

For $m = 1$ five groups of simple antisymmetry of the M^1-type are obtained:

$\{A, B\}\{C\}$
$\{E, E\}\{e_1\} \mapsto \{A, B\}\{C\}$
$\{e_1, e_1\}\{E\} \mapsto \{A, B\}\{C\}$
$\{e_1, e_1\}\{e_1\} \mapsto \{A, B\}\{C\}$
$\{E, e_1\}\{E\} \mapsto \{A\}\{B\}\{C\}$
$\{E, e_1\}\{e_1\} \mapsto \{A\}\{B\}\{C\}$.

In the first three cases AC remains unchanged, but in two other cases AC is transformed into the new $AC : \{A\}\{B\}\{C\}$. To continue the derivation of multiple antisymmetry grups of the M^m-type from the symmetry group **pm**, only the derivation of simple antisymmetry groups from $AC : \{A\}\{B\}\{C\}$ is indispensable. This AC is trivial and gives seven groups of simple antisymmetry. If $AC : \{A, B\}\{C\}$ is denoted by 3.2 and $AC : \{A\}\{B\}\{C\}$ by 3.1, then the result obtained can be denoted in a symbolic form by $3.2 \mapsto 2(3.1) + 3(3.2)$. Then we have

$N_1(\mathbf{pm}) = N_1(3.2) = 5 \quad N_1(3.1) = 7$

$N_2(\mathbf{pm}) = N_2(3.2) = 2N_1(3.1) + 3N_1(3.2) - 5 \cdot 1 =$

$= 2(N_1(3.1) - 1) + 3(N_1(3.2) - 1) = 2\underline{6} + 3\underline{4} =$

$= 2N_1(3.1) + 3N_1(3.2) - N_1(3.2) = 2N_1(3.1) + 2N_1(3.2) = 24.$

The meaning of every step in the mentioned computation is:

1) substruction of the number $N_1(3.2)$, i.e. of the five groups of uncomplete multiple antisymmetry of the $2M$-type;

2) every group of the M^1-type gives exactly one of these $2M$-type groups, so we obtain $2\underline{6} + 3\underline{4}$ groups of complete multiple antisymmetry of the M^2-type [5,8,10]. This step contains also essential data for the calculation of the number N_3: 6 groups mentioned possess AC 3.1, two of 4 groups mentioned possess AC 3.1 and two AC 3.1. Among five groups of uncomplete multiple antisymmetry of the $2M$-type there are three groups with AC 3.2 and two with AC 3.1;

3) by substitution $5 = N_1(3.2)$ we obtain $N_2(3.2)$ expressed by $N_1(3.1)$ and $N_1(3.2)$, i.e. $2N_1(3.1) + 2N_1(3.2)$. The sum of coefficients corresponding to the numbers N_1 in the last line gives $N_2(\mathbf{pm}) = 24$.

$N_3(\mathbf{pm}) = N_3(3.2) = 2 \cdot 6N_1(3.1) + 3 \cdot (2N_1(3.1) + 2N_1(3.2)) - 24 \cdot 3 =$

$= 18N_1(3.1) + 6N_1(3.2) - 24 \cdot 3 = 18(N_1(3.1) - 3) + 6(N_1(3.2) - 3) =$

$= 18\underline{4} + 6\underline{2} = 18N_1(3.1) + 6N_1(3.2) - 3(2N_1(3.1) + 2N_1(3.2)) =$

$= 12N_1(3.1) = 84 \quad (N_2(3.2)) = 12.$

Consequently, the method proposed makes possible complete reduction of the theory of multiple antisymmetry to the theory of simple antisymmetry. This refers not only to the possibility of computation of the numbers N_m and (N_{m-1}), but also to the possibility of applying the method of partial cataloguation of multiple antisymmetry groups of the M^m-type [8]. If we take the advantage of the suggested reduction, the use of this method is considerably simplified and demands only the catalogues of the simple antisymmetry groups of the M^1-type obtained from non-isomorphic AC.

4. Non-isomorphic AC formed by $1 \leq l \leq 4$ generators

As it is shown in §3 the theory of multiple antisymmetry can be reduced to the theory of simple antisymmetry. For that it is necessary to know all non-isomorphic AC formed by l generators. Non-isomorphic antisymmetry characteristics formed by $1 \leq l \leq 4$ generators are investigated in [9]. As the result of their study, the catalogue of that AC formed by $1 \leq l \leq 4$ generators, and the tables of the corresponding numbers N_m, are obtained. The completness of this catalogue is proved for $l \leq 2$, but for $l \geq 3$, having in mind a great number of possible cases which we must consider, the completness is not proved, and there is a possibility that some AC are not included into the catalogue.

In this catalogue for every AC is given a list of corresponding simple antisymmetry groups of the M^1-type, connections between AC in the case of transition from $m = 1$ to $m = 2$ and tables of the numbers N_m. The notation used and the method for obtaining results are the same as in the example of the symmetry group **pm** given in §3. In AC by parenthesis ( ) is denoted the obligation of cyclic permutation of appertaining elements, by [ ] the obligation of simultaneous commutation of elements; the elements in // parenthesis remain fixed on their places. AC obtained in all previous studies of the theory of simple and multiple antisymmetry for $1 \leq l \leq 4$ are included in this catalogue. The list is the following:

$l = 1$

1.1 $\{A\}$.

$l = 2$

2.1 $\{A\}\{B\}$;
2.2 $\{A, B\}$;
2.3 $\{A, B, AB\}$.

$l = 3$

3.1 $\{A\}\{B\}\{C\}$;
3.2 $\{A, B\}\{C\}$;
3.3 $(A, B, C, AB, AC, BC, ABC)$;
3.4 $\{A, B\}\{C, ABC\}$;
3.5 (A, B, C);
3.6 (A, B, C, ABC);
3.7 $\{A, B, C\}$;
3.8 $\{A, B\}, \{C, ABC\}\}$;
3.9 $\{A, B, C, ABC\}$;
3.10 $\{A, B, C, AB, AC, BC, ABC\}$.

$l = 4$

4.1 $\{A\}\{B\}\{C\}\{D\}$;
4.2 $\{A, B\}\{C\}\{D\}$;
4.3 $([A, B], [C, ABC], [D, ABD], [AC, BC], [AD, BD], [CD, ABCD], [ACD, BCD])$;
4.4 $\{A, B\}\{C, D\}\{AC, BD\}$;
4.5 $\{A\}\{B, C\}\{D, BCD\}$;
4.6 $\{A, B\}\{C, D\}$;
4.7 $\{B, AB\}\{C, AC\}\{D, AD\}$;

4.8 $\{A\}(B, C, D)$;

4.9 $(/A, B/, /C, ABC/, /D, ABD/, /ACD, BCD/)$;

4.10 $\{A, B, C\}\{D\}$;

4.11 $\{\{A, B, \{CA, CB\}\}\{D, CD\}$;

4.12 $\{[A, B], [C, D]\}$;

4.13 $\{\{B, AB\}, \{C, AC\}\}\{D, AD\}$;

4.14 (A, B, C, D);

4.15 $(C, A, CA)\{(B, C, ABC), (BD, BCD, ABCD)\}$;

4.16 $\{\{A, B\}, \{C, D\}\}$;

4.17 $(\{A, B\}, \{C, ABC\}, \{D, ABD\}, \{AC, BC\}, \{AD, BD\}, \{CD, ABCD\}, \{ACD, BCD\})$;

4.18 $\{A, B, AB\}\{C, D\}$;

4.19 $\{A, B, C, ABC\}\{D\}$;

4.20 $\{\{A, B\}, \{C, ABC\}\}\{\{D, ABD\}, \{ACD, BCD\}\}$;

4.21 $(\{A, AD\}, \{B, BD\}, \{C, CD\})$;

4.22 $\{A, B, C, D\}$;

4.23 $(\{A, B\}, \{C, ABC\}\}\{\{D, ABD\}, \{ACD, BCD\})$;

4.24 $\{\{B, AB\}\{C, AC\}, \{D, AD\}\}$;

4.25 $\{\{\{A, B\}, \{C, ABC\}\}, \{\{D, ABD\}, \{ACD, BCD\}\}\}$;

4.26 $\{A, B, C, ABC\}\{D, ABD, ACD, BCD\}$;

4.27 $\{\{A, B\}, \{C, D\}, \{AC, BD\}\}$;

4.28 $\{\{A, B\}, \{C, ABC\}, \{D, ABD\}, \{ACD, BCD\}\}$;

4.29 $\{A, B, C, D, ABC, ABD, ACD, BCD\}$;

4.30 $\{A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD\}$.

Besides all AC found in practice during previous studies of the theory of simple and multiple antisymmetry for $1 \leq l \leq 4$, in this catalogue there are some AC which are not found before.

Conjecture 1. *Every abstract algebraic term formed in accordance with Definition 1 is AC of some symmetry group.*

Most of the AC given in this catalogue, which are not found in earlier practice, satisfy Conjecture 1. For example, AC 4.22 corresponds to the symmetry group **mmmm** of the category G_{40}, and AC 4.30 corresponds to the symmetry group **P1111** of the category G_4.

If Conjecture 1 is valid, AC 4.21 and 4.22 are counter-examples of the supposition [1, pp. 138] that equality of the first and last members of the series $N_m(G)$ and $N_m(G')$ implies equality of the second members of these series.

Conjecture 2. *Every series N_m obtained from AC_l formed by l generators is identical with some series (N_{m+1}) obtained from corresponding AC_{l+1} formed by $l+1$ generators.*

As the examples of AC_{l+1} and AC_l which satisfy the Conjecture 2 for $1 \leq l \leq 4$, it is possible to notice the pairs of AC: 2.2 and 1.1, 3.4 and 2.1, 3.8 and 2.2, 3.9 and 2.3, 4.7 and 3.1, 4.13 and 3.2, 4.17 and 3.3, 4.20 and 3.4, 4.21 and 3.5, 4.23 and 3.6, 4.24 and 3.7, 4.25 and 3.8, 4.28 and 3.9, 4.19 and 3.10.

Conjecture 3. *Let AC_l formed by generators $A_1, \dots, A_l$ be given. Then by the substitution $A'_i = A_iA_{l+1}$, $i = \overline{1,l}$, can be obtained a new AC_{l+1}, such that AC_l and AC_{l+1} satisfy Conjecture 2.*

The study of particular non-isomorphic AC for $l > 4$ is almost a technical problem. However, a proof of completness of the catalogue of non-isomorphic AC for $l > 2$ is immensely important and one of the aims of future studies of the theory of simple and multiple antisymmetry must be the construction of an algorythm, which makes possible direct derivation of all non-isomorphic AC formed by l generators.

In many cases, especially for AC with a large number of generators, for the computing of numbers N_m it is possible to use the direct product of AC.

5. Direct product of AC

Definition 2. Let AC' and AC" with disjoint sets of generators be given. The new $AC = AC'AC$" obtained by adding in writing AC" to AC' is called the direct product of AC' and AC".

Theorem 3. *Let N_m, N'_m, N_m" be the series of numbers defined by AC, AC', AC" respectively. Then the relationship*

$$N_m = \sum_{\substack{k+l\geq m,\\ m\geq k,l\geq 0}} 2^{(m-k)(m-l)}C(m, m-k, m-l)N'_kN_m"$$

holds, where

$$C(l,k,m) = \frac{(2^l-1)(2^{l-1}-1)\dots(2^{l-k-m+1}-1)}{(2^k-1)(2^{k-1}-1)\dots(2-1)(2^m-1)(2^{m-1}-1)\dots(2-1)}.$$

As an illustration of the AC which satisfy Theorem 3, we are giving the following example

$AC' = 2.2 = \{A, B\}$ $N_1(2.2) = 2$ $N_2(2.2) = 3$

$AC" = 2.1 = \{C\}\{D\}$ $N_1(2.1) = 3$ $N_2(2.1) = 6$

$AC = \{A, B\}\{C\}\{D\} = 4.2.$

In accordance with Theorem 3,

$N_1(4.2) = 2{\cdot}3+2+3 = 11$ $N_2(4.2) = 3{\cdot}6+3+6+3{\cdot}2{\cdot}6+3{\cdot}3{\cdot}3+6{\cdot}2{\cdot}3 = 126$

$N_3(4.2) = 28\cdot 2\cdot 6 + 28\cdot 3\cdot 3 + 42\cdot 3\cdot 6 = 1344$

$N_4(4.2) = 560\cdot 3\cdot 6 = 10080.$

Other examples of $AC = AC'AC''$ from the catalogue of non-isomorphic AC for $1 \leq l \leq 4$ are $2.1 = (1.1)(1.1)$, $3.1 = (2.1)(1.1)$, $3.2 = (2.2)(1.1)$, $4.1 = (3.1)(1.1) = (2.1)(2.1)$, $4.2 = (3.2)(1.1)$, $4.6 = (2.2)(2.2)$, $4.8 = (3.5)(1.1)$, $4.10 = (3.7)(1.1)$, $4.18 = (2.3)(2.2)$, $4.19 = (3.9)(1.1)$.

6. Tables of numbers N_m

As the result we have the table survey of the numbers N_m for all nonisomorphic AC formed by $1 \leq l \leq 4$ generators:

$l = 1$

	N_1
1.1	1

$l = 2$

	N_1	N_2
2.1	3	6
2.2	2	3
2.3	1	1

$l = 3$

	N_1	N_2	N_3
3.1	7	42	168
3.2	5	24	84
3.3	4	24	96
3.4	4	15	42
3.5	3	14	56
3.6	3	12	42
3.7	3	10	28
3.8	3	9	21
3.9	2	4	7
3.10	1	1	1

$l = 4$

	N_1	N_2	N_3	N_4
4.1	15	210	2520	20160
4.2	11	126	1344	10080
4.3	9	120	1440	11520
4.4	9	108	1260	10080
4.5	9	84	756	5040
4.6	8	75	714	5040
4.7	8	63	462	2520
4.8	7	74	840	6720
4.9	7	66	672	5040
4.10	7	58	504	3360
4.11	7	54	420	2520
4.12	6	57	630	5040
4.13	6	39	252	1260
4.14	5	54	630	5040
4.15	5	44	448	3360

4.16	5	39	357	2520
4.17	5	36	264	1440
4.18	5	34	266	1680
4.19	5	28	168	840
4.20	5	27	147	630
4.21	4	23	154	840
4.22	4	22	147	840
4.23	4	21	126	630
4.24	4	19	98	420
4.25	4	18	84	315
4.26	4	16	63	210
4.27	3	21	210	1680
4.28	3	10	35	105
4.29	2	4	8	15
4.30	1	1	1	1

References

[1] A. M. Zamorzaev, *Teoriya prostoi i kratnoi antisimmetrii*, Shtiintsa, Kishinev (1976).

[2] A. M. Zamorzaev and A. F. Palistrant, *Antisymmetry, its generalizations and geometrical applications*, Z. Kristallogr. **151** (1980), 231-248.

[3] A. M. Zamorzaev, E. I. Galyarskij and A. F. Palistrant, *Colored symmetry, its generalizations and applications*, Kishinev, Shtiintsa (1978).

[4] A. M. Zamorzaev, Yu. S. Karpova, A. P. Lungu and A. F. Palistrant, *P-symmetry and its further development*, Shtiintsa, Kishinev (1986).

[5] H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups*, Springer Verlag, Berlin, Heidelberg, New York, 1980.

[6] S. V. Jablan, *Mackay Groups*, Acta Cryst **A 49** (1993), 132–137.

[7] S. V. Jablan, *A New Method of Generating Plane Groups of Simple and Multiple Antisymmetry*, Acta Cryst. **A 42** (1986), 209–212.

[8] S. V. Jablan, *A New Method of Deriving and Cataloguing Simple and Multiple Antisymmetry G_3^l Space Groups*, Acta Cryst **A 43** (1987), 326–337.

[9] S. V. Jablan, *Algebra of Antisymmetric Characteristics*, Publ.Inst. Math. **47 (61)** (1990), 39–55.

[10] S. V. Yablan and A. F. Palistrant, *Prostranstvennye gruppy prostoj i kratnoj tsvetnoj antisimmetrii*, Kristallografiya **38 2** (1993), 4–11.

[11] S. V. Jablan, *$(p2, 2^l)$-symmetry Three-dimensional Space Groups $G_3^{l,p2}$*, Acta Cryst. **A 48** (1992).

[12] S. V. Jablan, *Pawley Multiple Antisymmetry Three-dimensional Space Groups $G_3^{l,p'}$*, Acta Cryst. **A 49** (1993), 41–54.

Phylosophical Faculty, Department of Mathematics, 18000 Niš, Ćirila i Metodija 2, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 723–732
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

EUCLID – THE GEOMETRY THEOREMS PROVER

Predrag Janičić and Stevan Kordić

1. Introduction

Geometry is one of the mathematical disciplines demanding a big deal of the human intuition. That's why it is chalenging task to make a program solving a geometry problems. Program EUCLID proves theorems of geometry in a intuitive, geometrical way (*more geometrico*), and presents proves in a natural language form. Besides, the mechanism and the basic principles of the prover EUCLID led us to the new form of the foundation of geometry and the new classification of geometrical axioms.

Program EUCLID was written in Arity PROLOG, but essential machanism of the prover does not rely on PROLOG mechanisms. Despite the limited resources of Arity PROLOG, the program was written in PROLOG because of its suitable characteristics: flecsibility, mechanism of unification etc.

2. The Fundamentals of the Prover EUCLID

There are three modules in program EUCLID: the module of axioms, the knowledge-pool and the proving mechanism. Although these modules are independent they are built as a coherent system. Besides, these modules are related by internal language in which all knowledge and conclusions are expressed. The final output - proof of the certain theorem is written in a natural language form. Because of its importance, first of all, let us focus our attention at internal language.

3. Internal Language

The internal language L of the prover covers all objects and relations accuring in geometrical axioms. Also, theorem that is to be proved has to be expressed in the internal language, so the internal language is important for user, also. All relations of the internal language L (including unary relations defining objects) are shown in table 1.

predicate	we read
$t(a)$	a is a point
$l(b)$	b is a line
$p(c)$	c is a plane
$identical(a,b)$	a and b are identical
$non\_identical(a,b)$	a and b are not identical
$i(a,b)$	a and b are incident
$non\_i(a,b)$	a and b are not incident
$b(a,b,c)$	b lies between a and c
$non\_b(a,b,c)$	b does not lie between a and c
$c(a,b,c,d)$	pair (a,b) is congruent to pair (c,d)
$collinear(a,b,c)$	a,b and c are collinear
$non\_collinear(a,b,c)$	a,b and c are not collinear
$coplanar(a,b,c,d)$	a,b,c and d are coplanar
$non_{c}oplanar(a,b,c,d)$	a,b,c and d are not coplanar
$intersect(a,b)$	a intersects b
$non\_intersect(a,b)$	a does not intersect b

Table 1. Relations (primitive and defined) in system EUCLID

Internal representation of relations (except for unary relations) has one argument more then in a table, and a value of that argument is an index of relation in a knowledge-pool.

We denoted by $\mathcal{E}_L$ the class of all geometry theorems that can be expressed either as:

$$\forall x_1, \forall x_2, ...\forall x_n, \exists Y_1, \exists Y_2, ...\exists Y_m(\phi(x_1, x_2, ...x_n) \Rightarrow \psi(x_1, x_2, ...x_n, Y_1, Y_2, ...Y_m))$$

or as:

$$\forall x_1, \forall x_2, ...\forall x_n, \exists Y_1, \exists Y_2, ...\exists Y_m(\phi(x_1, x_2, ...x_n) \Rightarrow$$
$$\psi_1(x_1, x_2, ...x_n, Y_1, Y_2, ...Y_m) \vee \psi_2(x_1, x_2, ...x_n, Y_1, Y_2, ...Y_m) \vee ...$$
$$... \vee \psi_k(x_1, x_2, ...x_n, Y_1, Y_2, ...Y_m))$$

or as:

$$\forall x_1, \forall x_2, ...\forall x_n(\phi(x_1, x_2, ...x_n) \Rightarrow \psi(x_1, x_2, ...x_n))$$

or as:

$$\exists Y_1, \exists Y_2, ...\exists Y_m(\psi(Y_1, Y_2, ...Y_m))$$

where ϕ, ψ and ψ_i are conjunctions of L relations ranging over some of the arguments $x_1, x_2, ...x_n$ and $Y_1, Y_2, \dots Y_m$.

The first mentioned form we shall call *universal-existential form* ($\forall - \exists$), the second *universal-existential-disjunctive form* ($\forall - \exists - \vee$), the third *universal form* (∀) and the fourth *existential form* (∃). All of them we shall denote by $\mathcal{F}$.

We interpret the $\mathcal{E}_L$ theorems in program EUCLID in a appropriate PROLOG way. The class $\mathcal{E}_L$ will be the subject of the following text.

4. The Module of Axioms

Module of axioms consists of all geometrical axioms whithout continuity axioms and so called ADT module. ADT module consists of the following axiom, definitions and trivial theorems:

- The additional geometry axiom:

$$\forall a, \forall b, \forall c(i(a,b) \wedge i(b,c)) \Rightarrow i(a,c))$$

- The identity axioms:

$$\forall a\, identical(a,a)$$

$$\forall a, \forall b\,(identical(a,b) \Rightarrow identical(b,a))$$

$$\forall a, \forall b, \forall c\,(identical(a,b) \wedge identical(b,c) \Rightarrow identical(a,c))$$

- The substitution axioms:

$$\forall a_1, \ldots \forall a_i, \ldots \forall a_n, \forall b(\phi(a_1, \ldots a_i, \ldots a_n) \wedge identical(a_i, b)$$

$$\Rightarrow \phi(a_1, \ldots b, \ldots a_n))$$

  where ϕ is a L relation.
- The definitions:
  If points a, b and c and line l are such that $i(a,l), i(b,l), i(c,l)$, then we shall say that the points a, b and c are *collinear.*

  If points a, b, c and d and plane p are such that $i(a,p), i(b,p), i(c,p)$, $i(d,p)$, then we shall say that the points a, b, c and d are *coplanar.*

  If a is a point and $i(a,b), i(a,c)$, then we say that b intersects c.
- The trivial theorems:
  If points a, b and c and line l are such that $i(a,l), i(b,l), non\_i(c,l)$, then we shall say that the points a, b and c are non collinear.

  If points a, b, c and d and plane p are such that $i(a,p), i(b,p), i(c,p)$, $non\_i(d,p)$, then we shall say that the points a, b, c and d are not coplanar.

Let us note that all axioms, definitions and theorems just listed above are of one of the $\mathcal{F}$ forms. Also, each geometry axiom (excluding continuity axiom) can be put in one of the $\mathcal{F}$ forms [1] .

[1] Tarski wrote about these forms of geometry axioms in a different context.

We denoted by $\mathcal{E}$ the class of all $\mathcal{E}_L$ theorems which can be established by use of the elements from the module of axioms. It can be shown that program EUCLID can prove all $\mathcal{E}$ theorems.

Although prover EUCLID does not use any set-theory segment, class $\mathcal{E}$ is wide enough to cover great part of usual elementary geometry courses. Also, any geometry theorem could be included into the module of axioms for the sake of more efficient and simplier proving process.

According to the mechanism of the prover, order of the axioms is very important and determine the way of establishing of a theorem. Efficience of the prover is related to the order of axioms and this inspired us for the new classification of geometrical axioms. There are divided into five groups:

-"identity" axioms;

-"unproductive" axioms;

-"branching" axioms;

-"productive" axioms;

-"strongly productive" axioms.

Each group of axioms has a different status, and order of the axioms in a each group is also of the great importance. This order determines efficience of the prover.

Proving mechanism is the essential part of the program EUCLID and it is based upon so called "sentinel-principle". That principle enables proving of all E theorems in a finite number of steps.

5. The Knowledge-pool

All objects and knowledge which are used in prover EUCLID are expressed in the knowledge-pool. In the begining of the proving process for a certain theorem, knowledge-pool contains only datas about objects (denoted by letters) and relations given by theorem itself. During the proving process, all objects and relations inferenced upon module of axioms are being added to knowledge-pool with their unique (natural number) index. For unary predicates (defining objects) this index is their only argument and it is their identifier.

The state of the knowledge-pool is determined by value of the so-called sentinel. The sentinel determines the set of objects from the knowledge-pool that are accesable for certain axioms in a process of proving.

The current state of the knowledge-pool is determined by the proving mechanism. In case of branching (in process of theorem proving) parts of knowledge-pool related to disjunctive branches are independent and this saves integrity of the knowledge-pool as a knowledge base.

6. The Proving Mechanism

For sake of illustration, let us see the key PROLOG predicate of the proving mechanism:

```
proof :- contradiction.
proof :- proved.
proof :- adt(M),proof.
proof :- ax_u(M),proof.
proof :- assumption(P,NotP),
         index(B),
         retract(comments(true)),assert(comments(fail)),
         push(B,0),justified(P,IP),pop(B,0),
         push(B,0),justified(NotP,INotP),pop(B,0),
         retract(comments(fail)),assert(comments(true)),
         ((IP=true,INeP=true,proofp([P,NeP]));
         (IP=true,INeP=false,proofp([P]));
         (IP=false,INeP=true,proofp([NeP]))).
proof :- ax_b(M).
proof :- ax_p(M),proof.
proof :- sentinel(G),
         first_object(G,N),N1 is N+1,
         retract(sentinel(G)),
         assert(sentinel(N1)),
         proof.
proof :- ax_sp(M),proof.
```

In the very beginning of the program's work, there are to be given assumptions of a certain geometry theorem and its conclusion. Before activating the key PROLOG predicate in the proving mechanism - predicate "proof", knowledge-pool contains only datas about objects (denoted by letters) and relations given by theorem assumptions, and all that objects are accesable for the module of axioms.

The key part of the algorithm can be (unpresicely) defined as follows:

(1) Check if there is a contradiction in the knowledge-pool; if there is, report it and finish proving process in the current branch of a proof; if there is not, go to step (2);

(2) Check if there are enough knowledge in the knowledge-pool to conclude that the theorem is proved; If there are, report it, define objects and relations making conclusion of the given theorem and finish proving process in the current branch of a proof; if there are not, go to step (3);

(3) If possible, apply one of the ADT element and go to step (1); if not, go to step (4);

(4) If possible (according to current state of knowledge-pool and sentinel value), apply one of the unproductive axioms and go to step (1); if not, go to step (5);

(5) If possible, assume that some relation over some objects from current knowledge-pool holds and add this assumption to the knowledge-pool as a fact; similary, assume negation of that relation; make proves for both cases; if it is not possible to assume any ralation over objects from current knowledge-pool, go to step (6);

(6) If possible (according to current state of knowledge-pool and sentinel value), apply one of the branching axioms and make proves for all its branches; if not, go to step (8);

(7) If possible (according to current state of knowledge-pool and sentinel value), apply one of the productive axioms and go to step (1); if not, go to step (8);

(8) If possible (according to current state of knowledge-pool and sentinel value), apply one of the strongly productive axioms and go to step (1); if not, go to step (9);

(9) Select the object with the least index greater then current value of the sentinel; give the value of this index to the sentinel; go to step (1).

This algorithm can be modified in such a way to prove many theorems more efficiently, but that version of the prover can not prove all $\mathcal{E}$ theorems.

The sentinel has a key role in determining which objects from the knowledge-pool are accesable in certain moment of the proving process. It is a sentinel principle which guaranatee ability of proving all $\mathcal{E}$ theorems.

7. The Sentinel Principle

The sentinel value in each moment of the proving process is determining a set of accesable objects for geometry axioms. In the proving process all elements of ADT module could be applied no matter to the sentinel value (i.e. all objects from the knowledge-pool are accesable for them). Immidiate after entering the assumtions of the theorem which is to be proved, all objects occuring in these assumtions are accesable. During the proving process, the knowldge-pool is spreading (according to foregiven algorithm and by application of the geometry axioms and ADT elements). If step (9) of the algorithm is reached in the proving process, none axiom or ADT element could be applied according to current state of the knowledge-pool and the sentinel value. Then the set of accesable objects, i.e. the sentinel value is to be changed. The sentinel is getting a least index value of all objects which have indexes greater then current sentinel value. It means that the first object whose existency was established since the last change of the sentinel value will be added to the set of accesable objects. Then the proving process is continuing with application of axioms. That is how there is ensured inferencing of all possible conclusions for given set of relations and accesable objects. Also, that is how there is ensured inferencing of all concluslions relevant for the given theorem, and enabled occuring of any "infinite" branch in the proving process.

Let's point out (once again) that forementioned mechanism ensures proving of all $\mathcal{E}$ theorems in a finite number of steps (but, many theorems could not be proved because of the limited recources of the Arity PROLOG). Also, let's point out that it is not a difficult task to extend the program EUCLID in such a way to optimize its finished proves (i.e. to eliminate all unneccesary steps).

8. The EUCLID Axiomatic System of Elementary Geometry

In the axiomatic system of elementary geometry (geometry without axiom of continuity) inspired by the program EUCLID, as the primitive notions we take one fixed set $\mathcal{G}$ (*geometry objects set*) and seven primitive relations over geometry objects: three unary relations *is a point*, (denoted by t), *is a line* (l) and *is a plane* (p), two binary relations *identical* (denoted by *identical*) and *incidental* (i), one ternary relation *between* (denoted by b) and one quaternary relation *congruent* (denoted by c). (Instead of writing *For geometrical object a such that it holds t(a) ...* we shall write *For point a ....*) We also use negations of these relations (except for congruence - its negation does not occur in any axiom). As a defined relations, we use relations *colliner* and *coplanar* (with their negations) and relation *intersect* (we don't use definition for relation non-intersect, but we use it as a assumtion of a theorem or as a assumption during the proving process). We don't use any set-theory segment. We use "classical" geometry axioms, additional incidence axiom and "identity" axioms (see section 4) and use them according to rules of Gentzen's NK calculus. All axioms are divided into five groups:

-"identity" axioms (see section 4);

-"unproductive" axioms (axioms of the form $\forall$);

-"branching" axioms (axioms of the form $\forall - \exists - \vee$);

-"productive" axioms (axioms of the form $\forall - \exists$);

-"strongly productive" axioms (axioms of the form $\exists$).

(according to foregiven classification additional incidence axiom belongs to group of unproductive axioms)

In the forementioned modified version of the program EUCLID some of the productive axioms was put into the group of stronly productive axioms. That version of the program makes some proves more efficiently, but can't prove all $\mathcal{E}$ theorems.

During the proving proces prover EUCLID is denoting all new geometry objects by natural numbers and practicaly is making a model (or models) of elementary geometry in a set of natural number, or more precisely, part of that model sufficient to prove the given theorem. It means that, indepent of any concrete theorem, algorithm EUCLID could generate a model of elementary geometry in a set of natural numbers in an infinite (but recursive!) process.

Appendix

For the sake of traditional approach, a prover EUCLID gives proves according to traditional sense of relation incidence as a set relation.

Predicates occuring in the program's proves (except unary ones) have one additional argument and that is their unique index (see section 3).

Example 1.

Theorem: Just one plane passes throu two intersecting and distinct lines.

```
    ***** EUCLID v4.00 - Geometry Theorems Prover *****
    Enter list of assumtions:
[l(a),l(b),non_identical(a,b),intersect(a,b)].
    There exists point 3 such that 3 lies on lines a and b.
    Enter conclusion of the theorem:
p(X),e(a,X,_),e(b,X,_),unique(X). By axiom 1.1, there exists a point 6, distinct from 3, such that 6 lies on line a. Let us assume i(6,b).
        By axiom 1.3, since lines a and b have two common points (3 and
        6), then, lines a and b are identical.
        Conradiction: non_identical(a,b) and identical(a,b)!
Let us assume non_i(6,b).
        By axiom 1.1, there exists a point 14, distinct from 3, such that 3
        lies on line b.
        There must be non_identical(6,14) (oposite assumtion is obviously in
        contradiction to other assumtions).
        Since points 3 and 14 lie on line b and point 6 doesn't lie on line b,
        then, points 3, 6 and 14 are non_collinear.
        Points 3, 6 and 14 are non_collinear, thus point 14 doesn't lie on line
        a.
        By axiom 1.2, there exists a line 26, such that 26 passes throu points
        6 and 14.
```

Points 6, 14 and 3 are non_collinear, so point 3 doesn't lie on line 26.
By axiom 1.5, there exists a plane 30, such that points 3, 6 and 14 lie on plane 30.
By axiom 1.7, since points 3 and 6 lie on plane 30, line a lie on plane 30.
By axiom 1.7, since points 3 and 14 lie on plane 30, line b lie on plane 30.
Therefore there exists the plane we are seeking (plane 30).
Let us prove that this is only one such plane.
Let us assume the opposite - there exists one more such plane (36), planes 30 and 36 are non identical and let us this is in contradiction.
Point 3 lies on line a and line a lies on plane 36, thus point 3 lies on plane 36.
Point 6 lies on line a and line a lies on plane 36, thus point 6 lies on plane 36.
Since point 14 lies on line b and line b lies on plane 36, thus point 14 lies on plane 36.
By axiom 1.7, since points 6 and 14 lie on plane 30, thus line 26 lie on plane 30.
There must be non_identical(a,26) (opposite assumtion is obviously in contradiction to other assumtions).
There must be non_identical(b,26) (opposite assumtion is obviously in contradiction to other assumtions).
By axiom 1.6, since planes 30 and 36 have three commom non_collinear points (points 3, 6 and 14), planes 30 and 36 are identical.
Conradiction: non_identical(30,36) and identical(30,36)!
Therefore p(30),e(a,30,34),e(b,30,35),unique(30), QED.

Exaple 2.

Theorem: Relation of congruence is reflectiv one.

***** Euclid v4.00 - Geometry Theorems Prover *****

Enter list of assumptions:

[t(a),t(b),non_identical(a,b)].

Enter conclusions of the theorem:

c(a,b,a,b,_).

By axiom 3.2, it holds c(a,b,b,a).

By axiom 3.3, since c(a,b,b,a) and c(a,b,b,a), then c(b,a,b,a).

By axiom 3.2, it holds c(b,a,b,a).

By axiom 3.3, since c(b,a,a,b) and c(b,a,a,b), then c(a,b,a,b).

```
    Therefore, c(a,b,a,b,5), QED.
```

Example 3.

Theorem: Relation of non_itersection for lines lieing in plane is transitive one.

```
    ***** Euclid v4.00 - The geometry Theorems Prover *****
    Enter list of assumtions:
[p(alpha),l(a),l(b),l(c), non_identical(a,b), non_identical(a,c),
non_identical(b,c), i(a,alpha), i(b,alpha), i(c,alpha), non_intersect(a,b),
non_intersect(b,c)].
    Enter conclusions of the theorem:
non_intersect(a,c,_).
    Let us assume intersect(a,c)
            There exists point 16 such that 16 lies on lines a and c.
            Since point 16 lies on line a and line a lies on plane alpha, then, point
            16 lies on plane alpha.
            There must be non_i(16,b) (oppoisite assumtion obviously is in a
            contradiction with other assumtions).
            By axiom 4.1, there exists at most one line, such that it lies on plane
            alpha, passes throu point 16 and does not intersect line b, then lines
            a and c are identical.
            Contradiction: identical(a,c) and non_identical(a,c)!
Let us assume non_intersect(a,c).
            Therefore, non_intersect(a,c,24), QED.
```

University of Belgrade, Faculty of Mathematics, Studentski trg 16, 11000 Beograd

FILOMAT (Niš) 9:3 (1995), 733–742
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

AUTOMATIC THEOREM PROVING IN FIELD THEORY USING QUANTIFIER ELIMINATION

Aleksandar Jovanović and Žarko Mijajlović

ABSTRACT. In this paper we describe a new method of elimination of quantifiers for the theories of algebraically closed fields and theory of ordered real closed fields which may be used for the theorem provers for these theories. The method is based on the properties of resultants of polynomials.

1. Introduction

One could say that mathematics was introduced in logic by Tarski and Gödel while for Abraham Robinson and A. Malcev could be said that they introduced logic in mathematics. Namely, today probably most important applications of logic in other parts of mathematics (nonstandard analysis and model-theoretic algebra) originate in work of A. Robinson. First contributions of this kind in algebra were given by A. Malcev in 1936. The Robinson's solution of Seventeenth Hilbert problem by methods of mathematical logic, more precisely methods of model theory, represents an important contribution to the model-theoretic algebra. The solution is based on the method of elimination of quantifiers and notion of model completeness, the model-theoretic version of the elimination of quantifiers. Beside, this notion can be understood as a transfer principle, which is of significant importance for the applications in algebra.

Definition 1. A theory T in the first order predicate calculus admits elimination of quantifiers if for every formula φ or T there is a formula ψ in the language of T, without quantifiers such that: $T \vdash \varphi \Leftrightarrow \psi$.

Let us remind that the following theorem is basic for the model theoretic solution of the seventeenth Hilbert's problem.

We wish to thank to Professor Albert Dragalin for useful comments and remarks.

Theorem 1. (A. Tarski, 1948) *Theory of ordered real closed fields admits elimination of quantifiers.*

Theories which admit the elimination of quantifiers have this interesting property:

Every theory which admits eliminations of quantifiers is model complete.

In order to explain this notion, suppose that T is any first order theory in the language L. Let $\mathbf{A}$ and $\mathbf{B}$ be any models (i.e. operational-relational structures) of L. Model $\mathbf{A}$ is an elementary submodel of model $\mathbf{B}$, or $\mathbf{B}$ is an elementary extension of $\mathbf{A}$ if the following conditions are satisfied.

1. $\mathbf{A}$ is a submodel of $\mathbf{B}$.
2. For every formula $\varphi(\bar{x})$ of L and every $\bar{a} \in A$,

$$\mathbf{A} \models \varphi(\bar{a}) \qquad \text{if and only if} \qquad \mathbf{B} \models \varphi(\bar{a}).$$

The fact that $\mathbf{A}$ is an elementary submodel of $\mathbf{B}$ we denote by $\mathbf{A} \prec \mathbf{B}$.

Definition 2. Theory T is model complete iff for any two models $\mathbf{A}$ and $\mathbf{B}$ of theory T if $\mathbf{A}$ is a submodel of $\mathbf{B}$ then $\mathbf{A}$ is an elementary submodel of $\mathbf{B}$.

2. Quantifier elimination for the theory of algebraically closed fields

The axioms of the theory of algebraically closed fields are the axioms for fields and the following set of formulas, expressing that every polynomial of degree ≥ 1 has a root. Let T be the field theory and T^* the theory of algebraically closed fields. For example, the fields of complex numbers and algebraic numbers are models of the theory T^*.

Examples of quantifier elimination for theory T^* are known for long time in classical algebra. One of the best known, which will be used here is the Resultant Theorem.

Definition 3. Let $a(x) = \sum_{i \leq m} a_i x^i$, $b(x) = \sum_{j \leq n} x^j$ be complex polynomials. The resultant of polynomials a and b is the determinant

$$\operatorname{Res}(a,b) = \begin{vmatrix} a_0 & a_1 & \dots & a_m & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_m & \dots & 0 \\ & & & \vdots & & & \\ 0 & & \dots & a_0 & a_1 & \dots & a_m \\ b_0 & b_1 & \dots & b_n & 0 & \dots & 0 \\ 0 & b_0 & b_1 & \dots & b_n & \dots & 0 \\ & & & \vdots & & & \\ 0 & & \dots & b_0 & b_1 & \dots & b_n \end{vmatrix}$$

Hence, $\mathrm{Res}(a,b)$ is the $m+n$-degree determinant, where m and n are the degrees of polynomials a and b respectively. The main property of resultant is given in the following theorem.

Theorem 2. *The complex polynomials a and b have a common root in the field of complex numbers* $\mathbf{C}$ *iff* $\mathrm{Res}(a,b)=0$.

In other words, if a and b are polynomials of degrees m and n respectively, then

$$(\exists x)(a(x)=0\land b(x)=0)\Leftrightarrow \mathrm{Res}(a,b)=0 \tag{1}$$

The resultant of two polynomials in any field can be defined in the same way, thus within the theory T.

Let $a(x)=y_m+y_{m-1}x+\ldots+y_0x^m$, $b(x)=z_n+z_{n-1}x+\ldots+z_0x^n$ be polynomials, where $y_0,\ldots,y_m,z_0,\ldots,z_n$ are variables. Define polynomials $a_0(x)=a(x)$, $a_1(x)=y_m+\ldots y_1x^{m-1}$, $a_m(x)=y_m$ and similarly polynomials $b_i(x)$. Then, by Theorem 2, we have

$$\begin{aligned}
&(\exists x)(a(x)=0\land b(x)=0)\Leftrightarrow\\
&\bigvee_{\substack{i<m\\ j<n}}\Big(\deg(a_i)=m-i\land\deg(b_j)=n-j\land\mathrm{Res}(a_i,b_j)=0\Big)\lor\\
&\qquad\qquad\bigwedge_{i,j}(y_i=0\land z_j=0)\Leftrightarrow\\
&\bigvee_{\substack{i<m\\ j<n}}\Big(y_0=0\land\ldots\land y_{i-1}=0\land y_i\neq 0\land z_0=0\land\ldots\land z_{j-1}=0\land\\
&\qquad\qquad z_j\neq 0\land\mathrm{Res}(a_i,b_j)=0\Big)\lor\bigwedge_{i,j}(y_i=0\land z_j=0).
\end{aligned}\tag{2}$$

Consider other two simple quantifier elimination cases. Since every algebraically closed field is infinite (roots of the polynomial $(x-x_0)(x-x_1)\ldots(x-x_n)+1$ are different from $x_0,\ldots,x_n$), for the polynomial $a(x)=\sum_i y_ix^i$ we have

$$(\exists x)(a(x)\neq 0)\Leftrightarrow y_0\neq 0\lor\ldots\lor y_m\neq 0. \tag{3}$$

Now, let us show that quantifier elimination for the formula

$$(\exists x)(a(x)=0\land b(x)\neq 0). \tag{4}$$

is reduced to the case (2). First note that $b(x)\neq 0\Leftrightarrow(\exists y)(yb(x)-1=0)$ and that y is a factor of every member of the polynomial $yb(x)-1$, except in the free member. If we select the variable y so it is not a variable of the formula $a(x) = 0 \wedge b(x) \neq 0$, then

$$(\exists x)(a(x) = 0 \wedge b(x) \neq 0) \Leftrightarrow (\exists y)(\exists x)(a(x) = 0 \wedge yb(x) - 1 = 0).$$

By (2), formula $(\exists x)(a(x) = 0 \wedge yb(x) - 1 = 0)$ is equivalent to the disjunction $\varphi_1 \vee \ldots \vee \varphi_k$, which is quantifier free, and each of the formulas φ_j, $j < k$ is of the form

$$y_0 = 0 \wedge \ldots \wedge y_{i-1} = 0 \wedge y_i \neq 0 \wedge y_i \neq 0 \wedge z_0 y = 0 \wedge \ldots \wedge z_{j-1} y = 0 \wedge \\ z_j y \neq 0 \wedge \mathrm{Res}(a_i, b'_j) = 0.$$

for a polynomial b'_j. Since $\exists x \vee_i \varphi_i \Leftrightarrow \vee_i \exists x \varphi$ is a valid formula, it is sufficient to eliminate quantifiers of the formula $\exists y \varphi_i$. Now, observe that the following sentences are true in the field theory:

1° $\exists y(y = 0 \wedge \psi(y)) \Leftrightarrow \psi(0)$,
2° $\exists y(y = 0 \wedge \psi) \Leftrightarrow \psi$, if y does not occur in ψ,
3° $zy \neq 0 \Leftrightarrow z \neq 0 \wedge y \neq 0$,
4° $\exists y(zy = 0 \wedge \psi) \Leftrightarrow (z = 0 \wedge \exists y \psi) \vee \exists y(y = 0 \wedge \psi)$, if z is a variable different from y.

Therefore, it will suffice to eliminate existential quantifier of the formula

$$(\exists y)(y \neq 0 \wedge \mathrm{Res}(a_i, b'_j) = 0),$$

i.e. formula of the form $(\exists y)(y \neq 0 \wedge m(y) = 0)$ where $m(y)$ is a polynomial. Let $m(y) = m_0 + m_1 + \ldots + m_k y^k$. Then the following is obvious.

$$(\exists y)(y \neq 0 \wedge m(y) = 0) \Leftrightarrow \bigvee_{i<j} (m_i \neq 0 \wedge m_j \neq 0).$$

Now we consider the general case of quantifier elimination in the theory T^*. Let φ be any formula of the theory T. It is equivalent to a formula

$$(Q_1 x_1) \ldots (Q_n x_n) \psi$$

in the prenex normal form, where x is quantifier free. Using the equivalence

$$(\forall x)\alpha(x) \Leftrightarrow \neg(\exists x)\neg\alpha(x),$$

and the fact that the quantifier elimination for $\neg\varphi$ is done in the same way as for φ, we may assume that Q_n is the existential quantifier.

Further, by the theorem on the disjunctive normal form, there are formulas $\psi_1, \dots, \psi_k$ such that $\psi \Leftrightarrow \psi_1 \vee \dots \vee \psi_k$ and each formula ψ_i is a conjunction of formulas of the form $u = 0$, $v \neq 0$, since every algebraic expression of T is equal to a polynomial. Since $v_1 \neq 0 \wedge \dots v_m \neq 0 \Leftrightarrow v_1 v_2 \dots v_n \neq 0$ we may suppose that every disjunct ψ_i is of the form

$$a_1 = 0 \wedge \dots \wedge a_m = 0 \wedge b \neq 0.$$

Using (valid formula) $(\exists x) \bigvee_i \psi_i \Leftrightarrow \bigvee_i (\exists x) \psi_i$ it follows that it is sufficient to eliminate quantifiers for formulas of the form

$$(5) \qquad (\exists x)(a_1 = 0 \wedge \dots \wedge a_m = 0 \wedge b \neq 0).$$

Let us denote by θ the formula (5). Now we describe the recursive procedure of the quantifier elimination for θ. Let $\lambda_i x^{n_i}$ be the highest degree member of the polynomial $a_i(x)$, $i = 1, \dots, m$, and let $n_0 = m + \sum_i n_i$. We shall determine the formulas θ_1 and θ_2 of the form (5) such that $\theta \Leftrightarrow \theta_1 \vee \theta_2$ and $n_{\theta_1}, n_{\theta_2} < n_\theta$ if $n_\theta > 1$ and $m \geq 2$.

First suppose that $n_1 = 0$. Then

$$\theta \Leftrightarrow a_1 = 0 \wedge (\exists x)(a_2 = 0 \wedge \dots \wedge a_m = 0 \wedge b \neq 0).$$

So, assume that $n_1 > 0$ and $m > 1$. We can also suppose that $n_2 \leq n_1$. Let $a_1' = \lambda_2 a_1 - \lambda_1 x^{n_1 - n_2} a_2$, $a_2' = a_2 - \lambda_2 x^{n_2}$. Then

$$\begin{aligned}\theta \Leftrightarrow &(\lambda_2 = 0 \wedge (\exists x)(a_1 = 0 \wedge a_2' = 0 \wedge \dots \wedge a_m = 0 \wedge b \neq 0)) \wedge \\ &(\lambda_2 \neq 0 \wedge (\exists x)(a_1' = 0 \wedge a_2 = 0 \wedge \dots \wedge a_m = 0 \wedge b \neq 0)).\end{aligned}$$

Now it is clear that for θ_1 we can choose the first disjunct and for θ_2 the second disjunct of the right side of this equivalence. In this way, the recursive procedure of the quantifier elimination is defined which reduces the formula to the cases (2) and (3) whose solutions are described above.

Now, we can derive few corollaries for the theory of algebraically closed fields T^*.

1. Let φ be a sentence of the field theory and let ψ be the quantifier free formula such that $T^* \vdash \varphi \Leftrightarrow \psi$. Then ψ is variable free. Since the language of the field theory is $\{+, \cdot, 0, 1\}$, it is clear that for ψ we can take a Boolean combination of formulas of the form $n = 0$, where $n = 1 + \dots + 1$ (n times). If $p_1, \dots, p_k$ are all prime factors of n then $T^* \vdash n = 0 \Leftrightarrow p_1 = 0 \vee \dots \vee p_k = 0$. Further, for a formula φ of T^* and distinct primes p, q we have:

1° $T^* \vdash p = 0 \Rightarrow q \neq 0$. 2° $T^* \vdash p = 0 \vee q \neq 0 \Leftrightarrow q \neq 0$,

3° $T^* \vdash p = 0 \vee (p \neq 0 \wedge \varphi) \Leftrightarrow p = 0 \vee \varphi$,

4° $p = 0 \wedge q = 0$ is inconsistent with T^*.

Using DNF and the above listed properties, we see that $T^* \vdash \psi \Leftrightarrow \psi'$, where ψ' is true, false, or one of the formulas:

$p_1 = 0 \vee p_2 = 0 \vee \ldots \vee p_k = 0$,

finite disjunction of formulas of the form $q_1 \neq 0 \wedge q_2 = 0 \wedge \ldots q_l = 0$,

where $p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_l$ are distinct primes. As for a prime q in any field $\mathbf{F}$ of finite characteristic holds

$$q \neq 0 \Leftrightarrow \bigvee_{p \neq q} p = 0,$$

it follows that all complete extensions of theory T^* are the theories of the form $T_p = T^* \cup \{p = 0\}$, p is prime, (theories of algebraically closed fields of the characteristic p), and $T_0 = T^* \cup \{p_1 \neq 0, p_2 \neq 0, p_3 \neq 0, \ldots\}$, p_i are primes, (the theory of algebraically closed fields of the characteristic 0).

2. We have just described all complete extensions of T^*, and we see that there are countable many of them, and all of them can be listed in an effective and uniform way. Therefore, see e.g. Theorem 2.4.15, p. 57, [Mijajlović 1987], the theory of algebraically closed fields is decidable. Let us remind the reader that the field theory is not decidable.

3. Real closed fields

Artin-Schreier theory of real fields is used for the solution of seventeenth Hilbert's problem in the algebraic way. Beside it has applications in the other parts of mathematics, especially in algebraic geometry, as in Hilbert's the proof of Nullstellensatz, and nonstandard analysis. We note that every model of nonstandard analysis is a real algebraically closed field. On the other side, we used elements of this theory in the development of an algorithm for quantifier elimination for the theory of ordered real closed fields.

Sturm's algorithm

The quantifier elimination for the theory of algebraically closed fields can be done in somewhat another way. Let $\mathbf{F}$ be an algebraically closed field and let f and g be polynomials over $\mathbf{F}$ in the variable x. If f and g have a common root a, then the greatest common divisor of polynomials f and g is of degree ≥ 1 (since $x - a$ divides both f and g). Thus

$$(\exists x)(f(x) = 0 \wedge g(x) = 0) \Leftrightarrow \text{degree } \mathrm{GCD}(f, g) \geq 1.$$

The $\mathrm{GCD}(f, g)$ can be found bu use of the Euclid algorithm. For given polynomials f and g the algorithm ends in finally many steps, because its length depends essentially only on the degrees of f and g. Hence, we can easily see that this algorithm is described with the quantifier free formula, i.e. if

$$\begin{aligned} &f = q_1 g + m_2, \quad g = q_2 m_2 + m_3, \quad m_2 = q_3 m_3 + m_4, \dots, \\ (1) \qquad &m_{i-2} = q_{i-1} m_{i-1} + m_i, \quad m_{i-1} = q_i m_i \qquad \text{and} \\ &\deg(f) > \deg(g) > \deg(m_2) > \dots > \deg(m_i), \end{aligned}$$

then

$$(\exists x)(f(x) = 0 \land g(x) = 0) \Leftrightarrow f = q_1 g + m_1 \land \dots \land m_{i-1} = q_i m_i \land z \neq 0,$$

where z is the coefficient of the highest degree of variable x in m_i. Note that the right side of this equivalence is quantifier free. The other details of the proof are the same as in Section 2.

The method of the quantifier elimination for the theory of ordered real closed fields is similar to the previous procedure for algebraically closed fields. In fact, the procedure for the ordered real closed fields can be built on Sturm's algorithm in the way the above described algorithm is using the Euclid algorithm.

Sturm's Theorem. *Let $p(x)$ be a real polynomial and let $p_0, p_1, \dots, p_r$ be the sequence of real polynomials defined by:*

1. $p_0 = p$.
2. $p_1 = p'$, where p' is the first derivation of p.
3. For all i, $0 < i < r$, there is a polynomial q such that $p_{i-1} = p_i q_i - p_{i+1}$, where $p_{i+1} \neq 0$ and $\deg(p_{i+1}) < \deg(p_i)$. In other words q_i is the quotient, $-p_{i+1}$ is the reminder when p_{i-1} is divide by p_i.
4. $p_{r-1} = p_r q_r$.

Let $d(a)$ be the number of the sign changes in the sequence $p_0(a), \dots, p_r(a)$ (zeroes are ignored). Let a and b be real numbers which are not roots of p and let $a < b$. Then the number of roots of p (not counting the multiplicity of a root) in the interval $[a, b]$ is equal to $d(a) - d(b)$.

Now we give an illustration of Sturm's theorem application to the quantifier elimination on the example of a formula of the theory of ordered fields. Applying Sturm's theorem we get at once

$$(\exists x)(a < x \land x < b \land p(x) = 0) \Leftrightarrow d(a) > d(b).$$

Besides, similarly as for the formula (1), using Sturm's theorem, we can find quantifier free formulas ψ such that $d(a) > d(b) \Leftrightarrow \psi$. In this way the quantifier is eliminated from the formula $(\exists x)(a < x < b \land p(x) = 0)$.

Further reduction is obtained similarly to the procedure of algebraically closed fields. In this reduction the following equivalence is useful:

$$p_1 = 0 \wedge \ldots \wedge p_n = 0 \Leftrightarrow p_1^2 + \ldots + p_n^2 = 0$$

(note that this formula is not a theorem of the theory of algebraically closed fields).

Also, one can obtain in a similar way the following for the theory T of ordered real closed fields:

1. T is complete,
2. T is decidable.

4. Programming implementation

A group of students under our supervision implemented a prover for the theory of algebraically closed fields in the standard programming language C. The program is based on the procedures described in Section 2. It is running well on personal computers quickly solving problems stated in the language of the field theory. The input formula is proved or refuted by reducing it to a quantifier free formula.

The processing of sentences with more than a few quantifiers would be greatly accelerated with the introduction of fast calculators for long and very long disjunctive normal forms, and fast DNF transformers, which are suitable for parallelisation.

The prover for ordered real closed fields based on Sturm's theorem is being integrated. The plan is to optimize, accelerate and collect these procedures in one Elementary Mathematics problem solver, which might be expanded to other applications as well.

Let us mention just one possible application, namely we can apply the method of elimination of quantifiers in mathematical programming. Programming problem with algebraic constraints in several variables $x_1, \ldots, x_n$

$$f \longrightarrow \min, \quad p_1 = 0, \ldots, p_k = 0, \quad q_1 > 0, \ldots, q_m > 0$$

where $f, p_1, \ldots, p_k, q_1, \ldots, q_m$ are polynomials in variables $x_1, \ldots, x_m$ with rational coefficients, is easily stated in the theory of ordered fields as follows:

$$\begin{aligned}
&\exists x_1 \ldots x_n (y = f(x_1, \ldots, x_n) \wedge \\
&p_1(x_1, \ldots, x_n) = 0 \wedge \ldots \wedge p_k(x_1, \ldots, x_n) = 0 \wedge \\
&q_1(x_1, \ldots, x_n) > 0 \wedge \ldots \wedge q_m(x_1, \ldots, x_n) > 0) \wedge \\
&\forall x_1 \ldots x_n (p_1(x_1, \ldots, x_n) = 0 \wedge \ldots \wedge p_k(x_1, \ldots, x_n) = 0 \wedge \\
&q_1(x_1, \ldots, x_n) > 0 \wedge \ldots \wedge q_m(x_1, \ldots, x_n) > 0 \Rightarrow \\
&y \leq f(x_1, \ldots, x_n))).
\end{aligned}$$

Eliminating quantifiers from the above formula, we obtain a formula $\psi(y)$ of the theory of ordered fields in the single variable y. This formula is a finite disjunction of the formulas of the form $y > r$, $y < r$, $y = r$, where r is a rational number. Obviously this is a solution of the above stated mathematical programming problem. Observe that we, in fact, proved that problem of finding of solutions of mathematical programming problems with polynomial constraints is decidable.

5. Bibliographical and other remarks

First and most important step in the solution of the seventeenth Hilbert's problem was given in [Artin 1927]. Artin-Schreir theory of formally real fields is presented in detail in [Lang 1965]. The proof of Hilbert's basis theorem can be found in [Artin 1955]. Solution of Hilbert's seventeenth problem with described methods of mathematical logic is given in [Robinson 1955]. Elimination of quantifiers in the theory of algebraically closed fields and in the theory of ordered real closed fields with some detailed analysis evolving from these procedures, could be found in [Kreisel, Krivine 1971]. Here presented procedure of elimination of quantifiers differs from the last source, e.g. where we use the resultant of polynomials, in [Kreisel, Krivine 1971] one lemma which relates to divisibility of polynomials is used.

The proof of the theorem on resultant of polynomials could be found in any book on higher algebra, for example in [Kurepa 1965].

Complete solution of Hilbert's seventeenth problem based on Logic can be found in [Cherlin 1976] as well.

Problem of quantifier elimination can be treated in model theory in other way, too. In the other approach the diagrams of models, saturated models and elementary embedding have special importance. This approach is more complex but results are deeper, see ([Sacks 1972]).

References

[1] E. Artin, *Über die Zerlegung definiter Funktionen in Quadrate*, Abhandlungen aus dem mathematischen Seminar der Universitat Hamburg **5** (1927), 100-115.

[2] E. Artin, *Elements of Algebraic Geometry*, Courant Institute of Mathematical Scienes, New York University, New York, 1955..

[3] C.C. Chang, H.J. Keisler, *Model Theory*, North-Holland, Amsterdam, 1973.

[4] G. Cherlin, *Model-Theoretoc Algebra: Selected Topics, Lecture Notes in Mathematics 521*, Springer, Berlin, 1976.

[5] G. Kreisel, J.L. Krivine, *Elements of Mathematical Logic : Model Theory*, North-Holland, Amstedam, 1971.

[6] Đ. Kurepa, *Viša algebra I*, Školska knjiga, Zagreb, 1965.

[7] S. Lang, *Algebra*, Addison-Wesley, Reading Mass., 1965.

[8] Ž. Mijajlović, Z. Marković, K. Došen, *Hilbertovi problemi i logika*, Zavod za udžbenike i nastavna sredstv, Beograd, 1986..

[9] Ž. Mijajlović, *An Introduction to Model Theory*, Univ. of Novi Sad, Institute of Mathematics, Novi Sad, 1987.

[10] A. Robinson, *On ordered fields and definite functions*, Mathematishe Annalen **130** (1955), 257-277..

[11] G. Sacks, *Saturated Model Theory*, Benjamin, Reading Mass., 1972.

[12] J.R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading Mass, 1967.

Matematički fakultet, Studentski trg 11, 11000 Beograd, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 743–751
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

О ЗАДАЧЕ СИНТЕЗА ДЛЯ АВТОМАТОВ В ОДНОМ КЛАССЕ ЛАБИРИНТОВ

Г. Килибарда и Ш. Ушчумлич

Резюме. В работе дается более эффективный алгоритм обхода чем те, которые даны в работах [1], [2] и [3]. Также улучшены оценки времени обхода и количества состояний универсальных обходчиков лабиринтов, рассматриваемых в тех работах.

Все результаты и понятия, которые не приводятся здесь, можно найти в работе [3].

Пусть L — некоторый конечный плоский мозаичный лабиринт. Для любой конечной области $\Delta \in \mathrm{ar}(L)$ обозначим через $V(\Delta)$ множество всех вершин из L, которые лежат на границе этой области. Число

$$\mathrm{dc}(L) = \max\{\mathrm{diam}\, V(\Delta) \mid \Delta \text{ конечная область из } \mathrm{ar}(L)\}$$

назовем *циклическим диаметром* лабиринта L. Обозначим для любого $r \in \mathbf{R}^+$ через $\mathcal{L}^{\mathrm{m}}_{\mathrm{ОГ}}(r)$ класс всех плоских конечных мозаичных лабиринтов, таких, что $\mathrm{dc}(L) \leq r$. Ясно, что $\mathcal{L}^{\mathrm{m}}_{\mathrm{ОГ}}(r) = \emptyset$ при $0 < r < \sqrt{2}$.

Пусть τ — некоторая циклическая подстановка множества $\mathbf{D}$. Определим функцию $\hat{\nu}_\tau : \mathbf{D}^+ \to \mathbf{N}$, следующим способом:

а) $\hat{\nu}_\tau(\omega) = 0$, если $\omega \in \mathbf{D}$;

б) для любого слова $\omega\omega' \in \mathbf{D}^+$ имеет место

$$\hat{\nu}_\tau(\omega\omega') = \begin{cases} \nu_\tau(\omega\omega'), & \text{если } \nu_\tau(\omega\omega') \geq 0, \\ \nu_\tau(\omega\omega') + 3 & \text{в противном случае;} \end{cases}$$

в) $\hat{\nu}_\tau(\alpha) = \hat{\nu}_\tau(\omega_1\omega_2) +_3 \cdots +_3 \hat{\nu}_\tau(\omega_{k-1}\omega_k)$ для любого $\alpha = \omega_1 \ldots \omega_k \in \mathbf{D}^+$, $k \geq 2$.

Теорема 1. *Для всякого $r \in \mathbf{R}$, $r \geq \sqrt{2}$, существует инициальный автомат $\mathfrak{A}_{\text{ОГ}}(r)$, сильно обходящий класс $\mathcal{L}^{\text{m}}_{\text{ОГ}}(r)$, при этом число состояний $\mathfrak{A}_{\text{ОГ}}(r)$ не больше Cr, а время обхода лабиринта с n вершинами из этого класса не больше $C'r^2n$, где C и C' — константы.*

Возьмем произвольный вектор $\vec{v}$ в плоскости. Возьмем также некоторую прямую l вертикальную на $\vec{v}$. Пусть v — некоторая точка в плоскости. Тогда через $\mathbf{p}_l(v)$ обозначим нормальную проекцию точки v на прямую l. Ясно, что для любой конечной области Δ число $r_{\vec{v}}(\Delta) = \operatorname{diam}\{\mathbf{p}_l(v) \,|\, v \in V(\Delta)\}$ не зависит от l. Под $\vec{v}$-шириной лабиринта L понимаем число

$$\operatorname{wd}_{\vec{v}}(L) = \max\{r_{\vec{v}}(\Delta) \,|\, \Delta \text{ конечная область из } \operatorname{ar}(L)\}.$$

Обозначим для любого $r \in \mathbf{R}^+$ через $\mathcal{L}^{\text{m}}_{\vec{v}}(r)$ класс всех плоских конечных мозаичных лабиринтов, таких, что $\operatorname{wd}_{\vec{v}}(L) \leq r$. Ясно, если для некоторого $r \in \mathbf{R}^+$ имеет место $\operatorname{dc}(L) \leq r$, то $\operatorname{wd}_{\mathbf{e}}(L) \leq [\sqrt{r^2-1}]$. До того как докажем теорему 1 докажем следующую теорему.

Теорема 2. *Для любого $m \in \mathbf{N}$ существует универсальный обходчик класса $\mathcal{L}^{\text{m}}_{\mathbf{e}}(m)$, у которого $48m+52$ состояний и который любой лабиринт из $\mathcal{L}^{\text{m}}_{\mathbf{e}}(m)$ с n вершин обходит за время не больше $3n^2+2n-2$.*

Доказательство. Построим автомат $\mathfrak{A}_{\mathbf{e}}(m) = (A, Q, B, \varphi, \psi, q_{\mathbf{n}})$, обходящий все лабиринты из класса $\mathcal{L}^{\text{m}}_{\mathbf{e}}(m)$ следующим способом. В качестве множества состояний Q возьмем множество

$$\{(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5) \,|\, -m \leq \alpha_1 \leq 0, 0 \leq \alpha_2 \leq 2, 1 \leq \alpha_3 \leq 2, \\ \alpha_4 \in \{\mathbf{s},\mathbf{n}\}, \alpha_5 \in \{\mathbf{e},\mathbf{n},\mathbf{w},\mathbf{s}\}\} \cup Q_{\mathfrak{A}+}.$$

Функции φ и ψ определим следующим способом. На множестве $Q_{\mathfrak{A}+} \times A$ введем предикат $\mathbf{P}$ следующим способом: $\mathbf{P}(q,a) = 1$, если условие

$$\{\mathbf{w},\mathbf{s}\} \not\subseteq a \vee (q \neq q_{\mathbf{n}} \wedge \psi_{\mathfrak{A}+}(q,a) \neq \mathbf{s})$$

выполнено, и $\mathbf{P}(q,a) = 0$ в противном случае. Пусть $q \in Q$ и $a \subseteq \mathbf{D}$. Тогда, если $q \in Q_{\mathfrak{A}+}$ и $\mathbf{P}(q,a) = 1$, то положим $\psi(q,a) = \psi_{\mathfrak{A}+}(q,a)$ и $\varphi(q,a) = \varphi_{\mathfrak{A}+}(q,a)$. Если $q \in Q_{\mathfrak{A}+}$ и $\mathbf{P}(q,a) = 0$, то положим $\varphi(q,a) = (0,1,1,\alpha_4,\mathbf{s})$ и $\psi(q,a) = 0$, где $\alpha_4 = \mathbf{s}$ если $\psi(q,a) = \mathbf{s}$, а $\alpha_4 = \mathbf{n}$, если $q = q_{\mathbf{n}}$. Введем для любого $a \in A$ отображения $\theta^1_a : \mathbf{D} \to \mathbf{D}$ и $\theta^2_a : Q \backslash Q_{\mathfrak{A}+} \to Q$, таким способом, что

$$\theta^1_a(\omega) = \begin{cases} \omega, & \text{если } \omega \in a, \\ \sigma^+_a(\omega) & \text{в противном случае} \end{cases}$$

и

$$\theta_a^2(q) = \begin{cases} q, & \text{если } \mathbf{p}_5(q) \in a, \\ q_{\sigma_a^+(\mathbf{p}_5(q))} & \text{в противном случае.} \end{cases}$$

Если $q \notin Q_{\mathfrak{A}+}$, т.е. q является состоянием вида $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$, то предположим, что $\mathcal{L}_{\mathbf{e}}^{\mathrm{m}}(m)$ ведет себя согласно следующим правилам (любое нижеприведенное правило выполняется только в том случае, если выполнено соответствующее для этого правила условие):

1. Если $q = (0, 2, 1, \beta_4, \mathbf{n})$, то пусть $\varphi(q, a) = \theta_a^2[(-1, 2, 2, \beta_4, \mathbf{s})]$ и $\psi(q, a) = \theta_a^1(\mathbf{s})$.

2. Если $q = (0, 0, 1, \beta_4, \mathrm{n})$ и $\mathrm{e} \notin a$, то $\varphi(q, a) = \theta_a^2[(-1, 1, 2, \beta_4, \mathbf{s})]$ и $\psi(q, a) = \theta_a^1(\mathbf{s})$.

3. Если $q = (0, 0, 2, \beta_4, \mathrm{n})$, то положим, что

$$\varphi(q, a) = \begin{cases} q_{\theta_a^1(\mathbf{s})}, & \text{если } \beta_4 = \mathbf{s}, \\ \varphi_{\mathfrak{A}+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{n}, \end{cases}$$

и

$$\psi(q, a) = \begin{cases} \theta_a^1(\mathbf{s}), & \text{если } \beta_4 = \mathbf{s}, \\ \psi_{\mathfrak{A}+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{n}. \end{cases}$$

4. Если $q = (0, \beta_2, 1, \beta_4, \mathbf{e})$ и $\mathbf{s} \in a$, то $\varphi(q, a) = \theta_a^2[(-1, 2, 1, \beta_4, \mathbf{s})]$ и $\psi(q, a) = \theta_a^1(\mathbf{s})$.

5. Если $q = (0, 1, 1, \beta_4, \mathrm{e})$ и $a \subseteq \{\mathbf{w}, \mathbf{n}\}$, то возьмем $\varphi(q, a) = \theta_a^2[(0, 0, 2, \beta_4, \mathbf{w})]$ и $\psi(q, a) = \theta_a^1(\mathbf{w})$.

6. Если $q = (0, 1, 1, \beta_4, \mathrm{n})$, то возьмем $\varphi(q, a) = \theta_a^2[(-1, 2, 2, \beta_4, \mathbf{s})]$ и $\psi(q, a) = \theta_a^1(\mathbf{s})$.

7. Если $q = (0, 1, 2, \beta_4, \mathrm{n})$, то положим, что

$$\varphi(q, a) = \begin{cases} q_{\theta_a^1(\mathbf{s})}, & \text{если } \beta_4 = \mathbf{n}, \\ \varphi_{\mathfrak{A}+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{s}, \end{cases}$$

и

$$\psi(q, a) = \begin{cases} \theta_a^1(\mathbf{s}), & \text{если } \beta_4 = \mathbf{n}, \\ \psi_{\mathfrak{A}+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{s}. \end{cases}$$

8. Если не имеет место ни одно из предыдущих условий и автомат $\mathfrak{A}_{\mathbf{e}}(m)$ находится в состоянии $q = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, то он переходит в состояние $(\beta_1', \beta_2', \beta_3', \beta_4', \beta_5')$, где β_i', $1 \le i \le 5$, определено следующим образом. Прежде всего $\beta_4' = \beta_4$ и $\beta_3' = \beta_3$. Если

$$\beta' = \beta_1 + \mathrm{ch}[\psi_{\mathfrak{A}(\beta_3)}(q_{\beta_5}, a)] \geq -m,$$

то $\beta_1' = \beta'$, где

$$\mathrm{ch}(\omega) = \begin{cases} 1, \text{ если } \omega = \mathbf{n}, \\ -1, \text{ если } \omega = \mathbf{s}, \\ 0, \text{ если } \omega = \mathbf{e}, \mathbf{w}, \mathbf{0} \end{cases}$$

и $\mathfrak{A}(\beta_3) = \mathfrak{A}^+$, если $\beta_3 = 2$, а $\mathfrak{A}(\beta_3) = \mathfrak{A}^-$, если $\beta_3 = 1$; в противном случае

$$(\beta_1', \beta_2', \beta_3', \beta_4', \beta_5') = (-m + \mathrm{ch}(\overline{\beta}_5), \beta_2'', 2, \beta_4, \overline{\beta}_5) \quad \text{и} \quad \psi(q, a) = \overline{\beta}_5,$$

где

$$\beta_2'' = \begin{cases} 2, \text{ если } \beta_5 = \mathbf{w}, \\ 0, \text{ если } \beta_5 = \mathbf{s}, \\ 1, \text{ если } \beta_5 = \mathbf{e}, \mathbf{n}. \end{cases}$$

Далее, $\beta_5' = \psi_{\mathfrak{A}(\beta_3)}(q_{\beta_5}, a)$ и $\psi(q, a) = \psi_{\mathfrak{A}(\beta_3)}(q_{\beta_5}, a)$. Значение β_2 равно

$$\beta_2 +_3 \nu^{\mathrm{sm}(\beta_3)}(\beta_5 \psi_{\mathfrak{A}(\beta_3)}(q_{\beta_5}, a)),$$

где

$$\mathrm{sm}(\alpha_3) = \begin{cases} +, \text{ если } \alpha_3 = 2, \\ -, \text{ если } \alpha_3 = 1. \end{cases}$$

Смысл параметров $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ и α_5 будет следующим. Функции φ и ψ определены таким образом, что поведение автомата $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ в некотором произвольном лабиринте L класса $\mathcal{L}_{\mathbf{e}}^{\mathrm{m}}(m)$ будет разбито как бы на две части. Автомат $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ ведет себя как автомат $\mathfrak{A}^+$ пока не окажется в некоторой квазиособенной вершине v, и $q = q_{\mathbf{n}}$ или $\psi(q, [v]) = \mathbf{s}$. Если последнее имеет место, то $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ переходит в состояние вида $(0, 1, 1, \alpha_4, \alpha_5)$ и исследует вопрос, является ли вершина v особенной или нет. Пока он это делает он находится в состояниях вида $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ и ведет себя, если не имеет место некоторый из выше данных случаев 1–6, как автомат $\mathfrak{A}(\alpha_3)$, где $\mathfrak{A}(\alpha_3) = \mathfrak{A}^+$, если $\alpha_3 = 2$, а $\mathfrak{A}(\alpha_3) = \mathfrak{A}^-$, если $\alpha_3 = 1$; параметр α_5 обеспечивает такую работу автомата $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ Из работы [3] следует, что за это время автомат $\mathcal{L}_{\mathbf{e}}^{\mathrm{m}}(m)$ движется вокруг одной и той же самой области, на границе которой лежит вершина v. Параметр α_1 такой, что если автомат $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ оказался на некоторой вершине v', то $\alpha_1 = \mathbf{p}_2(v) - \mathbf{p}_2(v')$. Параметром α_2 автомат $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$ измеряет, на сколько он поворачивается, меняет направление, двигаясь вдоль грницы некоторой области. Легко убедиться, например, что если в течение некоторого времени, до момента t, когда автомат $\mathfrak{A}_{\mathbf{e}}^{\mathrm{m}}(m)$

оказался в v', параметр α_3 не менялся и $\alpha_3 = 1$ (аналогное утверждение имеет место и для $\alpha_3 = 2$), то в случае когда $\alpha'_1 = 0$, $\alpha'_2 = 2$ и $\alpha'_5 = \mathbf{n}$ выполнено, $\mathbf{p}_2(v) = \mathbf{p}_2(v')$ и $\mathbf{p}_1(v) < \mathbf{p}_1(v')$, а в случае когда $\alpha'_1 = 0$, $\alpha'_2 = 0$ и $\alpha'_5 = \mathbf{n}$ — $\mathbf{p}_2(v) = \mathbf{p}_2(v')$ и $\mathbf{p}_1(v') < \mathbf{p}_1(v)$, где $(\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4, \alpha'_5)$ — состояние автомата $\mathfrak{A}_{\mathrm{e}}^{\mathrm{m}}(m)$ в моменте t. Параметром α_4 данный автомат запоминает свое состояние в момент, когда последний раз оказался на квазиособенной вершине.

Ясно, что из данного описания автомата $\mathfrak{A}_{\mathrm{e}}^{\mathrm{m}}(m)$ следует, что вместо того, чтобы двигаться по лабиринту L, он движется как бы по лабиринту $\mathrm{Ct}(L)$. Но тогда получаем из [3], что этот автомат является универсальным обходчиком для класса всех лабиринтов из класса $\mathcal{L}_{\mathrm{OГ}}^{\mathrm{m}}(m)$.

Если лабиринт L содержит n вершин, то L не может иметь больше $n/2$ квазиособенных вершин (количество квазиособенных вершин v равно количеству вершин вида $v\mathbf{e}$). Ясно, что $|V(\Delta)| \leq n$ для любой конечной области $\Delta \in \mathrm{ar}(L)$. Из данного алгоритма обхода следует, что автомат задерживается у каждой квазиособенной вершины (двигаясь вокруг соответствующей области) на время не больше $3n$. Учитывая то, что каждую квазиособенную вершину проверяет на особеность два раза, то на это уходит время $3n^2$. В графе $\mathrm{Ct}(L)$ вершин $n + |\mathrm{ar}(L)| - 1$. Ясно, что в L не меньше $|\mathrm{ar}(L)| - 1$ особенных вершин. Время, которое уходит на обход этих вершин учтено выше при оценке времени обхода квазиособенных вершин. На самом деле оно входит в эту сумму умноженное на 2. С другой стороны на обход дерева с $n + |\mathrm{ar}(L)| - 1$ вершинами уходит время не больше $2(n + |\mathrm{ar}(L)| - 2)$. Из всего сказаного получается, что время обхода лабиринта L не больше $3n^2 + 2n - 2$.

Пусть $\mathcal{L}$ — некоторый класс лабиринтов. Через $\mathrm{Char}(\mathcal{L})$ обозначим число $\min\{Q_{\mathfrak{A}} \mid \mathfrak{A} \in \mathrm{Un}(\mathcal{L})\}$; если $\mathrm{Un}(\mathcal{L}) = \emptyset$, то положим, что $\mathrm{Char}(\mathcal{L}) = \infty$. Тогда результат теоремы 2. можем переформулировать следующим способом.

Теорема 3. *Имеет место следующая оценка*

$$\mathrm{Char}(\mathcal{L}_{\mathrm{e}}^{\mathrm{m}}(m)) \leq 48m + 52$$

дльа любого $m \in \mathbf{N}$.

Остается открытым вопрос улучшения данной оценки.

Доказательство теоремы 1. То что у автомата $\mathfrak{A}_{\mathrm{OГ}}^{\mathrm{m}}(r)$ не больше Cr состояний следует из теоремы 2. Осталось еще оценить время обхода. Пусть плоский мозаичный лабиринт L содержит n вершин.

Ясно, что $|V(\Delta)| \leq [r]^2$ для любой конечной области $\Delta \in \mathrm{ar}(L)$. Но тогда, как и выше, получаем, что время обхода не больше $3r^2n + 2n - 2$, т.е. данная в тереме оценка имеет место.

Построенный выше автомат не останавливается после обхода лабиринта. На вопрос, существует ли автомат, который обходит любой лабиринт из класса $\mathcal{L}^{\mathrm{m}}_{\mathrm{e}}(m)$ и останавливается после обхода, можно ответить отрицательно, поскольку уже в случае класса всех конечных шахматных лабиринтов без конечных дыр такой автомат не существует.

Плоский мозаичных лабиринт называется $k-x$-ограниченным ($k-y$-ограниченным), если он лежит между двумя параллельными прямыми, которые расположены на расстоянии $k \in \mathbf{N}$ одна от другой и паралелльны оси x (оси y). Обозначим класс всех $k-x$-ограниченных мозаичных лабиринтов через $\mathcal{L}^x_{||}(k)$, а всех $k-y$-ограниченных мозаичных лабиринтов — через $\mathcal{L}^y_{||}(k)$.

Следствие 1. *Для всех $k \in \mathbf{N}$ имеет место*

$$\mathrm{Char}(\mathcal{L}^x_{||}(k)) \leq 48k + 52, \quad \mathrm{Char}(\mathcal{L}^y_{||}(k)) \leq 48k + 52.$$

Вектор $\vec{v}$ в плоскости назовем целочисленным, если для некоторых $l, m \in \mathbf{Z}$ имеет место $\vec{v} = l\vec{i} + m\vec{j}$. Следующая теорема является обобщением теоремы 2. Она впервые доказана в [2]. Здесь дается более простое и короткое доказательство.

Теорема 4. *Для любого $d \in \mathbf{R}^+$ и любого целочисленного вектора $\vec{v}$ существует инициальный автомат, являющийся универсальным обходчиком класса $\mathcal{L}^{\mathrm{m}}_{\vec{v}}(d)$, у которого число состояний не больше Cd, где C — константа.*

Доказательство. Пусть $\vec{v} = m\vec{i} + l\vec{j}$ для некоторых $l, m \in \mathbf{Z}$. Рассмотрим случай, когда $l > 0$ и $m < 0$. Остальные случаи рассматриваются аналогичным образом. Ясно, что любая конечной область из $\mathrm{ar}(L)$ лежит в некоторой полосе ширины d, у которой тангенс угла наклона равен l/m. Пусть Δ — некоторая конечная область из $\mathrm{ar}(L)$. Говорим, что вершина $v \in V(\Delta)$, $v = (x_0, y_0)$, является Δ-(l,m)-*особенной*, если полуплоскость $-l/m(x - x_0) + (y - y_0) > 0$, не содержит ни одной вершины из множества $V(\Delta)$, и если на прямой $-l/m(x - x_0) + (y - y_0) = 0$ лежит вершина $v' \in V(\Delta)$, $v' = (x_0', y_0')$, то $x_0' < x_0$. Вершина $v \in V(L)$ является (l,m)-*особенной*, если существует $\Delta \in \mathrm{ar}(L)$, такая, что v является Δ-(l,m)-особенной вершиной. Заметим, что, если некоторая вершина $v \in V(L)$ (l,m)-особенная, то $\{\mathbf{w},\mathbf{s}\} \subseteq [v]$. Тогда, ясно, что (l,m)-особенную вершину надо искать среди квазиособенных вершин.

Длины проекций векторов $\mathbf{e}$ и $\mathbf{n}$ на нормаль полосы равны, соответственно, lh/ml и mh/ml. Пусть

$$\lambda_1 = \frac{ml}{h}\max\{mh/ml, lh/ml\} \text{ и } \lambda_2 = \frac{ml}{h}(-d - \max\{mh/ml, lh/ml\}).$$

Построим автомат $\mathfrak{A}_{\vec{v}}(d) = (A, Q, B, \varphi, \psi, q_{\mathbf{n}})$, обходящий все лабиринты из данного класса лабиринтов следующим способом. В качестве множества состояний Q возьмем множество

$$\{(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5) \,|\, [\lambda_2] \le \alpha_1 \le [\lambda_1], 0 \le \alpha_2 \le 2, 1 \le \alpha_3 \le 5,$$
$$\alpha_4 \in \{\mathbf{s},\mathbf{n}\}, \alpha_5 \in \{\mathbf{e},\mathbf{n},\mathbf{w},\mathbf{s}\}\} \cup Q_{\mathfrak{A}+}.$$

Функции φ и ψ определим следующим способом. Пусть $q \in Q$ и $a \subseteq \mathbf{D}$. Тогда, если $q \in Q_{\mathfrak{A}+}$ и если или $\{\mathbf{w},\mathbf{s}\}$ не является подмножеством множества a, или $q \neq q_{\mathbf{n}}$ и $\psi_{\mathfrak{A}+}(q,a) \neq \mathbf{s}$, то $\psi(q,a) = \psi_{\mathfrak{A}+}(q,a)$ и $\varphi(q,a) = \varphi_{\mathfrak{A}+}(q,a)$. Если $q \in Q_{\mathfrak{A}+}$, $\{\mathbf{w},\mathbf{s}\} \subseteq a$, и $q = q_{\mathbf{n}}$ или $\psi(q,a) = \mathbf{s}$, то $\varphi(q,a) = (0,0,1,\alpha_4,\mathbf{s})$ и $\psi(q,a) = 0$, где $\alpha_4 = \mathbf{s}$ если $\psi(q,a) = \mathbf{s}$, а $\alpha_4 = \mathbf{n}$, если $q = q_{\mathbf{n}}$. Как и выше ведем для любого $a \in A$ отображения $\theta_a^1 : \mathbf{D} \to \mathbf{D}$ и $\theta_a^2 : Q \backslash Q_{\mathfrak{A}+} \to Q$. Если $q \notin Q_{\mathfrak{A}+}$, т.е. q является состоянием вида $(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5)$, то предположим, что $\mathfrak{A}_{\vec{v}}(d)$ ведет себя согласно следующим правилам (любое нижеприведенное правило выполняется только в том случае, если выполнено соответствующее для этого правила условие):

1. Если $\alpha_1 > 0$, $\alpha_3 = 1$ или $\alpha_1 = 0$, $\alpha_3 = 1$ и

$$\alpha_5 = \mathbf{n} \vee (\alpha_5 = \mathbf{e} \wedge \mathbf{s} \notin a),$$

то $\varphi(q,a) = \theta_a^2[(\alpha_1 + \mathrm{ch}(\overline{\alpha}_5), \beta_2, 2, \alpha_4, \overline{\alpha}_5)]$ и $\psi(q,a) = \theta_a^1(\overline{\alpha}_5)$, где

$$\beta_2 = \begin{cases} 1, & \text{если } \alpha_5 = \mathbf{e} \wedge \alpha_2 = 2, \\ 2, & \text{если } \alpha_5 = \mathbf{n} \wedge \alpha_2 = 1, \\ 0, & \text{если } \alpha_5 = \mathbf{e} \wedge \alpha_2 = 0, \\ 1, & \text{если } \alpha_5 = \mathbf{n} \wedge \alpha_2 = 2. \end{cases}$$

2. Если $q = (0,0,2,\beta_4,\mathbf{n})$, то положим, что

$$\varphi(q,a) = \begin{cases} q_{\theta_a^1(\mathbf{s})}, & \text{если } \beta_4 = \mathbf{s}, \\ \varphi_{\mathfrak{A}+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{n}, \end{cases}$$

и

$$\psi(q,a) = \begin{cases} \theta_a^1(\mathbf{s}), & \text{если } \beta_4 = \mathbf{s}, \\ \psi_{\mathfrak{A}^+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{n}. \end{cases}$$

3. Если $\alpha_1 = 0$, $\alpha_2 = 2$. $\alpha_3 = 1$, $\alpha_5 = \mathbf{e}$ и $\mathbf{s} \in a$, то $\varphi(q,a) = \theta_a^2[(\alpha_1 + \mathrm{ch}(\mathbf{w}), 1, 2, \beta_4, \mathbf{w})]$ и $\psi(q,a) = \theta_a^1(\mathbf{w})$.

4. Если $\alpha_1 = 0$, $\alpha_2 = 0$. $\alpha_3 = 1$, $\alpha_5 = \mathbf{e}$ и $\mathbf{s} \in a$, то $\varphi(q,a) = \theta_a^2[(\alpha_1 + \mathrm{ch}(\mathbf{s}), 1, 1, \beta_4, \mathbf{s})]$ и $\psi(q,a) = \theta_a^1(\mathbf{s})$.

5. Если $q = (0,1,1,\beta_4,\mathbf{e})$, то возьем $\varphi(q,a) = \theta_a^2[(\mathrm{ch}(\mathbf{w}), 1, 2, \beta_4, \mathbf{w})]$ и $\psi(q,a) = \theta_a^1(\mathbf{w})$.

6. Если $q = (0,1,2,\beta_4,\mathbf{n})$, то положим, что

$$\varphi(q,a) = \begin{cases} q_{\theta_a^1(\mathbf{s})}, & \text{если } \beta_4 = \mathbf{n}, \\ \varphi_{\mathfrak{A}^+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{s}, \end{cases}$$

и

$$\psi(q,a) = \begin{cases} \theta_a^1(\mathbf{s}), & \text{если } \beta_4 = \mathbf{n}, \\ \psi_{\mathfrak{A}^+}(q_{\mathbf{n}}, a), & \text{если } \beta_4 = \mathbf{s}. \end{cases}$$

7. Если не имеет место ни одно из предыдущих условий, то автомат переходит в состояние $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, где β_i, $1 \le i \le 5$, определено следующим образом. Прежде всего $\beta_4 = \alpha_4$ и $\beta_3 = \alpha_3$. Если

$$\alpha' = \alpha_1 + \mathrm{ch}[\psi_{\mathfrak{A}(\alpha_3)}(q_{\alpha_5}, a)] \ge [\lambda_2],$$

то $\beta_1 = \alpha'$, где

$$\mathrm{ch}(\omega) = \begin{cases} l, \text{ если } \omega = \mathbf{e}, \\ -l, \text{ если } \omega = \mathbf{w}, \\ m, \text{ если } \omega = \mathbf{n}, \\ -m, \text{ если } \omega = \mathbf{s}, \\ 0, \text{ если } \omega = \mathbf{0} \end{cases}$$

и $\mathfrak{A}(\alpha_3) = \mathfrak{A}^+$, если $\alpha_3 = 2$, а $\mathfrak{A}(\alpha_3) = \mathfrak{A}^-$, если $\alpha_3 = 1$; в противном случае

$$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = \theta_a^2[(\alpha_1 + \mathrm{ch}(\overline{\alpha}_5), \beta_1'', 2, \alpha_4, \overline{\alpha}_5)] \quad \text{и} \quad \psi(q,a) = \theta_a^1(\overline{\alpha}_5),$$

где

$$\beta_2'' = \begin{cases} 2, \text{ если } \beta_5 = \mathbf{w}, \\ 0, \text{ если } \beta_5 = \mathbf{s}, \\ 1, \text{ если } \beta_5 = \mathbf{e}, \\ 2, \text{ если } \beta_5 = \mathbf{n}. \end{cases}$$

Далее, $\beta_5 = \psi_{\mathfrak{A}(\alpha_3)}(q_{\alpha_5}, a)$ и $\psi(q, a) = \psi_{\mathfrak{A}(\alpha_3)}(q_{\alpha_5}, a)$. Параметр β_2 равен

$$\alpha_2 +_3 \nu^{\mathrm{sm}(\alpha_3)}(\alpha_5 \psi_{\mathfrak{A}(\alpha_3)}(q_{\alpha_5}, a)),$$

где

$$\mathrm{sm}(\alpha_3) = \begin{cases} +, \text{ если } \alpha_3 = 2, \\ -, \text{ если } \alpha_3 = 1. \end{cases}$$

Смысл параметров α_1, α_2, α_3, α_4 и α_5 будет аналогичен смыслу тех же параметров в доказательстве теоремы 2.

Как и в теореме 2, легко убедится, что автомат $\mathfrak{A}_{\vec{v}}(d)$ удовлетворяет условиям данной теоремы. Если лабиринт L из рассматриваемого класса содержит n вершин, то время обхода такого лабиринта можно оценить, как в доказательстве теоремы 2, т.е. оно не больше $3n^2 + 2n - 2$. Этим теорема доказана.

Список литературы

[1] А. Н. Зырицхев, *О синтезе автомата, обходыасхцхего плоские лабиринты с ограницхенными дырами*, Дискретнаыа математика **3** (1991), no. 1, 105–113.

[2] А. А. Золотых, *Обход лабиринтов с ограницхенными в фиксированных направлениыах дырами*, Дискретнаыа математика **5** (1993), no. 1, 59–69.

[3] Г. Килибарда, *О сложности автоматного обхода лабиринтов*, Дискретнаыа математика **5** (1993), no. 3, 116–124.

Универзитет у Београду, ТМФ, Карнеджијева 4, 11000 Београд, Югославия

Универзитет у Београду, ТМФ, Карнеджијева 4, 11000 Београд, Югославия

FILOMAT (Niš) 9:3 (1995), 753–764
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

DISCRETE METHODS FOR VISUALIZING FRACTAL SETS

Ljubiša M. Kocić

ABSTRACT. A short summary of some known discrete visualizig models of fractal sets is given. A new algorithm, called graphical erosional algorithm, for visualising fractal sets from $\mathbf{R}^2$ is presented. Input parameters for the algorithm are functions from a hyperbolic iterated system. Beside visualizing, this algorithm permits estimation of fractal dimension for a set being visualized.

1. Introduction

The set of points from R^2, defined by $S = \{(x_i, y_i), i = 1, \dots n_x, j = 1, \dots n_y, n_x, n_y \in \mathrm{N}\}$ will be referred as *the picture support*. Let P be an arbitrary set such that $2 \leq Card(P) \leq n_c (\in \mathrm{N})$ and $\varphi : S \to P$ be any mapping. Then P is called *set of colors* and φ is *color function*. Under *discrete visualization* of an arbitrary set A one assumes the map $\phi : A \to S$, with a given color function ϕ. The triple (ϕ, S, φ) will be called a *discrete visual model* of A.

Discrete visual models are important for processing information by computer, especially when the plane set A has a complicated form, for example, when it is a fractal set or a chaotic attractor, like those in Figure 1. Actually, this figure illustrates orbits of two different dynamical systems. Namely, let $(\mathbf{X}, d)$ be a metric space and $f : \mathbf{X} \to \mathbf{X}$ be an arbitrary mapping. Then, $(\mathbf{X}, f)$ is a *dynamical system*. For any $x_0 \in \mathbf{X}$, the sequence $\{x_i\}_{i=0}^{+\infty}$, such that $x_{i+1} = f(x_i)$ is called the *orbit* of the point x_0. The limit of an orbit

This research was partly supported by Science Fund of Serbia, grant number 0401F, through Matematički institut SANU

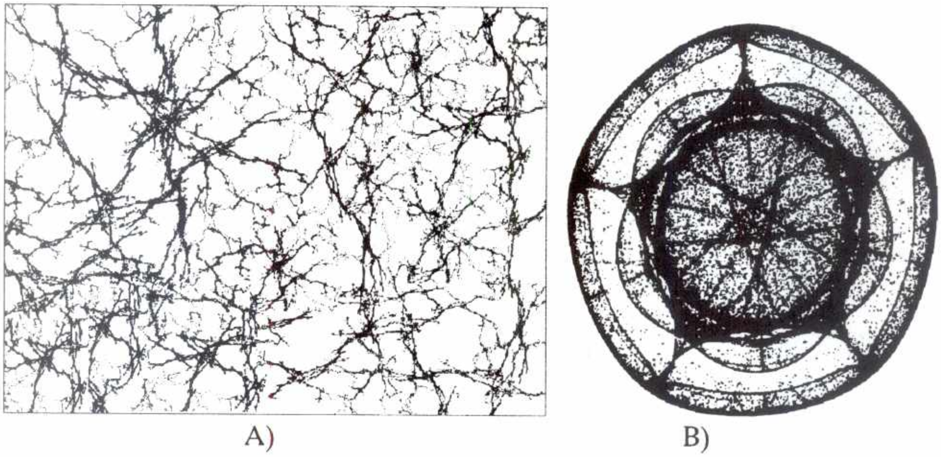


Figure 1. a) Neural network as fractal set; b) Pentagonal chaotic attractor of Chossat-Golubitsky mapping

can be a set A in $\mathbf{X}$ which "attracts" an orbit, so it is called *attractor* of dynamical system. In fact, the attractor

$$A = \lim_{n\to+\infty} x_n \;, \quad A \in \mathbf{X} \;,$$

is a fixed point of the mapping f and it does not depend on x_0. Interesting attractors usually have noninteger Hausdorf dimension, wherefrom the term "fractal" roots its name [5].

Example 1.1. Consider the mapping $f : \mathcal{H}\mathbf{R}^2 \to \mathcal{H}\mathbf{R}^2$ ($\mathcal{H}\mathbf{X}$ is the partitive set of $\mathbf{X}$), such that $f = f_1(B) \cup f_2(B)$, for any $B \subset \mathbf{R}^2$, where f_1 and f_2 are affine plane transformations defined by

$$\begin{aligned} f_1 &: \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0.1 & -0.7 \\ 0.7 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix} \;, \\ f_2 &: \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -0.3863 & 0.1562 \\ -0.3562 & -0.6863 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.4 \\ 0.4 \end{pmatrix} \;. \end{aligned} \tag{1.1}$$

The attractor A of the dynamical system $(\mathcal{H}\mathbf{R}^2, f)$ has the form of a neural cell (see Figure 5). By simple affine transformations of A the model of neural network, a fragment of neural tissue, displayed in Figure 1-A), is obtained.

Example 1.2. Let $\mathbf{C}$ be a complex plane and $f : \mathbf{C} \to \mathbf{C}$ be a mapping given by

$$f(z) = z(z^9 + 4z\overline{z} + \overline{z}^9 - 2.6) + \overline{z}^8 \;. \tag{1.2}$$

The orbit of the point $z_0 = 0$ tends to the attractor of dynamical system $(\mathbf{C}, f)$ shown in Figure 1-B). The mapping (1.2) is known as Chossat-Golubitsky formula [3], [6].

In these examples, two different algorithms are used for creating visual models of corresponding attractors. In both of them, the formula (1.1) or (1.2) are treated like continued expressions, but in computer enviroment, earlier or later, they are rounded off and the discrete values are used for creating visual model. In the next section, several methods that uses discrete tools for creating fractal visual models.

2. Discrete fractal structures

Hausdorf dimension is the most important number connected with a fractal attractor. It offers an estimation how "dense" this attractor occupies the metric space in which it is immerged. For an arbitrary $A \subset \mathbf{X}$, the *Hausdorf dimension** is defined by

$$D_H(A) = \inf_{\mu(A,p)=0} \{p\},$$

where, for $p \in \mathbf{R}$, $p \mapsto \mu(A, p)$ is the *Hausdorff p-dimensional measure* of A

$$\mu(A,p) = \sup_{\varepsilon > 0} \{\inf\{\sum_{i=0}^{+\infty} |K_i|^p\}\} .$$

Infimum is taken over all ε-covers $\mathcal{K} = \{K_i\}_{i=0}^{+\infty}$ of A. In the above formula, $|K_i|$ stands for the diameter of $K_i \subset \mathbf{X}$.

Example 2.1. Let consider the Pascal's triangle of binomoal coefficients [8]. Select these elements $p_{n,k} = \binom{n}{k}$ for which $p_{n,k}$ *mod* $2 = 1$ to obtain the set A. Choose the set of colors $P = \{\mathit{white}, \boldsymbol{black}\}$, and map A in the picture support S by replacing each element of A by a black point (see Figure 2). The visual model of A recognizes as a famous Sierpinski triangle. As it is shown in [8], Hausdorf dimension of A is $D_H = \log_2 3 = 1.58496...$ which is known to be the dimension of Sierpinski triangle [5].

Example 2.2. Many biological object possesses typical fractal properties. One of them, the neural tisue, is mentioned in Figure 1. Another one is the DNA chain, very important natural pattern that conways genetic information. DNA has a form of a double helix being composed of two strands that bind together by a specific base-pairing rule. *Adenine* (A) always pairs with

*Also known as *Hausdorf-Besicovitch* or *geometric* dimension

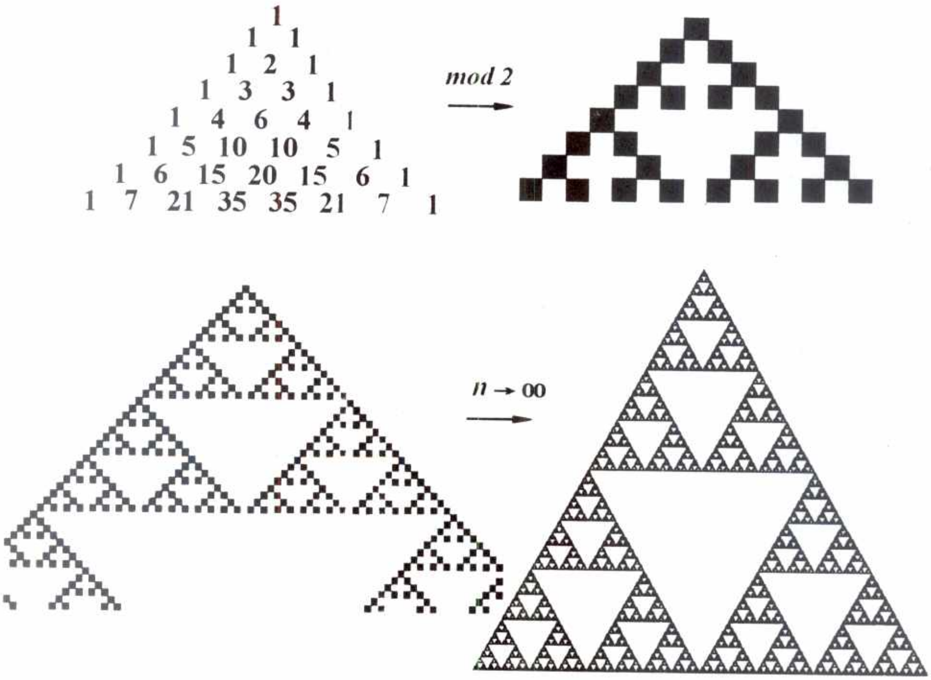


FIGURE 2. BINOMIAL COEFFICIENTS MOD 2 FORM THE SIERPINSKI TRIANGLE

thymine (T) whilst *cytosine* (C) always pairs with *guanine* (G), like in this fragment

$$\begin{array}{cccccccccccccccccccc}
- & A & A & C & T & G & G & G & A & T & A & T & A & T & T & T & G & G & G & - \\
 & | & | & | & | & | & | & | & | & | & | & | & | & | & | & | & | & | & | & \\
- & T & T & G & A & C & C & C & T & A & T & A & T & A & A & A & C & C & C & -
\end{array}$$

Each DNA strand can be connected with a Brownian motion path as follows: the pair AT corresponds to a particle being moved forward the x-axis for a given step. TA combination moves the particle in the oposite sense for the same step. CG and GC pair moves the particle along the y-axis or in the opposite direction. Alternation AC or CA with GT or TG directs along the $[(0,0)(1,1)]$ vector and the contrary, while AG/GA followed by CT/TC moves the particle along $[(0,0)(-1,-1)]$ or back.

Experiments, done by authors of [2], show that amino-bases of DNA, taken from GenBank has $D_H = 1.631 \pm 0.137$ which is significantly lower than Hausdorf dimension of the curve being a trajectory of a Brownian motion, which is $D_H = 2$. Figure 3 shows the path of Brownian motion (A) and the pseudorandom walks of two DNA (B) and (C).

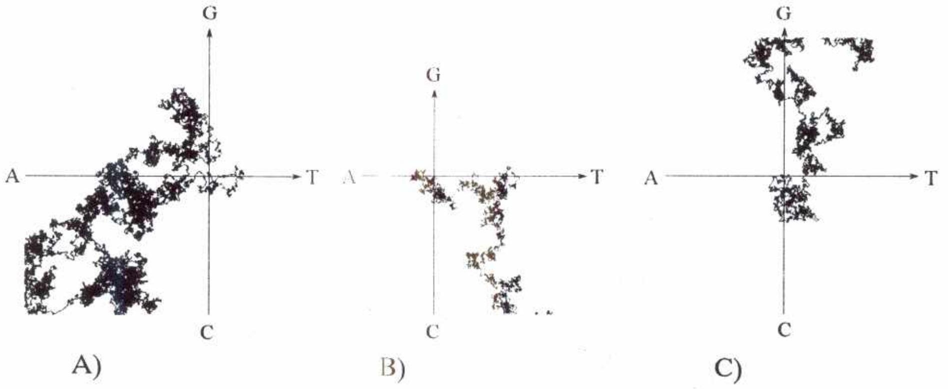


FIGURE 3. A) BROWNIAN MOTION, $H_D = 2$; B) DIMER OPSIN GENE, $D_H \approx 1.744$; C) ALPHA-1-GLYCOPROTEIN, $D_H \approx 1.671$

Example 2.3. Another discrete method for visualizing fractal sets is connected with tilling-patterns. Again start with genetic sequence of DNA taken from the human immunodeficiency virus type 1 (AIDS), and associate the Escherian* tile pattern shown in Figure 4-A) (above) in different orientation, depending on the letter A,T,C or G in the strand. An Escher-like tile is obtained (Fig. 4-A) bellow). Number of closed diamonds in the pattern, divided by the number of tiles, so called *diamond fraction*, characterize the randomnes of the data. If this fraction is about 0.05, the data are randomly distributed. Correlation appears if the fraction tends to zero.

Another type of patern, shown in Figure 4-B) is called Truchet patern, after Sebastian Truchet that studied such patterns in his paper from 1704. Diamond fraction is now replaced by the *dumbell fraction* which makes about 0.0125 for the random data. Increasind diagonal trend in the pattern reveals increasing correlation of the data. For more details see [6].

3. Graphical erosional algorithm

One of the most suitable ways to define and produce fractal sets is by (hyperbolic) Iterative Function System (IFS). This is a collection of contractive maps $(f_1, \ldots, f_n)$ that act in a metric space $(\mathbf{X}, d)$, i.e. $W = \{\mathbf{X}, f_1, \ldots, f_n\}$. The Lipschitz factor of W is $s = \max_i\{s_i\}$, where s_i is the contractive factor of f_i. Then, there exists a unique attractor A such that $A = F(A)$, where $F = \cup_{i=1}^{n} f_i$, the assertion known as the Hutchinson theorem. In other words, $A \in \mathcal{H}\mathbf{X}$ is a fixed point for the dynamical system $(\mathbf{X}, F)$.

*after Mauritus Cornelis Escher (1898–1971), duch artist

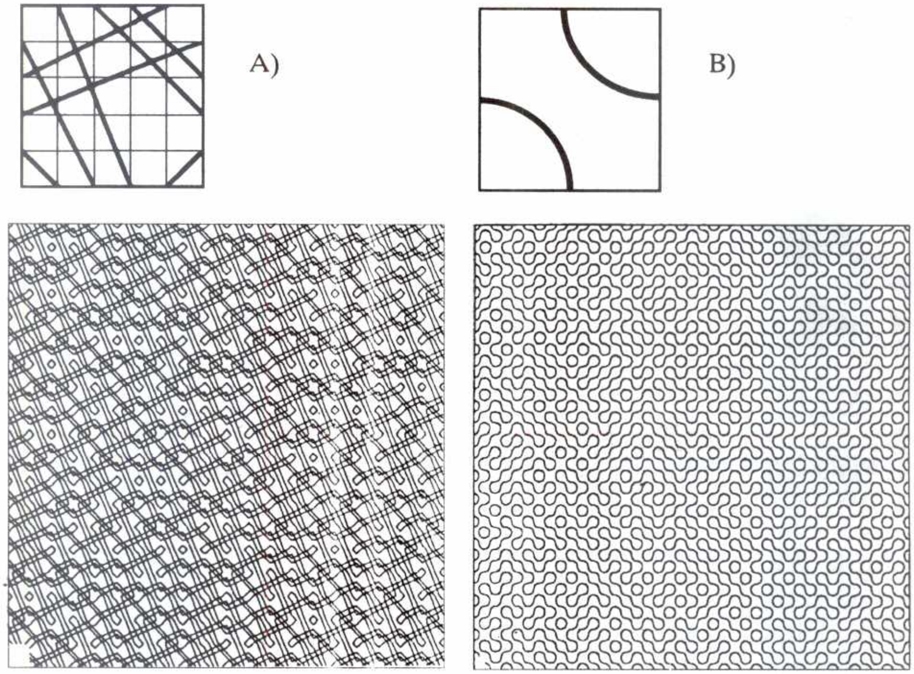


Figure 4. a) Escher-tile of random data (eschergram); b) Truchet-tile of random data

According to the literature ([4]), there are five different algorithms for calculating (and therefore for visualizing) fractal attractors in $(\mathbf{R}^2, d)$.

Algorithm A. *Based on the Hutchinson theorem, this algorithm starts with an arbitrary closed subset B from* $\mathbf{R}^2$ *and transforms it by F. More precisely,*

a) initialize $B_0 \in \mathcal{H}(\mathbf{R}^2)$,

b) calculate $B_{k+1} = \cup_i f(B_k)$, $k = 0, 1, \ldots,$

c) apply a discrete visualization $\phi : B_k \to B_k^s$.

Repeat b) and c) until $h(B_k^s, B_{k+1}^s) < \varepsilon$, *where h is Hausdorff metric and* ε *is the minimal distance between points in the picture support (picture norm).*

Good results are gain by choose B_0 *to be a singleton, typically a fixed point of one of contractions* f_i *from IFS.*

Algorithm B (Barnsley, Demko). *This algorithm uses a sequence of independent random variables* $\{\varphi_j\}_{j\in\mathbf{N}}$, *such that* $pr(\varphi_j = f_i) > 0$ *for any j and* $i = 1, \ldots n$.

a) Choose $r_0 \in \mathbf{R}^2$,

b) calculate $r_k = \varphi(x_{k-1})$, $k = 0, 1, \ldots,$

c) map each r_k *onto the picture support.*

Repeat b) and c) until the Hausdorff distance between two consecutive pictures become smaller than the picture norm.

The following three algorithms are given by Dubuc and Elqortobi [4].

Algorithm C (Based on Williams formula). *Let W^* be the set of all finite compositions of functions of W. Let for $g \in W^*$, $Fix(g)$ denotes the fixed point of g. Williams in* [7] has shown that the closure of $\cup_{g \in W^*} Fix(g)$ is invariant and there is no other closed bounded invariant sets for F.

Let $\varepsilon > 0$ and $W(\varepsilon)$ be a family of contractions. A function g is in $W(\varepsilon)$ if there is a finite sequence of functions of W, $f_1, f_2, \ldots, f_n$, such that:

(1) g is the composition $f_1 \circ f_2 \circ \cdots \circ f_n$ and the Lipschitz constant of g is $\leq \varepsilon$.

(2) If $k < n$, then the Lipschitz constant of $f_1 \circ \cdots \circ f_k$ is larger than ε.

Then the set $B = \{Fix(g) : g \in W(\varepsilon)\}$ is an approximation of A.

Algorithm D. *This is a variant of the Algorithm C. The attractor A is approximated by $B = \{h \circ R \circ g(x), h \in W(\varepsilon)\}$, where R is a rounding map of the metric space* $\mathbf{X}$.

Algorithm E (Graphical algorithm). *Let δ be a positive real number, and $M(\delta)$ be a subset of* $\mathbf{X}$ *such that*

(a) for any x in $\mathbf{X}$, $d(x, M(\delta)) < \varepsilon$;

(b) in any ball of $\mathbf{X}$, *there is just a finite number of points of $M(\delta)$.*

Let B_n and C_n are two sequences of subsets of $\mathbf{X}$ *for $n = 0, 1, \ldots$. Then,*

(1) *A point x is choosen in C_n;*

(2) *A temporary set T is initially empty. A loop over W is done, such that for each $f \in W$ for which $d(f(x), B_n \cap T) \geq \delta$, one chooses a point $x' \in M(\delta)$, such that $d(f(x), x') < \delta$, and is added to T.*

(3) $B_{n+1} = B_n \cup T$ *and* $C_{n+1} = C_n \cup T \backslash \{x\}$.

Probably the most important quantity connected with the fractal set is its dimension. There are many definitions of dimensions, but the way of calculating them may be an awquard question. The most popular method for experimental estimating the fractal dimension of an attractor in $\mathbf{R}^2$ is the box-counting method, which is based on the following theorem [1]:

Theorem 1 (Box Counting Theorem). *Let $A \in \mathcal{H}(\mathbf{R}^2)$, and Euclidean metric is used. Tile the plane* R^2 *by the square uniform mesh with the step 2^{-n}. Let $\mathcal{N}_n(A)$ denote the number of boxes from the mesh that intersect the attractor. If*

$$D_H = \lim_{n \to \infty} \left\{ \frac{\ln \mathcal{N}_n(A)}{\ln(2^n)} \right\} , \tag{2.1}$$

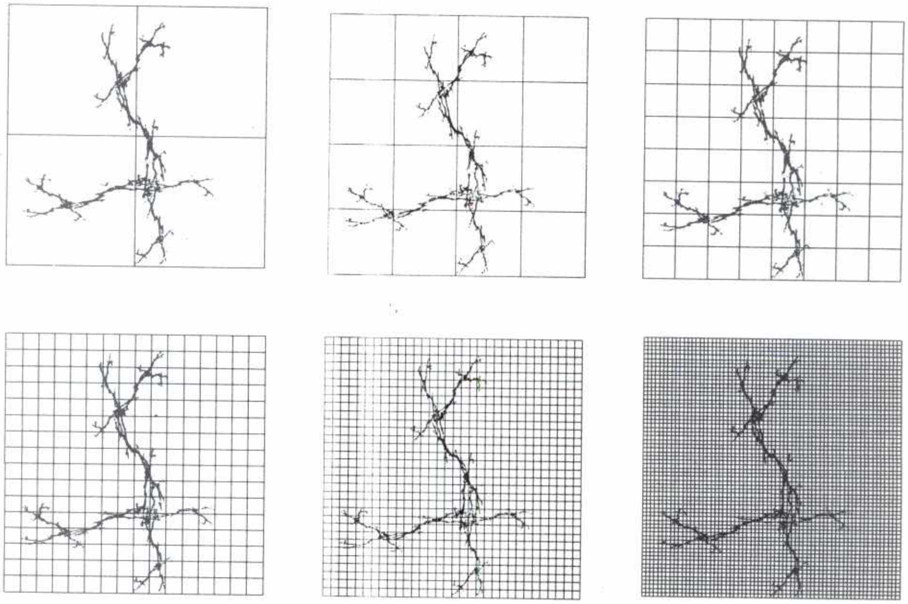


FIGURE 5. A "NEURAL CELL" AND THE BOX COUNTING METHOD

*exists, then A has fractal dimension** D_H.

The graphical erosional algorithm which will be described bellow, gives succesive approximations of fractal sets in **R** and calculates an approximation of its fractal dimension at same time.

Actually, let M_n be a uniform mesh as described in Theorem 1. Note that the scale plays no role in calculating D_H from (2.1). So, for a unit of measure one can take the side of a square that "nicely" framed the set A. Figure 5 shows a fractal attractor that resembles to the neural cell. It is framed by an apropriate square M_0 with the side lenght 1. It is divided into four subsquares which corresponds to the net M_1 (the upper left square in Figure 5). The process continues for $n = 3, 4, 5, 6$. Let $K_n(A)$ denotes 2^{-n}-cover of A. The following algorithm produces sets $K_n(A)$ and calculate the fractal dimension at the same time.

Graphical Erosional Algorithm (GEA). *Let $n \in \mathbf{N}$. The set of nodal points $\{(i/2^n, j/2^n)\}$ from $\mathbf{R}^2$ determines the standard orthogonl net M_n. Let $\Delta_{i,j}^n$ be a cell of M_n, i.e. the set of points (x, y) such that $i/2^n \leq x <$*

*also called *box dimension*

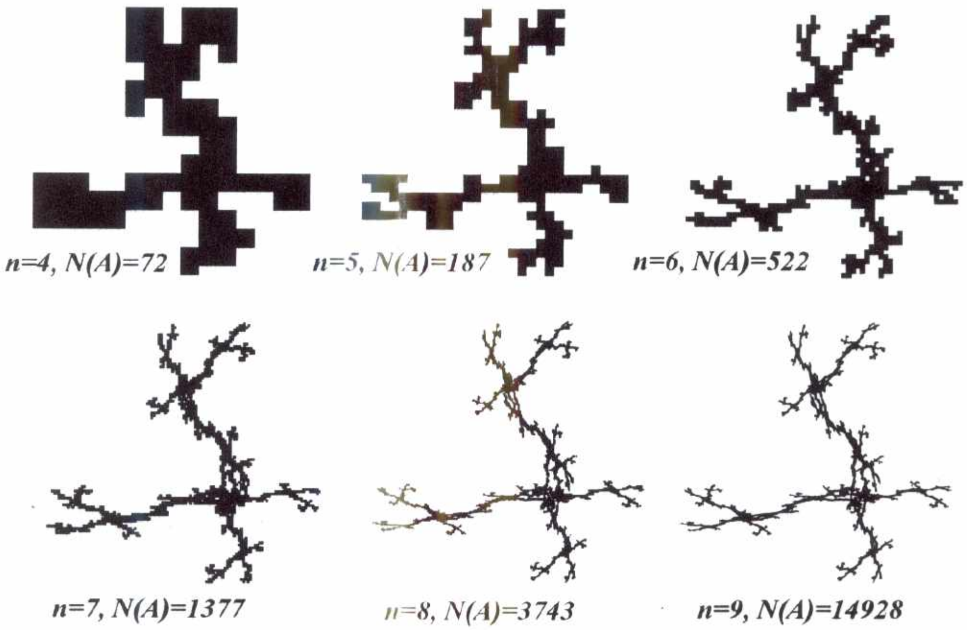


FIGURE 6. GRAPHICAL EROSIONAL ALGORITHM

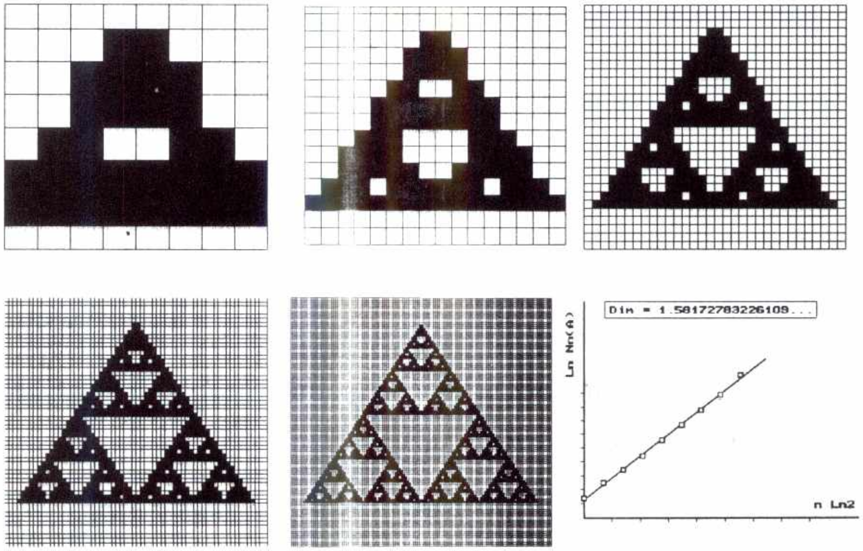


FIGURE 7. TESTING OF GRAPHICAL EROSIONAL ALGORITHM

$(i+1)/2^n$ and $j/2^n \leq y < (j+1)/2^n$. Let $S \in \mathbf{R}$ be the picture support with the norm $\delta(S) = \max\{d(x_i, x_{i+1}),\ d(y_j, y_{j+1})\}$, and $P = \{black, white\}$ be the set of colors. For any IFS, say $W = \{\mathbf{R}, f_1, \dots, f_m\}$ with the attractor A, the sequence of color functions $\phi_n : A \to S$ is associated with the net M_n according to the following steps:

(1) *Initialize $n = 1$ and $\mathbf{R}$ — $\{white\}$;*
(2) *Produce the point $p_k = (x_k, y_k)$ by the Barnsley-Demko algorithm;*
(3) *If $p_k \in \Delta^n_{i,j}$, then $\tilde{K}_n = S \cap \Delta^n_{i,j} \to \{black\}$;*
(4) *Count the number $\mathcal{N}_n(A)$ of Δ^n-cells in K_n;*
(5) *$n \leftarrow n+1$. If $\delta(S) > 2^{-n}$ then go to (1), otherwise go to (6).*
(6) *Calculate an approximation value of D_H, given by* (2.1), by fitting the data $\{(\ln(2^n), \ln \mathcal{N}_n(A))\}$ by a least-square affine function. The coefficient of the linear term is D_H. $D_H \approx \tan\alpha$.

This algorithm is illustrated in Figure 6 with A being a neural cell from Fig. 5. Note that the 'black' set, $\tilde{K}_n$, generated in step 3 approximates 2_n-covers of A. The number of black squares is denoted by $N(A)$.

Then the following theorem supports the algorithm:

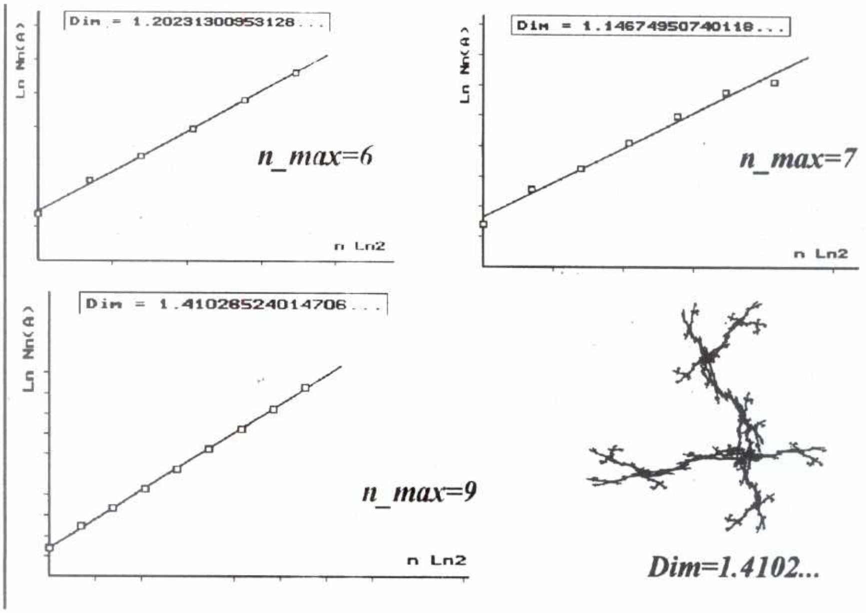


FIGURE 8. DIMENSION "BY HANDS' AND BY GEA

Theorem 2. *If $\delta(S) \to 0$ and $n \to \infty$, then the sequence $\{\tilde{K}_n\}_{n=1}^{\infty}$, generated by the GEA converges to A, in Hausdorff metric.*

Proof. Denote the Hausdorff distance between two sets X and Y from $\mathbf{R}^2$ by $h(X,Y)$. Suppose that this metric is induced by the Euclidean metric in $\mathbf{R}^2$. It is obvious that

$$h(S, A) \leq \frac{\sqrt{2}}{2}\delta(S) \; . \tag{3.1}$$

Let $r_k = (x_k, y_k)$ be the point produced by the Algorithm B. If $r_k \in \Delta_{i,j}^n$ then the cell $\Delta_{i,j}^n$ becomes a part of $K_n(A)$. As $h(r_k, \Delta_{i,j}^n) \leq \sqrt{2}2^{-n}$, then

$$h(K_n(A), A) \leq 2^{-n+\frac{1}{2}} \; , \tag{3.2}$$

As by definition, $\tilde{K}_n(A) = K_n(A) \cap S$, then by (3.1) and (3.2)

$$h(\tilde{K}_n(A), A) \leq \frac{\sqrt{2}}{2}(2^{-n+1} + \delta(S)) \; .$$

Thus, if $n \to \infty$ and $\delta(S) \to 0$ then $h(\tilde{K}_n(A), A) \to 0$, in Hausdorff metric.

Note that the proof holds if the Euclidean metric is replaced by any other metric in $\mathbf{R}^2$.

Calculation of fractal dimension follows from Theorem 1.

Since the algorithm constructs succesive $1/2^n$-covers of A, it resembles the process of erosion, which suggests the name. Algorithm is tested through several examples. Here, the results of applying GEA on Sierpinski triangle fractal attractor is shown in Figure 7. The estimated fractal dimension is 1.5817... which approximates the true dimension 1.58496... with accuracy 10^{-2}, which is a good result for PC computer where n can not exceed 9.

Fractal dimension of the "neural cell" is estimated to be 1.4102.... The data and the fitting line are shown in Figure 8 (left-bellow). Comparing with box-counting performed "by hands" for $n \leq 6$ and $n \leq 7$ (same figure, above), the data produced by GEA are much more regularly placed along the line. Note that accuracy fails for $n = 7$ due to the weakness of the human eye.

The GEA has one more advantage. It can be used for the rough estimation of the fractal attractor's shape, its dimensions and location in $\mathbf{R}^2$.

References

[1] M. F. Barnsley, *Fractals Everywhere*, Academic Press, 1988.

[2] C. L. Berthelsen, J. A. Glazier, M. H. Skolnick, *Global fractal dimension of human DNA sequences treated as pseudorandom walks*, Phys. Rew. A **45** (1992), 8902–8913.

[3] P. Chossat, M. Golubitsky. *Symmetry-increasing bifurcations of chaotic attractors*, Physica D **32** (1988), 423–426.

[4] S. Dubuc, A. Elqortobi, *Approximation of fractal sets*, J. Comput. Appl. Math. **29** (1990), 79–89.

[5] B. B. Mandelbrot, *The fractal geometry of nature*, Updated and augmented, W. H. Freeman, New York, 1983.

[6] C. A. Pickover, *Mathematics and beauty: Several short classroom experiments*, Notices AMS **38** (1991), 192–196.

[7] R. F. Williams, *Composition of contractions*, Bol. Soc. Brasil. Mat. **2** (1971), 55–59.

[8] S. Wolfram, *Geometry of binomial coefficients*, Amer. Math. Monthly **91** (1984), 566–570.

Department of Mathematics, Faculty of Electronic Engineering, P.O.Box 73, 18000 Niš

FILOMAT (Niš) 9:3 (1995), 765–781
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

IDENTITY, PERMUTATION AND BINARY TREES

Aleksandar Kron

ABSTRACT. Some extensions of the Anderson-Belnap (Dwyer-Powers) conjecture for $\mathbf{TW}_{\rightarrow}$ are applied to a set of binary trees.

A binary tree $\mathcal{T}$ is a tree with an origin and such that each node of $\mathcal{T}$ either has exactly two immediate successors or it is an end-node. A subtree $\mathcal{T}'$ of a tree $\mathcal{T}$ is a tree such that every node of $\mathcal{T}'$ is a node of $\mathcal{T}$ and the immediate successor relation in $\mathcal{T}'$ is the immediate successor relation in $\mathcal{T}$. A tree $\mathcal{T}$ is a *formula-like tree* (FLT) iff it has no proper subtree that is isomorphic to $\mathcal{T}$, and no proper subtree $\mathcal{T}'$ of $\mathcal{T}$ has a proper subtree $\mathcal{T}''$ isomorphic to $\mathcal{T}'$.

Every infinite FLT contains infinitely many (distinct) finite branches; every node of a FLT is a node of a finite branch. Hence, the maximal length of a branch of a FLT is ω.

If $\mathcal{T}_1$ and $\mathcal{T}_2$ are FLTs, then the tree obtained by adjoining a new origin $\mathcal{T}_1\mathcal{T}_2$ and such that $\mathcal{T}_1$ and $\mathcal{T}_2$ are immediate successors of $\mathcal{T}_1\mathcal{T}_2$, is a FLT.

With every node of a FLT $\mathcal{T}$ one of the numbers 0 or 1 is associated, as follows: 0 is associated with the origin of $\mathcal{T}$; if 0 (1) is associated with a node at level n, then 1 (0) is associated with its left hand immediate successor and 0 (1) is associated with its right hand immediate successor.

If 0 (1) is associated with the origin of a subtree $\mathcal{T}'$, then $\mathcal{T}'$ is a 0-subtree (1-subtree).

Let us consider the following operations on a FLT.

SU^0 Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_1\mathcal{T}_2$ be one of its 0-subtrees; then the subtree $\mathcal{T}_1\mathcal{T}_2$ can be cut off and a subtree $(\mathcal{T}_2\mathcal{T}_3)(\mathcal{T}_1\mathcal{T}_3)$ can be inserted in $\mathcal{T}$ instead, where $\mathcal{T}_3$ is any FLT.

PR^0 Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_2\mathcal{T}_3$ be one of its 0-subtrees; then the subtree $\mathcal{T}_2\mathcal{T}_3$ can be cut off and a subtree $(\mathcal{T}_1\mathcal{T}_2)(\mathcal{T}_1\mathcal{T}_3)$ can be inserted in $\mathcal{T}$ instead, where $\mathcal{T}_1$ is any FLT.

SUL^0 Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_1$ be one of its 0-subtrees; then the subtree $\mathcal{T}_1$ can be cut off and a subtree $(\mathcal{T}_1\mathcal{T}_2)((\mathcal{T}_2\mathcal{T}_3)\mathcal{T}_3)$ can be inserted in $\mathcal{T}$ instead, where $\mathcal{T}_2$ and $\mathcal{T}_3$ are any FLTs.

SU^1 Let $\mathcal{T}$ be a FLT and let $(\mathcal{T}_2\mathcal{T}_3)(\mathcal{T}_1\mathcal{T}_3)$ be one of its 1-subtrees; then the subtree $(\mathcal{T}_2\mathcal{T}_3)(\mathcal{T}_1\mathcal{T}_3)$ can be cut off and the subtree $\mathcal{T}_1\mathcal{T}_2$ can be inserted in $\mathcal{T}$ instead.

PR^1 Let $\mathcal{T}$ be a FLT and let $(\mathcal{T}_1\mathcal{T}_2)(\mathcal{T}_1\mathcal{T}_3)$ be one of its 1-subtrees; then the subtree $(\mathcal{T}_1\mathcal{T}_2)(\mathcal{T}_1\mathcal{T}_3)$ can be cut off and the subtree $\mathcal{T}_2\mathcal{T}_3$ can be inserted in $\mathcal{T}$ instead.

SUL^1 Let $\mathcal{T}$ be a FLT and let $(\mathcal{T}_1\mathcal{T}_2)((\mathcal{T}_1\mathcal{T}_3)\mathcal{T}_3)$ be one of its 1-subtrees; then the subtree $(\mathcal{T}_1\mathcal{T}_2)((\mathcal{T}_2\mathcal{T}_3)\mathcal{T}_3)$ can be cut off and the subtree $\mathcal{T}_1$ can be inserted in $\mathcal{T}$ instead.

PERM* Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_1(\mathcal{T}_2\mathcal{T}_3)$ be one of its 0-or-1-subtrees; then the subtree $\mathcal{T}_1(\mathcal{T}_2\mathcal{T}_3)$ can be cut off and the subtree $\mathcal{T}_2(\mathcal{T}_1\mathcal{T}_3)$ can be inserted in $\mathcal{T}$ instead.

The main theorem: starting with a FLT $\mathcal{T}$ and successively performing the operations SU^0, PR^0, SUL^0, SU^1, PR^1, SUL^1, and PERM* any finite number of times, in any order, and such that at least one of the first six rules is applied at least once, it is not possible to obtain $\mathcal{T}$ as a result.

1. Identity

When Alan Ross Anderson and Nuel D. Belnap were developing relevance logic, among numerous systems they have been considering there was an implicational fragment of a very weak logic called now **TW**. Since there is only one connective in such a fragment, namely →, we omit it and we write AB for $A \to B$. Also, we omit parentheses, whenever this causes no confusion. ABC stands for $(AB)C$ and $A.BC$ for $A(BC)$. Under this proviso, the implicational fragment of **TW** has modus ponens (MP) as the sole rule of inference and the following axiom-schemata:

ID	AA
ASU	$AB.BC.AC$
APR	$BC.AB.AC$

This fragment is now called $\mathbf{TW}_{\to}$. Let us write $A \equiv B$ iff both AB and BA are provable in $\mathbf{TW}_{\to}$. Then Anderson and Belnap have conjectured (cf. [1], p. 95) that

$$A \equiv B \text{ if and only if } A \text{ and } B \text{ are the same formula.}$$

We shall call this conjecture *Anderson - Belnap conjecture* (A-B).

By A-B the identity of formulas in the language of $\mathbf{TW}_{\rightarrow}$ is determined by logical means only – by provability in the very weak theory of implication $\mathbf{TW}_{\rightarrow}$.

Let $\mathbf{TW}_{\rightarrow}$–ID be the system obtained from $\mathbf{TW}_{\rightarrow}$ by omitting the axiom schema ID. Dwyer and Powers have shown that A-B is equivalent to the following claim:

NOID In $\mathbf{TW}_{\rightarrow}$–ID there is no theorem of the form AA.

NOID is a very strong claim. ID is a paradigm of a logical truth and there is hardly a descent logical theory where ID is not true. Nevertheless, in $\mathbf{TW}_{\rightarrow}$–ID not only there is a formula A such that AA does not hold for it, but, moreover, AA holds for no formula A.

The systems $\mathbf{TW}_{\rightarrow}$ and $\mathbf{TW}_{\rightarrow}$–ID have other formulations as well.

Let us consider a theory that has ASU and APR as axiom-schemata, but instead of MP it has the following rules of inference:

SU From AB to infer $BC.AC$
PR From BC to infer $AB.AC$
TR From AB and BC to infer AC

Let us call the new system $\mathbf{TRW}_{\rightarrow}$–ID. It is easy to see that the rules of $\mathbf{TRW}_{\rightarrow}$–ID are derived rules of $\mathbf{TW}_{\rightarrow}$–ID; hence, all theorems of $\mathbf{TRW}_{\rightarrow}$–ID are theorems of $\mathbf{TW}_{\rightarrow}$–ID. On the other hand, by an inductive argument it follows that $\mathbf{TRW}_{\rightarrow}$–ID is closed under MP (this was proved by Dwyer and Powers; cf. [4] and [2] for details. $\mathbf{TRW}_{\rightarrow}$–ID and $\mathbf{TRW}_{\rightarrow}$ are called in [4] **M** and **N**, respectively). Therefore, $\mathbf{TW}_{\rightarrow}$–ID and $\mathbf{TRW}_{\rightarrow}$–ID are equivalent.

Let us adjoin the axiom-scheme ID to $\mathbf{TRW}_{\rightarrow}$ – ID; the resulting theory is called $\mathbf{TRW}_{\rightarrow}$. It is clear that $\mathbf{TRW}_{\rightarrow}$ and $\mathbf{TW}_{\rightarrow}$ are equivalent.

Another equivalent formulation of NOID is the following one. Let us consider the theory $\mathbf{WTR}_{\rightarrow}$–ID, in the propositional language with → as the sole connective.

The rules of inference are SU, PR and TR, as in $\mathbf{TRW}_{\rightarrow}$–ID, but the axiom-schemata are

USA $(BC.AC).AB$
RPA $(AB.AC).BC$

The axioms of $\mathbf{WTR}_{\rightarrow}$–ID are not logical truths at all. By an inductive argument it can be shown that AB is a theorem of $\mathbf{TRW}_{\rightarrow}$–ID iff BA is a theorem of $\mathbf{WTR}_{\rightarrow}$–ID. Now NOID can be formulated as:

NOID' There is no formula provable both in $\mathbf{TRW}_{\rightarrow}$-ID and in $\mathbf{WTR}_{\rightarrow}$-ID.

Let $\mathbf{WTR}_{\rightarrow}$ be obtained from $\mathbf{WTR}_{\rightarrow}$-ID by adjoining ID; then, of course, NOID has the formulation

NOID" The only theorems common to $\mathbf{TRW}_{\rightarrow}$ and $\mathbf{WTR}_{\rightarrow}$ are all the formulas of the form AA.

A natural deduction formulation of $\mathbf{TRW}_{\rightarrow}$-ID

In the seventies some one-premiss natural deduction formulations of $\mathbf{TW}_{\rightarrow}$-ID and $\mathbf{TW}_{\rightarrow}$ have been elaborated in Belgrade by Božić, Došen and the present author.

Let us define *consequent* and *antecedent* occurrences of subformulas of a given formula A, as in [1], p. 93.

The formula A itself is a consequent occurrence of A in A.

If BC is a consequent (antecedent) occurrence of BC in A, then the displayed occurrence of B is an antecedent (consequent) occurrence of B in A, and the displayed occurrence of C is a consequent (antecedent) occurrence of C in A.

The logic $\mathbf{TRW}_{\rightarrow}$-ID has a neat formulation called $\mathbf{TRW}'_{\rightarrow}$-ID. There are no axioms in $\mathbf{TRW}'_{\rightarrow}$-ID and instead of SU, PR and TR we have the following four rules:

SU^0 Let AB have a consequent occurrence in a formula D; then we are allowed to substitute an occurrence of $BC.AC$ for that particular occurrence of AB in D, for any formula C;

PR^0 Let BC have a consequent occurrence in a formula D; then we are allowed to substitute an occurrence of $AB.AC$ for that particular occurrence of BC in D, for any formula C;

SU^1 Let $BC.AC$ have an antecedent occurrence in a formula D; then we are allowed to substitute an occurrence of AB for that particular occurrence of $BC.AC$ in D;

PR^1 Let $AB.AC$ have an antecedent occurrence in a formula D; then we are allowed to substitute an occurrence of BC for that particular occurrence of $AB.AC$ in D.

SU^0 and PR^0 are called *consequent* or 0-rules; SU^1 and PR^1 are called *antecedent* or 1-rules.

Let A and B be arbitrary formulas. Suppose that B is obtained from A by applying these four rules (at least one but not necessarily all of them) in any order; then we shall write $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B$ to denote this fact. Also, we shall write $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B \longrightarrow_{\mathbf{TRW}'\text{-ID}} C$ if $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B$ and $B \longrightarrow_{\mathbf{TRW}'\text{-ID}} C$. It is clear that the relation $\longrightarrow_{\mathbf{TRW}'\text{-ID}}$ is transitive.

If $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B$, then AB is called a *theorem* of $\mathbf{TRW}'_{\rightarrow}$-ID.

The theories $\mathbf{TRW}_{\rightarrow}$-ID and $\mathbf{TRW}'_{\rightarrow}$-ID are equivalent in the sense that they have the same set of theorems. This can be seen from the following considerations.

Let us define the *depth* of an occurrence of a subformula in a formula A as follows: A itself is at depth 0; if an occurrence of BC in A is at depth n, then the displayed occurrences of B and C in A are at depth $n+1$.

Theorem 1. *The theorems of* $\mathbf{TRW}'_{\rightarrow}$*-ID are theorems of* $\mathbf{TRW}_{\rightarrow}$*-ID.*

Proof. Suppose that $D \longrightarrow_{\mathbf{TRW}'\text{-ID}} E$. Proceed by induction on the number n of applications of 0-or-1-rules in the derivation $D \longrightarrow_{\mathbf{TRW}'\text{-ID}} E$ to show that DE is a theorem of $\mathbf{TRW}_{\rightarrow}$-ID.

Let $n = 1$. Suppose that E is obtained from D by one of the 0-or-1-rules. Proceed by another induction on depth at which the substitution takes place.

If the substitution takes place at depth 0, then DE is an instance of an axiom of $\mathbf{TRW}_{\rightarrow}$-ID.

Let $DE = D_1D_2.E_1E_2$. If $D \longrightarrow_{\mathbf{TRW}'\text{-ID}} E$ such that the substitution takes place at depth greater than 0, then either $D_1 = E_1$ and $D_2 \longrightarrow_{\mathbf{TRW}'\text{-ID}} E_2$ or $D_2 = E_2$ and $E_1 \longrightarrow_{\mathbf{TRW}'\text{-ID}} D_1$. In the first case, by induction hypothesis, D_2E_2 is a theorem of $\mathbf{TRW}_{\rightarrow}$. Hence, DE is obtained by PR. In the second case, by induction hypothesis, E_1D_1 is a theorem of $\mathbf{TRW}_{\rightarrow}$. Hence, DE is obtained by SU.

Let $n > 1$. Suppose that E' is obtained from D by $n-1$ applications of 0-or-1-rules, and that E is obtained from E' by a single application of a 0-or-1-rule; by induction hypothesis and the first part of the proof, DE' and $E'E$ are theorems of $\mathbf{TRW}_{\rightarrow}$-ID. Hence, by TR DE is a theorem of $\mathbf{TRW}_{\rightarrow}$-ID. □

Theorem 2. *The theorems of* $\mathbf{TRW}_{\rightarrow}$*-ID are theorems of* $\mathbf{TRW}'_{\rightarrow}$*-ID.*

Proof. It is easy to derive ASU and APR in $\mathbf{TRW}'_{\rightarrow}$-ID. Suppose that $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B$. This means that starting from A and applying the 0-or-1-rules we eventually obtain B. Let us start from BC; in this formula every consequent occurrence of a subformula in B is an antecedent occurrence in BC, and conversely, every antecedent occurrence of a subformula in B is a consequent occurrence in BC. It is easy to see that AC can be obtained from BC by applying the same rules that lead from A to B in reverse order. This means that AC is obtained from BC by applying a 0-rule instead of the corresponding 1-rule and a 1-rule instead of the corresponding 0-rule.

Hence, if AB is a theorem of $\mathbf{TRW}'_{\rightarrow}$-ID, so is $BC.AC$. In a similar way we can prove that $\mathbf{TRW}'_{\rightarrow}$-ID is closed under PR.

As to TR, it is trivial that if $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} B$ and $B \longrightarrow_{\mathbf{TRW}'\text{-ID}} C$, then $A \longrightarrow_{\mathbf{TRW}'\text{-ID}} C$. Hence, the set of theorems of $\mathbf{TRW}'_{\rightarrow}$-ID is closed under transitivity and all theorems of $\mathbf{TRW}_{\rightarrow}$–ID are theorems of $\mathbf{TRW}'_{\rightarrow}$–ID. □

Many logicians have tried to prove or disprove A-B, but it turned out that this was a very difficult task.

NOID and hence A-B has been proved true by R.K. Meyer and E. Martin (cf. [6]) who used a semantics developed for this purpose. Thus, indeed, the graphical identity of two formulas in a language with → as the sole connective is determined by purely logical means defined in a logical calculus in the same language.

A purely constructive proof of NOID has been obtained in [4] (cf. also [2]).

2. Permutation

In $\mathbf{TW}_{\rightarrow}$–ID there is almost no rule of permutation admissible. The next theorem seems to give the maximum of permutation allowed in $\mathbf{TW}_{\rightarrow}$–ID.

Theorem 3. *If $AB.CD$ is a theorem, then either (a) $A = C$ and BD is a theorem or else (b) $B = D$ and CA is a theorem or else (c) CA and BD are theorems or else (d) $C.ABD$ is a theorem.*

Proof. Consider $\mathbf{TRW}_{\rightarrow}$–ID and proceed by induction on theorems. □

Let $C.DE$ be a subformula of A; suppose that B is obtained from A by substitution of $D.CE$ for $C.DE$, at a single occurrence of $C.DE$ and let us write $A \sim B$ iff B can be obtained from A by a finite (possibly zero) number of substitutions of this kind. It is clear that $\sim$ is an equivalence relation. For any A by A^* we shall denote any formula B such that $A \sim B$.

Let us consider the following permutation rules.

PERM* From A to infer A^*.

RPERM If AB is a theorem, so is A^*B^*.

PERM If A is a theorem, so is A^*.

Here 'theorem' means 'theorem of the system under consideration'.

Let us adjoin RPERM to $\mathbf{TRW}_{\rightarrow}$–ID and let the resulting system be called $\mathbf{PTW}_{\rightarrow}$–ID.

If PERM is adjoined to $\mathbf{TRW}_{\rightarrow}$–ID, the resulting system is called $\mathbf{L}$; APR is then redundant.

Obviously, the theorems of $\mathbf{PTW}_{\rightarrow}$–ID are theorems of $\mathbf{L}$.

Theorem 4. $\mathbf{L} = \mathbf{PTW}_{\rightarrow}$*–ID + PASU + SUP + PRP, where* PASU *is the following axiom scheme* (ASU *with permutation)*

$$\text{PASU} \qquad A.AB.BCC$$

and SUP *and* PRP *are the following rules* (SU *and* PR *with permutation):*

SUP From AB to infer $A.BCC$

PRP From BC to infer $A.ABC$

Proof. It is clear that the theorems of $\mathbf{PTW}_{\rightarrow}$–ID + PASU + SUP + PRP are theorems of **L**, for PASU, SUP and PRP are obtained from ASU, SU, and PR by PERM, respectively.

On the other hand, by induction on theorems it can be shown that $\mathbf{PTW}_{\rightarrow}$–ID + PASU + SUP + PRP is closed under PERM. The only place in this proof that requires a little care is TR. Suppose that $A.CD$ is obtained in **L** from AB and $B.CD$ by TR, and that then $C.AD$ is obtained by PERM. By induction hypothesis, $C.BD$ is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID + PASU + SUP + PRP. On the other hand, from AB we obtain $BD.AD$ by SU; hence, $C.AD$ is a theorem, by TR. Therefore, the theorems of **L** are theorems of $\mathbf{PTW}_{\rightarrow}$–ID + PASU + SUP + PRP. □

It has been proved in [4] that NOID holds for **L** as well:

NOID(L) there is no theorem of **L** either of the form AA or of the form ABB or of the form $ABBA$ or of the form $A.ABB$.

This result was obtained by constructing a cut-free Gentzen-style formulation of **L** also called **L**. The structure of the proof is the following: it was obvious that pp is not derivable in **L**; if we assume that AA is derivable for some formula A, then there is a formula B of smallest degree such that BB is derivable. In considering the possible derivations of BB, there always was a formula C of degree smaller than the degree of B such that CC was derivable.

Neither $\mathbf{PTW}_{\rightarrow}$–ID nor **L** is closed under modus ponens. A counterexample provided in [5] can be used here as well. Let $A = pp.pp.pp$ and $B = (pp.pp)p.ppp$; AB is an instance of ASU. If $\mathbf{PTW}_{\rightarrow}$–ID were closed under MP, applying RPERM to $AB.Bp.Ap$ we would obtain $AB.A.Bpp$; hence, by MP applied twice, Bpp would be obtained in $\mathbf{PTW}_{\rightarrow}$–ID, contrary to NOID(L).

There are proper extensions of **L** closed under MP such that NOID still holds for them. Let **K** be the system defined by ASU, MP, PERM and the following assertion rule

ASS1 If A is a theorem of **K**, so is ABB.

There is a Gentzen-style formulation of **K** called in [5] **J**; it has been proved that

NOID(J) there is no theorem of **J** either of the form AA or else of the form $A.ABB$ or else of the form $ABBA$.

By ASU and MP **K** is closed under another assertion rule as well:

ASS2 If A and BC are theorems of **K**, so is ABC.

The connection between **K** and **L** is given in the next theorem.

Theorem 5. $\mathbf{K} = \mathbf{L} + \mathit{ASS1}$.

Proof. The rules of **L** + ASS1 are derived in **K**. We have to prove that **L** + ASS1 is closed under MP.

Suppose that (a) A_i and (b) $A_1\ \dots\ .A_{i-1}.A_i.A_{i+1}\ \dots\ .A_np$ are theorems of **L** + ASS1; we want to prove that (c) $A_1.\ \dots\ .A_{i-1}.A_{i+1}\ \dots\ .A_np$ is a theorem of **L** + ASS1. Proceed by induction on the combined weight of (a) and (b) (for the definition of combined weight cf. [7], p. 113).

Let (b) be an instance $A_3C.CD.A_3D$ of ASU, where $A_1 = A_3C$, $A_2 = CD$, and $D = A_4\ \dots\ .A_np$.

If $i < 3$, then (c) is obtained from A_i either by SU or by PR.

If $i = 3$, then from (a) we obtain (c') A_3CC by ASS1; hence, by (c') and SU we have $CD.A_3CD$; eventually, by PERM we derive (c).

Let $i > 3$ and $E = A_{i+1}\ \dots\ .A_np$; from A_i we obtain A_iEE by ASS1. Then we apply PR to derive $A_3D.A_3.A_4\ \dots\ A_{i-1}E$ and (c') $(CD.A_3D).CD .A_3.A_4\ \dots\ .A_{i-1}E$. But as an instance of ASU we have $A_3C.CD.A_3D$; hence, by using (c') and TR we obtain (c).

Let (b) be obtained by SU; if $i = 1$, then $A_1 = A_1'C$, where $C = A_3\ \dots\ .A_np$ and $A_2 = A_2'C$, and (b) is obtained from (b') $A_2'A_1'$. From (a) and (b') we obtain $A_2'C$ by TR, as required.

If $i = 2$, then (b) is $A_1'C.A_2C$ and it is obtained from (b') A_2A_1'. By (a), (b') and the induction hypothesis we have (c') A_1'; by (c') and ASS1 we obtain (c) $A_1'CC$.

Let $i \geq 3$ and let E be as before; then (b) is $A_1'(A_3\ \dots\ .A_{i-1}.A_iE).A_2.A_3\ \dots\ .A_{i-1}.A_iE$, and it is obtained from (b') A_2A_1', where, obviously, $A_1 = A_1'.A_3\ \dots\ .A_{i-1}.A_iE$.

From (a) A_i we obtain A_iEE by ASS1, and then (c') $A_1'(A_3\ \dots\ .A_{i-1}.A_iE) .A_1'.A_3\ \dots\ .A_{i-1}E$ by PR. By using PERM we have (c") $A_1'.A_1'(A_3\ \dots\ .A_{i-1} .A_iE).A_3\ \dots\ .A_{i-1}E$. Hence, by (b'), (c") and TR we obtain $A_2.A_1'(A_3\ \dots\ .A_{i-1}.A_iE).A_3\ \dots\ .A_{i-1}E$; by using PERM again, we obtain (c).

Let (b) be obtained by PR; if $i = 1$, then $A_1 = A_2C$ and (b) is obtained from (b') $C.A_3\ \dots\ .A_np$. From (a) and (b') we obtain $A_2.A_3\ \dots\ .A_np$ by TR, as required.

If $i = 2$, then (b) is $A_2C.A_2.A_3\ \dots\ .A_np$ and it is obtained from (b') $C.A_3\ \dots\ .A_np$. By (a) and ASS1 we obtain A_2CC, and then, by using (b') and TR we obtain (c).

Let $i \geq 3$ and let E be as before; then (b) is $A_2C.A_2.A_3\ \dots\ .A_{i-1}.A_iE$, and it is obtained from (b') $C.A_3\ \dots\ .A_{i-1}.A_iE.$, where, obviously, $A_1 = A_2C$. By induction hypothesis, $A_2C.A_2.A_3\ \dots\ .A_{i-1}E$ is a theorem; now (c) is obtained by PR.

If (b) is obtained by PERM, the use of induction hypothesis is straightforward.

Let (b) be obtained by ASS1; then $A_1 = A'_1.A_2\ \dots\ .A_np$ and (b) is obtained from (b') A'_1. If $i = 1$, then by (b'), (a) and the induction hypothesis, (c) $A_2\ \dots\ .A_np$ is a theorem.

If $i > 1$, let E be as before; from (b') we obtain (c') $A'_1(A_2\ \dots\ .A_{i-1}E).A_2\ \dots\ .A_{i-1}E$ by ASS1. On the other hand, from (a) we derive (c") $A_i(A'_1.A_2\ \dots\ .A_{i-1}E).A'_1.A_2\ \dots\ .A_{i-1}E$ by ASS1. Hence, by (c"), (c'), TR and PERM we obtain (c).

Let (b) be obtained from (b') A_1C and (b") $C.A_2\ \dots\ .A_np$ by TR. If $i = 1$, then by (a), (b') and the induction hypothesis we obtain C; hence, by C, (b") and the induction hypothesis we obtain (c).

If $i > 1$, let E be as before; by induction hypothesis, (c') $C.A_2\ \dots\ .A_{i-1}E$ is a theorem; hence, by (b'), (c') and TR we obtain (c).

This completes the proof of the theorem. □

The system **K** has an interesting property called NOE. It has been proved in [5] that **J** and hence **K** is closed under the following rule:

NOE $(A_1\ \dots\ .A_nB)B$ is a theorem of **K** iff so are $A_1, \dots, A_n$.

In particular, there is no theorem of **K** of the form $AABB$.

Natural deduction systems $\mathbf{L}'$, $\mathbf{L}''$ and $\mathbf{L}'''$

Let $\mathbf{L}'$ be the one-premiss natural deduction system obtained from $\mathbf{TRW}'_{\rightarrow}$ –ID by adjoining the rule PERM*.

Let A and B be arbitrary formulas. Suppose that B is obtained from A by applying SU^0, SU^1, PR^0, PR^1, or PERM* a finite number of times; then we shall write $A \longrightarrow_{\mathbf{L}'} B$ to denote this fact. If $A \longrightarrow_{\mathbf{L}'} B$ and one of the first four rules is applied at least once, then AB is called a theorem of $\mathbf{L}'$.

The restriction in the definition of theorems of $\mathbf{L}'$ is obvious; without it we have the following derivation: starting from $A.BC$ we apply PERM* twice and we obtain $A.BC$ again; hence, without the restriction, $A(BC).A.BC$ would be a theorem of $\mathbf{L}'$.

Theorem 6. $\mathbf{PTW}_{\rightarrow}$*-ID and* $\mathbf{L}'$ *have the same set of theorems.*

Proof. Let A^*B^* be a theorem of $\mathbf{L}'$ obtained from AB by RPERM; by induction hypothesis, $A \longrightarrow_{\mathbf{L}'} B$. Hence, $A^* \longrightarrow_{\mathbf{L}'} A \longrightarrow_{\mathbf{L}'} B \longrightarrow_{\mathbf{L}'} B^*$.

This shows that $\mathbf{L}'$ is closed under RPERM and it is easy to see that the theorems of $\mathbf{PTW}_{\rightarrow}$–ID are theorems of $\mathbf{L}'$.

Suppose that $D \longrightarrow_{\mathbf{L}'} E$ and proceed by induction on the number n of applications of 0-or-1-rules. Let $n = 1$ and let E be obtained from E' by an application of such a rule and proceed by another induction on depth. If the rule is applied at depth 0, then $E'E$ is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID by axioms. Obviously, $D \sim E'$ and DE is obtained by RPERM only.

Let $E'E = E_1'E_2'.E_1E_2$. If E is obtained from E' such that E_2 is obtained from E_2' by a 0-or-1-rule, then $E_1' = E_1$ and, by the second induction hypothesis, $E_2'E_2$ is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID. Hence, $E'E$ is obtained by PR. Again, we have $D \sim E'$ and we obtain DE.

If E is obtained from E' such that E_1' is obtained from E_1 by a 0-or-1-rule, then $E_2' = E_2$ and, by the second induction hypothesis, E_1E_1' is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID. Hence, $E'E$ is obtained by SU. Again, $D \sim E'$ and DE is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID.

Let $n > 1$. Suppose that E' is obtained from D by $n - 1$ applications of 0-or-1-rules, and that E is obtained from E' either by a single application of a 0-or-1-rule or by PERM*; by the first induction hypothesis DE' is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID. If E is obtained from E' by an application of a 0-or-1-rule, then $E'E$ is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID by the first part of this proof; hence, DE is a theorem of $\mathbf{PTW}_{\rightarrow}$–ID by TR. If E is obtained from E' by PERM*, then DE is obtained by RPERM from DE'.

This completes the proof of the theorem. □

Let $\mathbf{L}''$ be the one-premiss natural deduction system obtained from $\mathbf{L}'$ by adjoining the following two new 0-or-1-rules

SUL^0 Let A have a consequent occurrence in a formula D; then we are allowed to substitute an occurrence of $AB.BCC$ for that particular occurrence of A in D, for any formulas B and C;

SUL^1 Let $AB.BCC$ have an antecedent occurrence in a formula D; then we are allowed to substitute an occurrence of A for that particular occurrence of $AB.BCC$ in D.

Theorem 7. *The theorems of* $\mathbf{L}''$ *are theorems of* $\mathbf{L}$.

Proof. The proof of Theorem 6 can be extended in the case when either SUL^0 or SUL^1 is applied to E' at depth 0; then in $\mathbf{L}$ we can apply PERM to an instance of ASU. □

It is easy to derive ASU in $\mathbf{L}''$. Also, we can show that $\mathbf{L}''$ is closed under SU, PR, and TR. Hence, $\mathbf{L}''$ contains $\mathbf{TRW}_{\rightarrow}$–ID. However, there are theorems of $\mathbf{L}$ that are not theorems of $\mathbf{L}''$. In particular, $\mathbf{L}''$ is not closed under PERM: there is a theorem of $\mathbf{L}''$ of the form $A.BC$ such that $B.AC$ is not a theorem of $\mathbf{L}''$, as in the following example.

We have $(pp.ppp)p \longrightarrow_{\mathbf{L}'} pp$ by SUL^1, but not $p \longrightarrow_{\mathbf{L}'} (pp.ppp)pp$; the latter derivation is impossible in $\mathbf{L}''$. On the other hand, in $\mathbf{L}$ from the instance $pp.pp.pp$ of ASU we obtain $p.pp.ppp$ by PERM; then we apply SU to obtain $(pp.ppp)p.pp$; eventually, we use PERM to prove $p.(pp.ppp)pp$.

Let $\mathbf{L}'''$ be obtained from $\mathbf{L}''$ by adjoining PERM. This means that the set of theorems of $\mathbf{L}'''$ is defined as the smallest set of formulas satisfying the following two clauses: (1) if $A \longrightarrow_{\mathbf{L}'} B$ in $\mathbf{L}''$, then AB is a theorem of $\mathbf{L}'''$ and (2) if $A.BC$ is a theorem of $\mathbf{L}'''$, then $B.AC$ is a theorem of $\mathbf{L}'''$.

The definition of a theorem of $\mathbf{L}'''$ can be given by (1) and (2'): if $A \longrightarrow_{\mathbf{L}'} BC$ in $\mathbf{L}''$, then $B.AC$ is a theorem of $\mathbf{L}'''$.

It is not difficult to see that the definition using (1) and (2) is equivalent to the definition using (1) and (2'). Of course, $\mathbf{L}$ and $\mathbf{L}'''$ have the same set of theorems.

A natural deduction system $\mathbf{K}'$

Let us adjoin ASS1 to $\mathbf{L}'''$ and let the resulting system be called $\mathbf{K}'$. Hence, the set of theorems of $\mathbf{K}'$ is the smallest set satisfying the following conditions: (i) if AB is a theorem of $\mathbf{L}'''$, then AB is a theorem of $\mathbf{K}'$ and (ii) if A is a theorem of $\mathbf{K}'$, then ABB is a theorem of $\mathbf{K}'$.

The definition of a theorem of $\mathbf{K}'$ can be given by (i) and (ii'): if $A.BC$ is a theorem of $\mathbf{L}'''$, then $B.AC$ is a theorem of $\mathbf{K}'$.

It is easy to prove

Theorem 8. $\mathbf{K}$ *and* $\mathbf{K}'$ *have the same set of theorems.*

3. Binary trees

In [3] a connection between $\mathbf{TRW}'_{\rightarrow}$–ID and binary trees has been established.

By a *binary tree* we understand a tree such that (1) there is a unique element at level 0 called the *origin* of $\mathcal{T}$ and (2) each node of $\mathcal{T}$ is either an end-node or has exactly two immediate successors.

By a *subtree* $\mathcal{T}'$ of a binary tree $\mathcal{T}$ we understand a subset $\mathcal{T}'$ of nodes of $\mathcal{T}$ such that $\mathcal{T}'$ is a binary tree and the immediate successor relation in $\mathcal{T}'$ is the immediate successor relation in $\mathcal{T}$.

A subtree $\mathcal{T}'$ of $\mathcal{T}$ is *proper* iff $\mathcal{T}'$ is a subtree of $\mathcal{T}$, and $\mathcal{T}'$ and $\mathcal{T}$ are not identical. Obviously, a subtree $\mathcal{T}'$ of $\mathcal{T}$ is proper iff the origin of $\mathcal{T}'$ is distinct from the origin of $\mathcal{T}$.

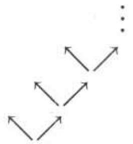
A binary tree $\mathcal{T}$ is finite (infinite) iff the set of nodes of $\mathcal{T}$ is finite (infinite).

After several conversations with Ilijas Farah in the period 1990 - 92 the concept of a *formula-like tree* (FLT) has emerged and the present author was able to represent the one-premiss natural deduction systems considered above as systems of operations on FLTs.

Let $\mathcal{T}'$ and $\mathcal{T}''$ be two binary trees. We shall say that they are *isomorphic* iff there is a mapping h from $\mathcal{T}'$ onto $\mathcal{T}''$ such that the following conditions are satisfied: (1) if x is the origin of $\mathcal{T}'$, then $h(x)$ is the origin of $\mathcal{T}''$ and (2) if the nodes y and z of $\mathcal{T}'$ are the left and the right immediate successor, respectively, of a node x in $\mathcal{T}'$, then $h(y)$ and $h(z)$ in $\mathcal{T}''$ are the left and the right immediate successor, respectively, of $h(x)$ in $\mathcal{T}''$.

If $\mathcal{T}'$ and $\mathcal{T}''$ are finite trees and one of them is a proper subtree of the other, they cannot be isomorphic. However, if they are infinite, it is possible that they are isomorphic and yet that one of them is a proper subtree of the other.

Let the *full binary tree* (FBT) be the infinite binary tree with no finite branch. In FBT every subtree is FBT. There are examples of binary trees that have isomorphic proper subtrees and are different from FBT. Here is one:



This is an infinite tree; each node in the infinite (rightmost) branch has an end-node as the left successor and a node in the infinite branch as the right successor. Any proper subtree with the origin in the infinite branch is isomorphic to the whole tree.

Let us call a tree $\mathcal{T}$ *formula-like tree* (FLT) iff (1) it has no proper subtree that is isomorphic to $\mathcal{T}$ and (2) no proper subtree $\mathcal{T}'$ of $\mathcal{T}$ has a proper subtree isomorphic to $\mathcal{T}'$.

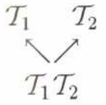
There is a trivial consequence of the above definition and the fact that being a subtree is a transitive relation.

Theorem 9. *A subtree of a* FLT *is a* FLT.

Every finite binary tree is a FLT, but there are infinite FLTs as well. For example, take the above infinite tree and extend each end-node by a finite tree that is different from all finite trees adjoined to previous end-nodes.

Every infinite FLT contains infinitely many (distinct) finite branches; every node of a FLT is a node of a finite branch. Hence, any branch of a FLT is at most of length ω. Therefore, the nodes of a FLT are arranged in levels and to each level there is attached a natural number. The number 0 is attached to the origin.

Theorem 10. *If $\mathcal{T}_1$ and $\mathcal{T}_2$ are FLTs, then*



is a FLT.

Proof. If the contrary is the case, then there is a proper subtree $h(\mathcal{T}_1\mathcal{T}_2)$ of $\mathcal{T}_1\mathcal{T}_2$ isomorphic to $\mathcal{T}_1\mathcal{T}_2$. By definition of isomorphism, $h(\mathcal{T}_1\mathcal{T}_2)=h(\mathcal{T}_1)h(\mathcal{T}_2)$ and $h(\mathcal{T}_1)$ and $h(\mathcal{T}_2)$ are isomorphic to $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. The origin of $h(\mathcal{T}_1\mathcal{T}_2)$ is either in the subtree $\mathcal{T}_1$ or in the subtree $\mathcal{T}_2$, say in $\mathcal{T}_1$. Now the origin of $h(\mathcal{T}_1\mathcal{T}_2)$ coincides either with the origin of $\mathcal{T}_1$ or with another node of $\mathcal{T}_1$. Since $h(\mathcal{T}_1\mathcal{T}_2) = h(\mathcal{T}_1)h(\mathcal{T}_2)$, the left successor of the node $h(\mathcal{T}_1\mathcal{T}_2)$ is $h(\mathcal{T}_1)$. But $\mathcal{T}_1$ and $h(\mathcal{T}_1)$ are isomorphic and $h(\mathcal{T}_1)$ is a proper subtree of $\mathcal{T}_1$, contrary to the assumption that $\mathcal{T}_1$ is a FLT.

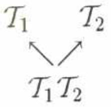
We proceed similarly if the origin of $h(\mathcal{T}_1\mathcal{T}_2)$ is in the subtree $\mathcal{T}_2$. □

Formulas and binary trees

Formulas of the propositional language with → as the sole connective are naturally connected with finite binary trees. The nodes of such a tree are subformulas of the formula A we are constructing the tree for. Thus, the formula A itself is at the origin of the tree. If an occurrence of a subformula BC of A is at a node at level n, then, at level $n+1$, the displayed occurrence of B is the left and the displayed occurrence of C is the right successor of the displayed occurrence of BC. The end-nodes of such formula trees are occurrences of propositional variables.

Suppose that in the propositional language that we are considering there is only one propositional variable, say p (this is sufficient to prove NOID, NOID(L) NOID(J) and NOE); then we can identify formulas and finite binary trees. Let p be the tree consisting of a single node. If A and B are finite binary trees, then AB is the tree obtained by taking a node as the origin of the tree such that A and B are the left and the right immediate successor of the origin.

In the sequel we shall interpret formulas as FLTs. For any propositional variable p we choose a FLT $\mathcal{T}$ and we interpret p as $\mathcal{T}$. Let A and B be any formulas and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be the FLTs such that A and B are interpreted as FLTs $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively; let us choose a new node called $\mathcal{T}_1\mathcal{T}_2$ as the origin of a new tree and take $\mathcal{T}_1$ and $\mathcal{T}_2$ to be the only immediate successors of $\mathcal{T}_1\mathcal{T}_2$, thus:



By Theorem 10, $\mathcal{T}_1\mathcal{T}_2$ is a FLT and we interpret AB as $\mathcal{T}_1\mathcal{T}_2$.

In denoting trees we shall use the conventions adopted in writing formulas.

Propositional formulas have a property called *substitution*; let us show that FLTs enjoy the same property. Suppose that $\mathcal{T}_1$ is a FLT and $\mathcal{T}_2$ a subtree of $\mathcal{T}_1$:

$$\begin{array}{c}\mathcal{T}_2\\ \vdots\\ \mathcal{T}_1\end{array}$$

then this occurrences of the FLT $\mathcal{T}_2$ in $\mathcal{T}_1$ can be cut off and a FLT $\mathcal{T}_3$ can be inserted instead:

$$\begin{array}{c}\mathcal{T}_3\\ \vdots\\ \mathcal{T}_1\end{array}$$

Theorem 11. *Let $\mathcal{T}_1$ be a FLT, let $\mathcal{T}_2$ be a subtree of $\mathcal{T}_1$ and let $\mathcal{T}_4$ be the tree obtained from $\mathcal{T}_1$ by cutting off $\mathcal{T}_2$ and by inserting a* FLT *$\mathcal{T}_3$ instead; then $\mathcal{T}_4$ is a* FLT.

Proof. Proceed by induction on levels. Let $\mathcal{T}_2$ be $\mathcal{T}_1$; then, obviously, $\mathcal{T}_4$ is $\mathcal{T}_3$. If $\mathcal{T}_1$ is $\mathcal{T}_1'\mathcal{T}_1''$, then the origin of $\mathcal{T}_2$ is at a certain level n in $\mathcal{T}_1$. If it is in, say, $\mathcal{T}_1'$, then in $\mathcal{T}_1'$ it is at level smaller than n and by induction hypothesis the result $\mathcal{T}_4'$ of substitution of $\mathcal{T}_3$ for $\mathcal{T}_2$ in $\mathcal{T}_1'$ is a FLT. By Theorem 9, $\mathcal{T}_1''$ is a FLT; hence, by Theorem 10, so is $\mathcal{T}_4'\mathcal{T}_1''$, i.e. $\mathcal{T}_4$. □

Natural deduction and FLTs

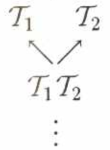
There is a connection between derivations in one-premiss natural deduction systems $\mathbf{TW}'_{\rightarrow}$–ID, $\mathbf{L}'$, and $\mathbf{L}''$ and FLTs. In order to explain this connection, let us show how the rules of $\mathbf{L}''$ can be interpreted as operations on FLTs.

To every node of a FLT $\mathcal{T}$ we associate one of the numbers 0 or 1, as follows: 0 is associated with the origin of $\mathcal{T}$; if 0 (1) is associated with a node at level n, then 1 (0) is associated with its left hand successor and 0 (1) is associated with its right hand successor at level $n+1$.

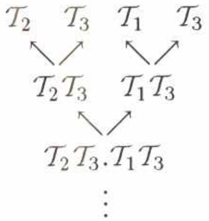
If 0 (1) is associated with a node of a tree, then we shall call it a 0-node (1-node).

Now the rules SU^0, PR^0, SU^1, PR^1, SUL^0, SUL^1, and PERM* can be represented as operations FLTs as follows.

SU^0 Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_1\mathcal{T}_2$ be one of its 0-nodes:



Then the subtree $\mathcal{T}_1\mathcal{T}_2$ can be cut off and a subtree $\mathcal{T}_2\mathcal{T}_3.\mathcal{T}_1\mathcal{T}_3$ can be inserted in $\mathcal{T}$ instead:


$\mathcal{T}_2$
$\mathcal{T}_3$
$\mathcal{T}_1$
$\mathcal{T}_3$
$\mathcal{T}_2\mathcal{T}_3$
$\mathcal{T}_1\mathcal{T}_3$
$\mathcal{T}_2\mathcal{T}_3.\mathcal{T}_1\mathcal{T}_3$

where $\mathcal{T}_3$ is any FLT. Let us call the new tree $\mathcal{T}'$.

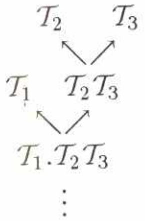
In a similar way we can represent the remaining 0-rules.

As to the 1-rules, let us represent SU^1. Suppose that $\mathcal{T}'$ is a FLT and let $\mathcal{T}_2\mathcal{T}_3.\mathcal{T}_1\mathcal{T}_3$ be one of its 1-nodes; then the subtree $\mathcal{T}_2\mathcal{T}_3.\mathcal{T}_1\mathcal{T}_3$ can be cut off and the subtree $\mathcal{T}_1\mathcal{T}_2$ can be inserted instead, producing thus the tree $\mathcal{T}$.

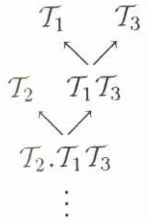
In a similar way we may represent the remaining 1-rules.

Now we represent PERM*.

PERM* Let $\mathcal{T}$ be a FLT and let $\mathcal{T}_1.\mathcal{T}_2\mathcal{T}_3$ be one of its 0-or-1-nodes:


$\mathcal{T}_2$
$\mathcal{T}_3$
$\mathcal{T}_1$
$\mathcal{T}_2\mathcal{T}_3$
$\mathcal{T}_1.\mathcal{T}_2\mathcal{T}_3$

Then the subtree $\mathcal{T}_1.\mathcal{T}_2\mathcal{T}_3$ can be cut off and the subtree $\mathcal{T}_2.\mathcal{T}_1\mathcal{T}_3$ can be inserted in $\mathcal{T}$ instead:


$\mathcal{T}_1$
$\mathcal{T}_3$
$\mathcal{T}_2$
$\mathcal{T}_1\mathcal{T}_3$
$\mathcal{T}_2.\mathcal{T}_1\mathcal{T}_3$

By Theorem 11, the result of an application of a 0-or-1-rule or PERM* to a FLT is a FLT. Suppose that these rules are applied to a finite binary tree; then NOID can be stated as follows:

NOID($\mathcal{T}$) (1) Starting from a FLT $\mathcal{T}$ and successively performing the operations SU^0, PR^0, SU^1, PR^1, SUL^0, SUL^1, and PERM^* any finite number of times and in any order, and such that one of the first six operation is performed at least once, it is not possible to obtain $\mathcal{T}'$ as a result, where $\mathcal{T}'$ is either $\mathcal{T}$ or $\mathcal{T}\mathcal{T}_1\mathcal{T}_1$;

(2) starting with a FLT $\mathcal{T}\mathcal{T}_1$ and successively performing the operations SU^0, PR^0, SU^1, PR^1, SUL^0, SUL^1, and PERM^* any finite number of times and in any order, and such that one of the first six operation is performed at least once it is not possible to obtain $\mathcal{T}_1$ as a result;

(3) starting with a FLT $\mathcal{T}\mathcal{T}_1\mathcal{T}_1$ and successively performing the operations SU^0, PR^0, SU^1, PR^1, SUL^0, SUL^1, and PERM^* any finite number of times and in any order, and such that one of the first six operation is performed at least once it is not possible to obtain $\mathcal{T}$ as a result.

If the rules are applied to a finite FLT $\mathcal{T}$, then NOID($\mathcal{T}$) is true, since we can identify formulas and finite binary trees.

Theorem 12. *NOID($\mathcal{T}$) is true for any FLT $\mathcal{T}$.*

Proof. An interpretation of a theorem of **L** in the set of all FLTs is a homomorphic image of a theorem of **L**; hence, it has a form of a theorem of **L**. By NOID(J) there is no theorem of **L** either of the form AA or of the form $A.ABB$ or of the form ABB; hence, there is no FLT either of the form $\mathcal{T}\mathcal{T}$ or of the form $\mathcal{T}_1.\mathcal{T}_1\mathcal{T}_2\mathcal{T}_2$ or of the form $\mathcal{T}_1\mathcal{T}_2\mathcal{T}_2$. If a FLT $\mathcal{T}_2$ can be obtained from a FLT $\mathcal{T}_1$ by a finite number of applications of 0-or-1-rules and PERM^*, then $\mathcal{T}_1\mathcal{T}_2$ is a homomorphic image of a theorem of **L**. Hence, NOID($\mathcal{T}$) is true. □

There is no natural interpretation of one-premiss natural deduction systems $\mathbf{L}'''$ or $\mathbf{K}'$ in terms of operations on FLTs, for there are theorems of these systems that cannot be obtained by performing such operations only.

References

[1] A.R. Anderson and N.D. Belnap, *Entailment, the Logic of Relevance and Necessity, Vol. I*, Princeton Univ. Press,, 1975.

[2] Y. Komori, *Syntactical Investigations into BI Logic and BB′I Logic*, Studia Logica **53** (1994), 397-416.

[3] A. Kron, *Identity and binary trees (in Hungarian), Majdanem nem let másként*, Tanulmńyok Vajda Mihály 60. szl̈etésnapjára, a szerzök (1995), 232 - 242..

[4] A. Kron, *A constructive proof of a theorem in relevance logic*, Zeitschrift für mathematische Logik und Grundlagen der Mathematik **Bd. 31** (1985), 423-430.

[5] A. Kron, *Identity and Permutation,*, Publications de l' Institut Mathématique (to appear).

[6] E. Martin and R.K. Meyer, *Solution to the P - W problem*, The Journal of Symbolic Logic **47** (1982), 869-887.

[7] R. Smullyan, *First-Order Logic*, Springer-Verlag,, Heidelberg – Berlin – New York, 1968.

Aleksandar Kron, Filozofski fakultet, 11000 Beograd, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 783–802
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

АВТОМАТЫ И ЛАБИРИНТЫ

В. Б. Кудрявцев, Ш. Ушчумлич и Г. Килибарда

Резюме. В работе развивается концепция поведения автоматов в лабиринтах за счет расширения толкования лабиринта до современного понимания дискретной геометрической среды, и толкования автомата — до иерархии абстрактных машин. Затем на примере конечныхавтоматов в плоских мозаичных лабиринтах очерчиваются контуры проблематики и приводется некоторые результаты.

1. Лабиринты. n-мерные лабиринты

Обозначим множество всех натуральных чисел через $\mathbf{N}$, целых чисел — через $\mathbf{Z}$, действительных чисел — через $\mathbf{R}$, множество $\mathbf{N} \cup \{0\}$ — через $\mathbf{N}_0$. Пусть $z_1, z_2 \in \mathbf{Z}$ и $n \in \mathbf{N}_0$, положим $\overline{n} = \{1, 2, \dots, n\}$, если $n > 0$, и $\overline{n} = \emptyset$, если $n = 0$, а также

$$(z_1, z_2) = \{z \in \mathbf{Z} | z_1 < z < z_2\}, \quad [z_1, z_2) = \{z \in \mathbf{Z} | z_1 \le z < z_2\},$$
$$(z_1, z_2] = \{z \in \mathbf{Z} | z_1 < z \le z_2\}, \quad [z_1, z_2] = \{z \in \mathbf{Z} | z_1 \le z \le z_2\}.$$

Всюду в последующем, если специально не оговорено, через i, j, k, l, m и n обозначены некоторые натуральные числа.

Пусть A — некоторый алфавит букв a. Обозначим через A^* множество всех слов α вида $a_1 \dots a_n$, где $a_i \in A$ $(1 \le i \le n)$, включая пустое слово Λ. Положим $A^+ = A^* \backslash \{\Lambda\}$.

Для слова $\alpha = a_1 \dots a_n \in A^+$ пусть $\mathbf{f}(\alpha) = a_1$, $\mathbf{l}(\alpha) = a_n$ и $|\alpha| = n$. Эти значения назовем, соответственно, левым и правым концом, и также длиной слова α. Любое слово α_1 вида $a_i \dots a_j$, $1 \le i < j \le n$, называем подсловом слова α, и обозначаем это $\alpha_1 < \alpha$. Пустое слово Λ является подсловом любого слова. Если $\alpha_1 < \alpha$, то при $\mathbf{f}(\alpha_1) = \mathbf{f}(\alpha)$, пишем $\alpha_1 \backslash \alpha$, а при $\mathbf{l}(\alpha_1) = \mathbf{l}(\alpha)$, пишем α / α_1.

Для любых $\alpha \in A^*$ и $n \in \mathbf{N}_0$. положим

$$\alpha^n = \underbrace{\alpha\alpha\dots\alpha}_{n},$$

если $n \neq 0$, и $\alpha^n = \Lambda$, если $n = 0$.

Пусть X — некоторое множество. Через $\mathcal{P}(X)$ обозначим множество всех подмножеств множества X, через $\mathcal{P}_0(X)$ — множество всех непустых подмножеств множества X, а через $|X|$ — мощность множества X.

Пусть $\{X_a : a \in A\}$ — некоторое индексированное семейство множеств X_a. Тогда для любого $a_0 \in A$ через $\mathbf{p}_{a_0}$ обозначим отображение проектирования произведения $\prod_{a\in A} X_a$ на сомножитель X_{a_0}. Если множество индексов A — конечное множество, то в последующем всегда будем предполагать, что $A = \{1, \dots, |A|\}$.

Пусть $\{f_j : j \in J\}$ — некоторое индексированное семейство функции, $f_j : A_j \to B$, таких, что $f_{j_1}(x) = f_{j_2}(x)$ для любых $x \in A_{j_1} \cap A_{j_2}$, если $j_1 \neq j_2$. Через $\cup_{j\in J} f_j$, $\cup_{j\in J} f_j : \cup_{j\in J} A_j \to B$, обозначим функцию, такую, что $\cup_{j\in J} f_j|_{A_{j'}} = f_{j'}$ для любого $j' \in J$. Для любого множества X обозначим через i_X, $i_X : X \to X$, его тождественную функцию, т.е такую, что $i_X(x) = x$ для всех x из X.

Пусть V и J — некоторые счетные множества, и $\Gamma \subseteq V \times V \times J$. Пусть $\bar{\cdot} : \Gamma \to \Gamma$ — некоторое частичное однозначное отображение, которое дуге $\gamma = (v_1, v_2, j)$ сопоставляет, если оно определено на γ, дугу $\overline{\gamma} = (v_2, v_1, j')$. Набор $(V, \Gamma, \bar{\cdot})$ называется *ориентированным графом* или *орграфом* . Элементы из V назовем его *вершинами*, а элементы из Γ — его *дугами*. Дугу $(v, v, j) \in \Gamma$ называем *петлей* в вершине v. В дальнейшем вместо $(V, \Gamma, \bar{\cdot})$, пишем (V, Γ), если специально не подчеркивается, о какой именно функции $\bar{\cdot}$ идет речь. Элементы из J считаем индексами дуг орграфа.

Для (V, Γ) и $v_1, v_2 \in V$ обозначим через $J_{v_1,v_2}(\Gamma)$ множество $\{j : (v_1, v_2, j) \in \Gamma\}$. Если $v_1 = v_2 = v$, то вместо $J_{v_1,v_2}(\Gamma)$ пишем $J_v(\Gamma)$. Если $J_v(\Gamma) = \emptyset$ для любого $v \in V$, т.е. в орграфе (V, Γ) нет петель, то (V, Γ) называется *орграфом без петель*. Если $|J_{v_1,v_2}(\Gamma)| \leq 1$ для любых $v_1, v_2 \in V$, то (V, Γ) называется *орграфом без кратных дуг*.

Если существует $\overline{\gamma}$ для любой $\gamma \in \Gamma$, то орграф (V, Γ) называется *графом*. В графе (V, Γ) для любой $\gamma \in \Gamma$ множество $< \gamma >= \{\gamma, \overline{\gamma}\}$ называется *ребром*. Если γ является петлей, то ребро $< \gamma >$ называется *петлей*. Через $< \Gamma >$ обозначим множество всех ребер графа (V, Γ). Граф (V, Γ) является графом без петель (без кратных ребер), если (V, Γ) — орграф без петель (без кратных дуг).

Если орграф (граф) (V, Γ) без кратных дуг (ребер), то вместо

(v_1, v_2, j)
$(< (v_1, v_2, j) >)$ пишем соответственно (v_1, v_2) $(< v_1, v_2 >)$.

Множество вершин V и множество дуг Γ орграфа G будем иногда обозначать соответственно через $V(G)$ и $\Gamma(G)$. Далее, пусть X и Y — некоторые множества. Если заданы некоторые функции $f : V \to X$ и $g : \Gamma \to Y$, то тройка (G, f, g) называется *нагруженным орграфом*, множества X и Y называются, соответственно, множествами *отметок* вершин и дуг орграфа G, g — *разметкой вершин* орграфа G и f — *разметкой дуг* орграфа G. Для любых $u \in V(G)$ и $\gamma \in \Gamma(G)$ значения $f(u)$ и $g(\gamma)$ функций f и g называются сооветственно *отметкой вершины* u и *отметкой дуги* γ в нагруженном орграфе (G, f, g). Через$\mathcal{G}(X, Y)$ обозначим класс всех нагруженных орграфов с множеством Xв качестве множества отметок вершин и Yв качестве множества отметок дуг этих орграфов. Если ясно, о каких функциях f и g идет речь, то для любых $u \in V(G)$ и $\gamma \in \Gamma(G)$ вместо $f(u)$ и $g(\gamma)$ пишем соответственно $|u|$ и $|\gamma|$, а если хотим подчеркнуть, о каком именно орграфе идет речь, то пишем соответственно, $|u|_G$ и $|\gamma|_G$. Для любого $v \in V(G)$ обозначим $[v]_G = \{|\gamma| \,|\, \mathbf{p}_1(\gamma) = v$ и $\gamma \in \Gamma(G)\}$.

Нагруженный орграф $(G_1, f_1, g_1) \in \mathcal{G}(X, Y)$ называется *частью* нагруженного орграфа $(G, f, g) \in \mathcal{G}(X, Y)$, если выполнены условия:

1) $V(G_1) \subseteq V(G)$ и $\Gamma(G_1) \subseteq \Gamma(G)$;

2) $f_1(u) = f(u)$ и $g_1(\gamma) = g(\gamma)$ для любых $u \in V(G_1)$ и $\gamma \in \Gamma(G_1)$.

Часть (G_1, f_1, g_1) нагруженного орграфа (G, f, g) называется *подорграфом*, если из $u, v \in V(G_1)$ и $(u, v, j) \in \Gamma(G)$ для некоторого $j \in J$ следует, что $(u, v, j) \in \Gamma(G_1)$.

Пусть I — закрытый единичнный интервал и X — некоторое топологическое пространство. Пусть Θ — счетное множество непрерывных функций $f : I \to X$, таких, что выполнены следующие условия:

1) для любых $f \in \Theta$ и $r_1, r_2 \in I$, если $f(r_1) = f(r_2)$, то $\{r_1, r_2\} = \{0, 1\}$.

2) для любой $f \in \Theta$ существует не более одной $g \in \Theta$, такой, что $f(I) = g(I)$;

3) для любых $f, g \in \Theta$, если $f(I) \neq g(I)$, то имеет место $f[(0,1)] \cap g[(0,1)] = \emptyset$;

4) для любых $f, g \in \Theta$, если $f \neq g$, $f(0) \neq f(1)$ и $f(I) = g(I)$, то $f(0) = g(1)$ и $f(1) = g(0)$;

5) для любых $f, g \in \Theta$, если $f \neq g$, $f(0) = f(1)$ и $f(I) = g(I)$, то у функции f и g "противоположные ориентации", т.е. существует убывающая функция $\alpha : I \to I$, такая, что, $f = g \circ \alpha$.

Определим орграф $(V(\Theta),\Gamma(\Theta))$ следующим способом. Пусть

$$V(\Theta)=\{f(0)\,|\,f\in\Theta\}\cup\{f(1)\,|\,f\in\Theta\},\quad \Gamma(\Theta)=\{(f(0),f(1),f)\,|\,f\in\Theta\}.$$

Если $\gamma=(f(0),f(1),f)$, то $\overline{\gamma}=(g(0),g(1),g)$, где g такое, что $g(I)=f(I)$; если такое g не существует, то $\overline{\gamma}$ не определено. Если $(V,\Gamma)\simeq(V(\Theta),\Gamma(\Theta))$, то Θ называется *X-реализацией* орграфа (V,Γ). В случае, когда $X=\mathbf{R}^2$, то вместо X-реализация говорим *плоская реализация.*Орграф называется *планарным*, если существует его плоская реализация. Набор $(V(\Theta),\Gamma(\Theta),\Theta)$ называется *X-орграфом*; если $X=\mathbf{R}^2$, то этот набор называется *плоским орграфом. Носителем* X-орграфа $(V(\Theta),\Gamma(\Theta),\Theta)$ называется множество $\cup_{f\in\Theta}f(I)$. Часто в последующем, если ясно из контекста или не хотим подчеркивать о каком именно множестве Θ идет речь, мы в обозначении плоского графа опускаем обозначение этого множества.

Пусть $G=(V,\Gamma)$ — некоторый орграф. Если для любых $v_1,v_2\in V$ существует последовательность дуг $\gamma_1,\gamma_2,\ldots,\gamma_m$, такая,что $\mathbf{p}_1(\gamma_1)=v_1$, $\mathbf{p}_2(\gamma_m)=v_2$ и $\mathbf{p}_2(\gamma_i)=\mathbf{p}_1(\gamma_{i+1})$ для любого i, $1\leq i\leq n-1$, то (V,Γ) назывется *связным орграфом.*

Для орграфа $G=(V,\Gamma)$ и $v\in V$ вводим следующие множества букв:

$$\mathbf{Con}_a(G)=\{a_\gamma\,|\,\gamma\in\Gamma\},\quad \mathbf{Con}_b(G)=\{b_\gamma\,|\,\gamma\in\Gamma\},$$
$$\mathbf{Con}_a^v(G)=\{a_\gamma\,|\,\mathbf{p}_1(\gamma)=v,\ \gamma\in\Gamma\},\quad \mathbf{Con}_b^v(G)=\{b_\gamma\,|\,\mathbf{p}_2(\gamma)=v,\ \gamma\in\Gamma\},$$
$$\mathbf{Con}(G)=\mathbf{Con}_a(G)\cup\mathbf{Con}_b(G),\quad \mathbf{Con}_v'(G)=\mathbf{Con}_a^v(G)\cup\mathbf{Con}_b^v(G)$$

Пусть Y — некоторое множество и $g:\Gamma\to Y\times Y$.Определим функции $\mathbf{Con}_a(g):\mathbf{Con}_a(G)\to Y$ и $\mathbf{Con}_b(g):\mathbf{Con}_b(G)\to Y$ следующим способом: $\mathbf{Con}_a(g)(a_\gamma)=\mathbf{p}_1(g(\gamma))$ и $\mathbf{Con}_b(g)(b_\gamma)=\mathbf{p}_2(g(\gamma))$ для любой $\gamma\in\Gamma$. Пусть

$$\mathbf{Con}(g)=\mathbf{Con}_a(g)\cup\mathbf{Con}_b(g).$$

На $\mathbf{Con}_v'(G)$ введем отношение эквивалентности $\overset{\mathrm{Con}}{\sim}$ следующим способом. Если $\mathbf{p}_1(\gamma)=v$ и для $\gamma\in\Gamma$ существует дуга $\overline{\gamma}$, то $a_\gamma\overset{\mathrm{Con}}{\sim}b_{\overline{\gamma}}$. Тогда $\mathbf{Con}_v(G)=\mathbf{Con}_v'(G)/\overset{\mathrm{Con}}{\sim}$. *Система следования* η в орграфе G есть множество $\{\eta_v:v\in V\}$, где η_v — бинарное отношение в $\mathbf{Con}_v(G)$ для любого $v\in V$. Обозначим $\Gamma_v^i=\{\gamma\in\Gamma|\mathbf{p}_i(\gamma)=v\}$ для любого $i=1,2$. Если для любого $\delta\in\mathbf{Con}_v(G)$ имеет место:

$$|\{\delta'\in\mathbf{Con}_v(G)|(\delta,\delta')\in\eta_v \text{ и } \delta'=[a_\gamma] \text{ для некоторого } \gamma\in\Gamma_v\}|=1,$$

то можно определить $\tilde{\eta}_v:\Gamma_v\to\Gamma_v$ таким способом, что $\tilde{\eta}_v(\gamma)=\gamma'$, где γ' такая, что $([a_\gamma],[a_{\gamma'}])\in\eta_v$. Если Γ_v конечно для любого $v\in V$, т.е.

орграф G является локально конечным, и $\tilde{\eta}_v$ определено и является циклической подстановкой для любой $v \in V$, то η называется *системой вращения.* В случае, когда G — граф, циклическая подстановка $\tilde{\eta}_v : \Gamma_v \to \Gamma_v$ полностью определяет η_v, и тем самым вращение η. На основании этого, с целью упрощения общей картины, мы в случае, когда G является графом, под вращением будем понимать семейство $\eta = \{\eta_v | v \in V\}$, такое, что η_v является циклической подстановкой множества Γ_v для любого $v \in V$. Пару (G, η) назовем *орграфом G с системой следования* η, а если η — система вращения, то — *орграфом G с системой вращения* η.

Пусть Ω, Ω_1, Σ и Σ_1 — конечные алфавиты букв, такие, что любое из этих множеств содержит пустой символ Λ, и (G, η) — некоторый связный орграф с системой следования. Далее, даны отображения $f : V(G) \to \Omega$, $g : \Gamma(G) \to \Sigma \times \Sigma$. Набор $L = (G, f, g, \eta, \Omega_1, \Sigma_1)$ называется (Ω_1, Σ_1)-*лабиринтом с множеством отметок вершин* Ω *и множеством отметок дуг* Σ. Если $\Omega_1 = \{\Lambda\}$, $\Sigma_1 = \{\Lambda\}$, то вместо (Ω_1, Σ_1)-лабиринта говорим просто *лабиринт* и вместо $L = (G, f, g, \eta, \{\Lambda\}, \{\Lambda\})$ пишем $L = (G, f, g, \eta)$, т.е. лабиринт просто рассматриваем как нагруженный орграф с системой следования. Под состоянием (Ω_1, Σ_1)-лабиринта подразумеваем любую пару вида (f_1, g_1), где $f_1 : V(G) \to \Omega_1$ и $g_1 : \Gamma(G) \to \Sigma_1 \times \Sigma_1$. Обозначим через $\mathcal{L}(\Omega, \Omega_1; \Sigma, \Sigma_1)$ класс всех (Ω_1, Σ_1)-лабиринтов, а через $\mathcal{L}(\Omega, \Sigma)$ — всех лабиринтов, с множеством отметок вершин Ω и множеством отметок дуг Σ. Если дан лабиринт $L = (G, f, g, \eta, \Omega_1, \Sigma_1)$, то G, $V = V(G)$, $\Gamma = \Gamma(G)$, f, g и η обозначаем соответственно через $G(L)$, $V(L)$, $\Gamma(L)$, f_L, g_L и η_L. Если специально не оговариваем, о каких f, g и η идет речь, то вместо $(G, f, g, \eta, \Omega_1, \Sigma_1)$ пишем (G, Ω_1, Σ_1); в случае, когда $\Omega_1 = \{\Lambda\}$ и $\Sigma_1 = \{\Lambda\}$, то вместо (G, Ω_1, Σ_1) пишем просто, что $L = (V(L), \Gamma(L))$. Как и в случае нагруженных орграфов, мы вместо $g(\gamma)$ и $g_1(\gamma)$ будем часто писать, соответственно, $|\gamma|$ и $||\gamma||$; $\gamma \in \Gamma(G)$. Если $g(\gamma) = (\sigma, \Lambda)$ $(g_1(\gamma) = (\sigma, \Lambda))$ для некоторых $\sigma \in \Sigma$ $(\sigma \in \Sigma_1)$ и $\gamma \in \Gamma(G)$, то будем писать $|\gamma| = g(\gamma) = \sigma$ $(||\gamma|| = g_1(\gamma) = \sigma)$.

В лабиринте L могут быть выделены два множества вершин V_1 и V_2 (эти множества могут быть и пустыми, но если первое пустое, то второе обязательно пусто). Вершины множества V_1 называем начальными, а V_2 — конечными. Лабиринт L в таком случае обозначаем $L_{V_1}^{V_2}$ или $(L; V_1, V_2)$. Часто в последующем будем рассматривать лабиринты с двумя выделенными (одной выделенной) различными вершинами, причем в таком случае первую из них называем *входом*, а другую *выходом* данного лабиринта (если выделена одна, то ее называем входом), т.е. будем рассматривать случай, когда $V_1 = \{v_1\}$ и $V_2 = \{v_2\}$ или $V_2 = \emptyset$. В таком случае вместо $L_{\{v_1\}}^{\{v_2\}}$ и $L_{\{v_1\}}$ будем писать соответственно $L_{v_1}^{v_2}$ и L_{v_1}, т.е. $(L; v_1, v_2)$ и $(L; v_1)$ (Здесь v_1 — вход, а v_2 — выход лабиринта $L_{u_1}^{v_1}$; в лабиринте L_{v_1} вершина v_1 является входом). Если в L выделены вход и выход, то иногда будем их обозначать через $v_s(L)$ и $v_f(L)$, соответственно.

Лабиринт $L_1 \in \mathcal{L}(\Omega, \Omega_1; \Sigma, \Sigma_1)$ называется *частью*(*подлабиринтом*) лабиринта $L \in \mathcal{L}(\Omega, \Omega_1; \Sigma, \Sigma_1)$, если орграф $G(L_1)$ является частью (подорграфом) орграфа $G(L)$, $f_{L_1} = f_L|_{V(L_1)}$, $g_{L_1} = g_L|_{\Gamma(L_1)}$ и $(\eta_{L_1})_v = (\eta_L)_v \cap (\Gamma_v^1(L_1) \times \Gamma_v^1(L_1))$.

Дадим сейчас определение одного класса лабиринтов, с которым последующем будем иметь дело. Это класс n-лабиринтов.

Обозначим через E^n множество $\{e_1, \dots, e_n\}$ базисных единичных векторов n-мерного евклидова пространства $\mathbf{R}^n$, а через $\tilde{E}^n$ обозначим множество

$$\{e_1, \dots, e_n, \overline{e}_1, \dots, \overline{e}_n\},$$

где $\overline{e}_i = e_i^{-1} = -e_i$, $1 \le i \le n$. В случае $n = 2$ и $n = 3$ вместо обозначений базисных векторов $\vec{\imath}, \vec{\jmath}, \vec{k}$ и векторов $\vec{\imath}, \vec{\jmath}, \vec{k}$ будем, соответственно, пользоваться обозначениями $\mathbf{e}$, $\mathbf{n}$, $\mathbf{u}$, $\mathbf{w}$, $\mathbf{s}$ и $\mathbf{d}$.

Лабиринт $L = (G, f, g, \eta) \in \mathcal{L}(\Omega, \Sigma)$, без кратных дуг и петель, называется *n-мерным лабиринтом* или *n-лабиринтом*, $n \ge 2$, если выполнены следующие условия:

1) $\Omega = \{\Lambda\}$ и $\Sigma = \tilde{E}^n \cup \{\Lambda\}$;
2) $G(L)$ является графом и $|\gamma| \in \tilde{E}^n$ для любой дуги $\gamma \in \Gamma(L)$;
3) для любого $\gamma \in \Gamma(L)$ имеет место $|\gamma| = |\overline{\gamma}|^{-1}$;
4) $\eta_v = \emptyset$ для любой $v \in V$.

В последующем, в обозначении n-лабиринта (G, f, g, η) будем опускать g, считая, конечно, что в любом конкретном случае g задана, а также будем опускать f и η, поскольку $f(v) = \Lambda$ и $\eta_v = \emptyset$ для любой $v \in V(L)$ у каждого n-лабиринта L, т.е. будем писать просто, что $L = (V(L), \Gamma(L))$. Также доопределим g на парах вида (v, v), $v \in V$, считая, что $|(v, v)| = \mathbf{0}$ ($\mathbf{0}$ — это нуль вектор).

Пусть M и N, $M \ne N$, — некоторые точки в R^n, и $\vec{MN} = \alpha_1 e_1 + \dots + \alpha_n e_n$. Будем говорить, что вектор $\vec{MN}$ *идет в направлении* e_i, если $\alpha_i > 0$ и $\alpha_j = 0$, и в *направлении* $\overline{e}_i$, если $\alpha_i < 0$ и $\alpha_j = 0$ для всех $j \ne i$, $1 \le j \le n$; $1 \le i \le n$. Множество T отрезков в R^n называется *n-конфигурацией*, если любые два разных отрезка из этого множества могут иметь не больше одной общей точки, причем, если она есть у них, то она обязательно является концевой для обоих отрезков.

n-лабиринт $L = (V, \Gamma)$, где $V \subseteq \mathrm{R}^n$, назовем *прямоугольным n-лабиринтом*, если

1) для любых $u, v \in V$из $(u, v) \in \Gamma$ следует, что $\vec{uv}$ идет в направлении $|(u, v)|$;

2) множество отрезков $T = \{\overline{uv} \,|\, (u, v) \in \Gamma\}$ является n-конфигурацией.

Для любой дуги $(u, v) \in \Gamma$, $u = (x_1, \ldots, x_n)$, $v = (y_1, \ldots, y_n)$, определим функцию $\{f_{u,v} : I \to \mathbf{R}^n\}$ такую, что

$$f_{u,v}(t) = (x_1 + (y_1 - x_1)t, \ldots, x_n + (y_n - x_n)t).$$

Ясно, что семейство $\{f_{u,v} | (u, v) \in \Gamma\}$ является $\mathbf{R}^n$-реализацией графа $G(L)$; назовем ее *линейной реализацией* данного прямоугольного лабиринта L. Фигура $\overline{L} = \bigcup_{(u,v) \in \Gamma(L)} \overline{uv}$ в R^n является носителем линейной реализации прямоугольного n-мерного лабиринта L.

Пусть $\mathbf{Z}^n$ — целочисленная решетка в R^n. Прямоугольный n-мерный лабиринт $L = (V, \Gamma)$ назовем *целочисленным* n-мерным лабиринтом, если $V \subseteq \mathbf{Z}^n$. n-мерный целочисленный лабиринт $L = (V, \Gamma)$ назовем *m-мозаичным* n-мерным лабиринтом, $m \in N$, если $T = \{\overline{uv} \,|\, (u, v) \in \Gamma\}$ — множество отрезков длины m. Вместо 1-мозаичный будем просто писать *мозаичный*.

Про вершину v m-мозаичного n-мерного лабиринта L говорим, что она *открыта в* L, если существует бесконечный n-мерный m-мозаичный лабиринт L_1 такой, что $\overline{L} \cap \overline{L}_1 = \{v\}$ и $v \in V(L_1)$. m-мозаичный n-мерный лабиринт $L_{u_1}^{v_1}$ называется *m-правильным* n-мерным лабиринтом, если вершина v_1 открыта в L. Вместо 1-правильный будем просто писать правильный.

Проведем через вершины $\mathbf{Z}^n$ все возможные прямые, параллельные осям координат. Ясно, что полученная фигура является носителем линейной $\mathbf{R}^n$-реализации некоторого прямоугольного n-мерного лабиринта, который обозначим через Z^n. Множество вершин этого лабиринта есть $\mathbf{Z}^n$. Очевидно, что мозаичный n-мерный лабиринт можно определить и как связную (нагруженную) часть лабиринта Z^n. Под *шахматным n-мерным лабиринтом* будем понимать любой подлабиринт лабиринта Z^n.

В дальнейшем вместо 2-мерный прямоугольный (мозаичный, целочисленный, правильный) лабиринт будем писать плоский прямоугольный (мозаичный, целочисленный, правильный) лабиринт.

Пусть $L = (V, \Gamma)$ — некоторый плоский прямоугольный лабиринт. Множество $\mathbf{R}^2 \backslash \overline{L}$ является открытым и в общем случае (если L не является деревом) несвязным. Лабиринт L назовем $k+1$-*связным*, если множество $\mathbf{R}^2 \backslash \overline{L}$ имеет k ограниченных компонентов связности. Пусть $L = (V, \Gamma)$ — некоторый плоский мозаичный лабиринт. Пусть $U_1, \ldots, U_k$ — все компоненты связности множества $\mathbf{R}^2 \backslash \overline{L}$. *Дырой* лабиринта L назовем любое непустое множество H вида $U_i \cap \mathbf{Z}^2$, $1 \leq i \leq k$. В случае, когда H конечно, дыру называем *конечной*; в противном случае *бесконечной* . Плоский мозаичный лабиринт L называем дырочно $k+1$-*связным*, если в нем точно k конечных дыр, $k \in \mathrm{N} = \{0,1,2,\ldots\}$; при $k=0$ лабиринт называем односвязным. В дальнейшем, когда речь идет о связности плоских мозаичных лабиринтов, мы имеем в виду число конечных дыр, если не оговорено иначе.

Класс всех плоских мозаичных лабиринтов обозначим через $\mathcal{L}_2$, класс всех конечных лабиринтов из $\mathcal{L}_2$ — через $\mathcal{L}_0$, а класс всех бесконечных лабиринтов из $\mathcal{L}_2$ — через $\mathcal{L}_1$.

2. Лабиринтные монстры. Допустимые автоматы в лабиринтах. Коллективы допустимых автоматов в лабиринтах

Основное предназначение лабиринта — запутать или поймать что-либо или кого-либо. Это что-либо в нашем случае мы будем называть лабиринтным монстром. Пусть $\mathcal{L} \in \mathcal{L}(\Omega, \Omega_1, \Sigma, \Sigma_1)$ — некоторый класс лабиринтов. $\mathcal{L}$-монстром, "многоруким", "многоголовым", "многоногим" существом, назовем машину:

а) у которой есть память, конечная внутренняя, которая может быть в сочетании с бесконечной внешней, организованной разными способами (в виде одного или нескольких магазинов, счетчиков, стеков, в виде одной или нескольких лент, таких, как у машины Тьюринга); состояние памяти в любой момент кодируется парой $q = (q', q'')$, где q' описывает состояние внутренней памяти, а q'' — состояние внешней, если она, конечно, существует, в противном случае q'' является пустым словом;

б) которая может присутствовать в нескольких вершинах лабиринта одновременно; она имеет n головок, которые могут находится в n различных вершинах некоторого лабиринта из данного класса лабиринтов;

в) которая является допустимой для данного класса $\mathcal{L}$ лабиринтов, т.е. может в любом лабиринте из данного класса "передвигаться" или, другими словами, этот лабиринт является ее "средой обитания" (в последующем разъясним о чем здесь идет речь);

г) которая может стирать или писать отметки из множества Ω' во всех вершинах лабиринта, в которых присутствует, или делать то же самое у любой инцидентной этим вершинам дуге (в ее конце или в начале в зависимости от того, является ли соответствующая вершина ее концом или началом соответственно) или в конце дуг, которых выбирает в качестве пути своего дальнейшего передвижения, с отметками из множества Σ';

д) которая может передвигать, брать с собою или оставлять в вершинах камни (они представляют своего рода конечную внешнюю память, их количество не увеличивается и не уменьшается, они сами по себе не передвигаются); среди этих камней могут быть такие, которыми можно "отмечать" только вершины (вершинные камни) или такие, которыми можно отмечать только дуги (ориентирные камни); некоторые из камней (вершинных, ориентирных или остальных) могут быть "одного и того же цвета" или могут быть "покрашены различными красками".

Таким образом, $\mathcal{L}$-монстр — это упорядоченный набор

$$\mathcal{A} = (A, Q, B, \mathcal{K}, H, \psi, \varphi, q_0),$$

у которого A — множество всех входных лабиринтных ситуаций, Q — множество всех кодов, описывающих состояние памяти, B — множество всех кодов, описывающих выходные действия, $\mathcal{K} = (K, K_1, K_2, \sim_{\mathcal{K}})$ — камневая структура, где множества K_1 и K_2, K_1, $K_2 \subseteq K$, $K_1 \cap K_2 = \emptyset$, являются, соответственно, множествами вершинных и ориентирных камней, и $\sim_{\mathcal{K}}$ — отношение эквивалентности в K, такое, что $[k] \subseteq K_i$ для любых $k \in K_i$ и $i = 1, 2$, множество $H = \{h_1, h_2, \ldots h_n\}$ — множество головок (эти головки не надо путать с головками внешней памяти, $\psi : Q \times A \to B$, $\varphi : Q \times A \to Q$ и q_0 — начальное состояние $\mathcal{L}$-монстра. Пусть L, $L = (G, f, g, \eta)$ — некоторый лабиринт из $\mathcal{L}(\Omega, \Omega_1, \Sigma, \Sigma_1)$, который находится в некотором состоянии (f', g') и даны некоторые отображения:

$$l_H : H \to V(L), \qquad l_{\mathcal{K}} : \mathcal{K} \to V(L) \cup \mathrm{Con}(L) \cup H.$$

где $l_{\mathcal{K}}(k) \in V(L) \cup H$ для любого $k \in K_1$ и $l_{\mathcal{K}}(k) \in \mathrm{Con}(L) \cup H$ для любого $k \in K_2$. Погруженным $\mathcal{L}$-монстром $\mathcal{A}$ в лабиринте L назовем тройку $(\mathcal{A}, l_H, L_{\mathcal{K}})$, а пару $(l_H, l_{\mathcal{K}})$ — его погружением.

Оказавшись в лабиринте L,для которого он является допустимым, т.ебудучи погружен в L, лабиринтный монстр $\mathcal{A}$собирает всевозможную ему доступную информацию, которая, конечно, зависит от погружения; обозначим ее через $a(l_H, l_{\mathcal{K}})$. На самом деле, пусть $v_i = f_H(h_i)$ для любого i, $1 \le i \le n$. Тогда, $a(l_H, l_{\mathcal{K}})$ содержит следующую информацию:

— об отметках вершин, в которых он присутствует, и об отметках дуг, инцидентных этим вершинам, т.е о множестве

$$\{(f(v_i), v_i) : 1 \le i \le n\} \cup \{(f'(v_i), v_i) : 1 \le i \le n\} \cup \{(g(x), x) :$$
$$x \in (\mathrm{Con}_{v_i}(L)), 1 \le i \le n\} \cup \{(g'(x), x) : x \in (\mathrm{Con}_{v_i}(L)) : 1 \le i \le n\};$$

— о наличии камней в вершинах, в которых он присутствует, о наличии камней на дугах, инцидентным этим вершинам, и наличии камней у себя, т.е. о множестве

$$\{(l_{\mathcal{K}}^{-1}(h), h) : h \in H\} \cup \{(l_{\mathcal{K}}^{-1}(c), c) : \\ c \in \mathrm{Con}_{v_i}(L), 1 \le i \le n\} \cup \{(l_{\mathcal{K}}^{-1}(v_i), v_i) : 1 \le i \le n\};$$

— взаимном расположении головок, т.е. определяется отношение эквивалентности $\rho(l_H)$ на множестве H следующим способом:

$$h_1\, \rho(l_H)\, h_2 \text{ тогда и только тогда, если } l_H(h_1) = l_H(h_2),$$

для любых $h_1, h_2 \in H$;

— о системе следования в вершинах, в которых он присутствует, т.е. о множестве

$$\{\eta_{v_i} \,|\, 1 \le i \le n\}.$$

Учитывая состояние внутренней памяти и состояние внешней памяти, $\mathcal{L}$-монстр $\mathcal{A}$ предпринимает следующие выходные действия:

1) решает, каким способом изменить временные отметки вершин (может даже их стирать, т.е отмечать вершину пустым символом Λ), в которых он присутствует, если они там существуют, или их туда записывать, если их там нет, т.е. на вершинах v_i, $1 \le i \le n$, изменяет значения функции f' и тем самым заменяет эту функцию на функцию $\hat{\psi}(f')$;

2) решает, какие камни возьмет (если они там есть), а какие камни оставит, в вершинах, в которых он присутствует, в дугах, которые инцидентны этим вершинам, и в конце дуг, которые он выбирает в качестве путей передвижения своих головок; т.е. таким способом меняет функцию $l_{\mathcal{K}}$ на функцию $\hat{\psi}(l_{\mathcal{K}})$;

3) изменяет состояния внутренней и внешней памяти, т.е из состояния q переходит в состояние $\varphi(q, a(l_H, l_{\mathcal{K}}))$;

4) выбирает дуги инцидентные вершинам, в которых он присутствует, в качестве путей дальнейшего передвижения своих головок. Пусть это дуги $\gamma_i = \psi_i(q, a(l_H, l_{\mathcal{K}}))$, $1 \le i \le n$; если для некоторых i, j имеет место $v_i = v_j$, то $\gamma_i = \gamma_j$. Тогда функцию l_H заменяет на функцию $\hat{\psi}(l_H)$, такую, что $\hat{\psi}(l_H)(h_i) = \mathbf{p}_2(\gamma_i)$.

Опишем еще допустимость данного $\mathcal{L}$-монстра $\mathcal{A}$,но прежде дадим некоторые определения.

Пусть дан набор

$$\mathbf{A} = (A_1, A_2, M, M', K; f, f', k; \sim_A, \rho, \sim_K),$$

где $A_1, A_2, A = A_1 \cup A_2, M, M', K$ — некоторые множества; $f : A \to M, f' : A \to M'$ — некоторые функции; $k : K \to A$ — некоторая частичная функция; $\sim_A$ — отношение эквивалентности в A, такое, что если $a_1 \sim a_2$, то

$$a_1 \in A_1 \text{ и } a_2 \in A_2, \text{ или } a_1 \in A_2 \text{ и } a_2 \in A_1,$$

$\sim_K$ — отношение эквивалентности в K и ρ — некоторое бинарное отношение в $A/\sim_A$. Автоморфизмом этой структуры назовем любую биекцию $i : A \to A$ такую, что

1) $f(i(a)) = f(a)$ и $f'(i(a)) = f'(a)$ для любого $a \in A$;
2) $|[p] \cap k^{-1}(a)| = |[p] \cap k^{-1}(i(a))|$ для любых $p \in K$;
3) $i(A_1) = A_1$ и $i(A_2) = A_2$;
4) $a_1 \sim_A a_2 \Rightarrow i(a_1) \sim_A i(a_2)$;
5) $([a_1], [a_2]) \in \rho \Rightarrow ([i(a_1)], [i(a_2)]) \in \rho$.

Стационарной точкой автоморфизма i назовем любую точку из A_1 для которой $i(a) = a$. Множество всех стационарных точек автоморфизма i обозначим через $\mathrm{St}(\mathbf{A}; i)$, а через $\mathrm{Aut}(\mathbf{A})$ — множество всех автоморфизмов структуры $\mathbf{A}$. Обозначим

$$\mathrm{St}(\mathbf{A}) = \bigcap_{i \in \mathrm{Aut}(\mathbf{A})} \mathrm{St}(\mathbf{A}; i).$$

Погружение (l_H, l_K) называется нетупиковым, если для любого i, $1 \le i \le n$, имеет место

$$\mathrm{St}(v_i) = \mathrm{St}[(\mathrm{Con}_a^{v_i}(L), \mathrm{Con}_b^{v_i}(L), \Sigma, \Sigma', K; f|_{\mathrm{Con}'_{v_i}(L)}, f'|_{\mathrm{Con}'_{v_i}(L)}, \\ l_K|_{\mathrm{Con}'_{v_i}(L)}; \overset{\mathrm{Con}}{\sim}, \eta_{v_i}(L), \sim_K)] \neq \emptyset$$

и $\psi_i(q, a(l_H, l_K)) \in \mathrm{St}(v_i)$. Пусть (l_H, l_K, q) и (l'_H, l'_K, q') — некоторые два погружения данного $\mathcal{A}$-монстра. Пишем

$$(l_H, l_K, q) \Rightarrow (l'_H, l'_K, q'),$$

если $q' = \varphi(q, a(l_H, l_K))$, $l'_H = \hat{\psi}(q, a(l_H, l_K))$ и $l'_K = \hat{\phi}(l_K)$. Обозначим множество всех нетупиковых погружений через $\mathbf{Emb}(\mathcal{A}, L)$. Если множество $\mathbf{Emb}(\mathcal{A}, L)$ закрыто по отношению к операции $\Rightarrow$, то этот автомат называется допустимым.

Пусть (l_H, l_K, q) и (l'_H, l'_K, q') — некоторые два погружения данного $\mathcal{A}$-монстра. Пишем

$$(l_H, l_K, q) \overset{*}{\Rightarrow} (l'_H, l'_K, q'),$$

если существует последовательность погружений $(l^i_H, l^i_{\mathcal{K}}, q_i)$, $1 \le i \le n$, такая, что

$$(l^1_H, l^1_{\mathcal{K}}, q_1) = (l^1_H, l^1_{\mathcal{K}}, q_1) \quad \text{и} \quad (l^n_H, l^n_{\mathcal{K}}, q_n) = (l'_H, l'_{\mathcal{K}}, q'),$$

и имеет место

$$(l^i_H, l^i_K, q_i) \Rightarrow (l^{i+1}_H, l^{i+1}_K, q_{i+1}),$$

для любого i, $1 \le i \le n-1$. Если для данного $\mathcal{L}$-монстра существует погружение (l_H, l_K, q), такое, что любое погружение (l'_H, l'_K, q'), удовлетворяющее условию $(l_H, l_K, q) \overset{*}{\Rightarrow} (l'_H, l'_K, q')$, является нетупиковым, то мы говорим, что данный монстр частично допустим.*Поведением* $\mathcal{L}$-монстра $\mathcal{A}$ в лабиринте L назовем последовательность погружений

$$(l^0_H, l^0_K, q_0), (l^1_H, l^1_K, q_1), \ldots,$$

такую, что $(l^0_H, l^0_K, q_0) = (l_H, l_K, q)$ и

$$(l^i_H, l^i_K, q_i) \Rightarrow (l^{i+1}_H, l^{i+1}_K, q_{i+1}),$$

для любого $i = 0, 1, \ldots$. Говорим, что $\mathcal{A}$ *обходит* лабиринт L, если

$$\bigcup_{i=1}^{\infty} l^i_H(H) = V(L).$$

Значит, лабиринтными монстрами могут быть и автомат, автомат с одним или несколькими магазинами, счетчиками, стеками, и машина Тьюринга, и даже коллективы таких машин. Ясно, что коллектив лабиринтных монстров является одним лабиринтным "сверхмонстром".

Абстрактным конечным автоматом называется набор $\mathfrak{A} = (A, Q, B, \varphi, \psi)$, где A, B и Q суть конечные алфавиты: входной, выходной и состояний соответственно; $\varphi : Q \times A \to Q$ и $\psi : Q \times A \to B$ суть функции переходов и выходов соответственно.

Посмотрим сейчас в качестве лабиринтных монстров автоматы, а в качестве лабиринтов класс $\tilde{\mathcal{L}}(\Omega, \Sigma)$, который определяется так: некоторый $L \in \mathcal{L}(\Omega, \Sigma)$ принадлежит классу $\tilde{\mathcal{L}}(\Omega, \Sigma)$ тогда и только тогда, если для любой $v \in V(L)$ и любых $\gamma_1, \gamma_2 \in \Gamma(L)$, таких, что $\mathbf{p}_1(\gamma_1) = \mathbf{p}_1(\gamma_2) = v$, следует, что $|\gamma_1| \ne |\gamma_2|$. Тогда картина дана в предыдущем параграфе упрощается и, поскольку этот случай для нас будет очень важным, мы ее здесь дадим в более конкретном виде.

Автомат $\mathfrak{A}$ назовем допустимым для класса лабиринтов $\tilde{\mathcal{L}}(\Omega, \Sigma)$, если его входной алфавит состоит из букв a вида $(\omega, \{\sigma_1, \ldots, \sigma_n\})$, где $\omega \in \Omega$ и $\{\sigma_1, \dots, \sigma_n\} \subseteq \Sigma$, и выходной алфавит есть $\Sigma \cup \{\kappa\}$, $\kappa \notin \Sigma$, при этом всегда $\psi(q,a) \in \mathbf{p}_2(a) \cup \{\kappa\}$. Обозначим класс всех таких автоматов через $\mathrm{A}(\Omega,\Sigma)$. Пусть $\mathfrak{A}_{q_0}$ — некоторый инициальный автомат из $\mathrm{A}(\Omega,\Sigma)$ и L_{v_0} — некоторый инициальный лабиринт из $\tilde{\mathcal{L}}(\Omega,\Sigma)$. Интерпретируем функционирование автомата $\mathfrak{A}_{q_0}$ в лабиринте L_{q_0} следующим образом. Автомат $\mathfrak{A}_{q_0}$ помещается в начальном моменте в вершину v_0 лабиринта L_{v_0} Предположим, что в какой-то момент автомат $\mathfrak{A}_{q_0}$ оказался в вершине v лабиринта L_{v_0} и в состоянии q. Считаем, что он обозревает нагруженную звезду, образованную исходящими из этой вершины дугами.Его входной буквой в этот момент является пара, образованная отметкой вершины и множеством отметок звезды. В следующий момент, если $\psi(q,a) \neq \kappa$, то автомат перемещается в вершину,в которую ведет дуга с отметкой $\psi(q,a)$, а если $\psi(q,a) = \kappa$, то остается на месте, и всегда переходит в состояние $\varphi(q,a)$. Этот процес продолжается далее. Таким образом автомат осуществляет движение по лабиринту, последовательно проходя некоторый путь. На самом деле функционирование автомата $\mathfrak{A}_{q_0}$ в лабиринте L_{q_0} можно определить как поведение автомата $\mathfrak{A}_{q_0}$ в лабиринте L_{v_0}

Последовательность пар

$$\pi(\mathfrak{A}_{q_0}; L_{v_0}) = (q_0, v_0), (q_1, v_1), \dots$$

называем *поведением автомата* $\mathfrak{A}_{q_0}$ *в лабиринте* L_{v_0}, если v_{i+1} есть вершина лабиринта L_{v_0}, в которую автомат, находясь в состоянии q_i, переходит из вершины v_i, а q_{i+1} есть состояние автомата $\mathfrak{A}_{q_0}$, в которое перейдет при этом автомат.Последовательность $\mathrm{Tr}(\mathfrak{A}_{q_0}; L_{v_0}) = v_0, v_1, \dots$ называем *траекторией* автомата $\mathfrak{A}_{q_0}$ в лабиринте L_{v_0}. Если для некоторого $u \in V(L_{v_0})$ существует $q \in Q_{\mathfrak{A}_{q_0}}$ такое, что пара (q,u) принадлежит $\pi(\mathfrak{A}_{q_0}; L_{v_0})$, то говорим, что $\mathfrak{A}_{q_0}$ обходит вершину u лабиринта L_{v_0}. Обозначим множество всех вершин, которые обходит $\mathfrak{A}_{q_0}$ в лабиринте L_{v_0}, через $\mathrm{Int}(\mathfrak{A}_{q_0}, L_{v_0})$.

Введем понятия допустимого автомата и его поведение в случае n-мерных прямоугольных лабиринтов более формально. Конечный автомат $\mathfrak{A}q_0 = (A, Q, B, \varphi, \psi, q_0)$ называется *допустимым*, если A — множество всех непустых подмножеств множества $\tilde{E}^n$, $B = \tilde{E}^n \cup \{0\}$, и $\psi(q,a) \in a \cup \{0\}$ для всех $q \in Q$ и $a \in A$. *Поведением* автомата $\mathfrak{A}_{q_0}$ в лабиринте $L = (V, \Gamma; v', v'')$ называем последовательность $\pi(\mathfrak{A}_{q_0}; L)$: $(q_0, v_0), (q_1, v_1), \dots$, где $v_0 = v'$, $(v_i, v_{i+1}) \in \Gamma$ или $v_i = v_{i+1}$, $q_{i+1} = \varphi(q_i, [v_i]_L)$ и $\psi(q_i, [v_i]_L) = |(v_i, v_{i+1})|$, $i = 0, 1, \dots$. Пару (q_i, v_i) из поведения $\pi(\mathfrak{A}_{q_0}; L)$ будем обозначать через $\pi_i(\mathfrak{A}_{q_0}; L)$.

Пусть $L_{v_0} \in \tilde{\mathcal{L}}(\Omega,\Sigma)$ и $\mathfrak{A}_{q_0} \in \mathrm{A}(\Omega,\Sigma)$. Если $\mathrm{Int}(\mathfrak{A}_{q_0}, L_{v_0}) = V(L_{v_0})$, то говорим, что $\mathfrak{A}_{q_0}$ обходит L_{v_0}; в противном случае L_{v_0} является *ловушкой* для $\mathfrak{A}_{q_0}$. Эти понятия сейчас можно расширить до любых сочетаний инициальных или неинициальных автоматов и лабиринтов. Чтобы легче описать все эти сочетания поступим так. Пусть $L \in \tilde{\mathcal{L}}(\Omega, \Sigma)$ и $\mathfrak{A} \in \mathrm{A}(\Omega, \Sigma)$ причем и L и $\mathfrak{A}$ могут быть как инициальными так и неинициальными. Рассмотрим понятия "$\alpha\beta$-обходит' и "$\beta\alpha$-ловушка", где $\alpha, \beta \in \{\mathrm{I,A,E}\}$. Если $\alpha = \mathrm{I}$ ($\alpha \neq \mathrm{I}$), то $\mathfrak{A}$ является инициальным (неинициальным) автоматом, а если $\beta = \mathrm{I}$ ($\beta \neq \mathrm{I}$), то L является инициальным (неинициальным) лабиринтом. Слово A указывает на то, что при этом берутся все вершины данного неинициального лабиринта L или все состояния данного неинициального автомата $\mathfrak{A}$, а слово E — на то, что берем только некоторую вершину данного неинициального лабиринта L или некоторое состояние данного автомата $\mathfrak{A}$. Так, например, $L_{v_0} \in \tilde{\mathcal{L}}(\Omega, \Sigma)$ является IA-*ловушкой* для $\mathfrak{A} \in \mathrm{A}(\Omega, \Sigma)$, если для всех $q \in Q_{\mathfrak{A}}$ лабиринт L_{v_0} является ловушкой для $\mathfrak{A}_q$. Автомат $\mathfrak{A} \in \mathrm{A}(\Omega, \Sigma)$ AA-*обходит* лабиринт $L \in \tilde{\mathcal{L}}(\Omega, \Sigma)$, если для всех $q \in Q_{\mathfrak{A}}$ и всех $v \in V$ автомат $\mathfrak{A}_q$ обходит лабиринт L_v. Если $\alpha, \beta \in \{\mathrm{I,E}\}$, то вместо $\alpha\beta$-обходит и $\beta\alpha$-ловушка говорим обходит и ловушка. Если $\alpha, \beta \in \{\mathrm{I,A}\}$, то вместо $\alpha\beta$-обходит и $\beta\alpha$-ловушка говорим *сильно обходит* и *сильная ловушка*.

Наряду с поведением автомата в лабиринте можно также рассмотреть поведение системы автоматов в лабиринте. Пусть $L_{v_1,\ldots,v_n} \in \tilde{\mathcal{L}}(\Omega, \Sigma)$ и задана система допустимых автоматов $\mathcal{A} = \{\mathfrak{A}^1_{q_1}, \ldots, \mathfrak{A}^n_{q_n}\}$. Если под поведением этой системы в $L_{v_1,\ldots,v_n}$ понимать множество поведений

$$\{\pi(\mathfrak{A}_{q_i}; L_{v_i}), \ldots, \pi(\mathfrak{A}_{q_n}; L_{v_n})\},$$

то эту систему называем *независимой*, а само поведение — *поведением независимой системы*. Если для некоторого i, $1 \leq i \leq n$, $\mathrm{Int}(\mathfrak{A}^i_{q_i}, L_{v_i}) = V$, то говорим, что $\mathcal{A}$ обходит $L_{v_1,\ldots,v_n}$, а если $\cup_{i=1}^n \mathrm{Int}(\mathfrak{A}^i_{q_i}, L_{v_i}) = V$, то говорим, что $\mathcal{A}$ A-*обходит* $L_{v_1,\ldots,v_n}$; в противном случае говорим, что $L_{v_1,\ldots,v_n}$ является ловушкой и соответственно A-ловушкой для независимой системы $\mathcal{A}$. Как и в случае одного автомата мы можем ввести аналогичным способом понятия $\alpha\beta$-обходит и $\beta\alpha$-ловушка ($\alpha\beta$-A-обходит и $\beta\alpha$-A-ловушка), где $\alpha, \beta \in \{\mathrm{I,A,E}\}$. Если $\alpha, \beta \in \{\mathrm{I,E}\}$, то вместо $\alpha\beta$-обходит и $\beta\alpha$-ловушка говорим обходит и ловушка. Если $\alpha, \beta \in \{\mathrm{I,A}\}$, то вместо $\alpha\beta$-A-обходит и $\beta\alpha$-A-ловушка говорим сильно обходит и сильная ловушка.

Рассмотрим теперь более сильный вариант поведения системы автоматов $\mathcal{A}$ допустимых для $\tilde{\mathcal{L}}(\Omega, \Sigma)$ в лабиринте $L_{v_1,\ldots,v_n} \in \tilde{\mathcal{L}}(\Omega, \Sigma)$.

Закодируем наши автоматы с помощью букв $u_1, \ldots, u_n$, считая, что u_i принимает в качестве значения то состояние, в котором находится $\mathfrak{A}^i_{q_i}$, или Λ. Если входной алфавит для автомата $\mathfrak{A}^i_{q_i}$, $1 \leq i \leq n$, состоит из букв a вида

$$(\omega, \{u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n\}, \{\sigma_1, \ldots, \sigma_m\}),$$

где $\omega \in \Omega$ и $\{\sigma_1, \ldots, \sigma_m\} \subseteq \Sigma$, а выходной алфавит есть множество $\Sigma \cup \{\kappa\}$, $\kappa \notin \Sigma$, и при этом всегда $\psi_i(q, a) \in \mathbf{p}_3(a) \cup \{\kappa\}$, $q \in Q_i$, то систему $\mathcal{A}$ назовем *коллективом.* Интерпретируем функционирование коллектива $\mathcal{A} = \{\mathfrak{A}^1_{q_1}, \ldots, \mathfrak{A}^n_{q_n}\}$ в $L_{v_1,\ldots,v_n}$ его движением в лабиринте $L_{v_1,\ldots,v_n}$ следующим образом. Автомат $\mathfrak{A}^i_{q_i}$ в начальный момент помещаем в вершину v_i лабиринта L, $1 \leq i \leq n$. Предположим, что в некоторый момент t автомат $\mathfrak{A}^i_{q_i}$ оказался в вершине v_i^t и в состоянии q_i^t. Считаем, что он обозревает нагруженную звезду, образованную исходящими из этой вершины дугами. Его входной буквой a_i^t в этот момент является тройка, образованная отметкой вершины, множеством кодов всех автоматов коллектива, находящихся в вершине v_i^t, кроме кода самого автомата $\mathfrak{A}^i_{q_i}$, и множеством отметок звезды. В следующий момент, если $\psi_i(q_i^t, a_i^t) \neq \kappa$, то автомат перемещается в вершину, в которую ведет дуга с отметкой $\psi_i(q_i^t, a_i^t)$, а если $\psi_i(q_i^t, a_i^t) = \kappa$, то остается на месте, и переходит в состояние $\varphi_i(q_i^t, a_i^t)$. Этот процесс продолжается далее. Таким образом автомат $\mathfrak{A}^i_{q_i}$ осуществляет движение по лабиринту, последовательно проходя некоторый путь. Последовательность пар

$$(q_i^0, v_i^0), (q_i^1, v_i^1), \ldots,$$

где $(q_i^0, v_i^0) = (q_i, v_i)$, v_i^{j+1} есть вершина, в которую переходит автомат из вершины v_i^j, находясь в состоянии q_i^j, а q^i_{j+1} — его новое состояние, называем *поведением автомата $\mathfrak{A}^i_{q_i}$ коллектива $\mathcal{A}$ в лабиринте $L_{v_1,\ldots,v_n}$;* при этом говорим, что $\mathfrak{A}^i_{q_i}$ обходит вершины $v_i^0, v_i^1, \ldots$ и обозначаем множество их через $\mathrm{Int}(\mathcal{A}, L_{v_1,\ldots,v_n}; i)$. Последовательность

$$\pi(\mathcal{A}, L_{v_1,\ldots,v_n}) = (q_1^0, \ldots, q_n^0, v_1^0, \ldots, v_n^0), (q_1^1, \ldots, q_n^1, v_1^1, \ldots, v_n^1), \ldots,$$

такая, что последовательность $(q_i^0, v_i^0), (q_i^1, v_i^1), \ldots$ является поведением автомата $\mathfrak{A}^i_{q_i}$ коллектива $\mathcal{A}$ в лабиринте $L_{v_1,\ldots,v_n}$, называется *поведением коллектива $\mathcal{A}$ в лабиринте $L_{v_1,\ldots,v_n}$*. Пусть $\mathrm{Int}(\mathcal{A}, L_{v_1,\ldots,v_n}) = \cup_{i=1}^{n} \mathrm{Int}(\mathfrak{A}_{q_i}, L_{v_i}; i)$. Если $\mathrm{Int}(\mathcal{A}, L_{v_1,\ldots,v_n}) = V$, то говорим что $\mathcal{A}$ *обходит* $L_{v_1,\ldots,v_n}$; в противном случае $L_{v_1,\ldots,v_n}$ является *ловушкой* для $\mathcal{A}$. Лабиринт L называем *сильной ловушкой* для $\mathcal{A}$, если для любых $v_1, \ldots, v_n \in V(L)$ лабиринт $L_{v_1,\ldots,v_n}$ является ловушкой для $\mathcal{A}$. Коллектив $\mathcal{A}$ *сильно обходит* лабиринт L, если для любых $v_1, \ldots, v_n \in V(L)$ коллектив $\mathcal{A}$ обходит лабиринт $L_{v_1,\ldots,v_n}$.

Набор $\mathcal{A} = (\mathfrak{A}_1, \ldots, \mathfrak{A}_m)$, где $\mathfrak{A}_i = (A_i, Q_i, B_i, \varphi_i, \psi_i, q_0^i)$, $i = 1, \ldots, m$, — некоторый автомат, называется *коллективом допустимых автоматов в* $\mathbf{R}^n$, если для любого $i = 1, \ldots, m$,

1) $A_i = \{a \in \mathcal{P}_0(\tilde{E}^n) \times [\prod_{i=1}^m (\theta \cup Q_i)] | \mathbf{p}_{i+1}(a) = \theta\}$;

2) $B_i = \tilde{E}^n \cup \{0\}$;

3) $\psi_i(q, a) \in \mathbf{p}_1(a) \cup \{0\}$, для любых $q \in Q_i$ и $a \in A$;

здесь предполагается, что θ — некоторый фиксированный элемент, не принадлежащий множеству $\cup_{i=1}^n Q_i$.

Пусть L — некоторый n-мерный лабиринт, v_i, $i = 1, \ldots, m$, — некоторые вершины n-мерного лабиринта L и q_i — некоторое состояние автомата $\mathfrak{A}_i$, $i = 1, \ldots, m$. Обозначим $\vec{v} = (v_1, \ldots, v_m)$ и $\vec{q} = (q_1, \ldots, q_m)$. Пусть

$$a_i(\vec{v}, \vec{q}) = ([v_i]_L, [a_i(\vec{v}, \vec{q})]_1, \ldots, [a_i(\vec{v}, \vec{q})]_m),$$

где

$$[a_i(\vec{v}, \vec{q})]_j = \begin{cases} q_j, & \text{если } v_i = v_j \text{ и } i \neq j; \\ \theta, & \text{если } v_i \neq v_j \text{ или } i = j. \end{cases}$$

Пусть $\vec{v}_0 = (v_0^1, v_0^2, \ldots, v_0^m)$ — некоторый набор вершин n-мерного лабиринта L. *Поведением коллектива* $\mathcal{A} = (\mathfrak{A}_1, \ldots, \mathfrak{A}_n)$ *в n-мерном лабиринте* $(L; \vec{v}_0)$ называется последовательность $\pi(\mathcal{A}; L, \vec{v}_0) = (\vec{v}_0, \vec{q}_0), \ldots, (\vec{v}_t, \vec{q}_t), \ldots$, где $\vec{v}_t = (v_t^1, \ldots, v_t^m)$ и $\vec{q}_t = (q_t^1, \ldots, q_t^m)$, такая, что для любых t, i, где $t = 0, 1, \ldots$ и $i = 1, \ldots, m$,

1) q_0^i — начальное состояние автомата $\mathfrak{A}_i$;

2) $(v_t^i, v_{t+1}^i) \in \Gamma$ или $v_t^i = v_{t+1}^i$;

3) $q_{t+1}^i = \varphi_i(q_t^i, a_i(\vec{v}_t, \vec{q}_t))$;

4) $\psi_i(q_t^i, a_i(\vec{v}_t, \vec{q}_t)) = |(v_t^i, v_{t+1}^i)|$.

Обозначим

$$\begin{aligned}
&\mathrm{Int}_j^i(\mathcal{A}; L, \vec{v}_0) = v_i^j \ (1 \leq j \leq m,\ i = 0, 1, \ldots),\\
&\mathrm{Int}^i(\mathcal{A}; L, \vec{v}_0) = \{v_i^1, \ldots, v_i^m\} \ (i = 0, 1, \ldots),\\
&\mathrm{Int}_j(\mathcal{A}; L, \vec{v}_0) = \cup_{i=1}^{\infty} \{v_i^j\},\\
&\mathrm{Int}(\mathcal{A}; L, \vec{v}_0) = \cup_{i=1}^{\infty} (\cup_{j=1}^m \{v_i^j\}),\\
&\mathrm{Fr}(\mathcal{A}; L, \vec{v}_0) = V \backslash \mathrm{Int}(\mathcal{A}; L, \vec{v}_0).
\end{aligned}$$

Если выше $v_0^1 = \cdots = v_0^m = v_0$, то в последующем всегда будем говорить не о поведении $\mathcal{A}$ в $(L; \vec{v}_0)$, а о поведении $\mathcal{A}$ в n-мерном лабиринте $(L; v_0)$, и во всех выше введенных обозначениях, в которых фигурирует $\vec{v}_0$, будем писать v_0 вместо $\vec{v}_0$.

Говорим, что коллектив допустимых автомат $\mathcal{A}$ *обходит* n-мерный лабиринт $(L; v_0)$, если $\mathrm{Fr}(\mathcal{A}; L, v_0) = \emptyset$. Говорим, что коллектив $\mathcal{A}$ *сильно обходит* n-мерный лабиринт L, если для любого $v \in V(L)$ коллектив $\mathcal{A}$ обходит $(L; v)$. n-мерный лабиринт $(L; v_0)$ называется *n-мерной ловушкой* для коллектива $\mathcal{A}$, если $\mathcal{A}$ не обходит $(L; v_0)$. Также, n-мерный лабиринт $(L; v_0, v_1)$ называется n-мерной ловушкой для $\mathcal{A}$, если $v_1 \in \mathrm{Fr}(\mathcal{A}; L, v_0)$ и $[v_1] \neq \tilde{E}^n$. n-мерный лабиринт L называется *сильной n-мерной ловушкой* дляд коллектива $\mathcal{A}$, если для любого $v \in V(L)$ коллектив $\mathcal{A}$ не обходит (L, v).

Отметим некоторые автоматы $\mathfrak{A}^{i_1}_{q_{i_1}}, \ldots, \mathfrak{A}^{i_m}_{q_{i_m}}$, $1 \leq i_1 < \cdots < i_m \leq n$, коллектива $\mathcal{A} = (\mathfrak{A}^1_{q_1}, \ldots, \mathfrak{A}^n_{q_n})$. Автоматы $\mathfrak{A}^{i_1}_{q_{i_1}}, \ldots, \mathfrak{A}^{i_m}_{q_{i_m}}$ называются *камнями в коллективе* $\mathcal{A}$, если имеют место следующие условия:

а) у автомата $\mathfrak{A}^{i_j}_{q_{i_j}}$, $1 \leq j \leq m$, только одно состояние — q_{i_j};

б) если для некоторого входа

$$a = (\omega, \{u_1, \ldots, u_{i_l-1}, u_{i_l+1}, \ldots, u_n\}, \{\sigma_1, \ldots, \sigma_m\})$$

автомата $\mathfrak{A}^{i_l}_{q_{i_l}}$, $1 \leq l \leq m$, имеет место $\psi_i(q, a) = \sigma_k$, $1 \leq k \leq m$, то существует $j \neq i_l$, $1 \leq j \leq n$, $1 \leq l \leq m$, такое, что $u_j \neq \Lambda$ и $\psi_j(q, a') = \sigma_k$, где $a' = (\omega, \{u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n\}, \{\sigma_1, \ldots, \sigma_m\})$ и u_i код сотояния q_i.

Коллектив $\mathcal{A}$ с m отмеченными автоматами $\mathfrak{A}^{i_1}_{q_{i_1}}, \ldots, \mathfrak{A}^{i_m}_{q_{i_m}}$, которые являются камнями, называется *коллективом из $n - m$ автоматов с m камней.*

3. Допустимые автоматы в плоских мозаичных лабиринтах

Пусть $\mathcal{L}$ — класс некоторых лабиринтов $(L; u)$. Если некоторый автомат $\mathfrak{A}_{q_0}$ сильно обходит любой лабиринт $(L; u)$ из $\mathcal{L}$, то мы говорим, что автомат $\mathfrak{A}_{q_0}$ является универсальным обходчиком для класса $\mathcal{L}$.

Предложение 3.1. *Существует универсальный обходчик для класса всех шахматных лабиринтов без конечных дыр.*

С другой стороны, имеет место нетривиальный результат: не существует конечный инициальный автомат, который обходит все конечные плоские мозаичные лабиринты. Это утверждение для конечных плоских шахматных лабиринтов фактически установлено в работе [3] с весьма громоздким обоснованием, использующим среди прочего и язык теории категорий. Элементарное и короткое доказательство этого утверждения дается в работах [17,18] (формальное отличие мозаичных и шахматных лабиринтов не является здесь существенным). Методически более наглядное доказательство этой теоремы содержится в [15], техника которого позволила решить некоторые другие и упростить уже решенные задачи типа задач обхода.

Пусть $\mathcal{L}$ — некоторый класс плоских мозаичных лабиринтов. На множестве всех пар $(i,j) \in \mathbf{N}^2$ определим частичный порядок $\leq$, пологая $(a,b) \leq (c,d)$ точно тогда, когда $a \leq c$ и $b \leq d$. Предикат $\mathbf{P}_{\mathcal{L}}(i,j)$ определим таким способом, что $\mathbf{P}_{\mathcal{L}}(a,b) = 1$, если существует коллектив типа (a,b), обходящий все лабиринты из $\mathcal{L}$, и $\mathbf{P}_{\mathcal{L}}(a,b) = 0$, если такой коллектив не существует. Нетрудно видеть, что предикат $\mathbf{P}_{\mathcal{L}}$ является монотонной функцией относительно этого частичного порядка. Точнее, пусть $(a,b) \leq (c,d)$. Тогда, если $\mathbf{P}_{\mathcal{L}}(a,b) = 1$, то $\mathbf{P}_{\mathcal{L}}(c,d) = 1$, а если $\mathbf{P}_{\mathcal{L}}(c,d) = 0$, то $\mathbf{P}_{\mathcal{L}}(a,b) = 0$. Пару (a,b) назовем нижней единицей для $\mathbf{P}_{\mathcal{L}}$, если $\mathbf{P}_{\mathcal{L}}(a,b) = 1$, а $\mathbf{P}_{\mathcal{L}}(c,d) = 0$ для любого (c,d), такого, что $(c,d) \leq (a,b)$ и $(c,d) \neq (a,b)$. Пусть $\mathbf{T}[\mathbf{P}_{\mathcal{L}}]$ — множество всех нижних единиц для $\mathbf{P}_{\mathcal{L}}$. Ясно, что задание $\mathbf{P}_{\mathcal{L}}$ однозначно определяется указанием $\mathbf{T}[\mathbf{P}_{\mathcal{L}}]$.

Теорема 3.1. *Имеет место равенство $\mathbf{T}[\mathbf{P}_{\mathcal{L}_0}] = \{(1,2,(2,0)\}$, при этом некоторые коллективы типа $(1,2)$ обходят лабиринты из n клеток класса $\mathcal{L}_0$ за время $O(n^3)$, а типа $(2,0)$ — за время $O(n^2)$, и останавливаются после обхода.*

В работе [10] показано, что $\mathbf{P}_{\mathcal{L}_0}(1,5) = 1$, причем существуют некоторые коллективы типа $(1,5)$, которые обходят и остановливаются после обхода любого лабиринта из $\mathcal{L}_0$. В работе [2] дан эскиз доказательства того, что $\mathbf{P}_{\mathcal{L}_0}(1,2) = \mathbf{P}_{\mathcal{L}_0}(2,0) = 1$ (полное доказательство можно найти, например, в работе [14]). И наконец, в работе [7] был приведен эскиз доказательства того, что $\mathbf{P}_{\mathcal{L}_0}(1,1) = 0$, а полное доказательство этого факта было дано в работе [8].

В работе [2] показано, что автомат со счетчиком обходит класс $\mathbf{P}_{\mathcal{L}_0}$, но за время $O(n^2)$. В работе [6] теорема 3.1 обобщается на случай плоских прямоугольных лабиринтов с соответствующими оценками времени обхода вида $O(n^2)$ и $O(n^3)$, а для автомата со счетчиком и одним камнем это время равно $O(n^2)$.

Следует заметить, что возможно "вложенное" расслоение класса $\mathcal{L}_0$, такое, что для каждого слоя уже найдется автомат с одним камнем, обходящий его.

Теорема 3.2 *Для любого $k \in \mathbf{N}$ существует коллектив типа $(1,1)$, сильно обходящий все конечные плоские (шахматные) мозаичные лабиринты, у которых не более k (компонент косвязности) дыр, при этом* *автомат имеет не более C^k состояний.*

В работе [12] было установлено, что существует коллектив типа (1,1), который сильно обходит все конечные плоские шахматные лабиринты, имеющие не более двух дыр. В работе [13] показано то же самое, но в случае, когда у лабиринта не больше трех дыр. Затем в [9] была доказана первая часть теоремы 3.2 для случая конечных плоских шахматных лабиринтов, а позже в [5] была установлена оценка для числа состояний автомата и упрощено доказательство первой части теоремы.

Возможности коллективов автоматов при обходе лабиринтов много шире, чем возможности независимых систем автоматов. Об этом свидетельствуют следующие утверждения, в которых речь идет о конечных и бесконечных плоских мозаичных лабиринтах.

Теорема 3.3. **[16]** *Имеет место соотношение* $\{(2,3),(3,2),(4,1),(5,0)\} \subseteq \mathbf{T}[\mathbf{P}_{\mathcal{L}_2}]$.

В работах [1,4] было установлено, что $\mathbf{P}_{\mathcal{L}_2}(1,7) = 1$, а в работе [11] — что $\mathbf{P}_{\mathcal{L}_2}(1,5) = 1$; наконец, с помощью достаточно общей конструкции была доказана теорема 3.3 [16] (очевидно, $\mathbf{P}_{\mathcal{L}_2}(0,j) = 0$). Доказательство теоремы 3.3 проводилось посредством конструирования соотвествующей ловушки для коллективов всех типов $(2,2),(3,1)$ и $(4,0)$. Затем строились примеры коллективов всех типов $(i,j) \in \mathbf{T}[\mathbf{P}_{\mathcal{L}_2}]$, которые обходят все плоские мозаичные лабиринты. Так же, как и для систем автоматов, интересно выяснить, какие достаточно широкие классы лабиринтов могут быть обойдены коллективами простых типов. Заметим, что если перейти к классу $\mathbf{P}_{\mathcal{L}'_2}$ всех плоских мозаичных лабиринтов, не содержащих бесконечных дыр, то останется справедливым утверждение, анологичное теореме 3.3.

Теорема 3.4. **[16]** *Для класса* $\mathcal{L}'_2$ *и предиката* $\mathbf{P}_{\mathcal{L}'_2}$ *имеет место соотношение* $\{(2,3),(3,2),(4,1),(5,0)\} \subseteq \mathbf{T}[\mathbf{P}_{\mathcal{L}'_2}]$.

В [16] показано, что $\mathbf{P}_{\mathcal{L}_2}(1,3) = 0$. Отсюда следует, что $\mathbf{T}[\mathbf{P}_{\mathcal{L}_2}]$ $(\mathbf{T}[\mathbf{P}_{\mathcal{L}'_2}])$ равно или $\{(1,5),(2,3),(3,2),(4,1),(5,0)\}$, или $\{(1,4),(2,3),(3,2),(4,1),(5,0)\}$. Существует гипотеза, что $\mathbf{T}[\mathbf{P}_{\mathcal{L}_2}] = \mathbf{T}[\mathbf{P}_{\mathcal{L}'_2}] = \{(1,5),(2,3),(3,2),(4,1),(5,0)\}$.

Список литературы

[1] M. Blum and W. Sakoda, *On the capability of finite automata in* 2 *and* 3 *dimensional space*, The Procedings of the 18th Annual Symposium on Foundations of Computer Science, 1977, pp. 147–161.

[2] M. Blum and D. Kozen, *On the power of the compass*, The Procedings of the 19th Annual Symposium on Foundations of Computer Science, 1978, pp. 132–142.

[3] L. Budach, *Automata and labirinths*, Math. Nachrichten **86** (1978), 195–282.

[4] Z. Habasinski and M. Karpinski, *A codification of Blum-Sakoda 7-pebbles algorithm*, ICS PAS Reports, vol. 448, 1981.

[5] A. Hemmerling, *1-pointer automata searching finite plane graphs*, Z. Math. Logik Grundlag. Math. **32** (1986), 245–256.

[6] A. Hemmerling, *Remark on the power of compass*, Lecture Notes in Computer Science, vol. 233, Springer-Verlag, 1986, pp. 405–413.

[7] F. Hoffmann, *One pebble does not suffice to search plane labyrinths*, Lecture Notes in Computer Science, vol. 117, 1981, pp. 433-444.

[8] F. Hoffman, *1-Kiesel-Automaten in Labirinthen*, Report R-Math-06/82., AdW der DDr, Berlin, 1982.

[9] K. Kriegel, *Universelle 1-kiesel automaten fur k-komponentige labirinthe*, Report R-Math-04/84, AdW der DDR, Berlin, 1984.

[10] N. A. Shah, *Pebble automata on arrays*, Computer graphics and Image Processing **3** (1974), 236–246.

[11] A. Szepietowski, *A finite 5-pebble automation can search every maze*, Information Processing Letters **15** (1982), no. 5, 199–204.

[12] A. Szepietowski, *On searching Plane Labyrinths by 1-pebble Automata*, EIK **19** (1983), no. 1/2, 79–84.

[13] A. Szepietowski, *Remarks on searching labirinths by automata*, Lecture notes in Computer Science, vol. 158, 1983, pp. 457–464.

[14] Г. Килибарда, *Об обходе конецхных лабиринтов системами автоматов*, Дискретнаыа математика **2** (1990.), no. 2, 71–82.

[15] Г. Килибарда, *Новое доказательство теоремы Будаха-Подколзина*, Дискретнаыа математика **3** (1991), no. 3, 135-146.

[16] Г. Килибарда, *О минимальных универсальных коллективах автоматов длыа плоских лабиринтов*, Дискретнаыа математика **6** (1994), no. 4, 133-153.

[17] В. Б. Кудрыавцев, А. С. Подколзин и И. Усхцхумлицх, *Введение в теориыу абстрактных автоматов*, М.: Изд-во МГУ, 1985.

[18] В. Б. Кудрыавцев, С. В. Алесхин и А. С. Подколзин, *Введение в теориыу автоматов*, М.: Наука, 1985.

Механико-математический факультет, МГУ, Москва, Россия

Универзитет у Београду, ТМФ, Карнеджијева 4, 11000 Београд, Југославия

Универзитет у Београду, ТМФ, Карнеджијева 4, 11000 Београд, Југославия

FILOMAT (Niš) 9:3 (1995), 803–807
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

OMITTING TYPES IN KRIPKE MODELS

Zoran Marković

ABSTRACT. When can a type be omitted in a Kripke model of some intuitionistic theory is investigated. As it is usual with intuitionistic systems, various classically equivalent formulations of the Omitting Types Theorem, become nonequivalent statements in the intuitionistic setting. Several such formulations are discussed in terms of whether they have the intended meaning in Kripke models, and several theorems are proved.

Classically, an Omitting Types Theorem states that an apparently weaker condition, concerning individual formulas from a type ("locally omitting"), suffices for the whole type to be omitted in some model. We will start by considering what meaning these expressions may have in the case of Kripke models of some intuitionistic theory T. A "type" should clearly be a type of an element of a Kripke model of T. If we restrict ourselves to Kripke models in which the frame, i.e., the partial ordering, has the least element the base node, this should be an element of the universe at the base node. A type for T can be defined as a set of formulas in the same language $\mathcal{L}(T)$ with one free variable, say x_0, consistent with T. Analogous definition may be given for n-types. If $\sum(x_0)$ is a type for T, we say that some Kripke model of T realizes $\sum$ if there is an element of the universe at the base node of this model, for which every formula from $\sum$ is forced. Dually, we say that some Kripke model of T omits $\sum$ if for every element of the universe at the base node of this model, there is some formula from $\sum$ which is not forced for this element. As for the "local omitting", we may consider the following four formulations:

(1) for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}(T)$ consistent with T, there exists some formula $\sigma(x_0) \in \sum$ such that the sentence $\exists x_0(\varphi(x_0) \wedge \neg\sigma(x_0))$ is consistent with T;

(2) for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}(T)$ consistent with T, there exists some formula $\sigma(x_0) \in \sum$ such that the sentence $\exists x_0 \neg(\varphi(x_0) \rightarrow \sigma(x_0))$ is consistent with T;
(3) for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}(T)$ consistent with T, there exists some formula $\sigma(x_0) \in \sum$ such that the sentence $\neg\forall x_0(\varphi(x_0) \rightarrow \sigma(x_0))$ is consistent with T;
(4) for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}(T)$ consistent with T, there exists some formula $\sigma(x_0) \in \sum$ such that $T \nvdash \forall x_0(\varphi(x_0) \rightarrow \sigma(x_0))$.

In intuitionistic predicate calculus it is easily provable that:

$$\exists x_0(\varphi(x_0) \wedge \neg\sigma(x_0)) \rightarrow \exists x_0 \neg(\varphi(x_0) \rightarrow \sigma(x_0))$$

and

$$\exists x_0 \neg(\varphi(x_0) \rightarrow \sigma(x_0)) \rightarrow \neg\forall x_0(\varphi(x_0) \rightarrow \sigma(x_0))$$

while neither of the reverse implications holds. Therefore, we have

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4).$$

However, the whole statement (2), even intuitionisticaly, implies (1), so we may dismiss it. It is easy to show that for the remaining there statements none of the reverse implications holds intuitionisticaly. The statement (4) is the most interesting, not only because it is the weakest of the four, but also because it is strong enough to prove that in the Lindenbaum algebra of consequences of T, $\sum$ generates a nonprincipal filter. We shall also use (1), mainly for technical reasons, while (3) does not seem to deserve much attention.

* * *

Let $\mathcal{L}$ be a countable first-order language, T a consistent (intuitionistic) theory in $\mathcal{L}$ and $\sum$ a set of formulas in $\mathcal{L}$ with at most x_0 free. In [3] the following theorem was proved.

Theorem 1. *If for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}$ consistent with T, there exists a formula $\sigma(x_0) \in \sum$ such that the sentence $\exists x_0(\varphi(x_0) \wedge \neg\sigma(x_0))$ is consistent with T then there exists a Kripke model of T with a countable universe at each node, such that for every element a of the universe at the base node, there exists a formula $\sigma(x_0) \in \sum$ such that $\neg\sigma[a]$ is forced at the base node.*

The proof is a Henkin-style argument along the lines of completeness proofs of [1] and [4]. T is gradually extended to an $\mathcal{L} \cup C$-saturated theory

(C being a countable set of new constants). At each stage three steps are made: for $n = 3k$ and $n = 3k + 1$ we work toward making the final theory saturated (we provide a "witness" from C for an existential consequence and add one of the disjuncts of a disjunction which is a consequence), while for $n = 3k + 2$ we add $\neg\sigma(c_k)$ for some appropriate $\sigma(x_0) \in \sum$.

It was observed later by Kripke that practically the same proof will prove the following dual theorem, which might be more useful for intuitionistic theories.

Theorem 2. (Intersecting Types Theorem) *If for any sentence $\exists x_0\varphi(x_0)$ in $\mathcal{L}$ consistent with T, there exists a formula $\sigma(x_0) \in \sum$ such that the sentence $\exists x_0(\varphi(x_0) \wedge \sigma(x_0))$ is consistent with T, then there exists a Kripke model of T with a countable universe at each node such that for every element a of the universe at the base node, there exists a formula $\sigma(x_0) \in \sum$ such that $\sigma[a]$ is forced at the base node.*

These results may be improved in two directions. One direction is to require T to be a saturated theory, i.e., a deductively closed consistent set of sentences satisfying the following two conditions:

- if $\exists x\varphi(x) \in T$ then $\varphi(c) \in T$ for some individual constant c from $\mathcal{L}(T)$
- If $\varphi \cup \psi \in T$ then $\varphi \in T$ or $\psi \in T$.

While Troelstra and Kreisel argue that we should not assume that all intuitionisticaly acceptable theories must be saturated (e.g. [5]), it is a fact that all major, naturally arising, examples of intuitionistic theories are saturated. Therefore, this is not an unreasonable requirement. In this case the condition (4) is sufficient for omitting.

Theorem 3. *If T is a saturated theory and for any sentence $\exists x_0\varphi(x_0)$ in $\mathcal{L}$ consistent with T there exists a formula $\sigma(x_0) \in \sum$ such that $T \nvdash \forall x_0(\varphi(x_0) \rightarrow \sigma(x_0))$ then there exists a Kripke model of T which omits $\sum$.*

Proof. Let $E = \{\exists x_0\varphi_0(x_0), \exists x_0\varphi_1(x_0), \dots\}$ be an enumeration of all existential sentences in $\mathcal{L}$ consistent with T. By the hypothesis of the Theorem, for each $i \in \omega$ there exists a Kripke model $\mathfrak{M}_i = ((S_i, O_i, \leq_i); \mathfrak{A}_s \colon s \in S_i)$ of T, a formula $\tau_i(x_0) \in \sum$ and an element $a \in A_{O_i}$ such that

$$O_i \Vdash T, \qquad O_i \Vdash \varphi_i[a] \quad \text{and} \quad O_i \nVdash \sigma[a].$$

Let $\mathfrak{M} = (\sum \mathfrak{M}_i)'$ be the collection of models $\mathfrak{M}_i (i \in \omega)$ (c.f. [4]). We shall prove the following two claims:

1° $\mathfrak{M} \models T$

2° $\mathfrak{M}$ omits $\sum$, i.e., if $\mathfrak{M} = ((S, O, \leq); \mathfrak{A}_s \colon s \in S)$ then for every $a \in A_0$ there exists $\sigma(x_0) \in \sum$ such that $O \nVdash \sigma[a]$.

For 1° it is enough to note that T is saturated and is, therefore, preserved under the operation of collection (cf. [4]). For 2°, we note that A_0 of $\mathfrak{M}$ consists, by definition, of individual constants occurring in T. Therefore, if $c \in A_0$ the sentence $\exists x_0(x_0 = c)$ will be a sentence of $\mathcal{L}$ consistent with T, so for some $i \in \omega$, φ_i in our enumeration E will be $(x_0 = c)$. Then, for some $a \in A_{O_i}$ we will have $O_i \Vdash a = c$ and $O_i \nVdash \sigma_i[a]$ and so $O_i \nVdash \sigma_i(c)$. As $O \leq O_i$ in $\mathfrak{M}$, we obtain $O \nVdash \sigma_i(c)$. □

Another direction in which we can improve Theorem 1. is to put some restriction on elements of $\sum$. We will show that in two such cases we can obtain the omitting types theorem in full strength, i.e., using (4) as the "locally omitting" condition.

Theorem 4. *Let $\sum$ be a set of negated formulas in $\mathcal{L}$, with at most x_0 free. If for any sentence $\exists x_0 \varphi(x_0)$ in $\mathcal{L}$ consistent with T, there exists a formula $\neg\sigma(x_0) \in \sum$ such that $T \nvdash \forall x_0(\varphi(x_0) \to \neg\sigma(x_0))$ then there exists a Kripke model of T omitting $\sum$.*

Proof. Consider $\sum' = \{\sigma : \neg\sigma \in \sum\}$. It is easy to prove that $T \nvdash \forall x_0(\varphi(x_0) \to \neg\sigma(x_0))$ if and only if $\exists x_0(\varphi(x_0) \wedge \sigma(x_0))$ is consistent with T. If $T \nvdash \forall x_0(\varphi(x_0) \to \neg\sigma(x_0))$, by completeness theorem (e.g. [1]) there exists a Kripke model $\mathfrak{M} = ((S, O, \leq); \mathfrak{A}_s : s \in S)$ of T in which for some $s \in S$ and $a \in A_s$ we have $s \Vdash \varphi[a]$ and $s \nVdash \neg\sigma[a]$ which means that for some $s' \in S$ we have $s \leq s'$ and $s' \Vdash \sigma[a]$. The truncation of $\mathfrak{M}$ at s', $\mathfrak{M}_{s'}$ will be a model of $T \cup \{\exists x_0(\varphi(x_0) \wedge \sigma(x_0))\}$. We may apply now the Intersecting Types Theorem (Theorem 4.) to $\sum'$ and obtain a Kripke model of T with a countable universe at each node which not only omits $\sum$ but in which actually for each element a of the universe at the base node there exists some $\neg\sigma(x_0) \in \sum$ such that $\sigma[a]$ is forced at the base node. □

Theorem 5. *Let $\sum$ be a set of formulas with at most x_0 free which are decidable in T, i.e., for each $\sigma(x_0) \in \sum$ we have $T \vdash \forall x_0(\sigma(x_0) \vee \neg\sigma(x_0))$. If for each sentence $\exists x_0 \varphi(x_0)$ consistent with T there exists a formula $\sigma(x_0) \in \sum$ such that $T \nvdash \forall x_0(\varphi(x_0) \to \sigma(x_0))$ then there exists a Kripke model of T omitting $\sum$.*

Proof. As in the proof of Theorem 4, $T \nvdash \forall x_0(\varphi(x_0) \to \sigma(x_0))$ implies that in some Kripke model $\mathfrak{M}$ of T for some s and $a \in A_s$ we have $s \Vdash \varphi[a]$ and $s \nVdash \sigma[a]$. As $s \Vdash T$ we get $s \Vdash \neg\sigma[a]$ and the truncated model $\mathfrak{M}_s$ is a model of $T \cup \{\exists x_0(\varphi(x_0) \wedge \neg\sigma(x_0))\}$. We may then apply Theorem 1. and obtain the model of T omitting $\sum$. □

References

[1] P. H. G. Aczél, *Saturated intuitionistic theories*, Contributions to Mathematical logic, ed H.A. Schmidt, K. Schute and H.J. Thiele, North Holland, Amsterdam, 1968), pp. 1–11.

[2] Kripke, S, *Semantical analysis of intuitionistic Logic I.*, Formal Systems and Recursive Functions (J.N. Crossley and M.A.E. Dummett, eds.), North-Holland Publ. Co., Amsterdam, 1965, pp. 92–130.

[3] Marković, Z., *An intuitionistic omitting types theorem*, Publ. Inst. Math. Belgrade **25 (40)** (1979), 167–169.

[4] Smorynski, C. A, *Applications of Kripke models*, [5], pp. 324-391.

[5] Troelstra, A. S, *Matemathematical Investigations of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics 344, Springer-Verlag, Berlin-Heidelberg, 1973.

Matematički institut, Kneza Mihaila 35, 11001 Beograd, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 809–822
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A CLASSIFICATION OF LOOPS ON AT MOST SIX ELEMENTS

Snežana Matić-Kekić and Dragan M. Acketa

ABSTRACT. Eight kinds of equivalence classes (five of which are new) within the family $L(n)$ of finite loops on n elements ($n \leq 6$) are considered. The classes arise by combining the operations of isotopy over $L(n)$ (with some of its specializations) and loop-parastrophy (parastrophy followed by a special isotopy, which returns the image to $L(n)$).

The used isotopies are triples of permutations of the ground-set (applied successively to rows, columns and elements of the associated Cayley table) which map $L(n)$ onto $L(n)$. Classical isotopy and isomorphic classes correspond to the triples of the form (p,q,r) and (p,p,p) respectively. Three new natural kinds of interclasses, denoted as C-, R- and E-classes, correspond to the triples of the form (q,p,p), (p,q,p) and (p,p,q) respectively. The combinations "isotopy over $L(n)$ + loop-parastrophy" and "isomorphism + loop-parastrophy" lead to the classical main classes and to a new kind of classes, denoted as Π-classes. Finally, a new kind of classes, called parastrophic closures, corresponds to the transitive closure of the loop-parastrophy operator.

Cardinalities, intersections and dualities for all the eight kinds of equivalence classes of loops are completely determined for $n \leq 6$. In addition, the following theorem, related to classical isomorphic, isotopy and main classes, is proved by using the new Π-classes: All the isotopy classes within a main class have the same family of cardinalities of their included isomorphic classes.

1. Introduction

Isotopy classes, isomorphic and main classes belong to the "folklore" of the theory of latin squares and loops. These classes were studied, for example, in [8], [6], [4], [7].

In particular, the figures 9408, 109, 22 and 12 of Table 1. were for the first time correctly determined in the papers [8], respectively [6]. These figures were confirmed by computer in [4]. A systematic tabulation of latin squares on at most six elements and of some their properties was given in [7]. An extensive review of the related results was given in the book [5], Sections 4.2 and 4.3.

In this paper are additionally considered five new ([2]) kinds od equivalence classes of loops: $C-$, $R-$, $E-$, $\Pi-$classes and parastrophic closures. The relationships among all the eight kinds of classes are studied in detail for the case of loops on at most six elements.

Isotopy and isomorphic classes of latin squares correspond to the isotopies determined by three and one permutation of the ground-set. $C-$, $R-$ and $E-$classes correspond to the cases when exactly two among the three permutations determining a loop-preserving isotopy – coincide.

A very small modification (abandoning of fixing the unit) of the algorithm for generating isomorphic classes of loops generates ([1]) $C-$ and $R-$classes. On the other hand, $C-$ and $R-$classes can be further used ([2]) for a construction of isotopy classes.

It is known ([3]) that iterative applications of parastrophic operators *within the class of loops* (to a fixed initial loop) – produce loops belonging to at most six different isomorphic classes. Parastrophic closures are obtained when the arising loops themselves are considered, instead of their isomorphic classes. The upper bound for the cardinality of parastrophic closures with loops of order n is equal ([2]) to $6 \cdot \max \text{g.c.d.}(s_1, \dots, s_k)$, where the maximum is taken over all the partitions $n - 1 = s_1 + \dots + s_k$.

The relationships between $\Pi-$classes and isomorphic classes are completely analogous to the relationships between main classes and isotopy classes.

The inclusion chart of the considered kinds of loop classes has the following outlook:

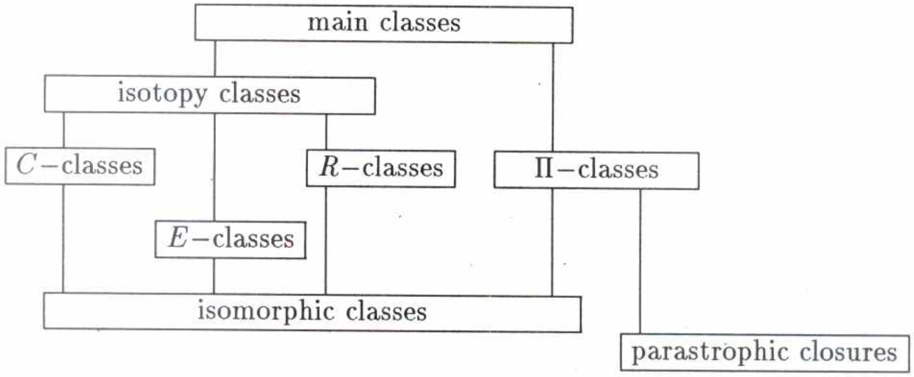


Figure 1.

In Table 1. are given some summary data for $n \leq 6$, which include cardinality of the family $L(n)$ of all loops of order n, as well as the number of all the above defined subclasses of $L(n)$. The figures for the three well-known kinds of classes can be also found in [5]:

n	≤ 3	4	5	6
cardinality of $L(n)$	1	4	56	9408
number of isomorphic classes in $L(n)$	1	2	6	109
number of $E-$classes in $L(n)$	1	2	5	103
number of $C-$ (also number of $R-$) classes in $L(n)$	1	2	3	40
number of isotopy classes in $L(n)$	1	2	2	22
number of $\Pi-$classes in $L(n)$	1	2	4	40
number of main classes in $L(n)$	1	2	2	12
number of parastrophic closures in $L(n)$	1	4	14	832

Table 1.

It would be hard to extend such results to larger values of n, since $|L(7)| = 16.942.080$ ([5]).

The classes were enumerated and analysed with the aid of a PC computer, by using algorithms given in [1]. Most of the running time was spent for the generation of parastrophic closures. This is a consequence of the fact that parastrophic closures are not superclasses of isomorphic classes.

The questions concerning the relationships among the considered classes of loops of order n are obviously trivial for $n \leq 3$. The full description will be given for $n = 6$, while the corresponding data for $n \in \{5, 4\}$ will be briefly listed in the last section.

Isomorphic classes of loops as well as their cardinalities are listed in the Appendix. These classes are basic constituents of all the considered classes of loops except for the parastrophic closures.

2. Definitions and denotations

Let $S(n)$ denote the set $\{1, \dots, n\}$.

A *latin square* of order n is an $n \times n$ matrix A with elements in $S(n)$, which satisfies that there are no two coinciding elements in the same row or in the same column of A.

A *loop* (with unit 1) of order n is a latin square A of order n, which additionally satisfies $A[i, 1] = A[1, i] = i$, for $1 \leq i \leq n$.

Let $L(n)$ denote the family of loops of order n.

We proceed with definitions of eight kinds of equivalence classes over $L(n)$.

Two loops X and Y of order n belong to the same *isotopy class* if there exists an *isotopy* , i.e., a triple $T = (p, q, r)$ of permutations of $S(n)$ satisfying $Y[p(i), q(j)] = r(X[i, j])$, for1 $\leq i, j \leq n$. In particular, if T is of the form

(p,p,p), (q,p,p), (p,q,p) or (p,p,q), then the loops X and Y are respectively said to belong to the same *isomorphic class*, *$C-$class*, *$R-$class* or *$E-$class*.

The *type of an isotopy class* is the family of cardinalities of the included isomorphic classes.

Let r_A and l_A respectively denote the permutations of $S(n)$ which produce the right and the left inverse elements of the loop A (thus $A[i, r_A(i)] = 1$ and $A[l_A(i), i] = 1$ for $i \in S(n)$).

Each loop A has six *loop-parastrophes* $A, \rho(A), \lambda(A), \tau(A), \lambda\tau(A), \rho\tau(A)$, associated to it, where τ is the transposition operator, while the operators ρ and λ have the following meaning (denotations ρ and λ are in accordance with the denotations used in [3]): $\rho(A)[r_A(i), A[i,j]] = j$, and $\lambda(A)[A[i,j], l_A(j)] = i$, for $1 \le i, j \le n$.

Two loops X and Y from $L(n)$ are said to belong to the same *main class* if there exists another loop $Z \in L(n)$, such that X and Z belong to the same isotopy class and Y is a loop-parastroph of Z. In particular, if the word "isotopy" in this definition is replaced by the word "isomorphic", then X and Y are said to belong to the same *$\Pi-$class*.

Two loops X and Y from $L(n)$ are said to belong to the same *parastrophic closure* if there exists a sequence $X = Z_1, Z_2, \dots, Z_k = Y$ of loops from $L(n)$, such that Z_{i+1} is a loop-parastroph of Z_i, for $1 \le i \le k-1$. The parastrophic closure, associated to a loop A, will be denoted by $PC(A)$.

The *order* O of a permutation p is the smallest natural number such that p^O is the identical permutation.

The ordinal numbers of isotopy classes will be followed by the letter "I". The ordinal numbers of $C-$ and $R-$ classes will be usually followed by the letters "C" and "R", respectively. No additional letters will be used with the ordinal numbers of isomorphic classes.

3. $C-$, $R-$, $E-$ and isotopy classes

Given a permutation p of $S(n)$, the permutations q of $S(n)$, such that the isotopies (q,p,p), (p,q,p), and (p,p,q) map $L(n)$ to $L(n)$ – are characterized in [2]. Although the definitions of $C-$, $R-$ and $E-$classes are analogous, it turns out, when consideration is restricted to the loops in $L(n)$, that $E-$classes have a special role.

Namely, isotopies **a)** (q,p,p) **b)** (p,q,p) **c)** (p,p,q) map a loop X from $L(n)$ to another loop in $L(n)$ if and only if ([2]) for $1 \le i \le n$:

a) $q(i) = p(X[i, p^{-1}(1)])$
b) $q(i) = p(X[p^{-1}(1), i])$
c) $q(X[p^{-1}(1), i]) = q(X[i, p^{-1}(1)]) = p(i)$

The *commutator* of a loop $X \in L(n)$ is the set of those elements

$k \in S(n)$, which satisfy that $X[k,j] = X[j,k]$, for each $j \in S(n)$. The number of commutators is ([2]) an invariant of an $E-$class. An abridged search for $E-$classes can be gained by partitioning (representatives of) isomorphic classes w.r.t. this number.

Those $E-$classes, the loops of which have more than one commutator, are listed (by means of their isomorphic subclasses) in the separate fields of 1., 3. and 5. column of Table 2. (the remaining $E-$classes necessarily coincide with isomorphic classes). Each represented $E-$class has in the next column to the right associated an expression of the form $(f(A) \cdot c(A) + \ldots)$, where:

- $c(A)$ is the cardinality of the isomorphic class determined by A
- $f(A)$ is the number of isotopies of the form **c**), which fix the loop A.

2 commutators		3 commutators		6 commutators	
3	(2·120)			1	$(12 \cdot 60)$
50	(2·120)			2	$(12 \cdot 60)$
92	(2·120)	4,79	$(6 \cdot 20 + 6 \cdot 40)$	39	$(120 \cdot 6)$
94	(2·120)	8,83	$(2 \cdot 60 + 2 \cdot 120)$	40,42	$(8 \cdot 60 + 8 \cdot 30)$
103	(2·120)	47,78	$(6 \cdot 20 + 6 \cdot 40)$	43,55	$(4 \cdot 120 + 4 \cdot 60)$
104	(2·120)	54,82	$(2 \cdot 60 + 2 \cdot 120)$	49	$(12 \cdot 60)$

Table 2.

The cardinality of the $E-$class determined by A is equal to

$$\frac{1}{f(A)} \cdot (n-1)! \cdot (\text{number of commutators of } A);$$

the numerator is equal to the number of isotopies of the form **c**).

The next two tables give the intersection and inclusion relationships among isotopy, $C-$, $R-$ and isomorphic classes over $L(6)$.

The denotations in the $x-$th row and the $y-$th column of Table 3. mean that the isomorphic class $10 \cdot x + y$ belongs to the intersection of the $C-$class C and the $R-$class R :

Each $C-$class has a non-empty intersection with each $R-$class within the same isotopy class ([2]). Consequently, each loop isotopy can be represented as a product of two special isotopies within $C-$classes and $R-$classes respectively. Isotopy classes of loops in $L(n)$ can be determined as the unions of those $R-$classes, which have non-empty intersections with the same $C-$class.

A further conclusion is that each $C-$class has at least one common isomorphic class with each $R-$class inside the same isotopy class. E.g., since the isotopy class $10I$ includes three $C-$ and three $R-$classes, it follows that

x	$y=0$	$y=1$	$y=2$	$y=3$	$y=4$	$y=5$	$y=6$	$y=7$	$y=8$	$y=9$
0		1C 1R	2C 2R	2C 2R	3C 3R	4C 4R	4C 5R	5C 4R	5C 5R	6C 3R
1	7C 6R	8C 7R	8C 7R	8C 8R	9C 9R	10C10R	11C 9R	12C 7R	13C11R	13C12R
2	14C13R	15C14R	16C15R	16C15R	14C16R	12C 8R	13C17R	16C15R	11C 9R	15C18R
3	17C16R	13C12R	17C19R	18C14R	14C19R	16C15R	18C18R	17C13R	19C10R	20C20R
4	21C21R	22C16R	23C22R	24C23R	25C24R	22C16R	26C25R	27C26R	28C27R	29C28R
5	24C23R	30C14R	22C13R	28C29R	31C27R	32C30R	26C24R	33C31R	34C 4R	31C29R
6	25C25R	33C32R	27C33R	28C34R	35C35R	36C 9R	22C19R	37C31R	33C36R	35C37R
7	37C36R	24C23R	29C28R	21C21R	32C30R	21C21R	20C20R	23C22R	27C38R	38C 3R
8	39C39R	29C28R	28C40R	40C 4R	24C23R	40C 4R	28C40R	28C34R	34C 4R	30C14R
9	30C18R	40C 5R	24C30R	26C25R	26C24R	24C30R	26C25R	22C16R	22C16R	34C 5R
10	26C25R	33C32R	37C32R	25C25R	32C23R	31C40R	21C22R	23C21R	31C34R	32C23R

Table 3.

the number of included isomorphic classes cannot be smaller than 9. This number is actually equal to 12; each one of the isomorphic classes 41, 45, 97 and 98 is included into the intersection of the classes $22C$ and $16R$.

Each row of Table 4. contains in order the ordinal number of an isotopy class, the included C–classes, the included R–classes and the set of included isomorphic classes:

$1I = 1C = 1R =$	$\{1\}$
$2I = 2C = 2R =$	$\{2,3\}$
$3I = 3C + 6C + 38C = 3R =$	$\{4,9,79\}$
$4I = 4C + 5C + 34C + 40C = 4R + 5R =$	$\{5,6,7,8,58,83,85,88,91,99\}$
$5I = 7C = 6R =$	$\{10\}$
$6I = 8C + 12C = 7R + 8R =$	$\{11,12,13,17,25\}$
$7I = 9C + 11C + 36C = 9R =$	$\{14,16,28,65\}$
$8I = 10C + 19C = 10R =$	$\{15,38\}$
$9I = 13C = 11R + 12R + 17R =$	$\{18,19,26,31\}$
$10I = 14C + 17C + 22C = 13R + 16R + 19R =$	$\{20,24,30,32,34,37,41,45,52,66,97,98\}$
$11I = 15C + 18C + 30C = 14R + 18R =$	$\{21,29,33,36,51,89,90\}$
$12I = 16C = 15R =$	$\{22,23,27,35\}$
$13I = 20C = 20R =$	$\{39,76\}$
$14I = 21C + 23C = 21R + 22R =$	$\{40,42,73,75,77,106,107\}$
$15I = 24C + 32C = 23R + 30R =$	$\{43,50,55,71,74,84,92,95,104,109\}$
$16I = 25C + 26C = 24R + 25R =$	$\{44,46,56,60,93,94,96,100,103\}$
$17I = 27C = 26R + 33R + 38R =$	$\{47,62,78\}$
$18I = 28C + 31C = 27R + 29R + 34R + 40R =$	$\{48,53,54,59,63,82,86,87,105,108\}$
$19I = 29C = 28R =$	$\{49,72,81\}$
$20I = 33C + 37C = 31R + 32R + 36R =$	$\{57,61,67,68,70,101,102\}$
$21I = 35C = 35R + 37R =$	$\{64,69\}$
$22I = 39C = 39R =$	$\{80\}$

Table 4.

4. $\Pi-$classes and main classes

$\Pi-$classes play a central role among the classes in Figure 1. They can be used for establishing a relationship among the well-known kinds of classes (isotopy, isomorphic and main):

Theorem 1. *All the isotopy classes within the same main class have the same type.*

The proof is based on the intermmediate notion of $\Pi-$class. It easily follows from the following three lemmas:

Lemma 1. *Each $\Pi-$class and each isotopy class within the same main class have non-empty intersection.*

Proof. Suppose that a main class contains a $\Pi-$class Π and an isotopy class IT s.t $\Pi \cap IT = \emptyset$. If $L_1 \in \Pi$ and $L_2 \in IT$, then by definition of main class, there exists an isotopy i and a loop-parastrophy π satisfying $L_1 = \pi i L_2$. Thus the loop iL_2 belongs to the classes IT and Π, contradicting $\Pi \cap IT = \emptyset$. □

Lemma 2. *Isomorphic classes within a $\Pi-$class have the same cardinality.*

Proof. Consider two isomorphic classes IM_1 and IM_2 within the same $\Pi-$class. Let $L \in IM_1$ and π be a loop-parastrophy satisfying $\pi(L) \in IM_2$. The function π maps IM_1 to IM_2 since $\pi i L = i \pi L \in IM_2$ for arbitrary $iL \in IM_1$. The operator π is expressed by means of the operators λ, ρ and τ. Since all these operators are involutive [2], it follows that there exists the inverse function π^{-1}. This implies that the function π is a bijection between IM_1 and IM_2. □

Lemma 3. *The intersections of a $\Pi-$class with distinct isotopy classes from the same main class – have the same number of included isomorphic classes.*

Proof. Analogously to the proof of previous lemma, one primarily proves that the intersections of isotopy classes with the same $\Pi-$class have the same cardinality (the proof remains valid when the isomorphism i is replaced by the isotopy). The application of Lemma 2 to the equicardinal intersections completes the proof. □

Proof of Theorem 1. Lemmae 1, 2 and 3 give that the intersections of two isotopy classes with each $\Pi-$class within a main class – consist of the same number of equicardinal isomorphic subclasses. □

It turns out that each two isotopy classes, taken from any two distinct main classes over $L(6)$ have different types (such a conclusion need not be valid for larger ground-sets). Therefore, main classes over $L(6)$ can be reconstructed by use of the relationships between isotopy and isomorphic classes.

According to the following Table 5., the isotopy classes of loops on 6 elements can be collected into 12 wholes (denoted by I,II,... ,XII) w.r.t. the type. The families of cardinalities of the included isomorphic classes are given in the third column of the table (e.g., the family $\{60, 60, 120, 120\}$ is written as $2 \cdot 60 + 2 \cdot 120$).

I	$1I$	$1 \cdot 60$
II	$2I$	$1 \cdot 60 + 1 \cdot 120$
III	$5I$	$1 \cdot 20$
IV	$6I$	$1 \cdot 60 + 4 \cdot 120$
V	$10I$	$4 \cdot 30 + 8 \cdot 120$
VI	$16I$	$9 \cdot 120$

VII	$22I$	$1 \cdot 40$
VIII	$3I,\ 17I,\ 19I$	$1 \cdot 20 + 1 \cdot 40 + 1 \cdot 60$
IX	$4I,\ 15I,\ 18I$	$2 \cdot 60 + 8 \cdot 120$
X	$7I,\ 9I,\ 12I$	$2 \cdot 60 + 2 \cdot 120$
XI	$8I,\ 13I,\ 21I$	$1 \cdot 6 + 1 \cdot 30$
XII	$11I,\ 14I,\ 20I$	$2 \cdot 30 + 2 \cdot 60 + 3 \cdot 120$

Table 5.

It follows from Theorem 1 and Table 5. that there are at least 12 main classes on 6 elements. The data from [5] confirm that each one of the 12 registered candidates is itself a main class. The same conclusion can be derived from Table 6; there are only 12 different collections of isotopy classes which have non-empty intersections with a $\Pi-$ class.

5. Duality

Loops L and $\tau(L)$ are said to be *dual* to each other. Two isotopy (isomorphic) classes are dual whenever they contain two mutually dual representatives. It easily follows from the definition that the dual of a $C-$class is an $R-$class within the same isotopy class, and conversely. On the other hand, $\Pi-$classes and main classes contain complete pairs of mutually dual isomorphic classes, since the duality operator is a special kind of a loop-parastroph operator.

Duality operator will be denoted by $\sim$; the denotation $\sim$ between two equicardinal sets of classes means that the underlying bipartite matching of mutually dual classes is not yet decided exactly.

The mutually dual pairs of isomorphic classes are (4, 47), (5, 53), (6, 59), (7, 48), (8, 54), (9, 62), (13,17), (14, 18), (15, 64), (16, 19), (21, 57),
(24, 52), (26, 65), (28,31), (29, 67), (30,66), (33, 68), (34, 37), (36, 70), (38, 69),
(51, 61), (56,60), (58, 63), (78, 79), (82, 83), (85, 86), (87, 88), (89, 101), (90, 102), (91, 105), (92, 104), (93, 96), (94, 103), (95, 109), (97, 98),
(99, 108), (106, 107), while the remaining 35 isomorphic classes are self-dual.

An abridged way to recognize duality of isomorphic classes would be to use dualities between $C-$ and $R-$, as well as between isotopy classes. Necessary data can be found in Tables 3, 4 and 9.

E.g., XII main class contains isotopy classes $11I$, $14I$ and $20I$. Consulting the numbers of included $C-$classes and $R-$classes, we conclude that $11I \sim 20I$ and that the isotopy class $14I$ is self-dual.

Let the class $14I$ be represented similarly as in Table 4. In addition, the isomorphic classes, as well as their cardinalities (in ( ) brackets) are listed in [ ] brackets after the corresponding $C(R)-$class:

$$\begin{aligned} 14I &= 21C[40(60), 73(120), 75(60), 106(120)] \\ &+ \quad 23C[42(30), 77(30), 107(120)] \\ &= 21R[40(60), 73(120), 75(60), 107(120)] \\ &+ \quad 22R[42(30), 77(30), 106(120)] \end{aligned}$$

Comparing the cardinalities of isomorphic classes included in distinct $C-$ and $R-$classes, it follows that $21C \sim 21R$ and $23C \sim 22R$. This implies (using also the cardinalities of isomorphic classes) that $\{40, 75\} \sim \{40, 75\}$, $\{42, 77\} \sim \{42, 77\}$ and $106 \sim 107$, which further gives that the isomorphic class 73 is self-dual.

It might be interesting to note that among the only six E-classes, which consist of two isomorphic classes each, there are two pairs of mutually dual $E-$classes: $(4, 79) \sim (47, 78)$ and $(8, 83) \sim (54, 82)$.

6. Parastrophic closures and $\Pi-$classes

Let r_A denote the permutation which produces the right inverse element of a loop A ($A[i, r_A(i)] = 1$ for each $i \in S(n)$). It can be proved that:

Theorem 2. *It is satisfed for each loop A from $L(n)$ that:*
$|PC(A)| \leq 6 \cdot order(r_A) \leq 6 \cdot max\ g.c.d.(s_1, \dots, s_k)$,
where the maximum is taken over all the partitions $n - 1 = s_1 + \dots + s_k$.

This is an analogue[1] to a statement ([3]) which claims that $PC(A)$ has non-empty intersections with at most six isomorphic classes for each loop A. Each loop from $PC(A)$ can be obtained from A by an application of transformations of the form $\lambda\rho\lambda\rho\dots$, when the order of r_A is odd, respectively of the form $\lambda\rho\lambda\rho\dots$ or $\tau\lambda\rho\lambda\rho\dots$, when the order of r_A is even.

Among all the 9408 loops in $L(6)$, only 5650 reach the above upper bound $6\cdot$ order(r_A) for $|PC(A)|$. More precisely, the bound is reached with all those loops $A \in L(6)$, which satisfy that $|PC(A)| > 12$, and only with 150 loops with smaller $|PC(A)|$ (120 with $|PC(A)| = 12$ and 30 with $|PC(A)| = 6$). We conjecture that $PC(A) = 6 \cdot$ order(r_A) whenever $|PC(A)| > 12$.

[1] when non-isomorphic loops are replaced by non-identical loops

The loops $A \in L(6)$ with $|PC(A)| = 10$ seem to be particularly interesting. All of them have order$(r_A) = 5$. In addition, 10 is the largest length that we know of a minimal $\lambda\rho\lambda\rho\ldots$ cycle which maps A to A, which is less than the theoretical maximum.

On the basis of tests with random loops, we conjecture that the length of the parastrophic closure of a loop $A \in L(n)$ for a larger n almost always coincides with $6 \cdot$ order(r_A). The minimal value of $|PC(A)|$ is, however, equal to 1 for each n (e.g., when A is the multiplication table of the cyclic group).

The following lemma claims that the above considerations may be raised to the level of $\Pi-$classes:

Lemma 4. *Parastrophic closures within a* $\Pi-$*class have the same cardinality.*

Proof. Let PC_1 and PC_2 denote two parastrophic closures within a $\Pi-$class. There exist two isomorphic loops L_1 and L_2 belonging to PC_1 and PC_2 respectively.

A parastrophic closure is determined by its any incident loop. Using commutative diagrams which connect isomorphism and loop-parastrophy operators, one easily concludes that the parastrophic closuures corresponding to L_1 and L_2 have the same cardinality. □

The first two columns of the following Table 6. contain the cardinality of parastrophic closures and the total number of parastrophic closures within $L(6)$ of a fixed cardinality. The denotation $X : Y$ is associated to the isomorphic class X, which is included into the isotopy class Y. $\Pi-$classes correspond to the ( ) brackets. The number of parastrophic closures within each $\Pi-$class[2] is given in [ ] brackets after ( ) brackets:

In particular, Table 6, can be used for an illustration of Theorem 1. For example, data from Table 6 give the structure of isomorphic classes within XII main class, distributed w.r.t. isotopy classes and Π-classes, given in Table 7. Note that the type of isotopy classes within XII main class is $2 \cdot 30 + 2 \cdot 60 + 3 \cdot 120$ (this can be also found in Table 5).

The second, the third and the fourth row of Table 7 correspond to isotopy classes, while all the columns, except for the first, correspond to Π-classes. For each Π-class are given three additional data. The cardinalities and the number of the included parastrophic closures are given in the 5th and the 6-th row of the table respectively. On the other hand, the first row of the table contains the cardinalities of the included isomorphic classes (taken from

[2] all the parastrophic closures within a $\Pi-$class have equicardinal intersections with all the isomorphic classes within the same $\Pi-$class

1)	[60]	$(1:1I)$	[60]		
2)	[40]	$(2:2I)$	[30]	$(10:5I)$	[10]
3)	[96]	$(9:3I,\ 49:19I,\ 62:17I)$	[60]	$(15:8I,\ 39:13I,\ 64:21I)$	[6]
		$(36:11I,\ 42:14I,\ 70:20I)$	[30]		
4)	[25]	$(25:6I)$	[15]	$(80:22I)$	[10]
6)	[240]	$(4:3I,\ 47:17I,\ 72:19I)$	[10]	$(14:7I,\ 18:9I,\ 22:12I)$	[30]
		$(8:4I,\ 54:18I,\ 55:15I)$	[30]	$(21:11I,\ 40:14I,\ 57:20I)$	[30]
		$(29:11I,\ 67:20I,\ 77:14I)$	[15]	$(32:10I,\ 34:10I,\ 37:10I)$	[15]
		$(33:11I,\ 68:20I,\ 75:14I)$	[30]	$(38:8I,\ 69:21I,\ 76:13I)$	[15]
		$(43:15I,\ 58:4I,\ 63:18I)$	[60]	$(20:10I)$	[5]
8)	[30]	$(3:2I)$	[15]	$(46:16I)$	[15]
10)	[48]	$(93:16I,\ 96:16I)$	[24]	$(97:10I,\ 98:10I)$	[24]
12)	[50]	$(6:4I,\ 59:18I,\ 74:15I)$	[15]	$(78:17I,\ 79:3I,\ 81:19I)$	[10]
		$(28:7I,\ 31:9I,\ 35:12I)$	[15]	$(12:6I)$	[10]
18)	[80]	$(7:4I,\ 48:18I,\ 71:15I)$	[20]	$(11:6I,\ 13:6I,\ 17:6I)$	[20]
		$(16:7I,\ 19:9I,\ 23:12I)$	[20]	$(24:10I,\ 41:10I,\ 52:10I)$	[20]
24)	[75]	$(5:4I,\ 50:15I,\ 53:18I)$	[15]	$(30:10I,\ 45:10I,\ 66:10I)$	[15]
		$(26:9I,\ 27:12I,\ 65:7I)$	[15]	$(44:16I,\ 56:16I,\ 60:16I)$	[15]
		$(51:11I,\ 61:20I,\ 73:14I)$	[15]		
30)	[48]	$(85:4I,\ 86:18I,\ 91:4I,\ 95:15I,\ 105:18I,\ 109:15I)$			[24]
		$(89:11I,\ 90:11I,\ 101:20I,\ 102:20I,\ 106:14I,\ 107:14I)$			[24]
36)	[40]	$(82:18I,\ 83:4I,\ 84:15I)$	[10]	$(94:16I,\ 100:16I,\ 103:16I)$	[10]
		$(87:18I,\ 88:4I,\ 92:15I,\ 99:4I,\ 104:15I,\ 108:18I)$			[20]

Table 6.

	30	60	30	60	120	120
$11I$	36	21	29	33	51	89, 90
$14I$	42	40	77	75	73	106, 107
$20I$	70	57	67	68	61	101, 102
	3)	6)	6)	6)	24)	30)
	[30]	[30]	[15]	[30]	[15]	[24]

Table 7.

Table 9 of Appendix). For example, the last column of the table corresponds to a Π-class having $3 \cdot 2 \cdot 120 = 30 \cdot 24 = 720$ loops.

7. Classifications on 5 and 4 elements

In this section are given the corresponding classifications of loops on 5 and 4 elements. Denotations in the tables are completely analogous to those on 6 elements, with the additional denotations ' and " for loops on 5 and 4 elements respectively.

$$n=5$$

Table 2'

2 commutators		2 commutators	
1', 2'	$(3\cdot 8+3\cdot 8)$	6'	$(20\cdot 6)$

Table 3'

$y=1$	$y=2$	$y=3$	$y=4$	$y=5$	$y=6$
$1R$':$1C$'	$2R$':$2C$'	$1R$':$2C$'	$2R$':$1C$'	$2R$':$1C$'	$3R$':$3C$'

Table 4'

$1I' = 1C' + 2C' = 1R' + 2R' = \{1',2',3',4',5'\}$

$2I' = 3C' = 3R' = \{6'\}$

Table 5'

I'	$1I'$	$1\cdot 2+3\cdot 8+1\cdot 24$

II'	$2I'$	$1\cdot 6$

Table 6'

1)	[6]	(6':$2I'$)	[6]
2)	[1]	(3':$1I'$)	[1]
6)	[4]	(1':$1I'$, 2':$1I'$, 4':$1I'$)	[4]
8)	[3]	(5':$1I'$)	[3]

$$n=4$$

Table 2"

4 commutators			
1"	$(24\cdot 1)$	2"	$(8\cdot 3)$

Table 3"

$y=1$	$y=2$
$1R$":$1C$"	$2R$":$2C$"

Table 4"

$1I" = 1C" = 1R" = \{1"\}$

$2I" = 2C" = 2R" = \{2"\}$

Table 5"

I"	$1I"$	$1\cdot 1$

II"	$2I"$	$1\cdot 3$

Table 6"

1)	[4]	(1":$1I"$)[1]	(2":$2I"$)[3]

References

[1] D.M. Acketa, S. Matić-Kekić, *An algorithm for generating finite loops, some of their subclasses and parastrophic closures*, Proccedings of the 13th International Conference on Information Technology Interfaces (ITI-91), pp. 549–554.

[2] D.M. Acketa, S. Matić-Kekić, *On some classes of finite loops*, Publications de l'Institut mathématique, Beograd, Nouvelle série, tome 55 **69** (1994), 1–8.

[3] V.D. Belousov, *Osnovi teorii kvazigrup i lup*, "Nauka", Moskva, 1967.

[4] B.F. Bryant, H. Schneider, *Principal loop isotopes of quasigroups*, Canad. J. Math. **18** (1966), 120–125.

[5] J. Dénes, A.D. Keedwell, *Latin squares and their applications*, (Akadémiai Kiadó, Budapest / English Universities Press, London / Academic Press, New York), 1974.

[6] R.A. Fisher, F. Yates, *The* 6 × 6 *latin squares, Proc.*, Camb. Phil. Soc. **30** (1934), 492–507.

[7] A. Sade, *Morphismes de quasigroupes. Tables.*, Univ. Lisboa Revista Fac. Ci. A (2) **13** (1970/71), no. 2, 149–172.

[8] E. Schönhardt, *Über lateinische Quadrate und Unionen*, J. Reine Angew. Math. **163** (1930), 183–229.

Appendix

The representatives of the 109 isomorphic classes within $L(6)$ are given in Table 8. Each one of these loops is represented by a 16-digit sequence; the four consecutive quadruples of the sequence contain the middle four elements of the 2nd, 3th, 4th and 5th row of the loop respectively:

1 = 1436456136526123	2 = 1436456136526124	3 = 1436456136526213
4 = 1436456156126123	5 = 1436456156236214	6 = 1436456161523624
7 = 1436456256136124	8 = 1436456256216213	9 = 1436456262513614
10 = 1436516262513614	11 = 1436516262513624	12 = 1436516262514613
13 = 1436516262514623	14 = 1436516262534621	15 = 1436516265134621
16 = 1436516462514623	17 = 1436516462533612	18 = 1436516462533621
19 = 1436516462534612	20 = 1436516465123621	21 = 1436516465124623
22 = 1436516465213612	23 = 1436516465233612	24 = 1436516465234612
25 = 1436526161523624	26 = 1436526161524623	27 = 1436526161533624
28 = 1436526165233614	29 = 1436526165234612	30 = 1436526461524613
31 = 1436526461533612	32 = 1436526465213612	33 = 1436526465214613
34 = 1436561261533264	35 = 1436561262533164	36 = 1436561461524263
37 = 1436562162533164	38 = 1436562462514163	39 = 1456416256136231
40 = 1456416256136234	41 = 1456416265233614	42 = 1456426156236132
43 = 1456426156236134	44 = 1456426165233612	45 = 1456426165233614
46 = 1456426165323614	47 = 1456456136126123	48 = 1456456136126234
49 = 1456456156326123	50 = 1456456156326213	51 = 1456456161323624
52 = 1456456162133624	53 = 1456456236216134	54 = 1456456236216213
55 = 1456456256316213	56 = 1456456261233614	57 = 1456456262133621
58 = 1456456262313614	59 = 1456461235616234	60 = 1456461265233164
61 = 1456461265313264	62 = 1456462135626134	63 = 1456462165323164
64 = 1456516462133621	65 = 1456526161323624	66 = 1456526165323614
67 = 1456526461233612	68 = 1456526465313612	69 = 1456561265313264
70 = 1456562165323164	71 = 1456612435614632	72 = 1456612435624631
73 = 1456621435614623	74 = 1456621435614632	75 = 1456651431624623
76 = 1456651432614623	77 = 1456652431624613	78 = 3156126456216432
79 = 3156126456236412	80 = 3156126465214632	81 = 3156126465234612
82 = 3156146256216234	83 = 3156146256236214	84 = 3156146265314623
85 = 3156426156236412	86 = 3156426156326413	87 = 3156426165231634
88 = 3156426165321624	89 = 3156456116236234	90 = 3156456116236432
91 = 3156456156236412	92 = 3156456256216413	93 = 3156456262311624
94 = 3156461215636234	95 = 3156461265231264	96 = 3156462115626234
97 = 3156462115636234	98 = 3156462115636432	99 = 3156462165321264
100 = 3156462165321463	101 = 3416156256216234	102 = 3416156456236132
103 = 3416156461524623	104 = 3416456216536124	105 = 3416456256316124
106 = 3416456261531624	107 = 3416526461521623	108 = 3416562115626234
109 = 3416562165321264		

Table 8.

Table 9. gives the number of loops within distinct isomorphic classes of $L(6)$. The set of (labels of) isomorphic classes which have cardinality c is denoted by "S_c".

$$\begin{aligned}
S_6 &= \{15, 39, 64\} \qquad S_{20} = \{4, 10, 47, 72\}\\
S_{30} &= \{20, 29, 32, 34, 36, 37, 38, 42, 67, 69, 70, 76, 77\} \qquad S_{40} = \{78, 79, 80, 81\}\\
S_{60} &= \{1, 2, 6, 8, 9, 14, 18, 21, 22, 25, 28, 31, 33, 35, 40, 49, 54, 55, 57, 59, 62, 68, 74, 75\}\\
S_{120} &= \{1, 2, \dots , 109\} - (S_6 \cup S_{20} \cup S_{30} \cup S_{40} \cup S_{60})
\end{aligned}$$

Table 9.

The corresponding tables for $L(5)$ and $L(4)$ are:

Table 8.' $1' = 145451523 \quad 2' = 145452513 \quad 3' = 145512351$
$4' = 145521352 \quad 5' = 315451523 \quad 6' = 345451512$

Table 9.' $S'_2 = \{3'\}, \quad S'_6 = \{6'\}, \quad S'_8 = \{1', 2', 4'\}, \quad S'_{24} = \{5'\}$

Table 8." $1" = 1441 \quad 2" = 1442$
Table 9." $S"_1 = \{1"\}, \quad S"_3 = \{2"\}$

Institute of Mathematics, 21 000 Novi Sad, Trg Dositeja Obradovića 4,
Yugoslavia

FILOMAT (Niš) 9:3 (1995), 823–829
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON DIMENSIONS OF CLASS SPACES

Žarko Mijajlović and Dušan Ćirić

ABSTRACT. In our previous papers, we have introduced the notion of a class space, i.e. topologies on proper classes, and we defined and studied there the main topological concepts on such spaces. In this paper we shall discuss the notion of dimensions of class spaces. Analogues of Ind and ind for class spaces are defined, and their properties are studied.

1. Introduction

In our previous papers [3], [4], we have introduced the notion of a class space, i.e. topologies on proper classes, and explained the reasons for studying so defined spaces. In this paper we shall discuss and study the notion of dimensions of class spaces.

First we shall review some notation. We shall use the notation and definitions introduced in [3], [4]. For example, by capital letters $X, Y, Z, \dots$ we denote classes, and by $x, y, z, \dots$ sets. Greek letters may stand both for classes and for sets. For our metatheory we shall take NBG class theory if not otherwise stated. Further, we shall assume the usual constructions and definitions from set theory and class theory. For example, we remind the reader that a class X is transitive if from $x \in y \in X$ it follows $x \in X$. Throughout the paper K will denote a transitive class. Now we review the axioms for class spaces as we shall often refer to them.

Let K be a class and τ and σ be classes of subsets of K. We call triple $\mathcal{K} = (K, \tau, \sigma)$ a topological class if the following axioms are satisfied:

0. $\emptyset \in \tau, \quad \emptyset \in \sigma$
1. $x, y \in \tau \Rightarrow x \cap y \in \tau$
2. For any i, and $\langle x_j |\ j \in i\rangle$, $(\forall j \in i\ x_j \in \tau) \Rightarrow \cup_j x_j \in \tau$
3. For any $a \in K$ there is $x \in \tau$ such that $a \in x$.
4. $\forall x \in \tau \forall y \in \sigma\ x - y \in \tau$.

1′. $x, y \in \sigma \Rightarrow x \cup y \in \sigma$.

2′. For any i, and $\langle x_j \mid j \in i \rangle$, $(\forall j \in i \ x_j \in \sigma) \Rightarrow \cap_j x_j \in \sigma$

3′. For any subset x of K there is $y \in \sigma$ such that $x \subseteq y$.

4′. $\forall x \in \tau \forall y \in \sigma \ y - x \in \sigma$.

Elements of τ are open subsets while elements of σ are closed subsets of $\mathcal{K}$. The following proposition from [3] states that σ is uniquely determined by τ, and vice versa.

Proposition 1.1. *Let $\mathcal{K} = (K, \tau, \sigma)$ and $\mathcal{K}' = (K, \tau, \sigma')$ be class spaces. Then $\sigma = \sigma'$.*

Various topological notions for class spaces, such as the continuity, the compactness, product of class spaces etc were introduced in our previous papers, and various results concerning these notions were proved. The most important result obtained is that the finite product of compact class spaces is also a compact class space. In this paper we shall discuss and develop the notion of dimensions of topological class spaces.

The functions Ind (Brower-Čech, or large inductive dimension), ind (Menger-Urysohn, or small inductive dimension), dim (see [1]) are the most important dimension functions for topological spaces. By use of these functions we can classify topological spaces according to their dimensions. Let us remind that the notion of the *compartment* plays the main role in the definitions of Ind and ind, while in the definition of dim this role have the notion of covering and the order of covering.

If X is a standard topological space, then B is a compartment between disjoint, closed subspaces P and Q if $X \backslash B = O_1 \cup O_2$, where O_1, O_2 are disjoint open subsets of X which contain P and Q respectively. Every compartment B in X defines a partition of X of the form $X = O_1 \cup B \cup O_2$. Now, suppose $\mathcal{K} = (K, \sigma, \tau)$ is a class space. If K is a proper class, then obviously there is no a partition of K into sets in the above form, neither there is a covering of K by a set-family of sets. Therefore, there is no straightforward way for defining of dimension functions. Our aim is to propose possible definitions of dimension functions on class spaces.

In the following, $\operatorname{cl} x$, $\operatorname{int} x$, $\operatorname{fr} x$, $\operatorname{acc} x$ denote respectively the closure, the interior, the boundary, and the set of accumulation points of a set $x \subseteq X$ in a class space $\mathcal{X}$. If $x \subseteq a \subseteq X$, then these terms in respect of the subspace a are denoted by $\operatorname{cl}_a x$, etc. If not otherwise stated, N denotes the set of non-negative integers.

2. Dimension functions

Let $\mathcal{K} = (K, \tau, \sigma)$ be a topological class space. A set $B \in \sigma$ is a compartment between sets $P, Q \in \sigma$ if there is $U \in \tau$ such that $P \cup Q \cup B \subset U$ and

$U\backslash B = O_1 \cup O_2$, where $O_1, O_2 \in \tau$, $P \subset O_1$, $Q \subset O_2$, and $O_1 \cap O_2 = \emptyset$. Let us notice that the notion of the compartment is well defined. Namely, by the Axiom 3, for every $x \in P \cup Q \cup B$ there is $U_x \in \tau$ such that $x \in U_x$ and $U = \cup_{x \in P \cup Q \in B} U_x \supset P \cup Q \cup B$. As we have $U \in \tau$, $B \in \sigma$, by Axiom 4 it follows $U\backslash B \in \tau$, and also $O_1 \cup O_2 \in \tau$. For the compartment B we shall say that it is a *thin compartment* if it has the empty interior, i.e. $\operatorname{int} B = \emptyset$.

Theorem 2.1. *For every compartment B between sets $P, Q \in \tau$ there is a thin compartment $B' \subset B$.*

Proof. As B is a compartment between P and Q, there is $U \in \tau$ such that $P \cup Q \cup B \subset U$, $U\backslash B = O_1 \cup O_2$, $P \subset O_1$, $Q \subset O_2$, $O_1, O_2 \in \tau$, and $O_1 \cap O_2 = \emptyset$. Let us choose $O_1^* = U\backslash \operatorname{cl} O_2$. As $\operatorname{cl} O_2 \in \tau$ and $U \in \tau$, it follows $O_1^* \in \tau$. Also $O_1^* \supset O_1$ and $O_1^* \cap O_2 = \emptyset$. Now we show that $\operatorname{int}(\operatorname{cl} O_2 \cap B) = \emptyset$. Suppose, in contrary, that $\operatorname{int}(\operatorname{cl} O_2 \cap B) \neq \emptyset$. Then there is $x \in U_x \subset \operatorname{cl} O_2 \cap B$, so $U_x \subset B$ and $U_x \cap O_2 \neq \emptyset$, i.e. $B \cap O_2 \neq \emptyset$, a contradiction. Then $B' = \operatorname{cl} O_2 \cap B$ is a thin compartment between P and Q. □

Theorem 2.2. *Let $\mathcal{K} = (K, \tau, \sigma)$ be a topological class space, $P, Q \in \sigma$ and B a compartment between P and Q. If $X_0 \in \sigma$ is such that $P \cap X_0 \neq \emptyset$ and $Q \cap X_0 \neq \emptyset$, then $B_0 = B \cap X_0$ is a compartment between sets $P_0 = X_0 \cap P$ and $Q_0 = X_0 \cap Q$ in the space X_0 with the topology induced by $\mathcal{K}$.*

Proof. As B is a compartment between P and Q in $\mathcal{K}$, and $X_0 \in \sigma$, we have $B_0 \in \sigma$, and also B_0 is a closed subset of X_0. Further, there is $U \in \tau$ such that $P \cup Q \cup B \subset U$, $U\backslash B = O_1 \cup O_2$, $P \subset O_1$, $Q \subset O_2$, and $O_1 \cap O_2 = \emptyset$. Let $O_1^* = (U \cup U^*) \cap O_1 \cap X_0$ and $O_2^* = (U \cup U^*) \cap O_2 \cap X_0$, where $U^* \in \tau$ such that $X_0 \subset U^*$ exist by the axioms for class spaces. Further, O_1^* and O_2^* are open in X_0 and $X_0\backslash B_0 = O_1^* \cup O_2^*$, thus B_0 is a compartment between P_0 and Q_0. □

Definition 2.3. Let $\mathcal{K} = (K, \tau, \sigma)$ be a topological class space where K is a transitive class. The function $\operatorname{Ind}_\sigma \colon \sigma \to N \cup \{-1\} \cup \{\infty\}$ is defined in the following way for $F \in \sigma$:

$\operatorname{Ind}_\sigma(F) = -1$ if and only if $F = \emptyset$.

Suppose that we have defined values $\operatorname{Ind}_\sigma(F) \leq n - 1$. Then $\operatorname{Ind}_\sigma(F) \leq n$ if for any disjoint and closed sets P, Q in F there is a compartment $B \in \sigma$ between P and Q in K such that $\operatorname{Ind}_\sigma(B) \leq n - 1$.

If $\operatorname{Ind}_\sigma(F) \leq n$ and there is a pair of disjoint and closed subsets of F such that for no compartment B between them in $\mathcal{K}$, $\operatorname{Ind}_\sigma(B) \leq n - 2$ the we shall say that $\operatorname{Ind}_\sigma(F) = n$. If for no $n \geq -1$, $\operatorname{Ind}_\sigma(F) \leq n$, then we put $\operatorname{Ind}_\sigma(F) = \infty$.

By use of Ind_σ we define $\mathrm{Ind}_C: K \to N \cup \{-1, \infty\}$, and $\mathrm{Ind}_C(K)$. For $X \in K$ we put $\mathrm{Ind}_C(X) = -1$ if and only if $X = \emptyset$. Suppose that we have defined values $\mathrm{Ind}_C(X) \leq n-1$, $X \in K$. Then for $X \in K$, $\mathrm{Ind}_C(X) \leq n$ if for any disjoint and closed sets P, Q in space X there is a compartment $B \in \sigma$ between P and Q in K such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Specially, $\mathrm{Ind}_C(X) \leq n$ if for any disjoint sets $P, Q \in \sigma$ there is a compartment $B \in \sigma$ such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Similarly we define $\mathrm{Ind}_C(K) \leq n$. Namely, $\mathrm{Ind}_C(K) \leq n$ iff for all disjoint $P, Q \in \sigma$ there is a compartment $B \in \sigma$ for P and Q such that $\mathrm{Ind}_C(B) \leq n-1$.

If $\mathrm{Ind}_C(X) \leq n$ and if in the space X there is a pair of closed, disjoint sets such that for every compartment B between these sets in $\mathcal{K}$, $\mathrm{Ind}_\sigma(B) \geq n-1$, then we say that $\mathrm{Ind}_C(X) = n$. If for no $n \geq -1$, $\mathrm{Ind}_C(X) \leq n$, then we put $\mathrm{Ind}_C(X) = \infty$. Similarly we define $\mathrm{Ind}_C(K) \leq n$.

Note 2.4 As K is a transitive class then $X \in K$ implies $X \subset K$, so $\mathcal{K}$ inherits a topological structure on X. Thus $\mathrm{Ind}_C(X)$ is well defined. But in general, elements of σ are not the elements of K, so $\mathrm{Ind}_C(F)$ is not necessarily defined for all $F \in \sigma$.

Note 2.5 $\mathrm{Ind}_C(K)$ is well-defined, and we see that $\mathrm{Ind}_C(K)$ is a numerical characteristic of $\mathcal{K}$ in respect to dimensions of elements of K. In this way we avoid the problem of defining of Ind_C on higher order classes (and type theory), at least for transitive topological class spaces.

Theorem 2.6. *Let $\mathcal{K} = (K, \tau, \sigma)$ be a class space. If $F, H \in \sigma \cap K$ and $F \subset H$, then $\mathrm{Ind}_C(F) \leq \mathrm{Ind}_C(H)$. Also, for all $F \in \sigma \cap K$, $\mathrm{Ind}_C(F) \leq \mathrm{Ind}_C(K)$.*

Proof. We prove $\mathrm{Ind}_C(H) \leq n \Rightarrow \mathrm{Ind}_C(F) \leq n$ by induction on n. If n=-1, then $H = \emptyset$, and so $F = \emptyset$, thus the inequality holds for $n = -1$. Suppose the inductive hypothesis for $n-1$, and let $\mathrm{Ind}_C(H) \leq n$. Suppose P and Q are disjoint, closed subsets of F. These sets are disjoint and closed subsets of H as well, so by the inductive hypothesis there is a compartment $B \in \sigma$ for these sets in K such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Then B is obviously a compartment for P and Q in F, thus $\mathrm{Ind}_C(F) \leq n$.

As disjoint and closed subsets of $F \in \sigma \cap K$ are members of σ, it follows $\mathrm{Ind}_C(F) \leq \mathrm{Ind}_C(K)$. □

Corollary 2.7. *If $\sigma \subset K$ then the following assertions are equivalent.*

(a) *For all $F \in \sigma$, $\mathrm{Ind}_C(F) \leq n$.*

(b) $\mathrm{Ind}_C(K) \leq n$.

Proof. The implication (b)⇒(a) follows from the above theorem. Now suppose (a). Then there are disjoint $P, Q \in \sigma$ such that $F = P \cup Q$. Then $F \in \sigma$. By the hypothesis $\mathrm{Ind}_C(F) \leq n$, so there is a compartment $B \in \sigma$ such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Therefore $\mathrm{Ind}_C(K) \leq n$. □

We note also the following statement.

Proposition 2.8. *If $\mathcal{K}$ is a topological class space and* $\mathrm{Ind}_C(\mathcal{K}) < \infty$, *then $\mathcal{K}$ is a normal class space.*

Theorem 2.9. *Let* Ind *be the large inductive dimension of (standard) topological spaces. Then for any class space $\mathcal{K}$ such that $\sigma \subset K$, and $F \in K$,* $\mathrm{Ind}(F) \leq \mathrm{Ind}_C(F)$.

Proof. We shall prove the statement of the theorem by induction on dimension. For $F = \emptyset$ the inequality obviously is true. Suppose the that the inequality holds holds for all natural numbers up to $n - 1$. Suppose $\mathrm{Ind}_C(F) \leq n$ and let P, Q be disjoint closed, subsets of F. As $F \in \sigma$ then $P, Q \in \sigma$, and as $\mathrm{Ind}_C(F) \leq n$ there is a compartment $B \in \sigma$ in $\mathcal{K}$ such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Then by Definition 2.3 it follows $\mathrm{Ind}_C|\sigma \cap K = \mathrm{Ind}_\sigma|\sigma \cap K$, thus $\mathrm{Ind}_C(B) \leq n-1$. Let $B_0 = B \cap F$. As $B_0 \in \sigma$, and σ is a subclass of K , by the inductive hypothesis it follows $\mathrm{Ind}(B) \leq \mathrm{Ind}_C(B)$. The dimension function Ind is monotonous on closed subsets, so $\mathrm{Ind}(B_0) \leq n - 1$, as by Theorem 2.2 for the compartment B_0 between closed sets P and Q in F we have $\mathrm{Ind}(F) \leq n$. □

Definition 2.10. Let K be a transitive class and $\mathcal{K} = (K, \tau, \sigma)$ be a class space. The function $\mathrm{ind}_\sigma \colon \sigma \to N \cup \{-1, \infty\}$ is defined in the following way: $\mathrm{ind}_\sigma(F) = -1$ if and only if $F = \emptyset$. Suppose that we have defined values $\mathrm{ind}_\sigma(F) \leq n - 1$, $F \in \sigma$. Now, we put $\mathrm{ind}_\sigma(F) \leq n$ if for any point $p \in F$ and any closed $Q \subset F$ such that $p \notin Q$ there is a compartment $B \in \sigma$ in $\mathcal{K}$ between p and Q such that $\mathrm{ind}_\sigma(B) \leq n - 1$. If $\mathrm{ind}_\sigma(F) \leq n$ and if there is $p \in F$ and closed $Q \subset F$ such that $p \notin Q$ so that for all compartment B in $\mathcal{K}$ between p and Q we have $\mathrm{ind}_\sigma(B) \geq n - 1$, then we put $\mathrm{ind}_\sigma(F) = n$. If for no integer $n \geq -1$, $\mathrm{ind}_\sigma(F) \leq n$, then we put $\mathrm{ind}_\sigma(F) = \infty$.

By use of ind_σ we define new dimension function $\mathrm{ind}_C \colon K \to N \cup \{-1, \infty\}$ and the value $\mathrm{ind}_C(K)$ as follows. If $X = \emptyset$ then we put $\mathrm{ind}_C(X) = -1$. Suppose that we have defined values $\mathrm{ind}_C(X) \leq n - 1$, $X \in K$. Then we put $\mathrm{ind}_C(X) \leq n$ if for any point p and any closed subset Q of space X such that $p \notin Q$ there is a compartment $B \in \sigma$ for p and Q in $\mathcal{K}$ such that $\mathrm{ind}_\sigma(B) \leq n - 1$. Similarly we define $\mathrm{ind}_C(K) \leq n$. Namely, $\mathrm{ind}_C(K) \leq n$ iff for any point $p \in K$ and closed $Q \in \sigma$ such that $p \notin Q$ there is a compartment $B \in \sigma$ for p and Q such that $\mathrm{ind}_C(B) \leq n - 1$.

In particular, $\mathrm{ind}_C(K) \leq n$ if for every point $p \in K$ and $Q \in \sigma$, $p \notin Q$, there is a compartment $B \in \sigma$ in $\mathcal{K}$ with $\mathrm{ind}_\sigma(B) \leq n - 1$.

If $\mathrm{ind}_C(X) \leq n$ and if there is $p \in X$ and closed $Q \subset X$ such that $p \notin Q$ so that for all compartment B in $\mathcal{K}$ between p and Q we have $\mathrm{ind}_C(X) \geq n-1$, then we put $\mathrm{ind}_C(X) = n$. If for no integer $n \geq -1$, $\mathrm{ind}_C(X) \leq n$, then we put $\mathrm{ind}_C(X) = \infty$.

Theorem 2.11. *Let $\mathcal{K}$ be a class space and $X, Y \in K$. Then $X \subset Y$ implies $\mathrm{ind}_C(X) \leq \mathrm{ind}_C(Y)$. Also for all $X \in K$, $\mathrm{ind}_C(X) \leq \mathrm{ind}_C(K)$.*

Proof. For $n = -1$, $Y = \emptyset$ implies $X = \emptyset$ so in this case the inequality holds. Suppose the inductive hypothesis, that the inequality holds for $\mathrm{ind}_C(Y) \leq n-1$. Suppose $\mathrm{ind}_C(Y) \leq n$, and p be a point and Q a closed subset of space X. Then there is Q^* in $\mathcal{K}$ such that $Q = Q^* \cap X \subset Q^* \cap Y$ and $p \notin Q^* \cap Y$. Since $\mathrm{ind}_C(Y) \leq n$ there is a compartment $B \in \sigma$ for p and Q^* such that $\mathrm{ind}_\sigma(B) \leq n-1$. Then B is a compartment between p and Q thus $\mathrm{ind}_C(X) \leq n$. In a similar way we prove $\mathrm{ind}_C(X) \leq \mathrm{ind}_C(K)$ for all $X \in K$. □

Corollary 2.12. *The following statements are equivalent:*

(a) *For all $X \in K$, $\mathrm{ind}_C(X) \leq n$.*
(b) $\mathrm{ind}_C(K) \leq n$.

Proof. The part (b)⇒(a) follows from the above theorem. Suppose (a). Let us choose $p \in K$ and $Q \in \sigma$ such that $p \notin Q$. Let us put $X = \{p\} \cup Q$. As $\mathrm{ind}_C(X) \leq n$ and Q is closed in X there is a compartment $B \in \sigma$ such that $\mathrm{ind}_C(B) \leq n-1$, and this means that $\mathrm{ind}_C(K) \leq n$. □

An immediate consequence of the definition of ind_C is the following assertion.

Proposition 2.13. *If* $\mathrm{ind}_C(\mathcal{K}) < \infty$ *then (K) is a regular topological class space.*

Proposition 2.14. *Let $\mathcal{K}$ be a class space. Then for every $X \in K$ we have* $\mathrm{ind}(X) \leq \mathrm{ind}_C(X)$, *where* $\mathrm{ind}(X)$ *is the small inductive dimension of X.*

Proof. The proof is by induction. This inequality is obviously true for $X = \emptyset$. Suppose the inductive hypothesis, that the inequality holds for all dimensions $\leq n-1$. Suppose $\mathrm{ind}_C(X) \leq n$ and let p be a point and Q a closed subset of X such that $p \notin Q$. As $\mathrm{ind}_C(X) \leq n$, there is a compartment B in $\mathcal{K}$ such that $\mathrm{ind}_\sigma(B) \leq n-1$. From Definition 1.10 it follows that $\mathrm{ind}_C|\sigma = \mathrm{ind}_\sigma$, so $\mathrm{ind}_C(B) \leq n-1$. Further, $B_0 = B \cap X$ is a compartment in X between the point p and the subset Q, and by the inductive hypothesis and Theorem 1.11 we have $\mathrm{ind}_C(B_0) \leq \mathrm{ind}_C(B) \leq n-1$ and $\mathrm{ind}(B_0) \leq \mathrm{ind}_C(B_0)$, hence $\mathrm{ind}(B_0) \leq n-1$. Therefore $\mathrm{ind}(X) \leq n$. □

Theorem 2.15. *Let $\mathcal{K}$ be a topological class space. Then for every $X \in K \cap \sigma$ we have* $\mathrm{ind}_C(X) \leq \mathrm{Ind}_C(X)$. *Specially, if $\mathcal{K}$ is a T_1 class space then* $\mathrm{ind}_C(K) \leq \mathrm{Ind}_C(K)$.

Proof. The proof is by induction. This inequality is obviously true for $X = \emptyset$. Suppose the inductive hypothesis, that the inequality holds for all dimensions $\leq n-1$. Suppose $\mathrm{ind}_C(X) \leq n$ and let p be a point and Q a closed subset of X such that $p \notin Q$. As $\mathcal{K}$ is T_1 class space then $\{p\}$ is a closed subset of X. As $\mathrm{Ind}_C(X) \leq n$ there is a compartment $B \in \sigma$ such that $\mathrm{Ind}_\sigma(B) \leq n-1$. Further, on $K \cap \sigma$ we have $\mathrm{Ind}_\Sigma = \mathrm{Ind}_C$ therefore $\mathrm{Ind}_C(B) \leq n-1$. By the inductive hypothesis $\mathrm{ind}_C(B) \leq \mathrm{Ind}_C(B) \leq n-1$, so $\mathrm{ind}_C(X) \leq n$ □

References

[1] П. С. Александров, Б. А Пасынков, *Введение в теорию Размерности*, "Наука", Москва, 1973.

[2] N. Bourbaki, *Topologie generale*, Hermann, Paris, 1971.

[3] D. Ćirić and Ž. Mijajlović, *Topologies on classes*, Math. Balkanica **777** (1990), 00-00.

[4] D. Ćirić and Ž. Mijajlović, *Class spaces*, Math. Balkanica (1994) (to appear).

[5] R. Engelking, *General topology*, PWN Warszawa, 1977.

[6] A. Fraenkel and Y. Bar-Hillel, *Foundations of set theory*, North-Holland Amsterdam, 1958.

[7] K. Kunen, *Set theory*, North-Holland, Amsterdam, 1983.

[8] A. Levy, *Basic set theory*, Springer-Verlag, Berlin, 1979.

University of Belgrade, Faculty of Science, Depart. of Mathematics, Studentski trg 16, 11000 Belgrade, Yugoslavia

University of Niš, Faculty of Philosophy, Ćirila i Metodija 2, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 831–837
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON THE NUMBER OF EXPANSIONS OF COUNTABLE MODELS OF FIRST ORDER THEORIES

Ž. Mijajlović and I. Farah

ABSTRACT. Let **A** be a countable model of a countable first-order language L, and T be a first-order theory of a countable expansion $L' \supseteq L$. Let $\mathcal{S}$ denote the set of all expansions of **A** to L' that are models of T. It is proved that $\mathcal{S}$ can be embedded into a metric Stone space as a G_δ subset, and therefore $k = |\mathcal{S}|$ satisfies CH, i.e. either $k \leq \aleph_0$ or $k = 2^{\aleph_0}$. Several examples that illustrates this theorem are presented, too.

Works of Kueker [5], Reyes [9], Barwise [1], Makkai [7] and others, show that certain sets of model-theoretic objects related to a countable model **A**, as Aut**A** for example, behave as analytic subsets of the Cantor discontinuum. This property can be proved in several ways, and we shall present here two methods. The first one is based on the coding of model-theoretic objects by real numbers (or characteristic functions of certain subsets of real numbers). The second one is based on the properties of Lindenbaum algebras, and it has more model-theoretic nature.

1. Coding by reals

We shall present this method by example, i.e. we shall illustrate it in the case of the Kueker's theorem:

Theorem 1.1. *Let* **A** *be a countable model of a countable language. Then CH holds for* Aut**A**, *i.e.*

$$|\mathrm{Aut}\mathbf{A}| \leq \aleph_0 \quad or \quad |\mathrm{Aut}\mathbf{A}| = 2^{\aleph_0}.$$

The proof of this theorem that we shall present, is based on the following well known facts:

Supported by Grant 0401A of RFNS through Math. Inst. SANU.

1.1° For Borel subsets of real numbers R, CH holds (M. Suslin), i.e. if $X \subseteq R$ then $|X| \leq \aleph_0$ or $|X| = 2^{\aleph_0}$.

1.2° Cantor's triadic set K is a closed subset of R and it has the cardinality of continuum.

1.3° Suppose X is a countable set. Then 2^X is homeomorphic to the Cantor space K. Here $2 = \{0, 1\}$ has the discrete topology, and 2^X has Tychonoff product topology.

Now we proceed to the proof of Theorem 1.1. For the simplicity of notation, we shall assume that $\mathbf{A}$ is grupoid, i.e. $\mathbf{A} = (A, \cdot)$, where $\cdot$ is a binary relation on domain A. Let $\mathcal{F}$ be the set of all mappings (characteristic functions) $k : A^2 \to 2$ such that:

(1) $\forall a, a', b, b' \in A \ \ (k(a, a') = 1 \wedge k(b, b') = 1 \Rightarrow k(a \cdot b, a' \cdot b') = 1)$,
(2) $\forall a, a', b \in A \ \ (k(a, a') = 1 \wedge b \neq a \Rightarrow k(b, a') = 0)$,
(3) $\forall b \in A \ \ \exists a \in A \ \ k(a, b) = 1$.
(4) $\forall a, b, b' \in A \ \ (k(a, b) = 1 \wedge k(a, b') = 1 \Rightarrow b = b')$

If $f \in \mathrm{Aut}\mathbf{A}$, then let $k_f \colon A^2 \to 2$ be defined by $k_f(a, b) = 1$ iff $b = f(a)$, $a, b \in A$. Then it is not difficult to see that k_f satisfies the properties (1), (2), (3), (4), and that to different automorphisms f and f' correspond different k_f and $k_{f'}$ respectively (for example, if $f(a) = b \neq f'(a)$ then $k_f(a, b) = 1$, while $k_{f'}(a, b) = 0$). On the other hand, if $f \in \mathcal{F}$ and $f \colon A \to A$ is defined by $b = f(a)$ iff $k(a, b) = 1$, $a, b \in A$, then f is a well-defined function and $f \in \mathrm{Aut}\mathbf{A}$. Therefore, if the map ϕ is defined by $\phi \colon f \mapsto k_f$, $f \in \mathrm{Aut}\mathbf{A}$, then $\phi : \mathrm{Aut}\mathbf{A} \xrightarrow[\text{1-1}]{\text{onto}} \mathcal{F}$, thus

(5) $|\mathrm{Aut}\mathbf{A}| = |\mathcal{F}|$.

Further, let $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ be sets of functions $k \colon A^2 \to 2$ that satisfy conditions (1), (2), (3) and (4) respectively. Then:

$$\begin{aligned}
\mathcal{F}_1 = \bigcap_{a,a',b,b' \in A} & (\{k \in 2^{A^2} \mid k(a, a') = 0\} \cup \\
& \{k \in 2^{A^2} \mid k(b, b') = 0\} \cup \{k \in 2^{A^2} \mid k(a \cdot b, a' \cdot b') = 1\}),
\end{aligned}$$

$$\mathcal{F}_2 = \bigcap_{a,a' \in A} (\{k \in 2^{A^2} \mid k(a, a') = 0\} \cup \bigcap_{b \in A, b \neq a} \{k \in 2^{A^2} \mid k(b, a') = 0\}),$$

$$\mathcal{F}_3 = \bigcap_{b \in A} \bigcup_{a \in A} \{k \in 2^{A^2} \mid k(a, b) = 1\}$$

$$\mathcal{F}_4 = \bigcup_{a,b} (\{k \in 2^{A^2} \mid k(a, b) = 0\} \cup \bigcap_{b \in A, b' \neq b} \{k \in 2^{A^2} \mid k(a, b') = 0\}$$

Let 2^{A^2} be the product topology, where 2 has the discrete topology. Then the set $\{k \in 2^{A^2} \mid k(a, b) = \alpha\}, a, b \in A$, $\alpha \in 2$, is a clopen set. Thus, $\mathcal{F}_1$,

$\mathcal{F}_2$ and $\mathcal{F}_4$ are closed, while $\mathcal{F}_3$ is a countable intersection of open sets. As $\mathcal{F} = \mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3 \cap \mathcal{F}_4$, it follows that $\mathcal{F}$ is a G_δ subset of the Cantor space 2^{A^2}, so by Fact 1.1° we have $|\mathcal{F}| \leq \aleph_0$ or $|\mathcal{F}| = 2^{\aleph_0}$. Hence, by (4) the theorem 1.1. is proved. □

2. Countable expansions in first-order logic

Now we shall present a proof of a general theorem, based on the properties of Lindenbaum algebras, that the set of all countable expansions of a countable model $\mathbf{A}$ of a first-order theory T satisfies CH. Let us introduce and review first some notation and terminology. Let L be a first-order language and $\mathbf{A}$ be a model of L. Then For_L denotes the set of all formulas of L, while Sent_L denotes the set of all sentences of L. Further, $L_A = L \cup \{\underline{a} \,|\, a \in A\}$, where $\underline{a}$ is the name of a, and $(A, a)_{a \in A}$ is the simple expansion of $\mathbf{A}$ to a model of L_A. By Th$\mathbf{A}$ we shall denote the set $\{\varphi \in \mathrm{Sent}_L \,|\, \mathbf{A} \models \varphi\}$. The Lindenbaum algebra of T over L is $\mathcal{L}(T, L) = \{[\varphi] \,|\, \varphi \in \mathrm{Sent}_L\}$, where $[\varphi] = \{\psi \in \mathrm{Sent}_L \,|\, T \vdash \varphi \leftrightarrow \psi\}$. The boolean operations $\{\cdot, +\}$ and constants $\{0, 1\}$ in $\mathcal{L}(T, L)$ are defined in the usual way: $[\varphi] \cdot [\psi] = [\varphi \wedge \psi]$, $[\varphi] + [\psi] = [\varphi \vee \psi]$, $[\varphi]' = [\neg\varphi]$, and $1 = [\theta]$, $0 = [\neg\theta]$, where θ is a tautology. In the following, we shall identify T with $\{[\varphi] \,|\, \varphi \in T\}$. If $T = \emptyset$ we shall write simply $\mathcal{L}(L)$ instead of $\mathcal{L}(T, L)$.

If $\mathbf{B}$ is an arbitrary Boolean algebra then $\mathbf{B}^*$ is the Stone space of $\mathbf{B}$, i.e. the set of all ultrafilters of $\mathbf{B}$ with clopen sets $a^* = \{p \in \mathbf{B}^* \,|\, a \in p\}$, $a \in \mathbf{B}$, as a topological basis. Thus the dual space $\mathcal{L}(T, L)^*$ of $\mathcal{L}(T, L)$ is the set of all complete consistent theories of L that extend theory T. We remind the reader that the Cantor space 2^N is the Stone space of the free Boolean algebra Ω_ω with countable many free generators. The dual of an ideal $I \subseteq \mathbf{B}$ is $I^* = \{p \in \mathbf{B}^* \,|\, p \cap I \neq 0\}$. Observe that $I^* = \cup_{a \in I} a^*$ is an open set. For the rest of notation and terminology, we shall follow [3].

Lemma 2.1. *Let* $\mathbf{B}$ *be a countable Boolean algebra. Then* $\mathbf{B}^*$ *can be embedded into* 2^N *as a closed subset.*

Proof. Since Ω_ω is a free Boolean algebra, there is a homomorphism

$$h : \Omega_\omega \xrightarrow{onto} \mathbf{B}.$$

If $I = \ker(h)$, then I is an ideal of Ω_ω, thus $\mathbf{B} \cong \Omega_\omega / I$ and $\mathbf{B}^* \cong \Omega_\omega^* - I^*$. Further, I^* is open, hence $\Omega_\omega^* - I^*$ is closed in 2^N. □

Remark that for above I and $\mathcal{F} = \{p \in \Omega_\omega \,|\, \bigwedge_{a \in p} ha > 0\}$ we have $\mathcal{F} = \Omega_\omega^* - I^*$.

Lemma 2.2. *Let* $S \subseteq 2^N$ *be closed, and* $H \subseteq S$ *be* G_δ *in* S. *Then* H *is* G_δ *in* 2^N.

Proof. There are open subsets V_i, $i \in N$, of S so that $H = \cap_{i \in N} V_i$. Hence there are open $U_i \subseteq 2^N$ such that $V_i = S \cap U_i, i \in N$. Thus $H = S \cap (\cap_i U_i)$, and as S is a countable intersection of open subsets of 2^N, it follows that H is G_δ in 2^N. □

Let L be in the following a first-order language, $L' \supseteq L$ an expansion of L, T a theory of L' and $\mathbf{A}$ an arbitrary model of L. By $\boldsymbol{S}(\mathbf{A}, T)$ we shall denote the set of all expansions $\mathbf{B}$ of the model $\mathbf{A}$ to L' such that $\mathbf{B} \models T$. Finally, let $k(\mathbf{A},T) = |\boldsymbol{S}(\mathbf{A}, T)|$.

Theorem 2.3. *Let* $\mathbf{A}$ *be a countable model of a countable first-order language* L, $L' \supseteq L$ *be a countable expansion, and* T *a consistent theory of* L. *Then the number* $k(\mathbf{A}, T)$ *satisfies CH, i.e. either* $k(\mathbf{A}, T) \leq \aleph_0$ *or* $k(\mathbf{A}, T) = 2^{\aleph_0}$.

Proof. Let $\mathbf{A}'$ be an expansion of $\mathbf{A}$ to L'_A such that $\mathbf{A}' \models T$. Let

$$P = P_{A'} = \{\varphi \in \mathrm{Sent}_{L'_A} \mid (\mathbf{A}', a)_{a \in A} \models \varphi\}$$

Then:

i. $T \cup \mathrm{Th}(\mathbf{A}, a)_{a \in A} \subseteq P$.
ii. If $\exists x \varphi x \in P$ then there is $a \in A$ such that $\varphi \underline{a} \in P$.
iii. P is a complete consistent theory.

Since $T \subseteq P$, by iii. we may assume that P is an ultrafilter of the Lindenbaum algebra $\mathcal{L}(T, L'_A)$.

Claim The correspondence $\Phi\colon \mathbf{A}' \mapsto P'_{\mathbf{A}}$ between expansions $\mathbf{A}'$ of $\mathbf{A}$ to L' such that $\mathbf{A}' \models T$, and ultrafilters of $\mathcal{L}(T, L'_A)$ satisfying conditions i.- iii. is one-to-one and onto.

Proof of Claim Suppose $\mathbf{A}'_1$and $\mathbf{A}'_2$ are different expansions of $\mathbf{A}$, and let P_1 i P_2 be the corresponding sets satisfying conditions i.- iii. Since $\mathbf{A}'_1 \neq \mathbf{A}'_2$ there is, for example, an n-ary relation symbol R of L' such that for some $a_1, a_2, \ldots, a_n \in \mathbf{A}$, $R_{\underline{a}_1} \ldots \underline{a}_n \in P_1$ while $\neg R \underline{a}_1 \ldots \underline{a}_n \in P_2$. Thus $P_1 \neq P_2$, so Φ is 1-1.

Now, let P be any set of sentences satisfying conditions i.-iii. Since P is consistent, there is a model $\mathbf{B}'$ of P, and without loss of generality we may assume $\mathbf{A} \prec \mathbf{B}$, where $\mathbf{B}$ is a reduct of $\mathbf{B}'$ to L. Further, define $\mathbf{A}'$, an expansion of $\mathbf{A}$ by

$$R^{\mathbf{A}'} a_1 a_2 \ldots a_n \quad \text{iff} \quad R \underline{a}_1 \underline{a}_2 \ldots \underline{a}_n \in P \qquad\qquad R \in \mathrm{Rel}_{L'},$$

$$F^{\mathbf{A}'} a_1 a_2 \ldots a_n = b \quad \text{iff} \quad (F \underline{a}_1 \underline{a}_2 \ldots \underline{a}_n = \underline{b}) \in P \qquad F \in \mathrm{Fnc}_{L'},$$

where $\mathrm{Rel}_{L'}$ is the set of all relation symbols of L', and $\mathrm{Fnc}_{L'}$ is the set of all function symbols of L'. The structure $\mathbf{A}'$ is well-defined. For example, if $F \in \mathrm{Fnc}'_L$ and $a_1, a_2 \dots a_n \in A$ then $\exists x(F\underline{a}_1\underline{a}_2 \dots \underline{a}_n = x) \in P$, so by the property ii. there is $b \in A$ such that $(F\underline{a}_1\underline{a}_2 \dots \underline{a}_n = b) \in P$.

Now we shall prove $\mathbf{A}' \prec \mathbf{B}'$. Really, suppose $\mathbf{B}' \models \exists x\varphi(x, \underline{a}_1, \underline{a}_2, \dots \underline{a}_n)$, where $a_1, a_2 \dots a_n \in A$ and $\exists x\varphi(x, \underline{a}_1, \underline{a}_2, \dots \underline{a}_n) \in \mathrm{For}_{L_{A'}}$. Since P is complete, and $\mathbf{B}'$ is a model of P, we have $\exists x\varphi(x, \underline{a}_1, \underline{a}_2, \dots \underline{a}_n) \in P$, so by the property ii. there is $b \in \mathbf{A}$ such that $\varphi(b, \underline{a}_1, \underline{a}_2, \dots \underline{a}_n) \in P$ i.e. $\mathbf{B}' \models \varphi(b, \underline{a}_1, \underline{a}_2, \dots \underline{a}_n)$. Therefore, by Tarski-Vaught's Theorem, it follows $\mathbf{A}' \prec \mathbf{B}'$. Hence $P = P_{A'}$ i.e. Φ is onto, and this proves the claim.

Now, by Lemma 2.1. we may assume that the Stone space $\mathcal{L}(T, L'_A)^*$ of the Lindenbaum algebra $\mathcal{L}(T, L'_A)$ is closed subset of 2^N. Let $\mathcal{Y}$ be the set of all ultrafilters of $\mathcal{L}(T, L'_A)$ satisfying properties i.-iii. Then $\mathcal{Y}$ is the intersection of the following sets:

1. $\mathcal{U} = \{p \in \mathcal{L}(T, L'_A)^* \mid \mathrm{Th}(\mathbf{A}, a)_{a\in A} \subseteq p\} = \bigcap\{[\varphi]^* \mid \varphi \in \mathrm{Th}(\mathbf{A}, a)_{a\in A}\}$
2. $\mathcal{V} = \bigcap\limits_{\exists x\varphi x \in \mathrm{For}_{L'_A}} ([\neg\exists x\varphi x]^* \cup \bigcup\limits_{a\in A} [\varphi\underline{a}]^*)$.

The set $\mathcal{U}$ is obviously closed and $\mathcal{V}$ is G_δ in $\mathcal{L}(T, L'_A)^*$ as a countable intersection of open sets. Observe, for example, that $[\neg\exists x\varphi x]^a st$ is open. Therefore, by Lemma 2.2. $\mathcal{Y}$ is G_δ in 2^N, and so $|\mathcal{Y}|$ satisfies CH. By Claim $|\boldsymbol{S}(\mathbf{A}, T)| = |\mathcal{Y}|$, so the theorem follows. □

In the case of finite expansions, the above theorem is a simple consequence of Perfect Subset Theorem in [7]. However, the presented proof of Theorem 2.3. relies on rather basic model theory, and besides it gives an estimate of the complexity of the set $\boldsymbol{S}(\mathbf{A}, T)$ in the analytic hierarchy (G_δ).

3. Examples

In this part we shall list some examples that are consequences of Theorem 2.3. In the following, **A** is a countable model of a countable language L.

Example 3.1. We revise the Kueker's example from Section 1: if a **A** is a countable model of a countable language L, then $|\mathrm{Aut}\mathbf{A}|$ satisfies CH. Really, let $L(F) = L \cup \{F\}$ where F is an unary function symbol, and T be a theory of $L(F)$ which states that F is an automorphism in respect to symbols of L. If F is a new (i.e. $F \notin L$) unary function symbol, then axioms of T are:

1° $F(G(x_1, x_2, \dots, x_n)) = G(F(x_1), F(x_2), \dots, F(x_n))$, $G \in L$ is an n-ary function symbol.

2° $R(x_1, x_2, \dots, x_n) \Leftrightarrow R(F(x_1), F(x_2), \dots, F(x_n))$, $R \in L$ is an n-ary relation symbol.

3° Axioms which says that F is one-to-one and onto function.

Then obviously there is one-to-one correspondence between expansions of $\mathbf{A}$ to $L \cup \{F\}$ that are models of T and automorphism of $\mathbf{A}$. Therefore $|\mathrm{Aut}\ \mathbf{A}|$ satisfies CH.

Example 3.2. (Burris and Kwatinetz, see [4, p.35]) Let $\mathbf{A}$ be a countable algebra of a countable language. The set of all subalgebras Sub$\mathbf{A}$, the set of all endomorphisms End$\mathbf{A}$ and the set of all congruences Con$\mathbf{A}$ satisfy CH. To prove the first assertion, let U be a new unary relation symbol. Then subalgebras of $\mathbf{A}$ can be described as interpretations $U^{\mathbf{A}}$ in the expansion $(\mathbf{A}, U^{\mathbf{A}})$ which satisfy the conditions:

$c^{\mathbf{A}} \in U^{\mathbf{A}}$, c is a constant symbol of L,
For all $x_1, x_2, \dots, x_n \in U^{\mathbf{A}})\, F(x_1, x_2, \dots, x_n) \in U^{\mathbf{A}}$, $F \in L$ is a function symbol,

i.e. the axioms

$U(c)$, c is a constant symbol of L,
$\forall x_1, x_2, \dots, x_n(U(x_1) \wedge U(x_2) \wedge \dots U(x_n) \Rightarrow U(F(x_1, x_2, \dots, x_n))$,

where F is a function symbol of L.

Other cases are described in a similar way.

Example 3.3. As in the previous example one can find that the set of submodels of $\mathbf{A}$ (or countable sequences of submodels) that satisfy certain first order properties, also satisfies CH. For example, with the same notation as in the previous example, the set of all elementary submodels $\mathcal{E}(\mathbf{A})$ of $\mathbf{A}$ are described with following sentences:

1° Axioms for Sub$\mathbf{A}$,
2° $(\forall x_1, x_2, \dots, x_n \in U)(\exists y\varphi \Rightarrow \exists y \in U\varphi)$, or more formally
$(\forall x_1, x_2, \dots, x_n)((U(x_1) \wedge U(x_2) \wedge \dots \wedge U(x_n)) \Rightarrow (\exists y\varphi \Rightarrow \exists y(U(y) \wedge \varphi)))$

By Tarski-Vaught theorem then easily follows that $U^{\mathbf{A}} \prec \mathbf{A}$ iff $(\mathbf{A}, U^{\mathbf{A}})$ satisfies the listed axioms.

Example 3.4. The set of all prime ideals of a countable commutative ring also satisfies CH. In other words, the Zariski space of a countable commutative ring satisfies CH. To see this observe that "I is a ring ideal" is a first-order property. It is described by universal closures of the following formulas in the language of rings $L = \{+, \cdot, 0\}$ with added unary predicate I which represent an ideal:

$$I(0),\ I(x) \wedge I(y) \Rightarrow I(x+y),\ I(x) \Rightarrow I(x \cdot y),\ I(xy) \Rightarrow I(x) \vee I(y).$$

In a similar way one can show that CH holds for the set of all maximal ideals of a countable ring.

Example 3.5. Let $\mathbf{P} = (P, \leq_P)$ be a countable, partially ordered set, $L = \{\leq\}$ and $L' = \{\leq, \preceq\}$. Taking for T the set of axioms of the linear ordering for $\preceq$ extending $\leq$, we find that the number of linear extensions of $\mathbf{P}$ satisfies CH. It is easy to design for each $0 < k \leq \aleph_0$ or $k = 2^{\aleph_0}$ a partially ordered set P which has exactly k linear extensions.

Some other families of subsets of $\mathbf{P}$ for which CH holds includes the set of all (maximal) chains, the set of all (maximal) antichains, and the set of all dense subsets of $\mathbf{P}$.

Example 3.6. If $\mathbf{A} = (A, G)$ is a planar graph, then by simple compactness argument one can show that $\mathbf{A}$ can be 4-colored, i.e. to elements of A can be assigned four colors so that the adjacent vertices are in different colors (assuming that the Four-coloring theorem for finite planar graphs is true). If $\mathbf{A}$ is infinite countable, let $a, b, c, d \in A$ be four distinct elements. Then every map $f \colon A \to \{a, b, c, d\}$ defines a coloring of $\mathbf{A}$. It is not difficult to write down first-order axioms which describes colorings of the above type. Thus, all 4-colorings of a countable planar graph satisfy CH.

References

[1] J. Barwise, *Admissible Sets And Structure*, Springer, Berlin, 1977.

[2] C. C. Chang, H. J. Keisler, *Model Theory*, North-Holland, Amsterdam, 1973.

[3] H. J. Keisler, *Model Theory For Infinitary Logic*, North-Holland, Amsterdam, 1971.

[4] R. N. McKenzie, G. F. McNulty, W. F. Tayler, *Algebras, Lattices, Varieties, Vol I*, Brooks/Cole Publ. Co., 1987.

[5] D. Kueker, *Definability, automorphisms and infinitary languages*, The Syntax and Semantics of Infinitary Logic edited by J Barwise, Lecture Notes in Mathematics, Vol. 72, Springer, Berlin, 1968, pp. 152-165..

[6] K. Kuratowski and A. Mostowski, *Set Theory*, PWN–Polish Scientific Publishers, Warsaw, 1967.

[7] M. Makkai,, *Admissible sets and infinitary logic*, Handbook of Mathematical Logic, ed J.Barwise, North-Holland, Amsterdam, 1977, pp. 233-282.

[8] Y. N. Moschovakis, *Descriptive Set Theory*, North-Holland, Amsterdam, 1980.

[9] G. E. Reyes, *Local definability theory*, Ann. Math. Logic **1** (1970), 95-137.

[10] S. Shelah, *Classification theory and the number of non-isomorphic models*, North-Holland, Amsterdam, 1978.

[11] S. Todorčević, I. Farah, *Some Application of the Method of Forcing*, Matematički Institut, Beograd, 1995.

University of Belgrade, Faculty of Sciences, Department of Mathematics,
Studentski trg 16, P.B.550, 11000 Belgrade, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 839–849
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A SET OF AXIOMS FOR EVALUATING THE MULTIPROCESSOR PERFORMANCES

I. Ž. Milovanović, E. I. Milovanović,
M. D. Mihajlović and M. K. Stojčev

Abstract. When designing a parallel computer it is very important that it has the predicted performances. The challenge for a computer designer is to discover the minimum organization and equipement necessary to achieve given level of performance. So, the developing analythical model for characterizing and understanding the parallel system performances is of a crucial interest. In order to avoid erroneous conclusions about the behaviour of parallel system a severe mathematical formulations should be involved. In this paper we give a survey of axioms that were proposed in the literature in order to introduce a scientific approach in studing the parallel system performances. Further we shall propose a modified and reduced set of axioms based on discrete mathematical apparatus.

1. Introduction

From the very beginning of digital computer development, the designers always storve to increase the speed of operations. There are number of possible ways to achieve this. An obvious approach is to improve the technology implemented in the realization of the computer components. There is of course a natural limitation in technology development: no signal can propagate faster than the speed of the light. Another way for increasing the speed of computation is by performing as many operations as possible simultaneosly, concurrently, in parallel, using parallel computers [8].

A parallel computer is one that consists of a collection of processing units, or processors, that cooperate to solve a problem by working simultaneously on different parts of that problem. The number of processors used can range from a few tens to several millions. As a result, the time required to solve a problem by a traditional uniprocessor computer is significantly reduced.

This approach is attractive for a number of reasons [1]. First, for many computational problems, the natural solution is a parallel one. Second, the cost and size of computer components have declined so sharply in recent years that parallel computers with a large number of processors have become feasible. And, third, it is possible in parallel processing to select the parallel architecture that is best suited to solve the problem or class of problems under consideration. Indeed, architects of parallel computers have freedom to decide how many processors are to be used, how powerful these should be, what interconnection network links them to one another, whether they share a common memory, to what extent their operations are to be carried out synchronously, and host of other issues. This wide range of choices has been reflected by many theoretical models of parallel computation proposed as well as by several parallel computers that were actually built. Since parallel computers are composed of multiple processors, interconnected to each other, and sharing the use of memory, input–output peripherials and other resources, estimating the performances of these systems is really complex. The fact that the same system behaves differently when solving various problems makes the performance evaluation even more difficult. Different problems have different possibilities for parallelization. Some problems can't be parallelized at all.

When designing parallel computer it is very important that the system has the predicted properties. It is also very important to design the algorithm that exploits both parallelism inherent in the problem and that available on the computer. The challenge for a computer designer is to discover the minimum organization and equipement necessary to achieve a given level of performance. By performance we mean the manner in which, or the efficiency with which, a computer system meats its goal. So, the developing analytical models for characterizing and understanding the parallel system performances is of a crucial interest. But, attempts to express some measure of performance as a explicit function of certain parameters were not successful always. Moreover, omitting one of the parameters leads to erroneous conclusions about the behaviour of parallel system. Thus, for example, in 1967 Amadahl (see [7]) made the observation that if s is the serial fraction in an algorithm, then its speedup is bounded by $1/s$, no matter how many processors are used. For example, if there are only 5% of the algorithm that can't be parallelized, then maximal speedup that can be achieved is 20, no matter how many processors are used. This statement, now popularly known as Amadahl's law, has been used by Amadahl and others to argue against the usefulness of large scale parallel computers. Fortunately, Amadahl was wrong. He missed the fact that the serial fraction, s, is a function of problem size, m. Moreover, for most scientific and technical applications it has property that $\lim_{m\to\infty} s(m) \to 0$ [4].

The above and other similar examples imply that in studing the performances of a parallel system, a severe mathematical formulations should be involved. Therefore the following should be developed:

- Explicit mathematical formulas that characterize the performances of parallel system
- Axioms for basic parameters.

The key for scientific approach in studing the performances lies in solving the above problems.

Some common performance measures of parallel algorithm running on a parallel computer are the execution time, the speedup, the efficiency, the scalability, etc.

2. Definitions and assumptions

In this section, we introduce some terminology used in the rest of the paper.

We assume the system of n identical processors interconnected in some way for the purpose of passing data and control information between the processors. Communication between the processors can be achieved via common memory modules or by message passing. Each processor is supplied by some amount of local memory. By a parallel system we mean a combination of a parallel algorithm and a parallel architecture comprising of identical processing units.

Definition 2.1. *The degree of parallelism* of a numerical algorithm is the number of operations in the algorithm that can be done in parallel.

Note that the degree of parallelism is independent of the number of processors in the system; it is an intristic measure of the parallelism in the algorithm.

Definition 2.2. *The average degree of parallelism* of an algorithm is the total number of operations in the algorithm divided by the number of stages. A stage is comprised of the operations that can be performed in parallel.

Consider a program for which execution time on a single processor is equal to $T(1)$. When this program runs on a multiprocessor, the execution time can be divided into two components:

- component with running time T_s that must be run sequentialy;
- component with sequential running time T_p that can be subdivided into parallel components running on different processors.

Note that

$$T(1) = T_s + T_p \,. \tag{1}$$

Since a small number of problems posses ideal intristic parallelism, T_s is greater than zero.

Assume that $T(n)$ is execution time when program is running on n–processor system. The $T(n)$ involves the following components:

- serial execution time T_s,
- parallel execution time, equal to T_p/n if the parallelizable part of the program can be partitioned into n parallel components of equal running time, and
- synchronization and communication overhead $T_0(n)$.

According to the previous, we have the following definition.

Definition 2.3. *The execution time of an algorithm* running on n–processor system is

$$T(n) = T_s + \frac{T_p}{n} + T_0(n) \,. \tag{2}$$

Definition 2.4. *The speedup* of parallel system is defined as

$$S(n) = \frac{T(1)}{T(n)} \,. \tag{3}$$

The parallel algorithm may not be the best algorithm on a single processor, so for $T(1)$ we take the execution time on a single processor of the fastest serial algorithm.

Definition 2.5. *The efficiency* of the parallel system is

$$E(n) = \frac{S(n)}{n} = \frac{T(1)}{nT(n)} \,. \tag{4}$$

Definition 2.6. *Parallel cost penalty* is defined as

$$C(n) = nT(n) \,. \tag{5}$$

Definition 2.7. *Relative parallel cost penalty* is

$$R(n) = \frac{nT(n) - T(1)}{n - 1}\,. \tag{6}$$

Definition 2.8. *The gain factor* of a parallel system is

$$G(n) = \frac{T(1) - T(n)}{T(1)} = 1 - \frac{1}{S(n)}\,. \tag{7}$$

It is not difficult to see from the above definitions, that the execution time is the primary measure of a parallel system performance which is used as a basis for estimating other characteristics of a system. So, it was natural to establish the set of axioms for this metric.

3. The set of axioms

In the text that follows we are going to give the survey of axioms that were proposed in literature in order to introduce a scientific approach in studing the parallel system performances. Further, we shall propose a modified set of axioms based on discrete mathematical apparatus.

As we have already mentioned, the execution time is the most important measure of parallel system performance. Its component $T_0(n)$ represents the influence of communication and synchronization between processors on execution time. The value of $T_0(n)$ directly affects the performance of the whole system. So, the analytical methods for characterizing and understanding this measure were developed. In [2] Flatt and Kennedy introduced the following axioms for $T_0(n)$:

F.1 $T_0(n)$ is continuous and twice differentiable in respect to n,
F.2 $T_0(1) = 0$,
F.3 $T_0'(n) > 0$ for all $n \geq 1$, hence $T_0(n)$ is nonnegative,
F.4 $nT_0''(n) + 2T_0'(n) > 0$ for all $n \geq 1$,
F.5 There exists $n_1 \geq 1$ such that $T_0(n_1) = T(1)$.

On the basis of the involved axioms, the authors have investigated the impact of synchronization and communication overhead on the performance of parallel systems. They have established upper bounds on the power of parallel processing in the presence of synchronization and communication overheads.

The pioniers work of Flatt and Kennedy has motivated researchers to investigate the following:

a) Is the set of axioms **F.1**–**F.5** the minimal one, or it can be reduced, and, can some conditions be weaker?

b) What is physical and/or geometrical meaning of **F.1**–**F.5**?

c) Why $T_0(n)$ and other measures are considered as real functions, if they are defined on the set of natural numbers, $\mathbb{N}$?

In [6] the problems a) and b) were considered. The axiom **F.5** is rejected as too strong, and instead of it the condition

$$\lim_{n\to\infty} T_0(n) = +\infty \tag{8}$$

is tested.

As a basic value in [6] the function $D(n) = nT_0(n)$, instead of $T_0(n)$, is taken. This enables author to introduce the following more geometricaly intuitive axioms:

D.1 $D(1) = 0$,

D.2 $D(n) \geq 0$,

D.3 $D(n)$ is strictly convex and differentiable.

The author has proved that any $T_0(n)$ satisfying **F.1** to **F.4** also satisfies **D.1** to **D.3**. The question c) was not addressed in this paper.

In [5] the problems a) and c) were addressed. Namely, the values $T_0(n)$ and $T(n)$ were considered as members of sequences $\{T_0(n)\}$ and $\{T(n)\}$, respectively. This enables authors to reduce the set of axioms, defined by Flatt and kennedy, from five to the following three:

P.1 $T_0(1) = 0,\ T_0(2) \geq 0$,

P.2 $(n+2)\Delta^2 T_0(n) + 2\Delta T_0(n) > 0$, for $n \geq 1$.

P.3 There exists n_1 such that $T_0(n_1) = T(1)$.

It was shown that performance evaluation can be carried out very efficiently using discrete mathematical apparatus.

Let us note that it is natural to use the discrete mathematical apparatus, since the number of processors in the system is an integer value. Besides, by utilizing this apparatus the condition **F.1** (that the function is continuous and twice differentiable) becomes needless. Also, the axiom **F.2** is expressed as natural and elemental condition $T_0(2) \geq 0$. The physical and/or geometrical meaning for **P.2** and **P.3** can't be given.

Inspired by the papers [6] and [5] and having in mind questions a), b) and c), we propose in this paper a new set of axioms for sequence $\{D(n)\}_{n\in\mathbb{N}}$, $D(n) = nT_0(n)$, as follows:

A.1 $D(1) = 0$,
A.2 $D(2) \geq 0$,
A.3 $\Delta^2 D(n) > 0$, for $n \geq 1$.

Usage of discrete apparatus enables us to propose somewhat weaker conditions for $D(n)$ compared with **D.2** and **D.3** from [6]. Namely, instead of $D(n) \geq 0$ we take a condition $D(2) \geq 0$ and for sequence $\{D(n)\}$ we assume to be convex instead of function $D(n)$ being strictly convex and differentiable. Further, instead of axiom **P.3** from [5] we shall take a weaker condition (8) as in [6].

Now, we shall prove the following result.

Theorem 3.1. *The set of axioms* **A.1**–**A.3** *is equivalent with* **P.1**–**P.2**.

Proof. Statement of the Theorem 3.1 directly follows from the equalities $D(1) = T_0(1)$, $D(2) = 2T_0(2)$ and $\Delta^2 D(n) = (n+2)\Delta^2 T_0(n) + 2\Delta T_0(n)$. □

The assumption of axiom **P.3** is not involved in **A.1**–**A.3**. Therefore, we are going to prove the main result for $T(n)$ from [5] using **A.1**–**A.3** and under assumption that (8) is satisfied. But, first, we shall prove two auxiliary results.

Lemma 3.1. *The sequence* $\{D(n)\}_{n \in \mathrm{N}}$, *is positive and monotone increasing.*

Proof. According to **A.3** it follows that

$$\sum_{k=1}^{n} k\Delta D(k+1) > \sum_{k=1}^{n} k\Delta D(k) .$$

From the above inequality it follows that $\dfrac{D(n+2)}{n+1} > \dfrac{D(n+1)}{n}$, i.e.

$$\frac{D(n+2)}{n+1} > \frac{D(n+1)}{n} > \cdots > \frac{D(2)}{1},$$

and according to **A.2** we have $D(2) \geq 0$, i.e. the sequence $\{D(n)\}$ is monotone increasing. □

Lemma 3.2. *The sequence* $\{T_0(n)\}_{n \in \mathrm{N}}$ *is monotone increasing.*

Proof. According to inequality $\Delta^2 D(k) > 0$, i.e. $\Delta D(k) > \Delta D(k-1)$, we have that

$$(k+1)\Delta T_0(k) > (k-1)\Delta T_0(k-1) .$$

From the above inequality it follows

$$\sum_{k=2}^{n} k(k+1)\Delta T_0(k) > \sum_{k=2}^{n}(k-1)k\Delta T_0(k)\,,$$

i.e.

$$n(n+1)\Delta T_0(n) > 1\cdot 2\Delta T_0(1) = 2T_0(2) = D(2) \geq 0\,,$$

which established the result. □

We now present the main result for $T(n)$.

Theorem 3.2. *If equality (8) is satisfied, then there exists an unique value, $n = n_0$, for which the sequence $\{T(n)\}$ reaches the minimum, i.e. the inequality*

$$T(n_0) \leq T(n)$$

for all $n \geq 1$, is valid. When the inequality $D(2) \leq T_p$ is valid, then n_0 is an unique solution to

$$\begin{aligned}(n_0-1)n_0\Delta T_0(n_0-1) - T_p &\leq 0\\ n_0(n_0+1)\Delta T_0(n_0) - T_p &\geq 0\,.\end{aligned}$$

Otherwise, $n_0 = 1$.

Proof. Let $D(2) \leq T_p$, i.e. $2T_0(2) \leq T_p$. According to (2) the equality

$$\Delta T(n) = \Delta T_0(n) - \frac{1}{n(n+1)}\,T_p = \frac{1}{n(n+1)}\,(n(n+1)\Delta T_0(n) - T_p) \tag{10}$$

is valid. From (10), for $n = 1$ we obtain the inequality

$$\Delta T(1) = \frac{1}{2}\,(D(2) - T_p) \leq 0\,.$$

Now, it is necessary to prove that the sequence $\{T(n)\}$ is not decreasing for all $n \geq 1$, but there exists n for which $\Delta T(n) \geq 0$. Assume the opposite, i.e. that for all $n \geq 1$ the inequality $\Delta T(n) \leq 0$ is valid. Then according to (10) we obtain

$$\Delta T_0(n) = \Delta T(n) + \frac{1}{n(n+1)}\,T_p \leq \frac{1}{n(n+1)}\,T_p\,. \tag{11}$$

According to (11) we have

$$\sum_{k=1}^{n-1} \Delta T_0(k) \leq \sum_{k=1}^{n-1} \frac{1}{k(k+1)}\,T_p\,,$$

i.e.

$$T_0(n) \leq T_p \left(1 - \frac{1}{n}\right) .$$

From the last inequality it follows that

$$\lim_{n \to +\infty} T_0(n) \leq T_p$$

which is in contradiction to the assumption that (8) is valid. Consequently, we conclude that the assumption $\Delta T(n) \leq 0$ for all $n \geq 1$, is not correct. Namely, there are values for n for which $\Delta T(n) \geq 0$, i.e. there is at least one value $n = n_0$ for which the inequalities

$$(12) \qquad \Delta T(n_0 - 1) \leq 0 , \qquad \Delta T(n_0) \geq 0$$

and (9) are valid.

Now, it is necessary to prove that n_0 is unique. Using **A.3** we obtain

$$(13) \qquad \Delta\left(n(n+1)\Delta T_0(n)\right) = (n+1)\Delta^2 D(n) > 0 .$$

From (13) it can be concluded that the sequence $\{n(n+1)\Delta T_0(n)\}$ is monotone increasing. Accordingly, there exist the unique value $n = n_0$ such that inequalities (9) and (12) are valid.

Now, we are going to prove that for $n = n_0$ the sequence $\{T(n)\}$ reaches a minimum. To prove this it is enough to show that $\Delta^2 T(n_0 - 1) > 0$ and that the sequence $\{T(n)\}$ is monotone increasing for $n \geq n_0$.

From (10) we obtain

$$(14) \qquad \Delta^2 T(n) = \Delta^2 T_0(n) + \frac{2}{n(n+1)(n+2)} T_p .$$

According to (10) and (14) it follows that

$$(15) \qquad \Delta^2 T(n) = \frac{1}{n+2}\left(\Delta^2 D(n)\right) - \frac{2}{n+2} \Delta T_0(n) .$$

Substituting $n = n_0 - 1$ in (15) and using inequality (12), the inequality

$$\Delta^2 T(n_0 - 1) \geq \frac{1}{n_0 + 1}\left(\Delta^2 D(n_0 - 1)\right) > 0$$

is obtained.

Since the sequence $\{n(n+1)\Delta T_0(n)\}$ is monotone increasing, and according to equality $n(n+1)\Delta T(n) = n(n+1)\Delta T_0(n) - T_p$, it follows that the sequence $\{n(n+1)\Delta T(n)\}$ is monotone increasing, also. On the other hand, since $n_0(n_0+1)\Delta T_0(n_0) - T_p \geq 0$ it follows that $n_0(n_0+1)\Delta T(n_0) \geq 0$. Now, according to inequality $n(n+1)\Delta T(n) \geq n_0(n_0+1)T(n_0) \geq 0$, we obtain that $\Delta T(n) \geq 0$, for all $n \geq n_0$.

If we assume that $D(2) \geq T_p$, then from (10) we obtain $\Delta T(1) = \frac{1}{2}(D(2) - T_p) \geq 0$. Further, since $\Delta T(n) \geq \Delta T(1) > 0$ for all $n \geq 1$, we conclude that in this case $n_0 = 1$. □

Remark. Theorem 3.2 have been proved in [5] under conditions **P.1** to **P.3**.

According to the results proved in Theorem 3.2 we are going to prove the following results for sequences $\{S(n)\}$ and $\{C(n)\}$, $n \in \mathbb{N}$.

Theorem 3.3. *Let the inequality (8) be satisfied. The sequence $\{S(n)\}$ has an unique maximum at $n_0 \geq 1$. Also, if $D(2) < T_p$ then*

$$\frac{T(1)}{\Delta C(n_0)} \leq S(n_0) \leq \frac{T(1)}{\Delta C(n_0 - 1)}. \tag{16}$$

Proof. In theorem 3.2 we have proved that, under certain conditions, the sequence $\{T(n)\}$, $n \in \mathbb{N}$, has the unique minimum at $n_0 \geq 1$. The following is also valid

$$\Delta T(n_0 - 1) \leq 0, \qquad \Delta T(n_0) \geq 0. \tag{17}$$

From (3) we have that

$$\Delta S(n) = T(1)\left(-\frac{\Delta T(n)}{T(n)\,T(n+1)}\right). \tag{18}$$

Combining (17) and (18) we obtain

$$\Delta S(n_0 - 1) \geq 0, \qquad \Delta S(n_0) \leq 0. \tag{19}$$

This means that the sequence $\{S(n)\}$, $n \in \mathbb{N}$, has the unique maximum at $n_0 \geq 1$.

Since $C(n) = nT(n)$, i.e. $C(n) = \dfrac{nT(1)}{S(n)}$, the following is also valid

$$\Delta C(n) = T(1)\,\frac{S(n) - n\,\Delta S(n)}{S(n)\,S(n+1)},$$

and

$$\Delta C(n) = T(1)\,\frac{S(n+1)-(n+1)S(n)}{S(n)\,S(n+1)}\,. \tag{21}$$

By substituting n with $n_0 - 1$ in (20) and n with n_0 in (21), the right and left parts of inequality (16) are obtained, respectively. □

Similarly as in [2], [5], some other properties of sequences defined by (2) to (7) can be proved. For the sake of illustration, we give some properties that directly follows from axioms **A.1** to **A.3**.

Theorem 3.4. *The sequence $\{C(n)\}$, $n \in \mathbb{N}$, is monotone increasing and convex.*

Theorem 3.5. *The sequence $\{E(n)\}$, $n \in \mathbb{N}$ is monotone decreasing, convex, and has a property $E(1) = 1$.*

Theorem 3.6. *The sequence $\{R(n)\}$, $n \in \mathbb{N}$, is monotone increasing for $n \geq 2$.*

Theorem 3.7. *If the equality (8) is valid, then the sequence $\{G(n)\}$, $n \in \mathbb{N}$, has an unique maximum at $n_0 \geq 1$.*

References

[1] S.G. Akl, *The Design and Analysis of Parallel Algorithms.*, Prentice–Hall Inc., New Jersey, 1989.

[2] H.P. Flatt and K. Kennedy, *Performance of Parallel Processors*, Parallel Comput. **12** (1989), 1–20.

[3] G. Golub and J.M. Ortega, *Scientific Computing: An Introduction with Parallel Computing*, Academic Press, Inc., 1993.

[4] J.L. Gustafsson, *The Scale–Sized Model: A Revision of Amadahl's Law*, Proc. 3rd Conf. on Supercomputing, Boston '88, 1988.

[5] I.Ž. Milovanović, E.I. Milovanović, M.K. Stojčev, *Discrete Timing Model for Parallel Processing*, Facta Universitatis (Niš) Ser. Math. Inform. **7** (1992), 123–144.

[6] D. Müller–Wichards, *Problem Size Scaling in the Presence of Parallel Overhead*, Parallel Comput. **17** (1991), 1361–1376.

[7] J.M. Ortega, *Introduction to Parallel and Vector Solution of Linear Systems*, Plenum Press, New York and London, 1988.

[8] D. Tabak, *Multiprocessors*, Prentice–Hall Inc., New Jersey, 1990.

Faculty of Electronic Engineering, Beogradska 14, P.O. Box 73, 18000 Niš, Serbia.

FILOMAT (Niš) 9:3 (1995), 851–855
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON ONE CONGRUENCE RELATION ON A GLOBAL SEMIGROUP

Žarko M. Mitrović

ABSTRACT. A group G end its global semigroup $Glb\,G$ is considered. Every congruence relation in the group G induces a congruence relation in the semigroup $Glb\,G$. These congruence relations and their classes are studied.

1. The induced equivalence relation.

1. Let ϱ be a binary relation defined on a set S and $Pow\,S$ the power set of S . The relation ϱ induces a series of relations on $Pow\,S$, defined by quantifiers. One of these relations is defined by

$$A\sigma B \iff (\forall a \in A, \exists b \in B) \quad a\varrho b\,, \tag{1}$$

where $A, B \in Pow\,S$.

Theorem 1. *If ϱ is a reflexive or transitive relation, then σ is also reflexive and transitive respectively.*

Proof. If ϱ is reflexive, then for every $a \in A$ there exists $a \in A$ such that $a\varrho a$ holds true. Therefore, $A\sigma A$ is valid for every $A \in Pow\,S$, so that σ is reflexive.

Let ϱ be transitive. If $A\sigma B$ and $B\sigma C$ hold true, then for every $a \in A$ and $b \in B$ there exist $b_1 \in B$ and $c \in C$ such that $a\varrho b_1$ and $b\varrho c$ hold true. Therefore, for $b_1 \in B$ there also exists $c_1 \in C$ such that $b_1\varrho c_1$. Now from $a\varrho b_1$ and $b_1\varrho c_1$ we get $a\varrho c_1$, i.e. for every $a \in A$ there exists $c_1 \in C$ such that $a\varrho c_1$ holds true. From this it follows that $A\sigma C$ holds true, so that σ is transitive. □

Using σ we can define a new induced relation δ on $Pow\,S$ by

$$\delta = \sigma \cap \sigma^{-1}\,. \tag{2}$$

In other words

$$A\delta B \iff [(\forall a_1 \in A, \exists b_1 \in B)\ a_1 \varrho b_1\ ] \ \wedge\ [(\forall b_2 \in B, \exists a_2 \in A)\ b_2 \varrho a_2\ ]\ ,$$

where $A, B \in Pow\, S$.

Theorem 2. *If* ϱ *is a reflexive and transitive relation on* S *, then* δ *is an equivalence relation on* $Pow\, S$ *.*

Proof. From the definition of δ we can directly see that δ is symmetric.

We shell use the following characteristics of relations ([3]) : If relations α and β are reflexive (transitive), then so are the following relations : $\alpha \cap \beta$ and α^{-1} . Thus, since ϱ is reflexive and transitive so is the relation δ . Therefore, δ is an equivalence relation. □

In what follows we shell assume that ϱ is an equivalence relation.

We shell say that δ is an *equivalence relation induced by* ϱ and denote it by $\hat{\varrho}$.

Let us introduce the notation

$$\varrho_X = \{\varrho_x \mid x \in X \in Pow\, S\}$$

(Elements of ϱ_X are equivalence classes.)

Theorem 3. *The following logical equivalence*

$$A\hat{\varrho}B \iff \varrho_A = \varrho_B \tag{3}$$

holds true, where $A, B \in Pow\, S$ *.*

Proof. Let $A\hat{\varrho}B$ be valid. If $\varrho_x \in \varrho_A$, then there exists $a \in A$ such that $\varrho_a = \varrho_x \in \varrho_A$. For this $a \in A$ there exists $b \in B$ such that $a\varrho b$ holds true, i.e. $\varrho_a = \varrho_b \in \varrho_B$. Therefore, $\varrho_x \in \varrho_B$, so that $\varrho_A \subseteq \varrho_B$ is valid. In exactly the same way we conclude that $\varrho_B \subseteq \varrho_A$ holds true. Therefore, $\varrho_A = \varrho_B$.

Conversely, let $\varrho_A = \varrho_B$ be valid. If $a \in A$, then $\varrho_a \in \varrho_A = \varrho_B$, so that there exists $b \in B$ such that $\varrho_a = \varrho_b \Leftrightarrow a\varrho b$ holds true. Similarly, from $b \in B$ it follows that there exists $a \in A$ such that $b\varrho a$. This means that $A\hat{\varrho}B$ is valid.

Therefore, (3) holds true. □

On account of Theorem 3 we conclude that we can describe equivalence classes for $\hat{\varrho}$ by

$$\hat{\varrho}_A = \{X \in Pow\, S \mid A\hat{\varrho}X\} = \{X \in Pow\, S \mid \varrho_A = \varrho_X\}\ .$$

Let us introduce the notation

$$[\hat{\varrho}_A] \;=\; \bigcup_{a\in A} \varrho_a \,. \tag{4}$$

Tho following logical equivalences

$$A\hat{\varrho}B \iff \hat{\varrho}_A = \hat{\varrho}_B \iff \varrho_A = \varrho_B \iff [\hat{\varrho}_A] = [\hat{\varrho}_B]$$

are now evident.

Let us now describe in detail the equivalence classes for $\hat{\varrho}$.

Theorem 4. $X \in \hat{\varrho}_A$ *if and only if* X *contains at least one element from every equivalence class of the family* ϱ_A *, but does not contain any element from any class out of* ϱ_A *. In other words*

$$X \in \hat{\varrho}_A \iff X \subseteq [\hat{\varrho}_A] \;\wedge\; [(\forall a \in A) \quad X \cap \varrho_a \neq \emptyset\,] \,.$$

Proof. From Theorem 3 we directly get the validity of that assertion. □

It is obvious that $[\hat{\varrho}_A]$ is the maximal element of $\hat{\varrho}_A$ (in the sense that there is no element of $\hat{\varrho}_A$ which contains $[\hat{\varrho}_A]$).

If $A \subseteq \varrho_b$, then $[\hat{\varrho}_A] = \varrho_b$, so that $X \in \hat{\varrho}_{\varrho_b} \Leftrightarrow X \subseteq \varrho_b$. This means that

$$\hat{\varrho}_{\varrho_b} \;=\; Pow\varrho_b \,. \tag{5}$$

2. Let $f : X \to Y$, then the relation kernel of f , $\kappa = kerf$, is an equivalence relation on X , defined by

$$(\forall a, b \in X) \qquad a\kappa b \iff f(a) = f(b) \,.$$

The equivalence classes of κ are defined by $\kappa_a = f^{-1}(f(a))$.

The kernel κ induces the equivalence relation $\hat{\kappa}$ on $Pow\,X$.

Since

$$\begin{aligned}[\hat{\kappa}_A] \;&=\; \bigcup_{a\in A} \kappa_a \;=\; \bigcup_{a\in A} f^{-1}(f(a)) \;=\; f^{-1}(\bigcup_{a\in A} f(a)) \\ &=\; f^{-1}\{f(a) \mid a \in A\} \;=\; f^{-1}(f(A))\end{aligned}$$

we have $[\hat{\kappa}_A] = f^{-1}(f(A))$, so that

$$A\hat{\kappa}B \iff f^{-1}(f(A)) = f^{-1}(f(B))$$

holds true.

2. The congruence relation on *Glb G*.

1. Let us now consider a group G . The set $Pow\,G$ with the global operations

$$AB = \{ab \mid a \in A, b \in B\}\ ,$$

$$A^{-1} = \{a^{-1} \mid a \in A\}$$

is the global monoid (i.e. semigroup with identity) with involution $^{-1}$ ([1], [2]).

This monoid we shall denote by $Glb\,G$. Suppose ϱ is a relation on G and σ and $\hat{\varrho}$ are the relations on $Pow\,G$, defined by (1) and (2) respectively.

Theorem 5. *If ϱ is a relation on $(G,.)$, compatible with the operations in G , then the relation σ is compatible with the global operations in $Glb\,G$.*

Proof. Let $A\sigma B$ and $X\sigma Y$ be valid $(A, B, X, Y \in Glb\,G)$. If $u \in AX$, then there exist $a \in A$ and $x \in X$ such that $u = ax$ holds true. In the other hand, from $A\sigma B$ $(X\sigma Y)$ it follows that for every $a \in A$ $(x \in X)$ there exists $b \in B$ $(y \in Y)$ such that $a\varrho b$ $(x\varrho y)$ is valid. Since ϱ is compatible with the binary operation, $ax\varrho by$ holds true. Thus, for any $u = ax \in AX$ there exists $v = by \in BY$ such that $ax\varrho by$ is valid. This means that $AX\sigma BY$ holds true, so that σ is compatible with the global binary operation.

Let $A\sigma B$ $(A, B \in Glb\,G)$ be valid, then for every $a \in A$ there exists $b \in B$ such that $a\varrho b$ holds true. Since ϱ is compatible with the unary operation $^{-1}$ in G , from $a\varrho b$ it follows $a^{-1}\varrho b^{-1}$. This means that for every $a^{-1} \in A^{-1}$ there exists $b^{-1} \in B^{-1}$ such that $a^{-1}\varrho b^{-1}$, i.e. $A^{-1}\sigma B^{-1}$ holds true. □ .

Theorem 6. *If ϱ is a congruence relation on G , then $\hat{\varrho}$ is also a congruence relation on $Glb\,G$.*

Proof. Using Theorem 5 and the well known result: If α and β are relations on G , compatible with operations in G , then so are the following relations: β^{-1} and $\alpha \cap \beta$, we can conclude that $\hat{\varrho}$ is a congruence relation on $Glb\,G$. □

2. If H is a subgroup of G , then the relation μ , defined by

$$(\forall x, y \in G) \quad x\mu y \quad \Longleftrightarrow \quad x^{-1}y \in H \quad \Longleftrightarrow \quad xH = yH\ ,$$

is an equivalence relation on G . The equivalence classes for μ are defined by $\mu_x = xH$.

Let us consider the equivalence relation $\hat{\mu}$ on $Glb\,G$ induced by μ. Since

$$[\hat{\mu}_A] = \bigcup_{a\in A} \mu_a = \bigcup_{a\in A} aH = AH ,$$

on account of (4) ,

$$A\hat{\mu}B \iff AH = BH$$

holds true.

Let us put $E = \{e\}$, where e is the identity of G . Then E is the identity of $Glb\,G$ ([4]). On account of $\mu_E = eH = H$ and (5) we have $\hat{\mu}_H = Pow\,H$ and $[\hat{\mu}_H] = H$.

Theorem 7. *Every normal subgroup H of G induces the congruence relation $\hat{\mu}$ on $Glb\,G$, defined by*

$$(6) \qquad (\forall A, B \in Glb\,G) \quad A\hat{\mu}B \iff AH = BH ,$$

where $\hat{\mu}_H = Glb\,H$.

Proof. If H is a normal subgroup of G , then μ is an equvalence relation on G , so that, account of Theorem 6, the induced equivalence relation $\hat{\mu}$ is a congruence relation on $Glb\,G$.

Since $\hat{\mu}_H = Pow\,H$ and H is a group, the equality $\hat{\mu}_H = Glb\,H$ must hold. □

References

[1] Chacron J., *Theorie des classes dans un demi-groupe involutif*, Semigroup Forum **2** (1971), 138–153.

[2] McCarthy D. J. and D. L. Hayes, *Subgroups of the power semigroup of a group*, J. Combin. Theory (A) **14** (1973), 173–186.

[3] Schreider Ju. A., *Equality, resemblance, and order*, Mir Publishers, Moscow, 1975.

[4] Tamura T. and J. Shafer, *Power semigroups*, Math. Japon. **12** (1967), 25–32.

University of Novi Sad, Technical Faculty "Mihajlo Pupin", 23000 Zrenjanin, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 857–865
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

THE INDUCED RELATIONS ON A POWER SET

Žarko M. Mitrović and Ivana F. Berković

ABSTRACT. Let ϱ be a relation on the set S . By using the quantifiers and the relations ϱ i ϱ^{-1} a series of relations on the power set $Pow\,S$ are defined. The characteristics of these relations are studied and their classification is made.

1. The elementary induced relations

Let ϱ be a binary relation defined on a set S and $Pow\,S$ the power set of S . The relation ϱ on S induces a series of relations on $Pow\,S$.

Induced relations determined only by quantifiers we shall call the *elementary induced relations*. There are four such relations and they are defined by:

$$(1)\qquad \begin{cases} A\varrho_{11}B & \Longleftrightarrow \quad (\exists a\in A, \exists b\in B) \quad a\varrho b\,, \\ A\varrho_{12}B & \Longleftrightarrow \quad (\forall a\in A, \exists b\in B) \quad a\varrho b\,, \\ A\varrho_{21}B & \Longleftrightarrow \quad (\exists a\in A, \forall b\in B) \quad a\varrho b\,, \\ A\varrho_{22}B & \Longleftrightarrow \quad (\forall a\in A, \forall b\in B) \quad a\varrho b\,. \end{cases}$$

If we introduce the notations

$$A\varrho B \quad \Longleftrightarrow \quad (\forall a\in A, \forall b\in B) \quad a\varrho b$$

and

$$a\varrho B \quad \Longleftrightarrow \quad \{a\}\varrho B\,,$$

then we can write

$$A\varrho_{21}B \quad \Longleftrightarrow \quad (\exists a\in A) \quad a\varrho B\,,$$

$$A\varrho_{22}B \iff A\varrho B\,.$$

It is obvious that

$$\varrho_{12},\,\varrho_{21} \subseteq \varrho_{11} \quad \textit{and} \quad \varrho_{22} \subseteq \varrho_{12},\,\varrho_{21} \tag{2}$$

hold true.

With the help of the operation $^{-1}$ we obtain the new four induced relations. For them, on account of (2),

$$\varrho_{12}^{-1},\,\varrho_{21}^{-1} \subseteq \varrho_{11}^{-1} \quad \textit{and} \quad \varrho_{22}^{-1} \subseteq \varrho_{12}^{-1},\,\varrho_{21}^{-1} \tag{2'}$$

hold true.

Theorem 1. *The inclusions*

$$\varrho_{21}^{-1} \subseteq \varrho_{12} \quad \textit{and} \quad \varrho_{21} \subseteq \varrho_{12}^{-1} \tag{3}$$

hold true.

Proof. If $A\varrho_{21}^{-1}B \Leftrightarrow B\varrho_{21}A$ is valid, then there exists at least one element of B which is ϱ-related to every element of A, so that every element of A is ϱ-related to at least one element of B, i.e. $A\varrho_{12}B$ holds true. Conversely, let $A\varrho_{12}B$ is valid, i.e. every element of A is ϱ-related to at least one element of B, but this does not guarantee that any two elements of A are ϱ-related to the same element of B. Therefore, the inclusion $\varrho_{21}^{-1} \subseteq \varrho_{12}$ holds true.

From the proved inclusion it directly follows that the second inclusion is also valid. □

With the help of the elementary induced relations and operations $\cap$ and $^{-1}$ with relations we can obtain new induced relations. On account of (2), (2') and (3) there are only ten new induced relations more.

Let us put

$$\alpha_{ij} = \varrho_{ij} \cap \varrho_{ij}^{-1} \qquad (i,j=1,2)\,,$$

$$\beta_1 = \varrho_{11} \cap \varrho_{12}^{-1}\,, \quad \beta_2 = \varrho_{22} \cap \varrho_{21}^{-1}\,, \quad \gamma = \varrho_{12} \cap \varrho_{21}\,.$$

The relations α are obviously symmetric.

2. $\tilde{}$-induced relations

Let us consider the elementary relations induced by ϱ^{-1} and put

$$\tilde{\varrho}_{ij} = (\varrho^{-1})_{ij} \quad (i,j=1,2)\,.$$

According to (1), using $a\varrho^{-1}b \Leftrightarrow b\varrho a$, we have

$$(1') \qquad \begin{cases} A\tilde{\varrho}_{11}B & \Longleftrightarrow \quad (\exists a \in A, \exists b \in B) \quad b\varrho a\,, \\ A\tilde{\varrho}_{12}B & \Longleftrightarrow \quad (\forall a \in A, \exists b \in B) \quad b\varrho a\,, \\ A\tilde{\varrho}_{21}B & \Longleftrightarrow \quad (\exists a \in A, \forall b \in B) \quad b\varrho a\,, \\ A\tilde{\varrho}_{22}B & \Longleftrightarrow \quad (\forall a \in A, \forall b \in B) \quad b\varrho a\,. \end{cases}$$

For these induced relations and their inverse relations the inclusions analogous to the inclusions (2), (2') and (3) hold true.

The relations $\tilde{\alpha}_{ij}$, $\tilde{\beta}_i$ and $\tilde{\gamma}$ $(i,j = 1,2)$ are defined like the corresponding relations from the point 1.

The relations induced by ϱ^{-1} we shall call ˜- *induced relations.*

Theorem 2. *The equalities*

$$(4) \qquad \varrho_{11}^{-1} = (\varrho^{-1})_{11} \qquad \textit{and} \qquad \varrho_{22}^{-1} = (\varrho^{-1})_{22}$$

hold true.

Proof. Since

$$\begin{aligned} A\varrho_{11}^{-1}B \quad &\Longleftrightarrow \quad B\varrho_{11}A \quad \Longleftrightarrow \quad (\exists b \in B, \exists a \in A) \quad b\varrho a \\ &\Longleftrightarrow \quad (\exists a \in A, \exists b \in B) \quad b\varrho a \quad \Longleftrightarrow \quad A\tilde{\varrho}_{11}B\,, \end{aligned}$$

we conclude that $\varrho_{11}^{-1} = \tilde{\varrho}_{11}$ is valid.

In a similar manner we can prove the second equality. □

3. The classification of the induced relations

Let σ be one of the induced relations, then the relations $\tilde{\sigma}$ and $\widetilde{\sigma^{-1}}$ we receive by using the conventions:

$$\widetilde{\mu \cap \nu} = \tilde{\mu} \cap \tilde{\nu}, \quad \widetilde{\mu^{-1}} = (\tilde{\mu})^{-1} \qquad \text{and} \qquad \tilde{\tilde{\mu}} = \mu\,.$$

For example, $\tilde{\beta}_1 = \tilde{\varrho}_{11} \cap (\tilde{\varrho}_{12})^{-1}$.

For every induced relation σ there exist three corresponding induced relations more: σ^{-1} , $\tilde{\sigma}$ and $(\tilde{\sigma})^{-1}$. We shall say that these four relations are *conjugate* to each other. If σ is an elementary induced relation, then we shall call all such relations the *elementary induced relations in the broader sense.*

Using the elementary induced relations in the broader sense and the operations $\cap$ and $^{-1}$ with relations, we receive the large number of new induced relations. There are 98 such relations. On account of inclusions (2), (2') and (3) and the corresponding inclusions for ˜-induced relations, as well as the equalities (4), these new induced relations may be formed by at the most four elementary induced relations in the broader sense. Because of this fact all the induced relations may be classified into four classes:

Class A. The elementary induced relations in the broader sense; there are 12 such relations.

Class B. The induced relations formed by using only two relations of the Class A ; there are 34 such relations.

Class C. The induced relations formed by using only three relations of the Class A ; there are 36 such relations.

Class D. The induced relations formed by using only four relations of the Class A ; there are 16 such relations.

All the induced relations may be grouped into groups of four conjugate relations each. There are groups which contain the same relations (for example, $\tilde{\varrho}_{11} = \varrho_{11}^{-1}$) .

In the Class A there are 2 groups of four and 2 groups of two relations each. Let the representatives of these groups be just the elementary induced relations. If we denote the groups by Ai $(i = 1,2,3,4)$, then those representatives are:

$$A1 : \varrho_{11} \ , \quad A2 : \varrho_{12} \ , \quad A3 : \varrho_{21} \ , \quad A4 : \varrho_{22} \ .$$

The groups $A1$ and $A4$ contain only two elements each.

In the Class B there are 5 groups of four, 6 groups of two and 2 groups of only one relations each. Let the representatives of the groups of the Class B be:

$$B1 : \varrho_{11} \cap \varrho_{11}^{-1} \ , \quad B2 : \varrho_{11} \cap \varrho_{12}^{-1} \ , \quad B3 : \varrho_{12} \cap \varrho_{21} \ , \quad B4 : \varrho_{12} \cap \varrho_{12}^{-1} \ ,$$

$$B5 : \varrho_{12} \cap \tilde{\varrho}_{12} \ , \quad B6 : \varrho_{12} \cap \tilde{\varrho}_{21} \ , \quad B7 : \varrho_{12} \cap (\tilde{\varrho}_{12})^{-1} \ ,$$

$$B8 : \varrho_{12} \cap (\tilde{\varrho}_{21})^{-1} \ , \quad B9 : \varrho_{21} \cap \varrho_{21}^{-1} \ , \quad B10 : \varrho_{21} \cap \varrho_{22}^{-1} \ ,$$

$$B11 : \varrho_{21} \cap \tilde{\varrho}_{21} \ , \quad B12 : \varrho_{21} \cap (\tilde{\varrho}_{21})^{-1} \ , \quad B13 : \varrho_{22} \cap \varrho_{22}^{-1} \ .$$

The groups $B4$, $B5$, $B7$, $B9$, $B11$ and $B12$ contain two and the groups $B1$ and $B13$ only one relations each.

In the Class C there are 8 groups of four and 2 groups of two relations each. Let the representatives of the groups of the Class C be:

$$C1 : \varrho_{11} \cap \varrho_{12}^{-1} \cap \tilde{\varrho}_{12} \ , \quad C2 : \varrho_{12} \cap \varrho_{21} \cap \tilde{\varrho}_{12} \ , \quad C3 : \varrho_{12} \cap \varrho_{21} \cap \tilde{\varrho}_{21} \ ,$$

$$C4 : \varrho_{12} \cap \varrho_{21} \cap (\tilde{\varrho}_{12})^{-1} , \quad C5 : \varrho_{12} \cap \varrho_{21} \cap (\tilde{\varrho}_{21})^{-1} , \quad C6 : \varrho_{12} \cap \varrho_{12}^{-1} \cap \tilde{\varrho}_{12} ,$$
$$C7 : \varrho_{12} \cap \varrho_{12}^{-1} \cap \tilde{\varrho}_{21} , \quad C8 : \varrho_{12} \cap \tilde{\varrho}_{21} \cap (\tilde{\varrho}_{21})^{-1} , \quad C9 : \varrho_{21} \cap \varrho_{21}^{-1} \cap \tilde{\varrho}_{21} ,$$
$$C10 : \varrho_{21} \cap \varrho_{22}^{-1} \cap (\tilde{\varrho}_{21})^{-1} .$$

The groups $C1$ and $C10$ contain only two relations each.

In the Class D there are 2 groups of four, 3 groups of two and 2 groups of one relation each. Let the representatives of the groups of the Class D be:

$$D1 : \varrho_{12} \cap \varrho_{21} \cap \tilde{\varrho}_{12} \cap \tilde{\varrho}_{21} , \quad D2 : \varrho_{12} \cap \varrho_{21} \cap (\tilde{\varrho}_{12})^{-1} \cap (\tilde{\varrho}_{21})^{-1} ,$$
$$D3 : \varrho_{12} \cap \varrho_{21} \cap \tilde{\varrho}_{12} \cap (\tilde{\varrho}_{12})^{-1} , \quad D4 : \varrho_{12} \cap \varrho_{21} \cap \tilde{\varrho}_{21} \cap (\tilde{\varrho}_{21})^{-1} ,$$
$$D5 : \varrho_{12} \cap \varrho_{12}^{-1} \cap \tilde{\varrho}_{12} \cap (\tilde{\varrho}_{12})^{-1} , \quad D6 : \varrho_{12} \cap \varrho_{12}^{-1} \cap \tilde{\varrho}_{21} \cap (\tilde{\varrho}_{21})^{-1} ,$$
$$D7 : \varrho_{21} \cap \varrho_{21}^{-1} \cap \tilde{\varrho}_{21} \cap (\tilde{\varrho}_{21})^{-1} .$$

The groups $D1$, $D2$ and $D6$ contain two relations and the groups $D5$ and $D7$ only one relation each.

4. Some characteristics of the elementary induced relations

The following characteristics of relations: reflexivity, antireflexivity, symmetry, antisymmetry and transitivity are said to be the fundamental characteristics of relations.

Lemma. *If relations σ and τ have one of the fundamental characteristics, then the relations σ^{-1} and $\sigma \cap \tau$ have the same characteristic* ([5]) .

If ϱ has one of the fundamental characteristics, then we shall find all the induced relations which have the same characteristic. We shall not be interested for such characteristics of induced relations which the relation ϱ does not have, i.e. which are not hereditary characteristics. For example, α are symmetric relations, but ϱ need not be.

Theorem 3. *If ϱ is a reflexive relation then ϱ_{11} and ϱ_{12} are also reflexive but ϱ_{21} and ϱ_{22} are not.*

Proof. Let ϱ be reflexive, i.e.

$$(\forall a \in S) \quad a \varrho a \tag{5}$$

and let $A \in Pow\, S$.

Since from (5) it follows $(\forall a \in A)\ a\varrho a$, we conclude that

$$(\exists a \in A) \quad a\varrho a \quad \Longleftrightarrow \quad A \varrho_{11} A ,$$

so that ϱ_{11} is reflexive.

From (5) we obtain $(\forall a \in A, \exists a \in A)\ a\varrho a$, i.e. $A\varrho_{12}A$ so that ϱ_{12} is reflexive.

However, it need not exist $a \in A$ such that $a\varrho A$ holds true. From this it follows that ϱ_{21} and ϱ_{22} are not reflexive. □

Theorem 4. *If ϱ is an antireflexive relation then ϱ_{21} and ϱ_{22} are also antireflexive, but ϱ_{11} and ϱ_{12} are not.*

Proof. Let ϱ be antireflexive, i.e.

$$(\forall a \in S) \quad (a,a) \notin \varrho$$

and let $A \in Pow\ S$.

It may exist $a_1, a_2 \in A$ such that $a_1 \varrho a_2$ holds true, so that it may be $A \varrho_{11} A$. Similarly, for every $a_1 \in A$ it may exist $a_2 \in A$ such that $a_1 \varrho a_2$ holds true, so that it may be $A \varrho_{12} A$. Therefore, ϱ_{11} and ϱ_{12} are not antireflexive.

Since $(\forall a \in A) \quad (a,a) \notin \varrho$, it may not exist $a_1 \in A$ such that, for every $a_2 \in A$, $a_1 \varrho a_2$ holds true. Hence, $(A,A) \notin \varrho_{21}$ and $(A,A) \notin \varrho_{22}$. □

Theorem 5. *If ϱ is a symmetric relation then ϱ_{11} and ϱ_{22} are also symmetric but ϱ_{12} and ϱ_{21} are not and the equalities $\varrho_{12} = \tilde{\varrho}_{12}$ and $\varrho_{21} = \tilde{\varrho}_{21}$ hold true.*

Proof. Let ϱ be symmetric, i.e.

$$a \varrho b \Rightarrow b \varrho a \tag{6}$$

and let $A, B \in Pow\ S$.

From (1) it is obvious that ϱ_{11} and ϱ_{22} are symmetric and that ϱ_{12} and ϱ_{21} are not symmetric.

Since

$$\begin{aligned} A \varrho_{12} B \quad &\iff \quad (\forall a \in A, \exists b \in B) \quad a \varrho b \\ &\implies \quad (\forall a \in A, \exists b \in B) \quad b \varrho a \quad \iff \quad A \tilde{\varrho}_{12} B\,, \end{aligned}$$

the equality $\varrho_{12} = \tilde{\varrho}_{12}$ holds true.

In exactly the same way we can prove the second equality. □

Theorem 6. *If ϱ is a transitive relation then ϱ_{11} is not transitive, but the other three elementary induced relations are transitive.*

Proof. Let ϱ be a transitive, i.e.

$$a \varrho b \wedge b \varrho c \implies a \varrho c. \tag{7}$$

If $A \varrho_{11} B \wedge B \varrho_{11} C$ holds true, then

$$[(\exists a \in A, \exists b_1 \in B) \quad a \varrho b_1\,] \quad \wedge \quad [(\exists b_2 \in B, \exists c \in C) \quad b_2 \varrho c\,]\,.$$

Since, generally speaking, $b_1 \neq b_2$, then $A\varrho_{11}C$ need not hold, so that ϱ_{11} is not transitive.

If $A\varrho_{12}B \wedge B\varrho_{12}C$ holds true, then

$$[(\forall a \in A, \exists b_1 \in B) \quad a\varrho b_1\,] \quad \wedge \quad [(\forall b \in B, \exists c \in C) \quad b\varrho c\,]\,.$$

Since, for every $b \in B$, there exists $c \in C$ such that $b\varrho c$, then for $b_1 \in B$ must exist $c_1 \in C$ such that $b_1\varrho c_1$. Now from $a\varrho b_1$ and $b_1\varrho c_1$ it follows $a\varrho c_1$. Thus, for every $a \in A$ there exists $c_1 \in C$ such that $a\varrho c_1$, so that $A\varrho_{12}C$ holds true.

From

$$A\varrho_{21}B \;\wedge\; B\varrho_{21}C \quad \Longleftrightarrow \quad [(\exists a \in A) \quad a\varrho B\,] \wedge [(\exists b \in B) \quad b\varrho C\,]$$

it obviously follows that $(\exists a \in A) \quad a\varrho C$, i.e. $A\varrho_{21}C$.

The logical equivalences

$$\begin{aligned} A\varrho_{22}B \;\wedge\; B\varrho_{22}C \quad &\Longleftrightarrow \quad A\varrho B \;\wedge\; B\,\varrho C \\ &\Longrightarrow \quad A\varrho C \quad \Longleftrightarrow \quad A\varrho_{22}C \end{aligned}$$

evidently hold true. □

If ϱ is an antisymmetric relation, then it is easy to prove that the elementary induced relations may not be antisymmetric.

Theorem 7. *If relation ϱ and the representative of a group of conjugate relations have one of the fundamental characteristics, then the same characteristic have all the relations of that group.*

Proof. The assertion of the Theorem follows directly from the Lemma. □

5. The hereditary characteristics of the induced relations

According to Theorem 7 we can conclude that:

I. If ϱ is a reflexive relation, then the relations of the following groups: $A1-2$, $B1-2$, $B4-5$, $B7$, $C1$, $C6$ and $D5$ are also reflexive.

II. If ϱ is an antireflexive relation, then the relations from the following groups: $A3-4$, $B9-13$, $C9-10$ and $D7$ are also antireflexive.

III. If ϱ is a symmetric relation, then $\varrho^{-1} = \varrho$ holds true, so that the ˜-induced relations coincide with the ordinary induced relations. According to (2), (2'), (3) and Theorem 4, we can conclude that:

(i) The Classes C and D do not exist.

(ii) In the Class B the groups:

(a) $B1$ coincides with the group $A1$;

(b) $B2$ and $B5$ coincide with the group $A2$;

(c) $B8$ and $B11$ coincide with the group $A3$;

(d) $B10$ and $B13$ coincide with the group $A4$.

(iii) The following groups mutually coincide: $B3$ and $B6$, $B4$ and $B7$, $B9$ and $B12$.

Therefore, in this case we have only the following induced relations:

$A1$: ϱ_{11} ,
$A2$: ϱ_{12} , ϱ_{12}^{-1} ,
$A3$: ϱ_{21} , ϱ_{21}^{-1} ,
$A4$: ϱ_{22} ,
$B3$: $\varrho_{12} \cap \varrho_{21}$, $\varrho_{12}^{-1} \cap \varrho_{21}^{-1}$,
$B4$: $\varrho_{12} \cap \varrho_{12}^{-1}$,
$B9$: $\varrho_{21} \cap \varrho_{21}^{-1}$.

The relations from $A1$ and $A4$ are symmetric relations if ϱ is symmetric, but the relations from the groups $B4$ and $B9$ are always symmetric. The other relations are not symmetric.

IV. If ϱ is a transitive relation, then the relations of the following groups: $A2-4$, $B3-13$, $C2-10$ and $D1-7$ are also transitive.

V. If ϱ is a quasi-order, i.e. a reflexive and transitive relation, then the relations from the following groups: $A2$, $B4-5$, $B7$, $C6$ and $D5$ are also quasi-orders. In particular, the relations from the group $B4$ are equivalence relations.

VI. If ϱ is a strict order, i.e. an antireflexive and transitive relation, then the relations of the following groups: $A3-4$, $B9-13$, $C9-10$ and $D7$ are also strict orders.

VII. If ϱ is a tolerance, i.e a reflexive and symmetric relation, then ϱ_{11} is tolerance. Moreover, if ϱ is only reflexive, then $\varrho_{12} \cap \varrho_{12}^{-1}$ is also a tolerance.

VIII. If ϱ is an equivalence relation, then only

$$\delta = \varrho_{12} \cap \varrho_{12}^{-1}$$

is an equivalence relation. Therefore, the unique equivalence relation on $Pow\, S$ induced by an equivalence relation on S is just the relation δ .

It is easy to show that if ϱ is the equality relation on S , then δ is the equality relation on $Pow\, S$.

Remark. We may obtain induced relations also by using the operations $\cup$ and $^{-1}$ or by using all the three operations $\cap$, $\cup$ and $^{-1}$, but we shall not do it in this paper.

6. Some applications

Let $\mathbb{K}^n$ be a Euclidean n-dimensional linear space and $Int\, \mathbb{R}$ be the set of all the closed real intervals.

In [1] the relations on $Pow\, \mathbb{K}^n$ induced by the orthogonality of vectors in $\mathbb{K}$ are considered. In [2] and [3] some relations on $Int\, \mathbb{R}$ induced by the relation $\leq$ (less or equal) on $\mathbb{R}$ are studied.

In [4] the congruence relations on global semigroup of a group G induced by a congruence relation on G is considered.

References

[1] I. F. Berković and Ž. M. Mitrović, *On orthogonality in interval analysis*, VII Conference on Applied Mathematics, Scitovski R., ed., University of Osijek, Osijek (1990), 13–19.

[2] Ž. M. Mitrović, *On the set-norm in the interval analysis*, Conference on Applied Mathematics 6, Jovanović B., ed., University of Beograd, Beograd (1989), 129–134.

[3] Ž. M. Mitrović and I. F. Berković, *Some induced relations on int* $\mathbb{R}$, Zbornik radova Tehničkog fakulteta "M. Pupin", Zrenjanin (1990), 61–65.

[4] Ž. M. Mitrović, *On one congruence relation on global semigroup.* (to appear).

[5] Ju. A. Schreider, *Equality, resemblance, and order*, Mir Publishers, Moscow, 1975.

University of Novi Sad, Technical Faculty "Mihajlo Pupin", 23000 Zrenjanin, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 867–872
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON MODIFICATIONS OF THE GALOIS GROUP

Boris V. Novikov

ABSTRACT. In this paper we define Doss modifications of groups and we describe such modifications of a simple cyclic group.

Let K be a field, L a finite normal extension of K with the Galois group G. Sweedler [4] has defined a Brauer monoid $M(G,L)$ which allows to classify so called strongly primary algebras. $M(G,L)$ is a semilattice of Abelian groups and its idempotents are in $1-1$ correspondence with some of partial orders on G (so called lower subtractive G-posets).

Another approach to studying the structure of the Brauer monoid was suggested in [3]. Let 0 be an element not contained in a finite multiplicative group G. We call *a modification* of G the monoid on $G\cup\{0\}$ with an operation $*$ such that $x*y$ is equal to xy or 0 for $x,y\in G$, and besides $1*x=x*1=x, 0*0=0*x=x*0=0$. It was shown in [3] that there exists a $1-1$ correspondence between the modifications of G and the idempotents of $M(G,L)$. Moreover, the describing of a Brauer monoid may be reduced to the studying of the modifications because their second 0-cohomology groups [2] are isomorphic to the group components of $M(G,L)$.

However, the enumerating of all modifications of a given group seems to be a difficult problem. In the general case there are known simple properties of modifications only such as [3]:

a) the modifications yields the condition of 0-cancellativity: if $x*z=y*z\neq 0$ or $z*x=z*y\neq 0$ then $x=y$;

b) the ideal of all non-invertible elements of a modification is nilpotent.

In this paper we describe a kind of modifications of a simple cyclic group.

Definition. The modification S of a group G is called *a Doss modification* (analogously to the well-known condition for embedding a semigroup into a group [1]) if

$$(\forall a, b \in S) \quad a * S \cap b * S \neq 0 \Rightarrow a \in b * S \ \vee \ b \in a * S.$$

In what follows we assume that $G = \langle a | a^p = 1 \rangle$ is a cyclic group of prime order p, S is its commutative Doss modification, different from $G^0(.)$ (i.e. $\exists x, y \in G \quad x * y = 0$).

Lemma 1. *S is a 0-direct union of monogenic semigroups (which are generated by indecomposable elements) with joined identity.*

Proof. If x, y are indecomposable elements then $x * S \cap y * S = 0$; in particular, $\langle x \rangle * \langle y \rangle = \langle y \rangle * \langle x \rangle = 0$. Since G has the prime order the subgroup of invertible elements is trivial. So every nonidentity element of S is divided by an indecomposable element and is in fact a power of it. Therefore $S - \{1\}$ is 0-direct union of the monogenic subsemigroups generated by the indecomposable elements. □

Remark. Evidently, the converse assertion is true too.

We shall describe the commutative Doss modifications with two or three generators. It will conveniently to denote an element $a^k \in G$ by b_k, when it is regarded as an element of S (hence $b_k^n = a^k * \dots * a^k$ n times). Accordingly to Lemma 1 $T = S - \{1\}$ is a semigroup which has the presentation of the form

$$T = \langle b_{i_1}, \dots, b_{i_r} | b_{i_k}^{\alpha_k} = 0, 1 \le k \le r \rangle$$

whence $\alpha_1 + \dots + \alpha_r = p + r - 1$.

Theorem 1. *If $r = 2$ then*

$$T = \left\langle b_k, b_{p-k} | b_k^m = b_{p-k}^{p-m+1} = 0 \right\rangle$$

where $1 \le k \le p - 1, 2 \le m \le p - 1$.

Proof. Let

$$T = \left\langle b_k, b_l | b_k^m = b_l^{p-m+1} = 0 \right\rangle$$

for some $1 \le k, l \le p - 1, k \neq l$. We may assume that $k < l$. Since

$$\{b_k, b_k^2, \dots, b_k^{m-1}, b_l, b_l^2, \dots, b_l^{p-m}\} = \{a, a^2, \dots, a^{p-1}\}$$

we obtain, regarding b_i^j as the element $a^{ij} \in G$:

$$k + 2k + \dots + (m-1)k + l + 2l + \dots + (p-m)l \equiv 1 + 2 + \dots + (p-1) \pmod p$$

or

$$m(m-1)k + m(m-1)l \equiv 0 \pmod p.$$

Therefore $k + l \equiv 0 \pmod p$ because $2 \le m \le p - 1$. □

Corollary. *The number of all distinct (but maybe isomorphic) 2-generated commutative Doss modifications is equal* $(p-1)(p-2)/2$.

The situation for $r = 3$ is more complicated. We need a number-theoretic lemma:

Lemma 2. *Let* k, α, x *be integers,* $\frac{p+2}{3} \leq \alpha \leq x \leq p-3$. *Denote the residue* $kx(\mathit{mod}\ p)$ *by* ρ_k *(i.e.* $0 \leq \rho_k \leq p-1$*). If* $\rho_k \geq \alpha$ *for all* $1 \leq k \leq (p-\alpha)/2$ *then* $p \equiv 1(\mathit{mod}\ 3), a = (p+2)/3$ *and* $x = (p-1)/2$.

Proof. If $x \neq (p-1)/2$ then either $x \leq (p-3)/2$ or $x \geq (p+1)/2$. So we consider three cases.

1. Let $x \leq (p-3)/2$ (then $(p+2)/3 \leq (p-3)/2$ whence $p \geq 13$).

We shall prove by induction by i that if $r_1, r_3, \ldots, r_{2i-1} \geq (p+2)/3$ then $\rho_{2i+1} = (2i+1)x - ip$.

For $i = 0$ it is evidently. Let

$$\rho_{2i-1} = (2i-1)x - (i-1)p \geq (p+2)/3.$$

Then

$$(2i+1)x - ip = \rho_{2i-1} + 2x - p \geq 0,$$

$$(2i+1)x - (i+1)p = \rho_{2i-1} + 2x - 2p < p + (p-3) - 2p < 0$$

and

$$\rho_{2i+1} = (2i+1)x - ip.$$

Therefore if it holds $\rho_{2k+1} \geq \alpha \geq (p+2)/3$ for all $k, 2k+1 \leq (p-\alpha)/2$, by the condition of Lemma 2 then $\rho_{2k+1} = (2k+1)x - kp$ for these values of k. However, if we chose $2k+1$ to be equal to the odd integer being between $(p-\alpha)/2 - 2$ and $(p-\alpha)/2$ then $k \geq (p-\alpha-5)/4$ and

$$\begin{aligned} \rho_{2k+1} &= x - k(p-2x) \leq x - \frac{p-\alpha-5}{4}(p-2x) = x\frac{p-\alpha-3}{2} - p\frac{p-\alpha-5}{4} \\ &\leq \frac{p-3}{2}\frac{p-\alpha-3}{2} - p\frac{p-\alpha-5}{4} = \frac{3\alpha-p+9}{4}. \end{aligned}$$

Since $(p+2)/3 \leq \alpha \leq (p-3)/2$ and $p \geq 13$, the last expression is less or equal $(\alpha+6)/4 < \alpha$.

2. Let $x \geq (p+1)/2$. We shall prove by induction that if $\rho_1, \ldots, \rho_{k-1} \geq (p+2)/3$ then $\rho_k = kx - (k-1)p$.

It is true for $k \leq 2 : \rho_1 = x$ and $\rho_2 = 2x - p$ because $2x > p$. So we may assume that $k \geq 3$.

Let $\rho_{k-1} = (k-1)x - (k-2)p \geq (p+2)/3$. Then it follows $kx < kp$ from $x \leq p-3$. On the other hand

$$\begin{aligned} kx &= \frac{k}{k-1}(\rho_{k-1} + (k-2)p) \\ &\geq \frac{k}{k-1}\left(\frac{p+2}{3} + (k-2)p\right) = (k-1)p + \frac{(k-3)p+2k}{3(k-1)}. \end{aligned}$$

The last summand is greater than $(k-1)p$ because $k \geq 3$. Hence $\rho_k = kx - (k-1)p$. Then we have for $k = [(p-\alpha)/3] + 1$:

$$\rho_k = p - k(p-x) < p - 3\frac{p-\alpha}{3} = \alpha.$$

3. Let $x = (p-1)/2$ and $k = 2[(p-\alpha+2)/4] - 1$. Then

$$kx = \left(\frac{p-\alpha+2}{4} - \frac{1}{2}\right)(p-1) \equiv \frac{p+1}{2} - \frac{p-\alpha+2}{4} (\mathrm{mod}\ p).$$

It is evidently that the right side of this congruence is equal to ρ_k. If $\rho_k \geq \alpha$ we obtain:

$$\alpha < \frac{p+1}{2} - \frac{p-\alpha+2}{4} + 1,$$

i.e. $\alpha < (p+4)/3$. Since α is an integer and $\alpha \geq (p+2)/3$, the assertion of Lemma is proved. □

Now we are able to describe the 3-generated modifications.

Theorem 2. *If $r = 3$ then under suitable choice of a generator of G*

$$T = \left\langle b_1, b_m, b_{p-1} | b_1^m = b_m^2 = b_{p-1}^{p-m} = 0 \right\rangle,$$

where $2 \leq m \leq p-2$.

Proof. Let

$$T = \left\langle b_k, b_m, b_n | b_k^\alpha = b_m^\beta = b_n^\gamma = 0 \right\rangle,$$

where

$$\alpha + \beta + \gamma = p + 2 \tag{1}$$

and $2 \leq \alpha, \beta, \gamma \leq p-2$ (whence $p \geq 5$). It is possible to choose the generator of G such that $k = 1$. Furthermore, we may assume that $\alpha \geq \beta, \gamma$ and $m < n$.

At first we shall show that $n = p - 1$. Let $n < p - 1$ on the contrary. Then it follows from (1) that $\alpha \geq (p+2)/3$ and either $\beta - 1$ or $\gamma - 1$ is $\geq (p - \alpha)/2$.

Let $\beta - 1 \geq (p - \alpha)/2$ for example. Since $b_k = b_1^k \notin \langle b_m \rangle$ for $k < \alpha$ all residues of $m, 2m, (\beta - 1)m$ are $\geq \alpha$ and we have from Lemma 2 $\alpha = (p+2)/3, m = (p-1)/2$. But then $\beta \geq (p+2)/3$ and $\alpha \geq \beta$ implies $\alpha = \beta = \gamma$. Again using Lemma 2 and taking into account that $n > m$ we obtain $n = p - 2$.

If $b_m^4 \neq 0$ then $b_m^4 = b_{p-2} = b_n$ what is impossible. So $\beta \leq 4$ whence $p = 7$ and

$$T = \langle b_1, b_3, b_5 | b_1^3 = b_3^3 = b_5^3 = 0 \rangle .$$

However one can check easily that such a modification doesn't exist.

The case $\gamma - 1 \geq (p - \alpha)/2$ is considered analogously.

Therefore $n = p - 1$ and

$$\begin{aligned} &\langle b_{p-1} \rangle = \{ b_{p-\gamma+1}, b_{p-\gamma+2}, \dots, b_{p-1} \} \cup 0 \\ &\langle b_m \rangle = \{ b_\alpha, b_{\alpha+1}, \dots, b_{p-\gamma} \} \cup 0. \end{aligned} \tag{2}$$

Lemma 3. *If* (2) *carries out then the inequalities*

$$km \geq (k-1)p$$

$$\frac{\alpha + (k-1)p}{k} \leq m \leq p - \gamma$$

are true for all $k \leq \beta - 1$.

Proof of Lemma. For $k = 1$ first inequality is evident and last one turns to the form

$$\alpha \leq m \leq p - \gamma$$

what follows from (2).

Let Lemma was proved for $k - 1$, i.e.

$$(k-1)m \geq (k-2)p \tag{3}$$

$$\frac{\alpha + (k-2)p}{k-1} \leq m \leq p - \gamma. \tag{4}$$

Suppose that $km < (k-1)p, k \geq 2$. Since $km > (k-2)p$ from (3) then

$$b_m^k = b_{km-(k-2)p}$$

whence

$$\alpha \le km-(k-2)p \le p-\gamma, \qquad \frac{\alpha+(k-2)p}{k} \le m \le \frac{(k-1)p-\gamma}{k}.$$

From here and (4) it follows

$$\frac{\alpha+(k-2)p}{k-1} \le \frac{(k-1)p-\gamma}{k}, \quad \text{i.e.} \quad k\alpha+(k-1)\gamma \le p.$$

In particular, for $k=2$ we have:

$$2\alpha+\gamma \le p = \alpha+\beta+\gamma-2, \qquad \beta \ge \alpha+2,$$

in contradiction with maximality of α.

Hence $km \ge (k-1)p$. But then it follows from $b_m^k = b_{km-(k-1)p}$ that

$$\alpha \le km-(k-1)p \le p-\gamma$$

whence

$$m \ge \frac{\alpha+(k-1)p}{k}. \qquad \square$$

Ending of the proof of Theorem 2. For $k=\beta-1$ we obtain consecutively:

$$\frac{\alpha+(\beta-2)p}{\beta-1} \le p-\gamma, \quad \beta\gamma-\beta \le 2\gamma-2, \quad \beta \le 2.$$

Therefore $\beta=2, \langle b_m\rangle = \{b_m, 0\}$ and $\alpha = p-\gamma = m$. $\square$

Finally we formulate two problems:

1. Find necessary and sufficient conditions under which the nilpotent semigroup with 0-cancellation is a modification of some finite group.

2. Let $\mathfrak{M}$ be the set of all modifications of a given group G, which are different from G^0(G with a joined zero). Let us define a partial order on $\mathfrak{M}: G(*) > G(\circ) \Leftrightarrow \forall x, y \in G(x*y=0 \Rightarrow x\circ y = 0)$. Describe the maximal elements of poset $\mathfrak{M}$ (at least for $G=\mathbb{Z}_p$).

References

[1] R. Doss, *Sur l'immersion d'une demi-groupe dans un groupe*, Bull.Sci. Math. **72** (1948), 139–150.

[2] B.V. Novikov, *On partial cohomologies of semigroups*, Semigroup Forum **28** (1984), 353–364.

[3] B.V. Novikov, *On the Brauer monoid*, Matem. Zametki (in Russian, submitted).

[4] M.E. Sweedler, *Weak cohomology*, Contemp. Math. **13** (1982), 109–119.

Saltovskoye shosse 258, apt. 20, 310178 Kharkov, Ukraine

FILOMAT (Niš) 9:3 (1995), 873–878
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

PARTIAL COMPLETIONS OF BOOLEAN ALGEBRAS

Žikica Perović

ABSTRACT. Similarly to notion of completeness and k-completeness we define a notion of partial completness with respect to a subalgebra C. We also give a categorial characterization of partially complete Boolean algebras, and a construction of partial completion.

Since the completion of a relatively small Boolean algebra could be very large, it is reasonable to consider the possibility of making a completion of just a subalgebra of given Boolean algebra. This makes even more sense for various types of lattices, where the completion of the whole lattice is not anymore in the same variety.

1. Partially complete Boolean algebras

To make a motivation for our definition, we will list some known facts on complete Boolean algebras.

Proposition 1.1. *(i) Boolean algebra B is complete iff for every partition $(b_i)_{i\in I}$, the mapping $\varphi : B \mapsto \prod_{i\in I}(b_i)$ defined by $\varphi(b) = (b \cdot b_i)_{i\in I}$ is an isomorphism. (b_i), as usually, denotes the principal ideal generated bu b_i.*

(ii) Let κ be a cardinal. Boolean algebra B is κ-complete iff every disjoint family D, of cardinality less than κ, could be extended to a partition $(b_i)_{i\in I}$ so that the mapping $\varphi : B \mapsto \prod_{i\in I}(b_i)$ defined by $\varphi(b) = (b \cdot b_i)_{i\in I}$ is an isomorphism.

Proof. (i) φ is trivialy a monomorphism, and it is onto since B is complete. On the other hand, let $(b_i)_{i\in I}$, be a disjoint family in B. Let now $(b_i)_{i\in I_1}$ be its extension to a maximal disjoint family which is a partition of one in B. Consider a sequence $(e_i)_{i\in I_1} \in \prod_{i\in I_1}(b_i)$ such that $e_i = b_i$ for $i \in I$, and 0 otherwise. The element corresponding to this sequence in the above isomorphism is $\sum_{i\in I} b_i$. Since every disjoint family in B has supremum, B is complete.

(ii) If B is κ-complete, then for every disjoint family of cardinality κ we just add the complement of its summ, to make the desired partition, and construct the isomorphism in the same way as in (i). For the other side, proof is just the same as in (i). □

Having this proposition in mind we define a notion of partial completeness of a Boolean algebra over its subalgebra.

Definition. Let B be a Boolean algebra, and C its subalgebra. B is partially complete with respect to C (shorter "C-complete") if for every partition $(c_i)_{i\in I}$ of C, the mapping $\varphi : B \mapsto \prod_{i\in I}(c_i)$, defined by $\varphi(b) = (b\cdot c_i)_{i\in I}$ is an isomorphism.

Proposition 1.2. *Let B be a Boolean algebra, and C its subalgebra so that B is C-complete.*

i) Let $b \in B$, $(b_i)_{i\in I} \subset B$. For every partition $(c_i)_{i\in I}$, in C, $b = \sum_{i\in I} b\cdot c_i$, and there exists $\sum_{i\in I} b_i \cdot c_i$.

ii) C is a regular subalgebra of B.

iii) for every $D \subset C$, there exists $supD$ in B.

Property i) is equivalent to B being C-complete.

Proof. i) Follows from the fact that the induced mapping $\varphi : B \cong \prod_{i\in I}(c_i)$ is an onto mapping.

ii) It is enough to prove the preservation for disjoint sums. So let $c = \sum_{i\in I} c_i$ in C. Then, $\{c_i : i \in I\} \cup \{c'\}$ is a partition in C, and the element in $\prod_{i\in I}(c_i) \times (c')$ having coordinates c_i, $i \in I$, and 0 on the coordinate c' corresponds to c.

iii) In the case of disjoint family it is just special case of i) for $b = 1$. In the general case we can disjointize it on every step. □

Definition. Let $C < B, S$. A homomorphism $\varphi : B \to S$ is C-complete if it preservs existing C-sums i.e. for every $\{c_i : i \in I\} \subset C$, $\varphi(\sum_{i\in I} c_i) = \sum_{i\in I} \varphi(c_i)$.

Proposition 1.3. *If B is C-complete then for any completion $\bar{C}$ of C there exists a C-complete embedding $m : \bar{C} \to B$, so that the following diagram commutes:*

$$\begin{tikzcd}
C \arrow[r, "e"] \arrow[dr, "i"'] & \overline{C} \arrow[d, "m"] \\
& B
\end{tikzcd}$$

Proof. Let $\bar{B}$ be a completion of B, and let $e_1 : B \to \bar{B}$ be the inclusion embedding. Since $f = e_1 \circ i$ is a complete embedding of C into complete Boolean algebra $\bar{B}$, by the Sikorski extension theorem, there exists a complete embedding $m : \bar{C} \to \bar{B}$, so that $f = m \circ e$.

$$\begin{tikzcd}
C \arrow[r, "e"] \arrow[d, "i"'] & \bar{C} \arrow[d, "m"] \\
B \arrow[r, "e_1"] & \bar{B}
\end{tikzcd}$$

Let us prove that $Im(m) \subset B$. Since every $d \in \bar{C}$ is of the form $a = \sum e(D)$, for some $D \subset C$, we have:

$m(a) = m(\sum e(D)) = \sum m(e(D)) = \sum f(D) = \sum D \in B$

Hence m is the embedding with the desired properties. □

We will prove the anologue of Sikorski's extension theorem.

Proposition 1.4. *Let B be a Boolean algebra, and C its subalgebra. If B is C-complete, then for every Boolean algebra A, and its subalgebra S and any homomorphism $f : S \to B$, such that $Im(f) \subset C$, there exists a homomorphism $g : A \to B$ satisfying $g \circ s = f$.*

Proof. Let us consider the following diagram:

$$\begin{tikzcd}
S \arrow[r, "s"] \arrow[d, "f_2"'] & A \arrow[dl, "g_1"] \\
\overline{C} \arrow[dr, "m"] & \\
C \arrow[u, "e"] \arrow[r, "i"] & B
\end{tikzcd}$$

Let f_1 denote f with the restricted codomen C, and $f_2 = e \circ f_1$. m is the mapping from the preceeding proposition. Since $\bar{C}$ is a complete Boolean algebra, by the Sikorski extension criterion, there exists homomorphism g_1, so that the upper diagram commutes. $g = m \circ g_1$. Let us prove that $g \circ s = f$. Really,

$g \circ s = m \circ g_1 \circ s = m \circ f_2 = m \circ e \circ f_1 = i \circ f_1 = f$

□

We define a completion of a Boolean algebra B over a subalgebra C analogously to definition of a completion.

Definition. Let $C < B < S$. B is C-dense in S, if for every element $a \in S$ there exists a partition $(c_i)_{i \in I}$ of 1 in C and a family $(b_i)_{i \in I} \subset B$, so that $a = \sum i \in I b_i \cdot c_i$.

Definition. Let $C < B$. Boolean algebra S is a C-completion of B if:
(i) S is a C-complete Boolean algebra.
(iii) B is C-dense in S.

2. Construction of partial completions

Constructing a partial completion of a Boolean algebra we will use sheafs over a subalgebra following [2]. Sheaf of Boolean algebras is a generalization of the notion of subdirect product of an indexed family of sets so that the index set and the members of the family have topological structure. Then Boolean algebra is represented as a set of continuous choice functions. We will just mention here the definition and the main representation theorem.

Definition. Let S and X be topological spaces, π a mapping $\pi : S \to X$ and $\mathcal{B} = (B_p)_{p \in X}$ a family of Boolean algebras indexed by a set X. $\mathcal{S} = (S, \pi, X, \mathcal{B})$ is a sheaf of Boolean algebras if it satisfies the following conditions:
(i) $(B_p)_{p \in X}$ is a partition of S.
(ii) π is a projection i.e. $\pi[B_p] = \{p\}$ and π is continuous, open and a local homeomorphism.
(iii) Let $u \subset X$ be an open subset, and $f_1, \ldots, f_n \in \prod_{p \in u} B_p$ continuous functions from u to S, and $t(x_1, \ldots, x_n)$ and $t_1(x_1, \ldots, x_n)$ Boolean terms. Then, the set

$$\{p \in u : t(f_1(p), \ldots, f_n(p)) = t_1(f_1(p), \ldots, f_n(p))\}$$

is open.

Boolean algebras B_p are called stalks of the sheaf $\mathcal{S}$. The set of sections over u is usually denoted by $\Gamma_u(\mathcal{S})$, and the set of global sections by $\Gamma(\mathcal{S})$. For two sections f, g over u, $\|f = g\|_u = \{p \in u : f(p) = g(p)\}$. It is easy to see that algebra of global sections is a subdirect product of the stalks.

Definition. Let B be a Boolean algebra, $C < B$ and $X = UltC$. Let for $p \in X$, $< p >^{fi}$ be filter in B generated by p, $B_p = B/\langle p\rangle^{fi}$. Let also, for $b \in B$, $f_b : X \to S$ be the mapping defined by $f_b(p) = b/\langle p\rangle^{fi}$ and $\pi_p : B \to B_p$ be the canonical homomorphism. Finally, let $S = \bigcup_{p\in X} B_p$ be the topological space having $\mathcal{D} = \{f_b[u] : b \in B, u \subset X, \text{open}\}$ as a base of topology, and $\pi : S \to X$ projection. $\mathcal{S} = (S, \pi, X, (B_p)_{p\in X})$ is called the sheaf of B over subalgebra C.

Theorem 2.1. *Let the notation be as in the preceeding definition. $\mathcal{S}$ is a sheaf of Boolean algebras. Boolean algebra B is isomorphic to Boolean algebra of global sections of $\mathcal{S}$.*

Following ideas from [1], we define an algebra of dense open sections over a subalgebra.

Definition. We say a function $f \in \prod B/P$ is a dense open section of B if the set of all points at which f is continuous is a dense open set of X. $\Gamma_D(B)$ is the set of dense open sections of B. For the congruence relation $\cong$ on $\prod B/P$ defined by: $f \sim g$ if they agree on a dense open subset of X, $\Gamma_D B/\sim$ will be denoted by $\mathcal{R}(B)$, the algebra of dense-open sections of B.

We summarize a few properties of $\mathcal{R}(B)$. s denotes the isomorphism from B onto $ClopUltB$ from the Stone duality.

Proposition 2.2. *(i) A function $f \in \prod B/P$ is continuous at point $P \in X$ iff there is some $c \in P$ and some $b \in B$ so that f agrees with f_b on $s(c)$.*

(ii) The mapping $\varphi : \Gamma(B) \mapsto \mathcal{R}(B)$ defined by $\varphi(f) = f/\sim$, is an embedding, into a dense subalgebra.

Proof. (i) It is easy to get from the definition the known fact that $f_b(u)$, $b \in B$, u open in X, constitute the base for the topology of S. Therefore, $\{f_b(x) | x \in P\}$ is a neighboorhood basis for the point a/P. Hence, if f agrees with f_a on $s(x)$ then f is continuous at each point of $s(x)$. On the other hand, if f is continuous at P, then $f(P)$ equals b/P, for some $b \in B$, hence $f(P) = f_b(P)$. Since $f_b[X]$ is an open neighboorhood of b/P, by continuity of f, there exists a neighboorhood u of P which is mapped into $f_b[X]$, meaning that f and f_b agree on u. c is then, any member of B such that $s[c] \subset u$.

(ii) It is obviously a homomorphism. Let us check that it is $1-1$. So suppose $\varphi(f) = 0/\sim$. This would mean that $\|f \neq 0\|$ is of first category, which is impossible since it is a non-empty open set. To prove that $Im(\varphi)$ is dense in $\mathcal{R}(B)$, suppose that f is a nonzero dense open section. By part (i) of this proposition, there exists $c \in C$ and $b \in B$, so that $f = f_b$ on $S(c)$. Then for the global section g defined as f_b on $S(c)$, and zero otherwise, we have $g \leq f$. □

In the sequal C-section will denote a memeber f of ΓB such that there exists $c \in C$ such that $f(P) = 1$ for $c \in P$, and zero otherwise. The set of C sections will be denoted by ΓC. By Theorem 2.1, C is isomorphic to ΓC, and we will identify them.

Proposition 2.3. *$\mathcal{R}(B)$ is C-complete.*

Proof. Consider $\{f_c : c \in D\}$, a family of C-sections. Let f be the dense open section defined by: $f(P) = 1$ iff $c \in P$ for some $c \in D$, 0 otherwise. It is really a dense open section, since , for $U = \bigcup\{S(c); c \in D\}$, it is continuous in every point of $U \cup intU^c$. □

Proposition 2.4. *$\Gamma B/\sim$ is C-dense in $\mathcal{R}(B)$.*

Proof. Let $f \in \mathcal{R}(B)$. Let it be continuous on a dense open set U. Let further $\{c_i : i \in I\}$ be a maximal disjoint family of clopen sets in U. Wlog, we can suppose that for each $i \in I$ there exists $b_i \in B$ so that for every $P \in c_i$, $f(P) = b_i/P$. If this is not the case, for every $P \in c_i$, we can find a neighbourhood so that for some $b \in B$, $f = f_b$ on that neighbourhood, and then find the finite subcover of every c_i, and finally take the refinement of $\{c_i : i \in I\}$. It is obviously $f/\sim = \sum_{i\in I} f_{b_i \cdot c_i}/\sim$. □

The folowing theorem directly follows from the preceeding propositions.

Theorem 2.5. *Let $C < B$. $\mathcal{R}(B)$ is a C-completion of B.*

References

[1] J. Harding, *Completions of orthomodular lattices II*, Order **10** (1993), 283-294.

[2] S. Koppelberg, *Handbook of Boolean algebras v.1 (J.D. Monk, R. Bonnet, eds.)*, North Holand, New York, 1989.

Žikica Perović, Filozofski fakultet, Ćirila i Metodija 2, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 879–886
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SOME CONGRUENCES ON AN AG^{**}-GROUPOID

Petar V. Protić and Milan Božinović

Abstract. Some congruences on AG^*-groupoids have been considered in [3]. In this paper we shall describe some congruences on AG^{**}-groupoids.

1. Preliminaries

If on groupoid S the following is true

$$(\forall a, b, c \in S)\ (ab)c = (cb)a, \tag{1}$$

then S is an *AG-groupoid* (Abel-Grassmann's groupoid), [4]. In an AG-groupoid, clearly, holds *medial law*

$$(ab)(cd) = (ac)(bd), \tag{2}$$

for every $a, b, c, d \in S$.

Let on AG-groupoid S the following is true

$$a(bc) = b(ac) \tag{3}$$

for every $a, b, c \in S$. This class of AG-groupoids we shall call an AG^{**}-groupoids. If S has left identity then S is an AG^{**}-groupoid, [8]. Let S be an AG^{**}-groupoid and $a, b, c, d \in S$, then

$$(ab)(cd) = c((ab)d) = c((db)a) = (db)(ca). \tag{4}$$

An AG-groupoid S is called an *inverse AG-groupoid* if for every $a \in S$ there exists $a' \in S$ such that $(aa')a = a, (a'a)a' = a'$ and a' is an *inverse* for a, [9]. As usually we shall denote by $V(a)$ the set of all inverses of $a \in S$. If $a, b \in S,\ a' \in V(a),\ b' \in V(b)$ then $a'b' \in V(ab)$, [9].

For undefined notions and notations we refer to [1],[2], [6] and [10].

2. The congruence σ

Lemma 2.1. *Let S be an AG^{**}-groupoid and $E(S) \neq \emptyset$, then $E(S)$ is a semilattice.*

Proof. If $e, f \in E(S)$, then by (4) and (2) it follows that

$$ef = (ee)(ff) = (fe)(fe) = (ff)(ee) = fe. \quad \square$$

The basic definitions of congruences on an AG-groupoid are given in [9] and those definitions are analogous with those in semigroup theory.

Theorem 2.1. *Let S be an AG^{**}-groupoid and let $E(S) \neq \emptyset$, then the relation σ defined on S with*

$$\sigma = \{(x,y) \in S \times S \mid (\exists e \in E(S))ex = ey\}$$

is a congruence relation on S and $e\sigma f$, for every $e, f \in E(S)$. Furthermore, $\sigma = \{(x,y) \in S \times S \mid (\exists e \in E(S))xe = ye\}$.

Proof. Clearly, σ is a reflexive and symmetric relation. Let $x\sigma y$, $y\sigma z$, then $ex = ey$, $fy = fz$ for some $e, f \in E(S)$. Now by (1), (2), (3) and (4) we have

$$\begin{aligned}(ef)x &= ((ee)f)x = (xf)(ee) = (ef)(ex) = (ef)(ey) = (ee)(fy)\\ &= (ee)(fz) = (ef)(ez) = (zf)(ee) = (zf)e = (ef)z.\end{aligned}$$

Since $ef \in E(S)$ we conclude $x\sigma z$ and σ is a transitive relation. Hence, σ is an equivalence relation.

Let $x\sigma y$ and $z \in S$, then $ex = ey$ for some $e \in E(S)$. Now we have

$$\begin{aligned}e(xz) &= (ee)(xz) = (ex)(ez) = (ey)(ez) = (ee)(yz) = e(yz),\\ e(zx) &= (ee)(zx) = (ez)(ex) = (ez)(ey) = (ee)(zy) = e(zy).\end{aligned}$$

Hence, σ is a congruence relation.

Let $e, f \in E(S)$, then since $E(S)$ is a semilattice we have

$$efe = eef = eff.$$

Hence, $e\sigma f$ for every $e, f \in E(S)$.

Let β be a relation defined on S with

$$\beta = \{(x,y) \in S \times S \mid (\exists e \in E(S))xe = ye\}.$$

If $(x,y)\in\beta$, then there exists $e\in E(S)$ such that $xe=ye$. Now, by (1) we have

$$ex=(ee)x=(xe)e=(ye)e=(ee)y=ey,$$

so $(x,y)\in\sigma$. Conversely, if $(x,y)\in\sigma$,then there exists $f\in E(S)$ such that $fx=fy$. Now, we have

$$\begin{aligned}xf&=x(ff)=f(xf)=(ff)(xf)=(fx)(ff)=(fy)(ff)\\&=(ff)(yf)=f(yf)=y(ff)=yf,\end{aligned}$$

so $(x,y)\in\beta$. Hence, $\beta\equiv\sigma$.

Now, if $(x,y)\in\sigma$ and $z\in S$, then for some $e\in E(S)$ hold $(xz)e=(yz)e$ and $(zx)e=(zy)e$. □

Corollary 2.1. *Let S be an AG-groupoid with left identity 1, then the relation σ is a smallest congruence on S with property that $e\sigma f$ for every $e,f\in S$.*

Proof. Let τ be an arbitrary congruence with above property, then for $a,b\in S$ from $a\sigma b\iff ea=eb$ we have $ea\tau=eb\tau$. Now, $e\tau a\tau=e\tau b\tau$ and since $1\in E(S)$ it follows that $1\tau a\tau=1\tau b\tau$, whence $a\tau=b\tau$. Hence, $\sigma\subseteq\tau$. □

3. The maximum idempotent-separating congruence μ

Lemma 3.1. *Let S be an inverse AG^{**}-groupoid, then $\mid V(a)\mid=1$ for each $a\in S$.*

Proof. Let $a\in S$ and $x,y\in V(a)$, then

$$xa=x((ay)a)=(ay)(xa)=(ax)(ya)=y((ax)a)=ya,$$

so

$$x=(xa)x=(ya)x=(xa)y=(ya)y=y.$$

Hence, $\mid V(a)\mid=1$. □

If S is an inverse AG^{**}-groupoid then unique inverse for $a\in S$ we denote with a^{-1}. Notice that aa^{-1} is not necessary idempotent.

Example 3.1. Let S be an AG-groupoid defined by the following Cayley table:

	1	2	3	4
1	2	2	4	4
2	2	2	2	2
3	1	2	3	4
4	1	2	1	2 .

Then S is an inverse AG^{**}-groupoid, $e, f \in E(S)$. Elements 1 and 4 are mutually inverse and $1 \cdot 4 = 4$, $4 \cdot 1 = 1$ are not idempotents.

Remark 3.1. We notice that if ρ is a congruence relation on AG^{**}-groupoid S, then S/ρ is an AG^{**}-groupoid. Also, if S is an inverse AG^{**}-groupoid, then S/ρ is an inverse AG^{**}-groupoid and if $(x, y) \in \rho$ then $(x^{-1}, y^{-1}) \in \rho$ and conversely.

Theorem 3.1. *Let S be an inverse AG^{**}-groupoid and $E(S) \neq \emptyset$, then the relation*

$$\mu = \{(a, b) \in S \times S \; a^{-1}a = b^{-1}b\}$$

is an idempotent-separating congruence on S. If on S holds $a^{-1}a \in E(S)$ for all $a \in S$ then μ is a maximum idempotent-separating congruence on S.

Proof. It is clear that μ is an equivalence.

If $a\mu b$, $c \in S$ and $e \in E(S)$ then

$$\begin{aligned}(ac)^{-1}(ac) &= (a^{-1}c^{-1})(ac) = (a^{-1}a)(c^{-1}c)\\ &= (b^{-1}b)(c^{-1}c)) = (b^{-1}c^{-1})(bc) = (bc)^{-1}(bc),\end{aligned}$$

so $ac\mu bc$. Similarly, $ca\mu cb$. Thus μ is a congruence.

Let $e, f \in E(S)$ and $e\mu f$ then $e = ee = ff = f$. Hence, μ is an idempotent-separating.

Let $a^{-1}a \in E(S)$ holds for all $a \in S$. If ρ is an idempotent-separating congruence on S, $a, b \in S$ and $a\rho b$ then from $a^{-1}\rho b^{-1}$ we have $a^{-1}a\rho b^{-1}b$. Since ρ is idempotent-separating it follows that $a^{-1}a = b^{-1}b$, whence it follows that $a\mu b$ and $\rho \subseteq \mu$. □

4. A congruence pair

Example 4.1. Let S be an AG-groupoid defined by the following Cayley table:

	1	2	3	4	5
1	2	1	1	1	1
2	1	2	2	2	2
3	1	2	4	5	3
4	1	2	3	4	5
5	1	2	5	3	4

Then S is an inverse AG^{**}-groupoid, $a = a^{-1}$ for every $a \in S$ and $aa^{-1} = a^{-1}a$.

In this section with S we shall denote the inverse AG^{**}-groupoid in which $aa^{-1} = a^{-1}a$ holds for every $a \in S$.

Lemma 4.1. *If $a \in S$ then $a^{-1}a \in E(S)$.*

Proof. Let $a \in S$ then

$$(a^{-1}a)(a^{-1}a) = a^{-1}((a^{-1}a)a) = a^{-1}((aa^{-1})a) = a^{-1}a \in E(S). \quad \Box$$

Hence, $E(S) \neq \emptyset$.

Definition 4.1. Let K be a subset of S, then: K is *full* if $E(S) \subseteq K$; K is *self-conjugate* if $x^{-1}(Kx) \subseteq K$ for every $x \in S$; K is *inverse closed* if from $x \in K$ it follows, $x^{-1} \in K$; K is *normal* if it is full, self-conjugate and inverse closed.

Let ρ be a congruence on S. The restriction $\rho_{|_{E(S)}}$ is the *trace* of ρ to be denoted by $tr\rho$, and the set $ker\rho = \{a \in S \mid (\exists e \in E(S))\ a\rho e\}$ is the *kernel* of ρ.

Lemma 4.2. *Let ρ be a congruence relation on S, then $ker\rho$ is a normal subgroupoid of S.*

Proof. If $a, b \in ker\rho$, then $a\rho e$, $b\rho f$ for some $e, f \in E(S)$. Now $ab\rho ef$ and since $ef \in E(S)$ we have that $ab \in ker\rho$. Hence, $ker\rho$ is a subgroupoid of S.

Clearly, $ker\rho$ is full.

Let $a \in S$, then $a(ker\rho \cdot a^{-1}) = \{a(ba^{-1}) \mid b \in ker\rho\}$. From $b \in ker\rho$ we have that $b\rho e$ for some $e \in E(S)$ so $a(ba^{-1})\rho a(ea^{-1})$. Since $a(ea^{-1}) = e(aa^{-1}) \in E(S)$, then $a(ba^{-1}) \in ker\rho$. Hence, $a(ker\rho \cdot a^{-1}) \subseteq ker\rho$ and $ker\rho$ is a self-conjugate subgroupoid of S.

If $x \in ker\rho$, then $x\rho e$ for some $e \in E(S)$ and $x^{-1}\rho e^{-1} = e$. Hence, $x^{-1} \in ker\rho$ and $ker\rho$ is inverse closed.

By above we conclude that $ker\rho$ is a normal subgroupoid of S. $\Box$

Definition 4.2. Let K be a normal subgroupoid of S and τ congruence on semilattice $E(S)$ such that

$$ea \in K,\ e\tau a^{-1}a \Longrightarrow a \in K \tag{5}$$

for every $a \in S$ and $e \in E(S)$. Then the pair (K, τ) is a *congruence pair* for S.

In such a case, we can define a relation $\rho_{(K,\tau)}$ on S by

$$a\rho_{(K,\tau)}b \Longleftrightarrow a^{-1}a\tau b^{-1}b,\ ab^{-1},\ ba^{-1} \in K. \tag{6}$$

Lemma 4.3. *For a congruence pair (K,τ) for S, we have*

$$e(ab)\in K,\ e\tau a^{-1}a \Longrightarrow ab\in K$$

for any $a,b\in S$, $e\in E(S)$,

Proof. Let $a,b\in S$, $e\in E(S)$, $e(ab)\in K$ and $e\tau a^{-1}a$, then

$$\begin{aligned} e(ab) &= (ee)(ab) = (be)(ae) = (((bb^{-1})b)e)(ae) = ((eb)(bb^{-1}))(ae) \\ &= (b((eb)b^{-1}))(ae) = (b((b^{-1}b)e))(ae) = (e((b^{-1}b)e))(ab) \\ &= ((b^{-1}b)e)(ab) = (e(b^{-1}b))(ab), \\ (ab)^{-1}(ab) &= (a^{-1}b^{-1})(ab) = ((ab)b^{-1})a^{-1} = ((b^{-1}b)a)a^{-1} = (a^{-1}a)(b^{-1}b) \\ &\quad \tau e(b^{-1}b). \end{aligned}$$

By above and (5) we have $ab\in K$. □

Theorem 4.1. *If (K,τ) is a congruence pair for S, then $\rho_{(K,\tau)}$ is the unique congruence ρ on S for which $ker\rho = K$ and $tr\rho = \tau$. Conversely, if ρ is a congruence on S, then $(ker\rho, tr\rho)$ is a congruence pair for S and $\rho_{(ker\rho,tr\rho)} = \rho$.*

Proof. Let (K,τ) be a congruence pair for S, and let $\rho=\rho_{(K,\tau)}$. Then ρ is reflexive since K is full, and it is symmetric since τ is symmetric. Let $a\rho b$ and $b\rho c$, so that $a^{-1}a\tau b^{-1}b\tau c^{-1}c$ and $ba^{-1}, bc^{-1}\in K$. Since K is inverse closed we have $(ba^{-1})^{-1} = b^{-1}a\in K$. Since K is a substructure we have

$$(b^{-1}a)(bc^{-1}) = (b^{-1}b)(ac^{-1})\in K.$$

From above and $b^{-1}b\tau a^{-1}a$, by Lemma 4.3, it follows that $ac^{-1}\in K$. Thus $a\rho c$ and ρ is transitive.

Next let $a\rho b$ and $c\in S$. Then

$$\begin{aligned} (ac)^{-1}(ac) &= (a^{-1}c^{-1})(ac) = (a^{-1}a)(c^{-1}c) \\ &\quad \tau(b^{-1}b)(c^{-1}c) = (b^{-1}c^{-1})(bc) = (bc)^{-1})(bc). \end{aligned}$$

Also,

$$\begin{aligned} (ac)(bc)^{-1} &= (ac)(b^{-1}c^{-1}) = (ab^{-1})(cc^{-1})\in K\cdot E(S)\subseteq K, \\ (bc)(ac)^{-1} &= (bc)(a^{-1}c^{-1}) = (ba^{-1})(cc^{-1})\in K\cdot E(S)\subseteq K. \end{aligned}$$

Hence, $ac\rho bc$. Similarly, $ca\rho cb$. Therefore ρ is a congruence on S.

If $a \in ker\rho$, then $a\rho e$ for some $e \in E(S)$. Now, $aa^{-1} = a^{-1}a\tau e$ and $ae,\ ea^{-1} \in K$ whence by (5) it follows that $a^{-1} \in K$. Since K is inverse closed we have $a \in K$. Conversely, if $a \in K$, then from $a = (aa^{-1})a \in K$ and $a^{-1}a\tau(a^{-1}a)(a^{-1}a)$ we have $a\rho a^{-1}a$ and $a \in K$. Consequently, $ker\rho = K$; and obviously $tr\rho = \tau$.

Now let λ be a congruence on S such that $ker\lambda = K$ and $tr\lambda = \tau$. Assume first that $a\lambda b$. Then $a^{-1}\lambda b^{-1}$ so that $a^{-1}a\lambda b^{-1}b$ and also $ab^{-1}\lambda bb^{-1} = b^{-1}b$, $ba^{-1}\lambda aa^{-1} = a^{-1}a$. This shows that $a^{-1}a\tau b^{-1}b$ and $ab^{-1},\ ba^{-1} \in ker\lambda = K$, which implies that $a\rho b$ and $\lambda \subseteq \rho$. Conversely, assume that $a\rho b$. Then $a^{-1}a\lambda b^{-1}b$ and $ab^{-1},\ ba^{-1} \in K = ker\lambda$. Now there exist $e, f \in E(S)$ such that $ab^{-1}\lambda e, ba^{-1}\lambda f$ whence $a^{-1}b\lambda e,\ b^{-1}a\lambda f$. From above and $a^{-1}a = aa^{-1},\ b^{-1}b = bb^{-1}$ we have

$$\begin{aligned} ab^{-1} &= ((aa^{-1})a)b^{-1} = (b^{-1}a)(aa^{-1})\lambda f(bb^{-1}),\\ ba^{-1} &= b((a^{-1}a)a^{-1}) = (a^{-1}a)(ba^{-1})\lambda(b^{-1}b)f\,. \end{aligned}$$

and since $E(S)$ is a semilattice it follows that

$$e\lambda = ab^{-1}\lambda = f(b^{-1}b)\lambda = ba^{-1}\lambda = f\lambda.$$

Now

$$\begin{aligned} a &= (aa^{-1})a\lambda(bb^{-1})a = (ab^{-1})b\lambda eb,\\ b &= (bb^{-1})b\lambda(aa^{-1})b = (ba^{-1})a\lambda ea \end{aligned}$$

and by (4)

$$\begin{aligned} a\lambda eb &= e(ea) = (ee)(ea) = (ae)(ee) = (ae)e = ((eb)e)e\\ &= (ee)(eb) = e(eb) = ea\lambda b. \end{aligned}$$

Hence, $\rho \subseteq \lambda$. Consequently, $\rho = \lambda$ which proves uniqueness.

Conversely, let ρ be a congruence on S. By Lemma 4.2 we have that $ker\rho$ is a normal substructure of S. For $a \in S,\ e \in E(S)$ let $ea \in ker\rho,\ e\, tr\rho a^{-1}a$, holds then $ea\rho f$ for some $f \in E(S)$. Now $a = (aa^{-1})a\rho ea\rho f$ and $a \in ker\rho$. Hence, statement (5) holds and $(ker\rho, tr\rho)$ is a congruence pair for S. From above it follows that $ker\rho_{(ker\rho,tr\rho)} = ker\rho,\ tr\rho_{(ker\rho,tr\rho)} = tr\rho$. Now the uniqueness just proved implies that $\rho_{(ker\rho,tr\rho)} = \rho$. □

References

[1] S. Bogdanović, *Semigroups with a system of subsemigroups*, Inst. of Math., Novi Sad, 1985.

[2] S. Bogdanović and M. Ćirić, *Polugrupe*, Prosveta, Niš, 1993.

[3] M. Božinǒvić and P. V. Protić, *Some congruences on a π-inverse LA^*-semigroups*, Facta Universitatis (to appear).

[4] J. Deneš and A. D. Keedwell, *Latin squares and their applications*, Akadémia Kiadó, Budapest, 1974.

[5] J. M. Howie, *The maximum idempotent-separating congruence on an inverse semigroup*, Proc. Edinburg Math. Soc. **14** (1964), 71-79.

[6] J. M. Howie, *An introduction to semigroup theory*, Academic Press, 1976.

[7] M. A. Kazim and M. Naseeruddin, *On almost semigroups*, The Aligarh Bull. Math. **2** (1972), 1-7.

[8] Q. Mushtaq and Q. Iqbal, *Decomposition of a locally associative LA-semigroup*, Semigroup Forum **41** (1990), 155-164.

[9] Q. Mushtaq and M. Iqbal, *Partial ordering and congruences on LA-semigroups*, Indian J. pure appl. Math. **22(4)** (1991), 331-336.

[10] M. Petrich, *Congruences on inverse semigroups*, J. of Algebra **55** (1978), 231-256.

[11] M. Petrich, *Inverse Semigroups*, John Wiley and Sons, New York, 1984.

Faculty of Civil Engineering, Beogradska 14, 18000 Niš, Yugoslavia

Technical Faculty, 19210 Bor, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 887–897
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

THE ALGORITHMS AND DATA STRUCTURES FOR FORMING SYMBOLIC MODELS OF THE ROBOTIC SYSTEMS

Miloš Racković and Dušan Surla

Abstract. One of the basic problems in forming mathematical models of the robotic systems in their symbolic form is the formation of calculation graph for the analytical expesions of robotic quantities. The first step in formation of the calculation graph is the splitting of the expression into the products of the other expressions and the remaining expression. In this paper the necessary and sufficient conditions for splitting of the analytical expressions into the set of products of two expressions are found, and the algorithm for solving this problem is described.

1. Introduction

Significant advancements in the development of mathematical models of robotic systems have been made by numeric-symbolic [1] and symbolic [1-3] methods. An algorithm has been constructed [2,3] to form the mathematical model of a simple kinematic chain in symbolic form, and the program environment SYM [4] has been implemented for modelling of the robotic systems. This algorithm has been modified for complex kinematic chains, using the programming package *Mathematica* [5].

In [6], a network model of database has been proposed for generating the mathematical models of robotic systems in symbolic form. The navigation through the database allows easy formation of the analytical expressions and obtaining numerical values for the corresponding robotic quantities.

The analytical expressions, obtained with the aid of the algorithm from [6] can be simplified by applying trigonometric identities. These expressions

The work was supported by the Ministry for Science and Technology of Serbia, grant no. 0413 through the Institute of Mathematics, Novi Sad

are of the following form:

$$Y = \sum_{i=1}^{N} S_i \tag{1}$$

where each of the addends is of the form:

$$S_i = k_i \cdot \prod_{j=1}^{L} x_j^{e_{ij}} \tag{2}$$

and where:

Y - is the robotic quantity to be calculated;
k_i - is a constant coefficient related to the i-th addend;
x_j -is one of the basic variables of the robotic system model represented by its name $(q, \dot{q}, \ddot{q}, \sin q, \cos q)$. For each addend the same sequence of variables x_j, $j = 1, ..., L$ is used.
e_{ij} - is the exponent of the j-th variable of the i-th addend. The algorithm for forming the mathematical model ensures that each of the exponents is a nonnegative integer number.

The main task is to form the calculation graph for the chosen analytical expressions of the type (1), with the minimum of mathematical operations. To obtain the maximal reduction in the number of mathematical operations this paper proposes splitting of the chosen expression in the form:

$$Y = \sum_{l=1}^{M} (Y_{l1} \cdot Y_{l2}) + Y_{M+1} \tag{3}$$

where Y_{l1}, Y_{l2}, $l = 1, ..., M$ and Y_{M+1} are also the expressions of the type (1).

The expressions Y_{l1}, Y_{l2}, $l = 1, ..., M$ have two addends at least and are determined in a way which maximize the reduction of the number of mathematical operations. Y_{M+1} represents the remainder of the expression Y which can not be split into products any more.

This paper gives the necessary and sufficiant conditions for reducing the expression for Y into that of type (3). The concept of structural matrices which is used in solving this problem, is briefly described in the next paragraph.

2. Structural matrices

The concept of structural matrices was introduced in [1] to represent analytical expressions of the robotic quantities.

Structural matrix S of the expression Y is represented with the vector of coefficients $K_S = \left[k_1^S, ..., k_N^S\right]^\top$, the vector of variables $X_S = \left[x_1^S, ..., x_L^S\right]^\top$ and the matrix of exsponents:

$$E_S = \begin{bmatrix} e_{11}^S & e_{12}^S & \dots & e_{1L}^S \\ e_{21}^S & e_{22}^S & \dots & e_{2L}^S \\ \dots & \dots & \dots & \dots \\ e_{N1}^S & e_{N2}^S & \dots & e_{NL}^S \end{bmatrix}$$

given in the previous paragraph.

In [1], the algebra of structural matrices is introduced, and here is described only the multiplication of the structural matrices because this is important for sloving the assigned problem.

Let us observe the multiplication of two structural matrices which have the same vestors of variables.

$$A = (K_A, X_A, E_A)$$
$$B = (K_B, X_B, E_B)$$
$$X_A = X_B$$

If the exponent matrices E_A i E_B are given with:

$$E_A = \begin{bmatrix} e_{11}^A & e_{12}^A & \dots & e_{1L}^A \\ e_{21}^A & e_{22}^A & \dots & e_{2L}^A \\ \dots & \dots & \dots & \dots \\ e_{I1}^A & e_{I2}^A & \dots & e_{IL}^A \end{bmatrix} E_B = \begin{bmatrix} e_{11}^B & e_{12}^B & \dots & e_{1L}^B \\ e_{21}^B & e_{22}^B & \dots & e_{2L}^B \\ \dots & \dots & \dots & \dots \\ e_{J1}^B & e_{J2}^B & \dots & e_{JL}^B \end{bmatrix}$$

then their product is the exponent matrix E_C where:

$$E_C = \begin{bmatrix} e_{11}^C & e_{12}^C & \dots & e_{1L}^C \\ e_{21}^C & e_{22}^C & \dots & e_{2L}^C \\ \dots & \dots & \dots & \dots \\ e_{M1}^C & e_{M2}^C & \dots & e_{ML}^C \end{bmatrix}$$

and where:

$$M = I \cdot J$$
$$e_{ml}^C = e_{il}^A + e_{jl}^B$$
$$m = (i-1) \cdot J + j$$
$$l = 1, ..., L; i = 1, ..., I; j = 1, ..., J.$$

Also, if the vectors of coefficients K_A and K_B are given by:

$$K_A = \begin{bmatrix} k_1^A \\ k_2^A \\ \dots \\ k_I^A \end{bmatrix}; K_B = \begin{bmatrix} k_1^B \\ k_2^B \\ \dots \\ k_J^B \end{bmatrix}$$

then, their product is the vector of coefficients K_C where:

$$K_C = \begin{bmatrix} k_1^C \\ k_2^C \\ \dots \\ k_M^C \end{bmatrix}$$

and where:

$$M = I \cdot J$$
$$k_m^C = k_i^A \cdot k_j^B$$
$$m = (i-1) \cdot J + j$$
$$i = 1, ..., I; j = 1, ..., J.$$

Thus the structural matrix C is obtained as a product of the structural matrices A and B.

$$C = (K_C, X_C, E_C)$$
$$X_C = X_A = X_B$$

In solving our problem we have an opossite situation. The structural matrix exists and two structural matrices whose product will be the given matrix, are to be found.

3. Splitting of the structural matrix in the form of products

The problem can be broken in two separate problems. The first is to split the exponent matrix and the second is to split the vector of coefficients.

Let us observe the equation

$$E_A \cdot E_B = E_C \tag{4}$$

where E_A, E_B i E_C are the exsponent matrices defined in the previous paragraph. The matrix E_C represent the exsponents of the expression of the type (1) which we want to write in the form of product of the expressions of the same type. The matrices E_A and E_B are the unknowns in this equation. If the solution of the system exists, then the matrices E_A and E_B which satisfy the equation will represent the exponent matrix of the expressions which form the product.

If we write equation (4) as a system we obtain an $I \cdot J \cdot L$ linear equations with $(I + J) \cdot L$ unknowns. The system is of the form:

$$\begin{gathered} e^A_{il} + e^B_{jl} = e^C_{ml} \\ m = (i-1) \cdot J + j \\ l = 1, ..., L; i = 1, ..., I; j = 1, ..., J. \end{gathered} \tag{5}$$

Let us denote vectors $\left[e^A_{i1}, ..., e^A_{iL}\right]^\top$, $\left[e^B_{j1}, ..., e^B_{jL}\right]^\top$ and $\left[e^C_{m1}, ..., e^C_{mL}\right]^\top$ (rows of the matrices E_A, E_B and E_C) with e^A_i, e^B_j and e^C_m respectively. Now, the system can be written in a vector form:

$$\begin{gathered} e^A_i + e^B_j = e^C_m \\ m = (i-1) \cdot J + j \\ i = 1, ..., I; j = 1, ..., J. \end{gathered} \tag{6}$$

where the addition of the vectors is defined in the usual way.

Let us choose $i1$ and $i2$ so that $i1, i2 \in \{1, ..., I\}$ and $i1 \neq i2$, and $j1$ and $j2$ so that $j1, j2 \in \{1, ..., J\}$ and $j1 \neq j2$. Then, we pick the four following equations from the system (6):

$$\begin{gathered} e^A_{i1} + e^B_{j1} = e^C_{m1} \\ m1 = (i1-1) \cdot J + j1 \end{gathered} \tag{7}$$

$$\begin{gathered} e^A_{i1} + e^B_{j2} = e^C_{m2} \\ m2 = (i1-1) \cdot J + j2 \end{gathered} \tag{8}$$

$$\begin{gathered} e^A_{i2} + e^B_{j1} = e^C_{m3} \\ m3 = (i2-1) \cdot J + j1 \end{gathered} \tag{9}$$

$$\begin{gathered} e^A_{i2} + e^B_{j2} = e^C_{m4} \\ m4 = (i2-1) \cdot J + j2 \end{gathered} \tag{10}$$

By summing equations (7) and (10) we obtain

$$e^A_{i1} + e^B_{j1} + e^A_{i2} + e^B_{j2} = e^C_{m1} + e^C_{m4} \tag{11}$$

Also, by summing equations (8) and (9) we obtain

$$e^A_{i1} + e^B_{j2} + e^A_{i2} + e^B_{j1} = e^C_{m2} + e^C_{m3} \tag{12}$$

It follows that following condition must be satisfied

$$e^C_{m1} + e^C_{m4} = e^C_{m2} + e^C_{m3} \tag{13}$$

or in another form

$$e^C_{m1} - e^C_{m3} = e^C_{m2} - e^C_{m4} \tag{14}$$

If the condition (13) is not fulfilled it follows that the system (6) has no solution.

If this procedure is repeated for each combinatin of $i1$, $i2$ and $j1$, $j2$ on the basis of transitivity law for equality, the $I \cdot J$ conditions are obtained which have to be satisfied to provide the solvability of the system (6).

If these conditions are satisfied it can be shown that the system has an infinite set of solutions in the set R^L. For solving the concrete problem where all the exponents which are found in analytical expressions of the robotic quantities have nonnegative integer values, we need to find our solutions of the system (6) in the set $(N \bigcup\{0\})^L$. The solution which belongs to the set $(N \bigcup\{0\})^L$ we will call the allowed solution. First, we can prove that there is at least one solution in the set Z^L.

Lemma 1. *If the conditions (13) are satisfied then exists at least one solution of the system (6) in the set* Z^L.

The lemma is proved by derivation of the solution.

Suppose that conditions (13) hold. Then, for e^A_1 we take the 0 vector. From J equations in which e^A_1 participates we have that:

$$\begin{gathered} e^B_j = e^C_j \\ j = 1, ..., J. \end{gathered} \tag{15}$$

Now for each $e^A_i, i \neq 1$ remains J equations from which follows that:

$$\begin{gathered} e^A_i = e^C_{m_1} - e^C_1 = ... = e^C_{m_J} - e^C_J \\ m_j = (i-1)\cdot J + j \\ i = 2, ..., I. \end{gathered} \tag{16}$$

Let us check the correctness of equation (16). We choose $i1 \in \{2, .., I\}$. For $i2$ we take 1, and for $j1, j2$ we take any combination from the set $\{1, ..., J\}$. Equation (16) follows from the condition (14) for this combination and transitivity law for equality.

Thus the solutions which are in the set Z^L are found for all e^A_i, $i = 1, ..., I$ and e^B_j, $j = 1, ..., J$ because all e^C_m, $m = 1, ..., M$ are from the set Z^L.

Now we can prove that exists an allowed solution.

Theorem 1. *If the conditions (13) are satisfied then exists at least one allowed solution of the system (6).*

Suppose that conditions (13) hold. For e_1^A we obtain each of the components so that $e_{1l}^A = min(e_{1l}^C, ..., e_{Jl}^C)$ for $l = 1, ..., L$. The obtained vector is allowed because all the e_m^C are allowed. From J equations in which e_1^A participates we have that:

$$(17) \qquad e_j^B = e_j^C - e_1^A$$
$$j = 1, ..., J.$$

These solutions are also allowed because $e_{jl}^C \geq e_{1l}^A$ for $j = 1, ..., J$ and $l = 1, ..., L$. Now for each $e_i^A, i \neq 1$ remains J equations from which follows that:

$$(18) \qquad e_i^A = e_{m_1}^C - e_1^C + e_1^A = ... = e_{m_J}^C - e_J^C + e_1^A$$
$$m_j = (i-1) \cdot J + j$$
$$i = 2, ..., I.$$

The correctness of equation (18) is proved in an analogous way as equation (16) from Lemma 1. Now we must prove that all the e_i^A for $i = 2, ..., I$ are alowed.

Suppose that there exists $i1$ and $l1$ for which $e_{i1l1}^A < 0$. Then, choose the $j1$ so that $e_{j1l1}^C = e_{1l1}^A = min(e_{1l1}^C, ..., e_{Jl1}^C)$. Now follows that $e_{j1l1}^B = e_{j1l1}^C - e_{1l1}^A = 0$. The vector e_{j1}^B paticipates in another $I - 1$ equations. From these equations we choose the one in which e_{i1}^A also participates. It follows that $e_{j1l1}^B = e_{m1l1}^C - e_{i1l1}^A = 0$ where $m1 = (i1 - 1) \cdot J + j1$. From this equation follows that $e_{m1l1}^C = e_{i1l1}^A$ and the contradiction is obtained with the assumption that $e_{i1l1}^A < 0$ because $e_{ml}^C \geq 0$, for $m = 1, ..., M$ and $l = 1, ..., L$.

In an analogous way we can solve the problem of splitting the vector of coefficients. Let us observe the equation

$$(19) \qquad K_A \cdot K_B = K_C$$

where K_A, K_B and K_C are the vectors of coefficients defined in the previous paragraph.

If we write equation (19) as a system we obtain $I \cdot J$ equations with $(I + J)$ unknowns. The system is of the form:

$$(20) \qquad k_i^A \cdot k_j^B = k_m^C$$
$$m = (i-1) \cdot J + j$$
$$i = 1, ..., I; j = 1, ..., J.$$

By a procedure analogous to the one for splitting the exponent matrices we obtain the following conditions:

$$k_{m1}^{C} \cdot k_{m4}^{C} = k_{m2}^{C} \cdot k_{m3}^{C} \tag{21}$$

or in another form

$$\frac{k_{m1}^{C}}{k_{m3}^{C}} = \frac{k_{m2}^{C}}{k_{m4}^{C}} \tag{22}$$

which are analogous to the conditions (13) and (14).

The following theorem can be proved:

Theorem 2. *If the conditions (21) are satisfied, then exists at least one solution of the system (20) in a set of real numbers.*

Suppose that conditions (21) hold. For k_1^A we take the value 1.00. From J equations in which k_1^A participates we have that:

$$\begin{gathered} k_j^B = k_j^C \\ j = 1, ..., J. \end{gathered} \tag{23}$$

Now, for each $k_i^A, i \neq 1$ remains J equations from which follows that:

$$\begin{gathered} k_i^A = \frac{k_{m_1}^{C}}{k_1^C} = ... = \frac{k_{m_J}^{C}}{k_J^C} \\ m_j = (i-1) \cdot J + j \\ i = 2, ..., I. \end{gathered} \tag{24}$$

The correctness of equation (24) is proved in an analogous way as equation (16) from Lemme 1.

Now, the complete problem can be formulated. Let us observe the equation

$$A \cdot B = C \tag{25}$$

where A, B and C are the structural matrices. The goal is to obtain the matrices A and B so that equation (25) holds.

Theorem 3. *The structural matrices A and B which satisfy the equation (25) exist if and only if conditions (13) and (21) are fulfilled.*

If the conditions (13) and (21) hold, the existence of the matrices A and B is proved by applying Theoreme 1 and Theoreme 2, and assemblying the corresponding exponent matrix and vector of coefficients into a structural matrix. The theorem in the opposite direction obviously follows from the procedure for obtaining the conditions (13) and (21). In the case when the conditions (13) are not satisfied the contradiction is obtained in the system (6), and if the conditions (21) are not satisfied the contradiction is obtained in the system (20).

4. The algorithm

When the conditions (13) and (21) are derived two rows from the matrix E_A and E_B, i.e. two elements of the vectors K_A and K_B, are chosen. Also, by choosing four rows from the matrix E_C, four elements of the vector K_C are chosen. This means that these four addends from the expression described by the exponent matrix E_C and the vector of coefficients K_C, are obtained by multiplying two expressions each containing two addends which are described by the chosen parts of the matrices E_A and E_B and vectors K_A and K_B. Thus every condition of the type (13) and (21) represent a "2×2" multiplication.

In the same way equations (16) and (18) describe a "$2 \times J$" multiplication because the two rows of the matrix E_A and all J rows of matrix E_B are chosen as well as two elements of the vector K_A and all J elements of the vector K_B. Also are chosen the $2 \cdot J$ rows of the matrix E_C and the same number of elements of the vector K_C.

The algorithm task is to reduce the starting expression Y to the type (3) by splitting the structural matrix S into the products.

The first step is to obtain all possible "$2 \times n$" multiplications, where n is determined as large as possible for each multiplication. The description of this part of the algorithm, written in pseudocode, is given bellow.

```
for ($i1 = 1$ ; $i1 < N$ ; $i1++$)
 {
  for ($i2 = i1 + 1$ ; $i2 \leq N$ ; $i2++$)
   {
    if (NotMem(i1,i2))
     {
      MemPair(i1,i2,mu);
      $s = e^S_{i1} - e^S_{i2}$;
      $q = \frac{k^S_{i1}}{k^S_{i2}}$;
      for ($i3 = 1$ ; $i3 < N$ ; $i3++$)
       {
        for ($i4 = i3 + 1$ ; $i4 \leq N$ ; $i4++$ )
         {
          if ($s == e^S_{i3} - e^S_{i4}$ && $q == \frac{k^S_{i3}}{k^S_{i4}}$)
            MemPair(i3,i4,mu);
          else if ($s == e^S_{i4} - e^S_{i3}$ && $q == \frac{k^S_{i4}}{k^S_{i3}}$)
            MemPair(i4,i3,mu);
         }
       }
```

```
            }
          MemMul(mu);
          FreeMul(mu);
        }
    }
```

Function **NotMem** checks if the given pair is memorised in the knowledge database. If it is the function returns **0** because of the transitivity law for equality there is no sense to check the conditions for that pair.

Procedure **MemPair** memorises the given pair in variable **mu** which will contain the complete multiplication.

Procedure **MemMul** memorises complete multiplication **mu** in the knowledge database.

Procedure **FreeMul** frees the variable **mu** for a new multiplication.

After applying this part of the algorithm we have memorised in the knowledge database all possible "$2 \times n$" multiplications over the structural matrix S. Now, the manipulating over the knowledge database is used to provide such a selection of multiplications which enable the reducing the starting expression Y in the type (3) with a minimum of mathematical operations. Here will be explained only how to produce the "$m \times n$" multiplications using $m-1$ "$2 \times n$" multiplications.

If $m-1$ "$2 \times n$" multiplications (or their parts) are found in the knowledge database having the same first elements of each pair and all different second elements of the each pair, it is easy to prove that a "$m \times n$" multiplying is constructed. From the transitivity law for equality follows that all the conditions for the existing solution of the systems (6) and (20) are satisfied when $I = m$ and $J = n$.

6. Conclusion

The algorithms for forming symbolic mathematical models of complex robotic systems have been developed. With the aim of reducing the number of mathematical operations, the investigations are directed towards the development of the algorithms for symplification of analytical expressions and the formation of a corresponding calculation graph.

A theorem is proved with the necessary and sufficient conditions for splitting the analytical expressions into the products of two expressions. On the basis of this theorem an algorithm for forming the set of candidates for splitting the expressions into products is given. This algorithm is implemented and tested on an example of the robotic mechanism with six rotational degrees of freedom. Of the set of obtained candidates we can choose the ones which reduce the number of mathematical operations. On the concrete example significant reduction of analytical expressions has been achieved.

Further investigations will be concerned with the development of algorithms for grouping the expression candidates memorised in the knowledge database which reduce the number of mathematical operations.

References

[1] M. Vukobratović and N. Kirćanski, *Real-Time Dynamics of Manipulation Robots*, Springer-Verlag, 1985.

[2] A. Timčenko, *A Program Package for Generating Symbolic Models of Manipulation Robots*, M. Sc. Thesis, Beograd, 1989. (In Serbian)

[3] A. Timčenko, N. Kirćanski and M. Vukobratović, *A Two-Step Algorithm for Generating Efficient Manipulator Models in Symbolic Form*, Proceedings of the 1991 IEEE International Conference on Robotics and Automation, Sacramento, California - April 1991, pp. 1887–1892.

[4] A. Timčenko, N. Kirćanski, D. Urošević and M. Vukobratović, *SYM - Program Environment for Manipulator Modeling, Control and Simulation*, Proceedings of the 1991 IEEE International Conference on Robotics and Automation, Sacramento, California - April 1991, pp. 1122–1127.

[5] M. Racković and D. Surla, *Generating the Analytic Expressions for Driving Moments of the Complex Kinematic Chain*, XXXVIII Yugoslav Conference of ETRAN, Niš 1994., pp. 105–106. (In Serbian)

[6] D. Surla and M. Racković, *The Implementation of Symbol Operations into Symbolic Modelling of Robotic Systems*, IX Conference on Aplied Mathematics PRIM'94, Budva 1994, pp. 371–377.

Institute of Mathematics, 21000 Novi Sad, Trg Dositeja Obradovića 4, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 899–910

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON HARTE'S THEOREM FOR REGULAR BOUNDARY ELEMENTS

Vladimir Rakočević

Abstract. This paper is a paraphrase and extension on my talk given at the conference *Algebra, Logic and Dicrete Mathematics*, Niš, April 14-16, 1995, and it is inspired by Harte's theorem (Proc. Amer. Math. Soc. 99(1987), 328-330). In this paper we would like to present some results and problems connected with Harte's theorem.

1. Introduction.

Let S be a semigroup (ring) with identity. The element $a \in S$ is (*(von Neumann) regular* if $a \in aSa$. That is, there is a solution of the equation $axa = a$. These solutions are usually called *inner* or 1-*inverses* of a, and will be denoted by a^-. If in addition, $xax = x$, then we call x a *reflexive* inverse of a, and denote it by a^+. The set of all regular elements in S will be denoted by $\widehat{S}$ and it obviously includes the invertible group S^{-1} and the idempotents $S^\bullet = \{a \in S : a^2 = a\}$. An element a is *unit regular* or *decomposably regular* provided there is $b \in S^{-1}$ such that $aba = a$ ([1], [2]). It is easy to prove that

$$S^{-1}S^\bullet = S^\bullet S^{-1} = \{a \in A : a \in aS^{-1}a\}. \tag{1.0.1}$$

When A is a Banach algebra with the identity 1, Harte [14, Theorem 1.1] (see also [15], [26]) has shown that the decomposably invertible elements are the intersection of the regular elements with the closure of the invertibles (for a subset M of A let δM and $cl\, M$ denote, respectively, the boundary and the closure of M) :

$$A^{-1}A^\bullet = \widehat{A} \cap cl(A^{-1}). \tag{1.0.2}$$

Supported by Grant 0401C of RFNS through Math. Inst. SANU.

Let us remark that the left side in the equality (1.0.2) is purely algebraic, while the right side in (1.0.2) depends on metric properties of A. Hence, the remarkable characteristics of Harte's theorem is that it proves the equality of two different quantities. In this paper we would like to present some results and problems connected with Harte's theorem.

2. Harte's type theorems

In this section A denotes a Banach algebra with identity 1.

Theorem 2.1. *Let A be a Banach algebra with identity 1, and S be a multiplicative semigroup of A, such that $A^{-1} \subset S \subset \widehat{A}$. Then*

$$SA^\bullet = \widehat{A} \cap cl(S) \Longleftrightarrow SA^\bullet \subset \widehat{A}. \tag{2.1.1}$$

Proof. It is enough to prove $\Longleftarrow$. If $a \in \widehat{A} \cap cl(S)$ then there are $a^+ \in A$ and $b \in S$ such that $1+(b-a)a^+ = c \in A^{-1}$. Hence $a+(b-a)a^+a = ca$, i.e., $a = (c^{-1}b)a^+a$, and $a \in SA^\bullet$. To prove '$\subset$', suppose that $a \in SA^\bullet$. Hence, $a \in \widehat{A}$, and there are $c \in S$ and $p \in A^\bullet$ such that $a = cp$. Set $p_n = p - 1/n$, $n = 2, 3, \ldots$, and $a_n = cp_n$. It is clear that $p_n \in A^{-1}$, $a_n \in S$ and $a_n \to a$. Hence $a \in cl(S)$. □

Corollary 2.2. *Let A be a Banach algebra with identity 1, and S be a multiplicative semigroup of A, such that $A^{-1} \subset S \subset \widehat{A}$. Then*

$$A^\bullet S = \widehat{A} \cap cl(S) \Longleftrightarrow A^\bullet S \subset \widehat{A}. \tag{2.2.1}$$

Proof. By the proof of Theorem 2.1; let us only remark that now if $a \in \widehat{A} \cap cl(S)$ then there are $a^+ \in A$ and $b \in S$ such that $1+a^+(b-a) = c \in A^{-1}$. Hence $a + aa^+(b-a) = ac$, i.e., $a = aa^+(bc^{-1})$, and $a \in A^\bullet S$. □

Let A_l^{-1} (A_r^{-1}) denotes the semigrop of all left (right) invertible elements of A. Now we have

Corollary 2.3. *Let A be a Banach algebra with identity 1. Then*

$$A_l^{-1}A^\bullet = \widehat{A} \cap cl(A_l^{-1}), \tag{2.3.1}$$

$$A^\bullet A_r^{-1} = \widehat{A} \cap cl(A_r^{-1}), \tag{2.3.2}$$

$$A_l^{-1}A^\bullet \cap A^\bullet A_r^{-1} = \widehat{A} \cap cl(A_l^{-1}) \cap cl(A_r^{-1}). \tag{2.3.3}$$

Proof. By Theorem 2.1 and Corollary 2.2; let us only remark that $A^{-1} \subset A_l^{-1} \subset \widehat{A}$, $A_l^{-1}A^\bullet \subset \widehat{A}$, $A^{-1} \subset A_r^{-1} \subset \widehat{A}$, and $A^\bullet A_r^{-1} \subset \widehat{A}$. □

Remark 2.4. Let us remark that

$$A_l^{-1}A^{\bullet} \subset \{a \in A : a \in aA_l^{-1}a\} \subset A^{\bullet}A_l^{-1} \subset \widehat{A}, \tag{2.4.1}$$

and

$$A^{\bullet}A_r^{-1} \subset \{a \in A : a \in aA_r^{-1}a\} \subset A_r^{-1}A^{\bullet} \subset \widehat{A}. \tag{2.4.2}$$

Only for a special semigroup $S \subset \widehat{A}$, say S is a subgroup of A^{-1}, one can has

$$SA^{\bullet} = \{a \in A : a \in aSa\} = A^{\bullet}S \subset \widehat{A}. \tag{2.4.3}$$

With this observation, we now come to Harte's theorem [14, Theorem 1.1].

Corollary (Harte's theorem) 2.5. *Let A be a Banach algebra with identity* 1. *Then*

$$A^{-1}A^{\bullet} = \widehat{A} \cap cl(A^{-1}). \tag{2.5.1}$$

Proof. By Theorem 2.1. □

Corollary 2.6. *Let A be a Banach algebra with identity 1, $a \in \widehat{A}$ and S be an open multiplicative semigroup of A, such that $A^{-1} \subset S \subset \widehat{A}$ and $SA^{\bullet} \subset \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta S$,
(ii) $a = sp$, $s \in S$, $p \in A^{\bullet}$ *and* $sp \notin S$.

Proof. By Theorem 2.1. □

Corollary 2.7. *Let A be a Banach algebra with identity 1, $a \in \widehat{A}$ and S be an open multiplicative semigroup of A, such that $A^{-1} \subset S \subset \widehat{A}$ and $A^{\bullet}S \subset \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta S$,
(ii) $a = ps$, $s \in S$, $p \in A^{\bullet}$ *and* $ps \notin S$.

Proof. By Corollary 2.2. □

Corollary 2.8. *Let A be a Banach algebra with identity 1, $a \in \widehat{A}$ and S be an open multiplicative semigroup of A, such that $A^{-1} \subset S \subset \widehat{A}$, $A^{\bullet}S \subset \widehat{A}$ and $A^{\bullet}S \subset \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta S$,
(ii) $a = s_1p_1 = p_2s_2$, $s_i \in S$, $p_i \in A^{\bullet}$ $(i = 1, 2)$, $s_1p_1 \notin S$ *and* $p_2s_2 \notin S$.

Proof. Clear. □

Corollary 2.9. *Let A be a Banach algebra with identity 1 and $a \in \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta A_l^{-1}$,
(ii) $a = sp$, $s_i \in A_l^{-1}$, $p \in A^\bullet$ *and* $p \neq 1$.

Proof. Clear. □

Corollary 2.10. *Let A be a Banach algebra with identity 1 and $a \in \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta A_r^{-1}$,
(ii) $a = ps$, $s_i \in A_r^{-1}$, $p \in A^\bullet$ *and* $p \neq 1$.

Proof. Clear. □

Corollary 2.11. *Let A be a Banach algebra with identity 1 and $a \in \widehat{A}$. Then the following conditions are equivalent:*

(i) $a \in \delta A^{-1}$,
(ii) $a = s_1 p_1 = p_2 s_2$, $s_i \in A^{-1}$, $p_i \in A^\bullet$ *and* $p_i \neq 1\,(i = 1, 2)$.

Proof. Clear. □

Recall that the generalised exponential, $\mathrm{Exp}(A)$, [15, Theorem 7.11.4] form the connected component of 1 in A^{-1};

$$\mathrm{Exp}(A) = \{e^{c_1}e^{c_2}\ldots e^{c_k} : c_i \in A,\, i = 1, \ldots k\}.$$

It is well known that $\mathrm{Exp}(A)$ is an open subset of A and a closed normal subgroup of A^{-1}. Also, (see [19, (5.5)]

$$\mathrm{Exp}(A)A^\bullet = \{a \in A : a \in a\mathrm{Exp}(A)a\} = A^\bullet\mathrm{Exp}(A) \subset \widehat{A}.$$

For the proof of the next result see [19, Theorem 6]

Theorem (Harte-Raubenheimer) 2.12. *Let A be a Banach algebra with identity 1. Then*

$$Exp(A)A^\bullet = \widehat{A} \cap cl\, Exp(A). \tag{2.12.1}$$

Recall that $\mathrm{Exp}(A)$ is the unique open subset of A^{-1} which is a connected subgroup of A^{-1} [21, Theorem 4.4.2]. In addition to Theorem 2.12 we have

Theorem 2.13. *Let A be a Banach algebra with identity 1, and S be an open subset of A^{-1} and subgroup of A^{-1}. Then*

$$SA^\bullet = \widehat{A} \cap cl\,(S) \Longleftrightarrow A^\bullet \subset cl\, S. \tag{2.13.1}$$

Proof. It is enough to prove the $\Longleftarrow$. From $A^\bullet \subset cl\, S$ we have $SA^\bullet \subset cl\, S$. Now (2.13.1) follows from the proof of Theorem 2.1. □

Corollary 2.14. *Let A be a Banach algebra with identity 1, $a \in \widehat{A}$, S be an open subset of A^{-1} and subgroup of A^{-1}, and $A^{\bullet} \subset cl\, S$. Then the following conditions are equivalent:*

(i) $a \in \delta S$,
(ii) $a = s_1 p_1 = p_2 s_2$, $s_i \in S$, $p_i \in A^{\bullet}$ *and* $p_i \neq 1\,(i = 1, 2)$.

Proof. By Theorem 2.13. □

Recall that an element a in A is *hermitian* if $\|\exp(ita)\| = 1$ for all real t [28]. Let us denote the set of all hermitian idempotents in A by $A_h^{\bullet}$. In connection with the Moore-Penrose generalized inverse, Rakočević [22] (see also [6], [17], [18], [23], [25]) has studied the set of elements a in A for which there exists an x in A satisfying the following conditions:

(2.14.1) $\qquad axa = a,$

(2.14.2) $\qquad xax = x,$

(2.14.3) $\qquad ax$ is hermitian,

(2.14.4) $\qquad xa$ is hermitian.

By [22, Lemma 2.1] there is at most one x such that equations (2.14.11), (2.14.12), (2.14.13) and (2.14.14) hold. The unique x is denoted by $a^{\dagger}$ and colled the Moore-Penrose inverse of a. Let $A^{\dagger}$ denote the set of all elements in A which have Moore-Penrose inverses. Clearly $A^{\dagger} \subset \widehat{A}$, and if A is a C^*-algebra then $A^{\dagger} = \widehat{A}$ [18, Theorem 6].

For the proof of the next two results see [22, Theorem 2.5, Corollary 2.6]

Theorem (Rakočević) 2.15. *Let A be a Banach algebra with identity 1. Then*

$$A^{-1} A_h^{\bullet} \cap A_h^{\bullet} A^{-1} = A^{\dagger} \cap cl\,(A^{-1}). \tag{2.15.1}$$

Corollary 2.16. *Let A be a Banach algebra with identity 1 and $a \in A^{\dagger}$. Then the following conditions are equivalent:*

(i) $a \in \delta A^{-1}$,
(ii) $a = s_1 p_1 = p_2 s_2$, $s_i \in A^{-1}$, $p_i \in A_h^{\bullet}$ *and* $p_i \neq 1\,(i = 1, 2)$.

3. Semigroups in $B(X)$.

Now we shall describe others semigroups which obey condition (2.4.3). Let X be an infinite-dimensional complex Banach space and denote the set of bounded (compact) linear operators on X by $B(X)\,(K(X))$. The fact that

$K(X)$ is a closed two-sided ideal in $B(X)$ enables us to define the *Calkin* algebra over X as the quotient algebra $C(X) = B(X)/K(X)$. $C(X)$ is itself a Banach algebra in the quotient algebra norm

$$\|T + K(X)\| = \inf_{K \in K(X)} \|T + K\|. \tag{3.0.1}$$

We shall use π to denote the natural homomorphism of $B(X)$ onto $C(X)$; $\pi(T) = T + K(X)$, $T \in B(X)$. Throughout this paper $N(T)$ and $R(T)$ will denote, respectively, the null space and the range space of T. Set $\alpha(T) = \dim N(T)$ and $\beta(T) = \dim X/R(T)$. An operator $T \in B(X)$ is *Fredholm* if $R(T)$ is closed, and both $\alpha(T)$ and $\beta(T)$ are finite. If $T \in B(X)$ and $R(T)$ is closed, it is said that T is *semi-Fredholm* operator if either $\alpha(T) < \infty$ or $\beta(T) < \infty$. Set

$$\Phi_+(X) = \{T \in B(X) : R(T) \quad \text{is closed and} \quad \alpha(T) < \infty\}, \tag{3.0.2}$$

and

$$\Phi_-(X) = \{T \in B(X) : R(T) \quad \text{is closed and} \quad \beta(T) < \infty\}. \tag{3.0.3}$$

It is clear that $\Phi(X) = \Phi_+(X) \cap \Phi_-(X)$. Let us mention that $\Phi(X)$, $\Phi_+(X)$ and $\Phi_-(X)$ are multiplicative open semigroup in $B(X)$ ([7], [15]) and by Atkinson's theorem ([7, Theorem 3.2.8], [15, Theorem 6.4.3]) we have

$$\Phi(X) = \pi^{-1}(C(X)^{-1}). \tag{3.0.4}$$

The index of an operator $T \in B(X)$ is defined by $i(T) = \alpha(T) - \beta(T)$, if at least one of $\alpha(T)$ and $\beta(T)$ is finite. It is well known that $B(X)^{-1} + K(X) \subset \Phi(X)$, and that $T \in B(X)^{-1} + K(X)$ if and only if $T \in \Phi(X)$ and $i(T) = 0$. Set

$$\Phi_0(X) = \{T \in \Phi(X) : i(T) = 0\}, \tag{3.0.5}$$

$$\Phi_l(X) = \pi^{-1}(C(X)_l^{-1}), \tag{3.0.6}$$

$$\Phi_r(X) = \pi^{-1}(C(X)_r^{-1}). \tag{3.0.7}$$

It is well-known that $\Phi_0(X)$, $\Phi_l(X)$ and $\Phi_r(X)$ are open semigroups in $B(X)$ ([7], [15]). Further, $T \in \Phi_l(X)$ if and only if $T \in \Phi_+(X)$ and there exists a bounded projection of X onto $R(T)$; $T \in \Phi_r(X))$ if and only if $T \in \Phi_-(X)$ and there exists a bounded projection of X onto $N(T)$ ([6], [7], [15]). Recall that an operator T is regular, i. e., $T \in \widehat{B(X)}$, if and only if $N(T)$ and $R(T)$ are closed, complemented subspaces of X ([6], [15], [26]). Let us mention that Gonzalez [11, Theorem] has proved [14, Theorem 1.1] for operators. His proof was based on a theorem of Caradus [6, Chapter 5, Theorem 13] involving two kinds of "gap" between the subspaces (see a good comment [14, pp. 329], and for further related results see e.g. [3], [4], [5], [10], [27]).

Theorem (Gonzalez) 3.1. *Let X be a Banach space, $T \in \widehat{B(X)}$ and P is a projection in $B(X)$ with $N(P) = N(T)$. Then the following conditions are equivalent.*

(i) *There is a sequence $\{U_n\}$ in $B(X)^{-1}$ with $\{\|U_n^{-1}\|\}$ bounded, such that $\|T - U_nP\| \to 0$.*
(ii) *There is $U \in B(X)^{-1}$ such that $T = UP$.*
(iii) *$T \in \delta B(X)^{-1}$.*
(iv) *$N(T)$ is isomorphic to a complement of $R(T)$.*

Theorem 3.2. *If X is a Banach space, then*

$$\Phi_l(X)B(X)^\bullet = \widehat{B(X)} \cap cl(\Phi(X)_l), \tag{3.2.1}$$

$$B(X)^\bullet\Phi_r(X) = \widehat{B(X)} \cap cl(\Phi_r(X)), \tag{3.2.2}$$

$$\Phi_l(X)B(X)^\bullet \cap B(X)^\bullet\Phi_r(X) = \widehat{B(X)} \cap cl(\Phi(X)_l) \cap cl(\Phi_r(X)). \tag{3.2.3}$$

Proof. By [6, p. 132, Theorem 2] we have that $\Phi_l(X)B(X)^\bullet \subset \widehat{B(X)}$ and $B(X)^\bullet\, \Phi_r(X) \subset \widehat{B(X)}$. Hence the proof followes by Corollary 2.3. □

Corollary 3.3. *Let X be a Banach space and $A \in \widehat{B(X)}$. Then the following conditions are equivalent:*

$$T \in \delta\Phi_l(X), \tag{3.3.1}$$

$$T = PB, \quad P \in B(X)^\bullet \setminus \Phi_l(X) \quad \text{and} \quad B \in \Phi_l(X), \tag{3.3.2}$$

Proof. By (3.2.1) and the fact that $\Phi_l(X)$ is an open subset of $B(X)$. □

Corollary 3.4. *Let X be a Banach space and $A \in \widehat{B(X)}$. Then the following conditions are equivalent:*

$$T \in \delta\Phi_r(X), \tag{3.4.1}$$

$$T = CQ, \quad Q \in B(X)^\bullet \setminus \Phi_r(X) \quad \text{and} \quad C \in \Phi_r(X), \tag{3.4.2}$$

Proof. By (3.2.2) and the fact that $\Phi_r(X)$ is an open subset of $B(X)$. □

Let us mention that it has been proved in [24, (3.5)] that

$$\{A \in B(X) : A \in A\Phi(X)A\} = B(X)^\bullet\Phi(X) = \Phi(X)B(X)^\bullet. \tag{3.4.3}$$

The following three results are from [24].

Theorem (Rakočević) 3.5. *If X is a Banach space then*

$$B(X)^{\bullet}\Phi(X) = \widehat{B(X)} \cap cl\,\Phi(X). \tag{3.5.1}$$

Corollary 3.6. *Let X be a Banach space and $A \in \widehat{B(X)}$. Then the following conditions are equivalent:*

$$A \in \delta\Phi(X), \tag{3.6.1}$$

$$A = PB, \quad P \in B(X)^{\bullet} \setminus \Phi(X) \quad \text{and} \quad B \in \Phi(X), \tag{3.6.2}$$

$$A = CQ, \quad Q \in B(X)^{\bullet} \setminus \Phi(X) \quad \text{and} \quad C \in \Phi(X), \tag{3.6.3}$$

For any Hilbert space X, let $\dim_H X$ denote the Hilbert dimension of X, that is the cardinality of an orthonormal basis of X. We set $\mathrm{nul}_H(T) = \dim_H N(T)$ and $\mathrm{def}_H(T) = \dim_H R(T)^{\perp}$ for $T \in B(X)$. If X is a separable Hilbert space, then with connection according to Theorem 3.5 we have

Theorem 3.7. *Let X be a separable Hilbert space. Then*

$$\begin{aligned} &\widehat{B(X)} \cap cl\,\Phi(X) \\ &\quad = \Phi(X) \cup \{T \in B(X) : nul_H(T) = def_H(T) \quad \text{and} \quad R(T) \quad \text{closed}\}. \end{aligned} \tag{3.7.1}$$

Theorem 3.8. *If X is a Banach space then*

$$B(X)^{\bullet}\Phi_0(X) = \Phi_0(X)B(X)^{\bullet} = \widehat{B(X)} \cap cl\,\Phi_0(X). \tag{3.8.1}$$

Proof. By [6, p. 132, Theorem 2] we have that $\Phi_0(X)B(X)^{\bullet} \subset \widehat{B(X)}$ and $B(X)^{\bullet}\Phi_0(X) \subset \widehat{B(X)}$. Hence we can apply Theorem 2.1 and Corollary 2.2. □

Corollary 3.9. *Let X be a Banach space and $A \in \widehat{B(X)}$. Then the following conditions are equivalent:*

$$A \in \delta\Phi_0(X), \tag{3.9.1}$$

$$A = PB, \quad P \in B(X)^{\bullet} \setminus \Phi_0(X) \quad \text{and} \quad B \in \Phi_0(X), \tag{3.9.2}$$

$$A = CQ, \quad Q \in B(X)^{\bullet} \setminus \Phi_0(X) \quad \text{and} \quad C \in \Phi_0(X), \tag{3.9.3}$$

Proof. By Theorem 3.8, Corollary 2.7 and Corollary 2.9. □

Remark 3.10. Let X be a Banach space. By Theorem 3.8 we have

$$B(X)^{\bullet}\Phi_0(X) = \Phi_0(X)B(X)^{\bullet}. \tag{3.10.1}$$

From the proof of [24, Theorem 3, (3.3)] we can conclude that

$$\{A \in B(X) : A \in A\Phi_0(X)A\} \subset B(X)^{\bullet}\Phi_0(X). \tag{3.10.2}$$

Now we have the following question (problem): If X is a Banach space, must we have

$$B(X)^{\bullet}\Phi_0(X) = \Phi_0(X)B(X)^{\bullet} = \{A \in B(X) : A \in A\Phi_0(X)A\}? \tag{3.9.3}$$

Recall that by Atkinson's theorem a bounded linear operator on a Banach space is Fredholm if and only if it has an invertible coset in the Calkin algebra. Motivated by this Harte ([12], [13], [15], [16], [19]) has associated (and has investigated) "Fredholm" elements of a Banach algebra A with an arbitrary homomorphism $T : A \mapsto B$; (A and B are complex Banach algebras with identity $1 \neq 0$, T is bounded with $T(1) = 1$). An element $a \in A$ is *Fredholm* (more precise *T-Fredholm*) iff $T(a) \in B^{-1}$. The set of all T-Fredholm elements of A is denoted by $\Phi_T(A)$. Recall that the homomorphism $T : A \mapsto B$ is *finitely regular* if

$$T^{-1}(0) \subset \widehat{A},$$

and an ideal I of A is *inessential* if the set of accumulation points of the spectrum of $x \in I$ is a subset of $\{0\}$ for each $x \in I$.

Recently Djordjević ([8], [9]) has investigated regular and T-Fredholm elements and, among other things, he has proved

Theorem (Djordjević) 3.11. *Suppose that the inessential ideals I_i, $i = 1, 2$, of A have the same sets of idempotents, I_2 is a closed subset of A, and let $P_i : A \mapsto A/I_i$ be the natural homomorphisms of A onto A/I_i, $i = 1, 2$. Now, if P_1 is a finitely regular, then*

$$A^{\bullet}\Phi_{P_1}(A) = \widehat{A} \cap cl\,(\Phi_{P_2}(A)). \tag{3.8.1}$$

Djordjević has got Theorem 3.5 as a corollary of Theorem 3.11. (The proof is based on the facts that the ideal of finite-rank operators in $B(X)$, $F(X)$, and $K(X)$ have the same sets of idempotents and $F(X) \subset \widehat{B(X)}$, and then applying Theorem 3.11 with $B(X)$ in place of A, $F(X)$ in place of I_1 and $K(X)$ in place of I_2.).

4. Partial order and regular boundary elements

Recall that in semigroup S the relation

$$e \le f \Longleftrightarrow e = ef = fe, \quad e, f \in S^{\bullet}, \tag{4.0.1}$$

is well-known standard partial ordering relation on the set of idempotents, if any. Hartwig [20] has introduced the following, so colled *plus-relation.*

Definition (Hartwig) 4.1. Let S be a semigroup. For $a, b \in S$ set $a \le b$, if

$$\text{(4.1.1)} \quad \begin{array}{ll} \text{(i)} & a \quad \text{is regular, and} \\ \text{(ii)} & \text{there is some} \quad a^{+} \in S, \quad \text{such that} \quad a^{+}a = a^{+}b,\ aa^{+} = ba^{+}. \end{array}$$

It is well known [20, Theorem 1] that the plus-relation of (4.1.1) defines a partial-order on S. This partial order is called *plus-partial order*, shortly +-order, and for idempotents the standard order (4.0.1) coincides with the +-order.

Remark 4.2. Let $(G, \le)$ be a partially ordered set. By a *closed interval* in G we shall mean any subset of the form $\{x \in G : a \le x \le b\}, \{x \in G : x \ge a\}$, or $\{x \in G : x \le a\}$, where a and b are arbitrary elements of G. There are many known ways of using the order properties of G to define a topology on G. Recall that a base for the open set in the well-known *interval topology* of G consists of all subsets of the form $\cap\{C_i : i = 1, 2, \ldots, n\}$, where each C_i is the complement of a closed interval. We let $\mathcal{I}$ denote the interval topology on G. It is natural to set the following question (problem):

If S is a semigroup, $(S, \le)$ is a partial ordered set with the plus-partial order and $cl_{\mathcal{I}}(S^{-1})$ the closure of S^{-1} in interval topology on S, must we have

$$S^{-1}S^{\bullet} = \widehat{S} \cap cl_{\mathcal{I}}(S^{-1})? \tag{4.2.1}$$

Clearly, instead of interval topology, we can consider other topologies defined by plus-partial order (or other partial order) on S, and set the similar question to (4.2.1).

If we specialize to the case where $S = R$ is a ring with unity, then we have

Theorem 4.3. *Let R be a ring with unity, and $L(R^{-1}) = \{y \in R : y \le x \quad \text{for some} \quad x \in R^{-1}\}$ be the set of predecessors of R^{-1}, where $\le$ is the plus-partial order. Then we have*

$$R^{-1}R^{\bullet} = \widehat{R} \cap L(R^{-1}) = L(R^{-1}). \tag{4.3.1}$$

Proof. By [20, Proposition 3, (i), (v)] and (1.0.1). □

Acknowledgement

I am grateful to Professor Robin Harte for the opportunity to see his results before publication.

References

[1] Bogdanović S., *Semigroups with a system of subsemigroups*, Math. Monography Inst. of Math., Univ. Novi Sad, 1985.

[2] Bogdanović S. and Ćirić M., *Polugrupe*, Prosveta, Niš, 1993.

[3] Bouldin R., *The essential minimum modulus*, Indiana Univ. Math. J. **30** (1981), 513–517.

[4] Bouldin R., *Closure of invertible operators on a Hilbert space*, Proc. Amer. Math. Soc. **108** (1990), 721–726.

[5] Burlando L., *Distance formulas on operators whose kernel has fixed Hilbert dimension*, Rendiconti di Matematica, Serie VII, Roma **10** (1990), 209–238.

[6] Caradus R. S., *Generalized Inverses and Operator Theory*, Queen s Papers in Pure and Applied Mathematics no 50, Queen s University, Kingston, Ontario, 1978.

[7] Caradus R. S., Pfaffenberger E. W. and Yood B., *Calkin Algebras and Algebras of Operators on Banach Spaces*, Dekker, New York, 1974.

[8] Djordjević D., *Regular and T-Fredholm elements in Banach algebras*, Publ. Inst. Math. **56(70)** (1994), 90–94.

[9] Djordjević D., *Harteov doprinos Fredholmovoj teoriji*, Magistarska teza, Filozofski fakultet, Matematika, Univerzitet u Nišu, 1995 (to appear).

[10] Feldman J. and Kadison V. R., *The closure of the regular operators in a ring of operators*, Proc. Amer. Math. Soc. **5** (1954), 909–916.

[11] Gonzalez M., *A perturbation result for generalized Fredholm operators in the boundary of the group of invertible operators*, Proc. R. Ir. Acad. **86 A** (1986), 123–126.

[12] Harte R., *Fredholm theory relative to a Banach algebra homomorphism*, Math. Z. **179** (1982), 431–436.

[13] Harte R., *Fredholm, Weyl and Browder theory*, Proc. R. Ir. Acad. **85** (1985), 151–176.

[14] Harte R., *Regular boundary elements*, Proc. Amer. Math. Soc. **99** (1987), 328–330.

[15] Harte R., *Invertibility and Singularuty for Bounded Linear Operators*, Marcel Dekker, Inc., New York and Basel, 1988.

[16] Harte R., *Fredholm, Weyl and Browder theory II*, Proc. R. Ir. Acad. **91 A** (1991), 79–88.

[17] Harte R., *Polar decomposition and the Moore-Penrose inverse*, Panamerican Mathematical Journal **2(4)** (1992), 71–76.

[18] Harte R. and Mbekhta M., *On generalized inverses in C^*-algebras*, Studia Math. **103** (1992), 71–77.

[19] Harte R. and Raubenheimer H., *Fredholm, Weyl and Browder theory III (to appear)*.

[20] Hartwig E. R., *How to partially order regular elements*, Math. Japonica **25** (1980), 1–13.

[21] Hille E. and Phillips S. R., *Functional analysis and semi-groups*, Amer. Math. Colloq. Publ. Vol. 31, Revised edition, Providence, R.I., American Mathematical Society, 1957.

[22] Rakočević V., *Moore-Penrose inverse in Banach algebras*, Proc. R. Ir. Acad. **88 A** (1988), 57–60.

[23] Rakočević V., *On the continuity of the Moore-Penrose inverse in Banach algebras*, Facta Universitatis (Niš), Ser. Math. Inform. **6** (1991), 133–138.

[24] Rakočević V., *A note on regular elements in Calkin algebras*, Collect. Math. **43** (1992), 37–42.

[25] Rakočević V., *On the continuity of the Moore-Penrose inverse in C^*-algebras*, Mathematica Montisnigri **2** (1993), 89–92.

[26] Rakočević V., *Funkcionalna analiza*, Naučna knjiga, Beograd, 1994.

[27] Treese W. G. and Kelly P. E., *Generalized Fredholm operators and the boundary of the maximal group of invertible operators*, Proc. Amer. Math. Soc. **67** (1977), 123–128.

[28] Vidav I., *Eine metrische Kennzeichnung der selbstadjungieten Operatoren*, Math. Z. **66** (1956), 121–128.

Department of Mathematics, University of Niš, Faculty of Philosophy,
Ćirila and Metodija 2, 18000 Niš, Serbia, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 911–913
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A NONSTANDARD PROOF OF STEINHAUS'S THEOREM

Miodrag Rašković

ABSTRACT. We give an intuitive and easy proof of the well known Steinhaus's theorem by use of nonstandard analysis.

In [5] Hugo Steinhaus proved the following very useful result.

Theorem 1. (Hugo Steinhaus, 1920) *Let A be a set with positive Lebesgue measure $\lambda(A) = m > 0$. Then, there is an interval $[a,b]$, such that $[a,b] \subseteq A - A = \{x - y | x, y \in A\}$.*

Before we prove the theorem we introduce some notation from nonstandard analysis and prove the lemma. We make use of Loeb measure [1].

We can suppose that $A \subseteq [0,1]$. Let $T_H = \{0, \frac{1}{H}, \frac{2}{H}, \dots, 1\}$ (for $H \in {}^*N \backslash N$) be a hyperfinite interval, ${}^*\mathcal{P}(T_H)$ a set of hyperfinite subsets of T_H, $\mu_H(A) = \frac{|A|}{H}$ (for $A \in {}^*\mathcal{P}(T_H)$) counting measure, $L(\mu_H)$ Loeb measure, $[a,b]_H = \{x \in T_H | a \leq x \leq b\}$ (for $a, b \in T_H$) and $st_H \colon T_H \to [0,1]$ standard part map. Let $\alpha * \beta = \frac{[\alpha, \beta H]}{H}$, where $[\alpha]$ is a integer part of $\alpha \in {}^*R_{\text{fin}}$.

Lemma. *Let A be a set with positive Lebesgue measure $\lambda(A) = m > 0$. Let B be a hyperfinite set, $B \subseteq st_H^{-1}(A)$ and $\mu_H(B) > \frac{3}{10}m$. Then, there are $a \in T_H$ and $n \in N$ such that*

$$(*) \qquad \mu_H\left(B \bigcap \left[a, a + \frac{1}{n}\right]_H\right) > \frac{3}{4n}$$

Proof. Suppose that there are not a and n such that $(*)$ holds.

The set $S = \left\{n \in {}^*N \,|\, (\forall a \in T_H)\left(\mu_H\left(B \cap \left[a, a + \frac{1}{n}\right]_H\right) \leq \frac{3}{4n}\right)\right\}$ is internal and $N \subseteq S$. By over shpil there is $K \in {}^*N \backslash N$ such that $K \in S$ and $\frac{K}{H} \approx 0$.

Let $B' = \{\frac{S}{K} | B \cap [\frac{S}{K}, \frac{S+1}{K})_H \neq \emptyset\}$. Then $st_K(B') \subseteq A$, $B' \subseteq st_K^{-1}(A)$ and $\mu_K(B') \leq L(\mu_K)(st_K^{-1}(A)) = m$. According to the fact that k $\in S$ we have

$$\begin{aligned} |B| &= \sum_{B\cap[\frac{S}{K},\frac{S+1}{K})_H \neq \emptyset} \left| B \cap \left[\frac{S}{K}, \frac{S+1}{K}\right)_H \right| \\ &\leq |B'| \max_S \left| B \cap \left[\frac{S}{K}, \frac{S+1}{K}\right)_H \right| \leq |B'| \frac{3}{4} \frac{H}{K}. \end{aligned}$$

It follows then that

$$\frac{9}{10} m \leq \mu_H(B) \leq \frac{|B'| \frac{3}{4} \frac{H}{K}}{H} = \frac{3}{4} \frac{|B'|}{K} = \frac{3}{4} \mu_K(B') \leq \frac{3}{4} m$$

a contradiction □

Proof of the theorem. Let a and n be as in lemma above. Let $I = [-\frac{1}{4n}, \frac{1}{4n}]_H$ and $C = B \cap [a, a + \frac{1}{n}]_H$. Then $I \subseteq C - C$. Otherwise, $C \cap (C + x) = \emptyset$ for some $x \in I$ and $\frac{5}{4n} \geq \mu_H(C) + \mu_H(C + x) \geq \frac{6}{4n}$. Contradiction. Finally, $st_H I \subseteq st_H(C - C) = st_H(C) - st_H(C) \subseteq A - A$. □

Now, we shall give an application of Steinhaus's theorem and method from [3] and [4].

First, we give the following definitions and theorems.

Definition 1. A function $f \colon R \to R$ is *measurable on* $A \subseteq R$ iff for each $r \in R \cup \{\infty\}$ a set $\{x \in R | f(x) \leq r\} \cap A$ is a measurable.

Let $L(\mu)$ be the Loeb measure obtained from counting measure μ on

$$T_{K,H} = \left\{-K, -K + \frac{1}{H}, \ldots, K - \frac{1}{H}, K\right\}$$

(for $k \in N$).

Let f be a map from $[-K, K]$ into R and let F be a interval map from $T_{K,H}$ into ${}^*R$.

Definition 2. The function F is a lifting of the function f if and only if

$$L(\mu)\{x \in T_{K,H} | st_H(F(x)) \neq f(st_H(x))\} = 0$$

Definition 3. The function F is a uniform lifting of the function f if and only if $st_H(F(x)) = f(st_H(x))$ for each $x \in T_{K,H}$.

The following two theorems are of great importance for example in probability theory (see [1]).

Theorem 2. (see [1]) *The function f is Lebesgue measurable if and only if it has a lifting function F.*

Theorem 3. (see [1]) *The function f is continuous if and only if it has a uniform lifting function F.*

The proofs of theorems can be found in [1].

Theorem 4. *Let $f^i(x-y) = g_i(f^1(x), f^1(y), \dots f^m(x), f^m(y), x, y)$ $(i = 1, \dots, m)$ be a system of functional equations, such that $g_i\colon R^{2m+2} \to R$ (for $i = 1, \dots, m$) are continuous functions. Let $A \subseteq R$ be a set of positive Lebesgue measure. Then, if all solutions f^i are measurable in A if follows that its are continuous at zero.*

Proof. Let B, n and H be as in proof of the last theorem. Let $h_i(x) = \left\{\begin{smallmatrix} f^i(x) & x\in A \\ 0 & x\in R\setminus A \end{smallmatrix}\right.$ and let H_i be a lifting of h_i (Theorem 2). Then, using Theorem 1, we can define an improved lifting function, such that for each $x \in \left[-\frac{1}{4n}, \frac{1}{4n}\right]_H \subseteq B - B$

$$F^i(x) = \min\{{}^*g(H_1(y), H_1(z), \dots, H_m(y), H_m(z), y, z) | x = y - z, x, y \in B\}.$$

Then, for some $y_0, z_0 \in B$

$$\begin{aligned} stF^i(x) &= st^*g_i(H_1(y_0), H_1(z_0), \dots, H_m(y_0), H_m(z_0), y_0, z_0) \\ &= g_i(stH_1(y_0), stH_1(z_0), \dots, stH_m(y_0), stH_m(z_0), sty_0, stz_0) \\ &= g_i(h_1(sty_0), h_1(stz_0), \dots, h_m(sty_0), h_m(stz_0), sty_0, stz_0) \\ &= f^i(sty_0 - stz_0) = f^i(st(y_0 - z_0)) = f^i(stx) \end{aligned}$$

(where, we write st instead of st_H).

Hence F^i is a uniform lifting function for f^i on the interval $\left[-\frac{1}{4n}, \frac{1}{4n}\right]$ and by Theorem 3 f^i is continuous on the same interval. □

References

[1] N. J. Cutland, *Nonstandard measure theory and its applications*, Bull. London Math. Soc. **15** (1983), 529–589.

[2] P. A. Loeb, *Conversion from non-standard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), 113–122.

[3] M. D. Rašković, *An application of nonstandard analysis to functional equations*, Publ. Inst. Math. N.S. **37(51)** (1985), 23–24.

[4] M. D. Rašković, *Which measurable solutions of functional equations are continuous*, (to appear).

[5] H. Steinhaus, *Sur les distances des points des ensembles de measure positive*, Fundamenta Mathematicae I (1920), 93–104.

Faculty of Sciences, Department of Mathematics , 21000 Kragujevac, Serbia

FILOMAT (Niš) 9:3 (1995), 915–924

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

DECOMPOSITION OF COEQUALITY RELATION ON THE CARTESIAN PRODUCT OF SETS WITH APARTNESSES

Daniel Abraham Romano

ABSTRACT. A coequality relation on a set with an apartness is defined by standard way as a consistent, symmetric and cotransitive relation. The coequality relation on the product $\prod X_i$ of sets with apartnesses is called decomposible if it is determined by its special projections on X_i respectively. This paper contains some theorems about characterization of decomposible coequality on the Cartesian product $\prod X_i$ of sets X_i with apartnesses which are generalization of result of the main theorem in a paper by the author. As application of these theorems, we give the exact description of coideals of the commutative rings $\prod_{i=1}^{n} X_i$ and $\prod_{i=1}^{\infty} X_i$.

Introduction

This paper continues the program of [7,8,9,10,12] to develop of *coequality* relations from the constructive mathematics ([1],[6],[14],[15]). The author is introduced the notion of coequality relation on sets with *apartnesses* in his papers [7],[8] and describes their basic properties in his papers [8] and [12]. Coequality relations on the Cartesian product of sets with apartnesses plays a central role in the developments [9] and [11].

At the beginning of the seventies, there appeared a number of papers dealing with decomposible congruences on the direct product of algebras, see e.g. papers [3] and [16]. In them it is used the well-known concept of diagonal operation ([2],[4]). The notion of compatibility of relation and the operation on the set given in the classical book [5]. If C is a coequality relation on the set $(X,=,\neq)$, and if w is an internal binary operation on X, then we say ([8],[10],[11]) that they are *compatible* if and only if

$$(\forall x,x',y,y' \in X)((w(x,y),w(x',y')) \in C \Rightarrow (x,x') \in C \vee (y,y') \in C.$$

In this paper we give the necessary and sufficient conditions for decomposibility relation on the Cartesian product $\prod_{i=1}^{n} X_i$ and $\prod_{i=1}^{\infty} X_i$ which are generalizations of the main theorem in the paper [13].

A coequality relation C on a commutative ring $(R, =, \neq, +, \cdot)$ ([6],[10],[14]) with an apartness is a *cocongruence* on R ([10]) if it is a coequality relation compatible with the operations in R and if holds

$$(\forall x, x' \in R)((xx', 0) \in C \Rightarrow (x, 0) \in C \wedge (y, 0) \in C).$$

We get, as applications of the main theorems that every cocongruences on the commutative rings $\prod_{i=1}^{n} X_i$ and $\prod_{i=1}^{\infty} X_i$ with apartnesses are decomposible. If C is a cocongruence on the commutative ring R, then ([10]) the set $S = \{x \in R : (x, 0) \in C\}$ is a *coideal* ([10],[14],[15]) of the ring R. It is a *strongly extensional* subset ([14],[15]) of R such that

$$\neg(0 \in S), x \in S \Rightarrow -x \in S, x + y \in S \Rightarrow x \in S \vee y \in S, xy \in S \Rightarrow x \in S \wedge y \in S.$$

As the last we give a description of coideals of the commutative rings $\prod_{i=1}^{n} X_i$ and $\prod_{i=1}^{\infty} X_i$ using the coideals of X_i.

Results I

Theorem 1. *Let C be a coequality relation on the Cartesian product $\prod_{i=1}^{n} X_i$ of sets $X_i, (i = 1, \ldots, n)$ with apartnesses. Then relations q_i on $X_i, (i = 1, \ldots, n)$, defined by $(x_i, x_i') \in q_i \Leftrightarrow$*

$$(\forall j = \{1, \ldots, n\} - \{i\})(((x_1, \ldots, x_i, \ldots, x_n), (x_1, \ldots, x_i', \ldots, x_n)) \in C)$$

is a coequality relation on $X_i, (i = 1, \ldots, n)$.

Proof. We give the proof for q_1. For $q_2, \ldots, q_n$ the proofs are analogous.

$$\begin{aligned}(x_1, x_1') \in q_1 &\Leftrightarrow (\exists x_2 \in X_2) \ldots (\exists x_n \in X_n)(((x_1, \ldots, x_n), (x_1', x_2, \ldots, x_n)) \in C) \\ &\Rightarrow ((x_1, x_2, \ldots, x_n), (x_1', x_2, \ldots, x_n)) \neq ((a, x_2, \ldots, x_n), (a, x_2, \ldots, x_n)) \\ &\Leftrightarrow (x_1, x_2, \ldots, x_n) \neq (a, x_2, \ldots, x_n) \vee (x_1', x_2, \ldots, x_n) \neq (a, x_2, \ldots, x_n) \\ &\Rightarrow x_1 \neq a \vee x_1' \neq a \\ &\Leftrightarrow (x_1, x_1') \neq (a, a); \\ (x_1, x_1') \in q_1 &\Leftrightarrow (\exists x_2 \in X_2) \ldots (\exists x_n \in X_n)(((x_1, \ldots, x_n), (x_1', x_2, \ldots, x_n)) \in C) \\ &\Leftrightarrow (\exists x_2 \in X_2) \ldots (\exists x_n \in X_n)(((x_1', x_2, \ldots, x_n), (x_1, x_2, \ldots, x_n)) \in C) \\ &\Leftrightarrow (x_1', x_1) \in q_1.\end{aligned}$$

$$\begin{aligned}&(x_1, x_1'') \in q_1 \Leftrightarrow (\exists x_2 \in X_2) \ldots (\exists x_n \in X_n)(((x_1, x_2, \ldots, x_n), (x_1'', x_2, \ldots, x_n)) \in C) \\ &\Rightarrow ((x_1, x_2, \ldots, x_n), (x_1', x_2, \ldots, x_n)) \in C \vee ((x_1', x_2, \ldots, x_n), (x_1'', x_2, \ldots, x_n)) \in C) \\ &\Leftrightarrow (x_1, x_1') \in q_1 \vee (x_1', x_1'') \in q_1. \qquad \square\end{aligned}$$

Using the strongly extensional and embedding bijection

$$f : \left(\prod_{i=1}^{n} X_i\right)^2 \ni ((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \longmapsto ((x_1, y_1), \ldots, (x_n, y_n)) \in \prod_{i=1}^{n} X_i^2,$$

we have the following

Corollary 1.1. *Let $C \subseteq \left(\prod_{i=1}^{n} X_i\right)^2$ be a coequality relation. Then*

$$f(C) \subseteq \bigcup_{i=1}^{n} \left(\prod_{j=1}^{i-1} X_j^2 \times q_i \times \prod_{j=i+1}^{n} X_j^2 \right).$$

Proof. $((x_1, y_1), \ldots, (x_n, y_n)) \in f(C) \iff$

$((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \in C \Rightarrow ((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) \in C$
$\vee\, ((y_1, x_2, \ldots, x_n), (y_1, y_2, x_3, \ldots, x_n)) \in C \vee \ldots$
$\vee\, ((y_1, \ldots, y_{n-1}, x_n), (y_1, y_2, \ldots, y_n)) \in C \Rightarrow$
$(x_1, y_1) \in q_1 \vee (x_2, y_2) \in q_2 \vee \cdots \vee (x_n, y_n) \in q_n \Rightarrow$
$((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) \in (q_1 \times {X_2}^2 \times \ldots {X_n}^2)$
$\bigcup ({X_1}^2 \times q_2 \times {X_3}^2 \times \cdots \times {X_n}^2) \bigcup \cdots \bigcup ({X_1}^2 \times \cdots \times {X_{n-1}}^2 \times q_n) \Rightarrow$

$$((x_1, y_1), \ldots, (x_n, y_n)) \in \bigcup_{i=1}^{n} \left(\prod_{j=1}^{i-1} X_j^2 \times q_i \times \prod_{j=i+1}^{n} X_j^2 \right). \quad \square$$

Definition 1. A coequality relation C on the Cartesian product $\prod_{i=1}^{n} X_i$ is *decomposible* if and only if

$$f(C) = \bigcup_{i=1}^{n} \left(\prod_{j=1}^{i-1} X_j^2 \times q_i \times \prod_{j=i+1}^{n} X_j^2 \right).$$

Theorem 2. *Let C be a coequality relation on the product $\prod_{i=1}^{n} X_i$ of sets with apartnesses. Then C is decomposible if and only if C is compatible with the diagonal operation d defined by*

$$d : \left(\prod_{i=1}^{n} X_i \right)^n \ni (x^1, \ldots, x^n) \longmapsto (x_1^1, \ldots, x_n^n) \in \prod_{i=1}^{n} X_i.$$

Proof. (1) Let C be a decomposible on the Cartesian product $\prod_{i=1}^{n} X_i$. We have

$$\begin{aligned}
&(d(x^1,x^2,\ldots,x^n), d(y^1,y^2,\ldots,y^n)) \in C\\
&\quad \Leftrightarrow ((x_1^1,x_2^2,\ldots,x_n^n),(y_1^1,y_2^2,\ldots,y_n^n)) \in C\\
&\quad \Leftrightarrow ((x_1^1,y_1^1),(x_2^2,y_2^2),\ldots,(x_n^n,y_n^n)) \in f(C)\\
&\quad \Leftrightarrow (\exists i = 1,\ldots,n)((x_i^i,y_i^i) \in q_i)\\
&\quad \Rightarrow (\exists i = 1,\ldots,n)(((x_1^i,y_1^i),\ldots,(x_n^i,y_n^i)) \in f(C))\\
&\quad \Leftrightarrow (\exists i = 1,\ldots,n)(((x_1^i,x_2^i,\ldots,x_n^i),(y_1^i,y_2^i,\ldots,y_n^i)) \in C)\\
&\quad \Leftrightarrow (\exists i = 1,\ldots,n)((x^i,y^i) \in C).
\end{aligned}$$

(2) Let C be a coequality relation on $\prod_{i=1}^n X_i$ compatible with the diagonal operation d. Let $((a_1,b_1),\ldots,(a_n,b_n))$ be an arbitrary element of

$$\bigcup_{i=1}^{n}\left(\prod_{j=1}^{i-1} X_j^2 \times q_i \times \prod_{j=i+1}^{n} X_j^2\right).$$

Then there exists $i = 1,\ldots,n$ such that $(a_i,b_i) \in q_i$, i.e there exists $i = 1,\ldots,n$ and there exists $x_1 \in X_1,\ldots,x_{i-1} \in X_{i-1}, x_{i+1} \in X_{i+1},\ldots,x_n \in X_n$ such that $((x_1,\ldots,x_{i-1},a_i,x_{i+1},\ldots,x_n),(x_1,\ldots,x_{i-1},b_i,x_{i+1},\ldots,x_n)) \in C$. Therefore

$$((a_1,\ldots,a_n),(b_1,\ldots,b_n)) \in C. \quad \square$$

Theorem 3. *Let X_i be commutative rings with apartnesses and let S be a coideal of the ring $\prod_{i=1}^n X_i$. Then there exists coideal s S_i of $X_i (i = 1,\ldots,n)$ such that*

$$S = \bigcup_{i=1}^{n}\left(\prod_{j=1}^{i-1} X_j \times S_i \times \prod_{j=i+1}^{n} X_j\right).$$

Proof. Let S be a coideal of the ring $\prod_{i=1}^n X_i$. Then there exists cocongruemce C on $\prod_{i=1}^n X_i$ such thet

$$(x,y) \in C \Leftrightarrow x - y \in S.$$

The cocongruence C is compatible with the diagonal operation d because the diagonal operation d can be expressed as follows

$$\begin{aligned}
d(x^1,\ldots,x^n) &= (x_1^1,\ldots,x_n^n)\\
&= (x_1^1,0,\ldots,0) + \cdots + (0,\ldots,0,x_n^n)\\
&= (x_1^1,x_2^1,\ldots,x_n^1)(1,0,\ldots,0) + \cdots + (x_1^n,x_2^n,\ldots,x_n^n)(0,0,\ldots,1).
\end{aligned}$$

Therefore, if we put $(1,0,\ldots,0)=e^1,\ldots,(0,\ldots,0,1)=e^n$,

$$\begin{aligned}&(d(x^1,\ldots,x^n),(y^1,\ldots,y^n))\in C\\ &\quad\Leftrightarrow\left(\sum_{i=1}^{n}x^ie^i,\sum_{i=1}^{n}y^ie^i\right)\in C\Rightarrow\bigvee_{i=1}^{n}((x^ie^i,y^ie^i)\in C)\\ &\quad\Rightarrow\bigvee_{i=1}^{n}((x^i,y^i)\in C)(\text{because }(\forall i=1,\ldots,n)\neg((e^i,e^i)\in C)).\end{aligned}$$

By Theorem 2, the cocongruence C is decomposible such that

$$f(C)=\bigcup_{i=1}^{n}\left(\prod_{j=1}^{i-1}X_j^2\times q_i\times\prod_{j=i+1}^{n}X_j^2\right).$$

It is easy to prove that q_i is a cocongruence on $X_i,(i=1,\ldots,n)$. Thus, by Proposition 2.5 in [10], there exists the coideal S_i of $X_i,(i=1,\ldots,n)$ such that

$$(x_i,x'_i)\in q_i\iff x_i-x'_i\in S_i.$$

Now, we have

$$\begin{aligned}&x\in S\Leftrightarrow(x,0)\in C\\ &\Leftrightarrow((x_1,\ldots,x_n),(0,\ldots,0))\in f(C)=\bigcup_{i=1}^{n}\left(\prod_{j=1}^{i-1}X_j^2\times q_i\times\prod_{j=i+1}^{n}X_j^2\right)\\ &\Leftrightarrow(\exists i,1\le i\le n)((x_i,0)\in q_i)\Leftrightarrow(\exists i,1\le i\le n)(x_i\in S_i)\\ &\Leftrightarrow x\in\prod_{j=1}^{i-1}X_j\times S_i\times\prod_{j=i+1}^{n}.\quad\square\end{aligned}$$

Results II

Theorem 4. *Let C be a coequality relation on the Cartesian product $\prod_{i=1}^{\infty}X_i$. Then the relation q_i on $X_i(i\in\mathbf{N})$, defined by*

$$\begin{aligned}&(x,y)\in q_i\Leftrightarrow(\exists a,b\in\textstyle\prod_{j=1}^{\infty}X_j)\\ &\qquad(a(i)=x\wedge b(i)=y\wedge(\forall k\in\mathbf{N}-\{i\})(a(k)=b(k)\wedge(a,b)\in C),\end{aligned}$$

is a coequality relation on $X_i(i\in\mathbf{N})$.

Proof. (1) Let x,y be elements of X_i such that $(x,y)\in q_i$ and let u be an arbitrary element of X_i. Then there exist $a,b,c\in\prod_{i=1}^{\infty}X_i$ such that

$a(i) = x \wedge b(i) = y \wedge (\forall k \in \mathbb{N} - \{i\})(a(k) = b(k) \wedge (a,b) \in C)$, and $c(i) = u \wedge (\forall k \in \mathbb{N} - \{i\})(c(k) = a(k))$. From here, we have

$$\begin{aligned}(a,b) \in C \Rightarrow (a,c) \in C \vee (c,b) \in C &\Rightarrow a \neq c \vee c \neq b \\ &\Rightarrow x \neq u \vee u \neq y \Leftrightarrow (x,y) \neq (u,u).\end{aligned}$$

(2) $(x,y) \in q_i \Leftrightarrow (\exists a,b \in \prod_{j=1}^{\infty} X_j)(a(i) = x \wedge b(j) = y \wedge (\forall k \in N - \{i\})(a(k) = b(k)) \wedge (a,b) \in C) \Leftrightarrow (y,x) \in q_i$.

(3) Let x,y,z be elements of X_i such that $(x,z) \in q_i$. Then there exists $a,b,c \in \prod_{j=1}^{\infty} X_j$ such that $a(i) = x \wedge c(i) = z$ and $(\forall k \in N - \{i\})(a(k) = b(k)) \wedge (a,c) \in C$ and

$$b(i) = y \wedge (\forall k \in N - \{i\})(b(k) = a(k) = c(k)).$$

Therefore

$$\begin{aligned}(a,c) \in C &\Rightarrow (a,b) \in C \vee (b,c) \in C \\ &\Rightarrow (x,y) \in q_i \vee (y,z) \in q_i. \qquad \square\end{aligned}$$

Using the strongly extensional and embedding bijection

$$f : \left(\prod_{j=1}^{\infty} X_j\right)^2 \ni (a,b) \longmapsto \{(a(i), b(i)) : i \in N\} \in \prod_{j=1}^{\infty} X_j^2,$$

we have the following

Corollary 4.1. *Let $C \subset (\prod_{j=1}^{\infty} X_j)^2$ be a coequality relation. Then*

$$f(C) \subseteq \bigcup_{i=0}^{\infty} \left(\prod_{j=1}^{i} X_i^2 \times q_{i+1} \times \prod_{j=i+2}^{\infty} X_j^2\right).$$

Proof. Let a,b be elements of $\prod_{j=1}^{\infty} X_j$ such that $(a,b) \in C$. If we put

$$a^t \in \prod_{j=1}^{\infty} X_j (t \in \{0\} \cup N \cup \{\infty\})\ a^0 = a, a^{\infty} = b,$$

$$(\forall j \in N)(j \leq t \Rightarrow a^t(j) = a(j) \wedge t > j \Rightarrow a^t(j) = b(j)),$$

we have

$$\begin{aligned}(a,b)\in C &\Rightarrow \bigvee_{i=0}^{\infty}((a^i,a^{i+1})\in C)\\ &\Rightarrow \bigvee_{i=0}^{\infty}((a(i),a(i+1)\in q_{i+1})\\ &\Rightarrow \bigvee_{i=0}^{\infty}\left((a,b)\in\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right)\\ &\Leftrightarrow (a,b)\in\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right).\end{aligned}$$

Therefore

$$f(C)\subseteq\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right). \quad \Box$$

Definition 2. A coequality relation C on the Cartesian product $\prod_{i=1}^{\infty}X_i$ is *decomposible* if only if

$$f(C)=\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right).$$

Theorem 5. *Let $C\subset\left(\prod_{i=1}^{\infty}X_i\right)^2$ be a coequality relation on the Cartesian product $\prod_{i=1}^{\infty}X_i$ of sets with aparthnesses. Then C is decomposible if and only if C is compatible with the diagonal operation d on $\prod_{i=1}^{\infty}X_i$ defined by*

$$d:\left(\prod_{i=1}^{\infty}X_i\right)^{\mathbb{N}}\ni F\longmapsto\{F_n(n)\in X_n:n\in\mathbb{N}\}\in\prod_{i=1}^{\infty}X_i.$$

Proof. (1)Let C be a decomposible relation on the Cartesian product $\prod_{i=1}^{\infty}X_i$ of sets and let $F\equiv\{\{F_n(j)\in X_j:j\in\mathbb{N}\}:n\in N\}$ and $G\equiv\{\{G_n(j)\in X_j:j\in\mathbb{N}\}:n\in\mathbb{N}\}$ be arbitrary elements of $\left(\prod_{i=1}^{\infty}X_i\right)^{\mathbb{N}}$. Then

$$\begin{aligned}&(d(F),d(G))\in C\\ &\quad\Leftrightarrow(\{F_n(n)\in X_n:n\in\mathbb{N}\},\{G_n(n)\in X_n:n\in\mathbb{N}\})\in C\\ &\quad\Leftrightarrow\{(F_n(n),G_n(n))\in X_n^2:n\in\mathbb{N}\}\in f(C)=\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right)\\ &\quad\Leftrightarrow(\exists n\in\mathbb{N})((F_n(n),G_n(n))\in q_n)\\ &\quad\Rightarrow(\exists n\in\mathbb{N})(\{(F_n(i),G_n(i))\in X_i^2:i\in\mathbb{N}\}\in f(C))\\ &\quad\Leftrightarrow(\exists n\in\mathbb{N})((F_n,G_n)\in C.\end{aligned}$$

(2) Let C be a coequality relation on $\prod_{i=1}^{\infty} X_i$ compatible with the diagonal operation d and let $\{(a_i, b_i) : i \in \mathbf{N}\}$ be an arbitrary element of

$$\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i} X_j^2 \times q_{i+1} \times \prod_{j=i+2}^{\infty} X_j^2\right).$$

Then there exists n in $\mathbf{N}$ such that $(a_n, b_n) \in q_n$, i.e. there exists $n \in \mathbf{N}$ and there exist $x, y \in \prod_{j=1}^{\infty} X_j$ such that

$$(x, y) \in C \wedge x(n) = a_n \wedge y(n) = b_n \wedge (\forall k \in \mathbf{N} - \{n\})(x(k) = y(k) = x_k \in X_k).$$

Let we define $x^i, y^i \in \prod_{j=1}^{\infty} X_j (i \in \mathbf{N})$ such that

$$x^n = a \wedge y^n = b \wedge (\forall i \in \mathbf{N} - \{n\})(\forall j \in \mathbf{N})(x^i(j) = x_i = y^i(j)).$$

Then, by compatibility of d, we have

$$\begin{aligned}(x, y) \in C &\Leftrightarrow \bigvee_{j=1}^{\infty}((x^i, y^i) \in C)\\ &\Rightarrow (x^n, y^n) \in C (\text{because } (\forall j \in \mathbf{N} - \{n\})\neg((x^j, y^j) \in C))\\ &\Leftrightarrow (a, b) \in C \Leftrightarrow \{(a(i), b(i)) \in X_i^2 : i \in \mathbf{N}\} \in f(C). \quad \square\end{aligned}$$

Theorem 6. *Let $X_i (i \in N)$ be commutative ring with an apartness and let S be a coideal of the ring $\prod_{i=1}^{\infty} X_i$. Then there exists the coideal S_i of $X_i (i \in N)$ such that*

$$S = \bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i} X_j \times S_{i+1} \times \prod_{j=i+2}^{\infty} X_j\right).$$

Proof. Let S be a coideal of the ring $\prod_{i=1}^{\infty} X_i$. Then there exists the cocongruence C on $\prod_{i=1}^{\infty} X_i$ such that $(x, y) \in C \Leftrightarrow x - y \in S$. The cocongrunce C is compatible with the diagonal operation d because the diagonal operation d can be expressed as follows,if we put $e^i = (0, \dots, 0, 1, 0, \dots)(i \in N)$

$$d(F) = \{F_n(n) \in X_n : n \in N\} = \sum_{n=1}^{\infty} F_n e^n.$$

Therefore

$$\begin{aligned}(d(F),d(G))\in C\Leftrightarrow&\left(\sum_{n=1}^{\infty}F_ne^n,\sum_{n=1}^{\infty}G_ne^n\right)\in C\\ &\Rightarrow\bigvee_{n=1}^{\infty}((F_ne^n,G_ne^n)\in C)\\ &\Rightarrow\bigvee_{n=1}^{\infty}((F_n,G_n)\in C).\end{aligned}$$

By Theorem 5, the cocongruence C is decomposible such that

$$f(C)=\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right).$$

It is easy to prove that q_i is a cocongruence on $X_i(i\in N)$. Then, by Proposition 2.5 in [10], there exists the coideal S_i of $X_i(i\in N)$, such that $(x,y)\in q_i\Leftrightarrow x-y\in S_i(i\in N)$. Now, we have

$$\begin{aligned}a\in S&\Leftrightarrow(a,0)\in C\\ &\Leftrightarrow\{(a(i),0)\in X_i^2:i\in N\}\in f(C)=\bigcup_{i=0}^{\infty}\left(\prod_{j=1}^{i}X_j^2\times q_{i+1}\times\prod_{j=i+2}^{\infty}X_j^2\right)\\ &\Leftrightarrow(\exists n\in N)((a(n+1),0)\in q_{n+1})\\ &\Leftrightarrow(\exists n\in N)(a(n+1)\in S_{n+1})\\ &a\in\bigcup_{n=0}^{\infty}\left(\prod_{j=1}^{n}X_j\times S_{n+1}\times\prod_{j=n+2}^{\infty}X_j\right).\quad\square\end{aligned}$$

References

[1] Bishop, E., *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1967.

[2] Duda, J., *Directly Decomposible Compatible Relations*, Glasnik mat. **19 (39)** (1984), 225–229.

[3] Fraser, G. A. and Horn, A., *Congruence Relations in Direct Products*, Proc. Amer. Math. Soc. **26** (1970), 390–394.

[4] Grätzer, G., *Universal Algebra*, Van Nostand, Amsterdam, 1979.

[5] Kuratowski, K. and Mostowski, A., *Set Theory*, North-Holland, Amsterdam, 1970.

[6] Mines, R., Richman, F. and Ruitenburg, W., *A Course of Constructive Algebra*, Springer, New York, 1988.

[7] Romano, D. A., *Equality and Diversity Relations in Constructive Mathematics*, Publ. VTŠ, Ser. Math. **1** (1985), 1–14.

[8] Romano, D. A. and Božić, M., *Relations, Function's Relations and Operations in Constructive Mathematics*, Publ. VTŠ, Ser. Math. **2** (1985), 24–39.

[9] Romano, D. A., *Construction of the Compatible Relations on the Cartesian Products of Sets*, Radovi mat. **3 (1)** (1987), 85–92.

[10] Romano, D. A., *Rings and Fields, a Constructive View*, Z. Math. Logik Grundl. Math. **34 (1)** (1988), 25–40.

[11] Romano, D. A., *Equality and Coequality Relations on the Cartesian Product of Sets*, Z. Math. Logik Grundl. Math. **34 (1)** (1988), 471–480.

[12] Romano, D. A., *Some Remarks on Coequality Relations*, Proceeding of the 7th conference "Algebra and Logic", Maribor 1989, Znanstv.rev., vol. 2 (1), 1990, pp. 111–116.

[13] Romano, D. A., *Decomposition of Coequality Relations*, Zbornik radova PMF (Kragujevac) **15** (1994), 45–52.

[14] Ruitenburg,W., *Intuitionistic Algebra*, Ph.D.Thesis, Utrecht, 1982.

[15] Troelstra, A. S. and D. van Dalen, *Constructivism in Mathematics, An Introduction*, North-Holland, Amsterdam yr 1988.

[16] Werner, H., *Congruences on Product of Algebras and Functionaly Complete Algebras*, Algebra Universalis **4** (1979), 99–105.

University of Banja Luka, Faculty of Philosophy, Department of Mathematics, 78000 Banja Luka, Bana Lazarevica 1, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 925–930

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SEMIDIRECT PRODUCTS OF SOME SEMIGROUPS

Blagoje Stamenković

ABSTRACT. Regular semidirect products of semigroups have been studied by G. B. Preston [5], orthodox by T. Saito [6] and E-inversive by F. Cabro and M. Micoli [3] e.t.c. In the present paper we study semidirect products belonging to one of the following classes of semigroups: π-regular, semilattices of Archimedean semigroups, Archimedean, left Archimedean, right Archimedean and other. At the end of the paper we give a new proof of the Wilkinsons theorem.

Let T and S be semigroups and let $\theta : S \to EndT$ be an antimorphism of S into the endomorphism semigroup of T. For $s \in S$, $t \in T$ we denote $t(s\theta)$ by t^s. If $t, t_1 \in T$, $s, s_1 \in S$ then $(tt_1)^s = t^s t_1^s$ and $(t^s)^{s_1} = t^{s_1 s}$. By semidirect product of T and S with structural mapping θ we mean the set $T \times S$ with the following multiplication:

$$(t,s)(t_1,s_1) = (tt_1^s, ss_1), \text{B. H. Neumann, [4].}$$

This product will be denoted by $T_\theta \times S$.

By $\mathbf{Z}^+$ we denote the set of all positive integers. If $a, b \in S$, then $a \underset{l}{|} b$ if $xa = b$ for some $x \in S^1$. A semigroup S is π-regular if for every $a \in S$ there exist $n \in \mathbf{Z}^+$ such that $a^n \in a^n S a^n$. By $E(S)$ we denote the set of all idempotents of S. A semigroup S is Archimedean (left Archimedean, right Archimedean) if for all $a, b \in S$ there exists $n \in \mathbf{Z}^+$ such that $a^n \in SbS$ $(a^n \in Sb, a^n \in bS)$. For undefined notions and notations we refer to [1].

Regular semidirect products of semigroups have been studied by G. B. Preston [5], orthodox by T. Saito [6] and E-inversive by F. Cabro and M. Micoli [3] e.t.c. In the present paper we study semidirect products belonging to one of the following classes of semigroups: π-regular, semilattices of Archimedean semigroups, Archimedean, left Archimedean, right

Archimedean and other. At the end of the paper we give a new proof of the Wilkinsons theorem.

Theorem 1. *The following conditions are equivalent on the semigroup $U = T_\theta \times S$.*

(i) *U is a π-regular,*

(ii) *S is a π-regular, and for all $t \in T$ and $s \in S$ there exist $x \in T$, $y \in S$ and $m \in \mathbf{Z}^+$ such that:*

$$tt^st^{s^2}\dots t^{S^{m-1}}x^{s^m}(tt^st^{s^2}\dots t^{S^{m-1}})^{s^m_y} = tt^st^{s^2}\dots t^{S^{m-1}}, \tag{1}$$

(iii) *S is a π-regular, and for every $s \in S$ there exists $y \in S$ and $m \in \mathbf{Z}^+$ such that for $e = s^m y \in E(S)$ and for every $t \in T$ it holds:*

$$tt^st^{s^2}\dots t^{S^{m-1}} \in T(tt^st^{s^2}\dots t^{S^{m-1}})^e \tag{2}$$

and

$$(tt^st^{s^2}\dots t^{s^{m-1}})^e \in (tt^st^{s^2}\dots t^{S^{m-1}})^e t^e (tt^st^{s^2}\dots t^{S^{m-1}})^e. \tag{3}$$

Proof. (i) ⇒ (ii) Let U be π-regular. Then for $(t,s) \in U$ there exist $(x.y) \in U$ and $m \in \mathbf{Z}^+$ such that:

$$(tt^s\dots t^{s-1}x^{s^m}(tt^s\dots t^{s^{m-1}})^{s^m y}, s^m y s^m) = (tt^s\dots t^{s^{m-1}}, s^m) \tag{4}$$

whence it follows (ii).

(ii) ⇒ (iii) Let (ii) hold. By (1) it follows (2), since for $s \in T$ there exist $y \in S$ and $m \in \mathbf{Z}^+$ such that $s^m y s^m = s^m$ and $s^m y = e \in E(S)$. Moreover, acting on (1) with $e = s^m y$ gives

$$(tt^s\dots t^{S^{m-1}})^e = (tt^s\dots t^{s^{m-1}})^e (x^{s^m})^e (tt^s\dots t^{S^{m-1}})^e,$$

so (3) holds.

(iii) ⇒ (i). Suppose that (iii) hold and let $(t,s) \in U$. Then there exist $y \in S$ and $m \in \mathbf{Z}^+$ such that $s^m y s^m = s^m$. Also by (3) there exists $u \in T$ such that:

$$(tt^s\dots t^{S^{m-1}})^e = (tt^s\dots t^{s^{m-1}})^e u^e (tt^s\dots t^{S^{m-1}})^e$$

and by (2) we obtain

$$(tt^s\dots t^{S^{m-1}})^e = v(tt^s\dots t^{s^{m-1}})^e$$

for some $v \in T$. Let $x = u^y$. Then

$$\begin{aligned}
&tt^s \dots t^{s^{m-1}} x^{s^m} (tt^s \dots t^{s^{m-1}})^{s^m y} = \\
&v(tt^s \dots t^{s^{m-1}})^e (u^y)^{s^m} (tt^s \dots t^{s^{m-1}})^e = \\
&v(tt^s \dots t^{s^{m-1}})^e u^e (tt^s \dots t^{s^{m-1}})^e = \\
&v(tt^s t^{s^2} \dots t^{s^{m-1}})^e = \\
&tt^s \dots t^{s^{m-1}}.
\end{aligned}$$

This proves (4) and completes th proof of π-regularity of U. □

Note that for $m = 1$ the condition (2) becomes $t \in Tt^e$, and (3) means that T^e is a regular semigroup, so as a consequence we obtain Theorem 6 [5].

Corollary 1. (G. B. Preston [5]) *Let $U = T_\theta \times S$ be a semidirect product of semigroups. Then U is regular if and only if:*

(i) *S is regular, and*
(ii) *for all s in S there exists $y \in S$ such that for $e = sy \in E(S)$ T^e is regular and for every t in T, $t \in Tt^e$.*

Theorem 2. *Let $U = T_\theta \times S$ be semidirect product of semigroups. Then U is a semilattice of Archimedean semigroups if and only if next conditions holds:*

(i) *S is a semilattice of Archimedean semigroups and*
(ii) $(\forall s, s_1 \in S)(\forall t, t_1 \in T)(\exists u \in T)(\exists y, v \in S)(\exists n \in \mathbf{Z}^+)((ss_1)^n = ys^2v \Rightarrow (tt^s u^{s^2})^y \underset{l}{\mid} \prod_{i=0}^{n-1} (tt_1^s)^{(ss_1)^i}$.

Proof. Let U be semilattice of Archimedean semigroups. Then for all (t, s), $(t_1, s_1) \in U$ there exist $(x, y), (u, v) \in U$ and $n \in \mathbf{Z}^+$, such that:

$$((t, s)(t_1, s_1))^n = (x, y)(t, s)^2(u, v).$$

From this we can obtain (i) and (ii).

Conversely, suppose that conditions (i) and (ii) holds. Then by (i) and Theorem 1 from [2] it follows that for arbitrrary $s, s_1 \in S$ exist $y, v \in S$ and $n \in \mathbf{Z}^+$ such that

$$(ss_1)^n = ys^2v \tag{5}$$

From (ii) and (5) we can conclude tat there exists $x \in T$ such that:

$$tt_1^s(tt_1)^{ss_1} \dots (tt_1)^{(ss_1)^{n-1}} = x(tt^s u^{s^2})^y = x(tt^s)^y u^{ys^2}. \tag{6}$$

Consequently, for all $(t,s),(t_1,s_1) \in U$ there exist $(x,y),(u,v) \in T$ and $n \in \mathbf{Z}^+$ such that

$$\begin{aligned}((t,s)(t_1,s_1))^n &= (tt_1^s, ss_1)^n \\ &= ((t,t_1^s)(tt_1^s)^{ss_1} \dots (tt_1^s)^{(ss_1)^{n-1}}, (ss_1)^n) \\ &= (x(tt^s)^y u^{ys^2}, ys^2v) \quad \text{from (5) and (6)} \\ &= (xy)(t,s)^2(u,v)\end{aligned}$$

which together with Theorem 1 from [2] implies that U is a semilattice of Archimedean semigroups. □

Theorem 3. *Let $U = T_\theta \times S$ semidirect product of semigroups. Then U is an Archimedean semigroup if and only if next conditions are fulfield:*

(i) *S is an Archimedean semigroup, and*

(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists u \in T)(\exists y, v \in S)(\exists k \in \mathbf{Z}^+)(s_1^k = ysv \Rightarrow (tu^s)^y \underset{l}{\mid} \prod_{i=o}^{k-1} t_1^{s_1^i})$.

Proof. Let U be an Archimedean semigroup. Then for all $(t,s),(t_1,s_1) \in U$ there exist $(x,y),(u,v) \in U$ and $k \in \mathbf{Z}^+$ such that:

$$(t_1,s_1)^k = (x,y)(t,s)(u,v).$$

From this immediately follows that conditions (i) and (ii) holds.

Conversely, suppose that conditions (i) and (ii) are fulfield. Then from (i) we obtain that for all $s, s_1 \in S$ there exist $y, v \in S$ and $k \in \mathbf{Z}^+$ such that

$$s_1^k = ysv, \tag{7}$$

hence, from (ii), we conclude that for all $t, t_1 \in T$ there exist $x, u \in T$ such that:

$$\prod_{i=0}^{k-1} t^{s_1^i} = t_1 t_1^{s_1} \dots t_1^{s_1^{m-1}} = x(tu^s)^y. \tag{8}$$

Consequently, for all $(t,s),(t_1,s_1) \in U$ there exist $(x,y),(u,v) \in U$ and $k \in \mathbf{Z}^+$ such that:

$$\begin{aligned}(t_1,s_1)^k &= (t_1 t_1^{s_1} \dots t_1^{s_1^{m-1}}, s_1^m) \\ &= (x(tu^s)^y, ysv) = (xt^y u^{ys}, ysv) \quad \text{by(7)and(8)} \\ &= (x,y)(t,s)(u,v),\end{aligned}$$

which means that U is an Archimedean semigroup. □

From Theorem 3, putting $k = 1$, we obtain the next corollary.

Corollary 2. *Semidirect product of semigroups $U_\theta \times S$ is a simple semigroup if and only if next conditions are fulfield:*

(i) *S is a simple semigroup, and*
(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists u \in T)(\exists y, v \in S)(s_1 = ysv \ \Rightarrow \ (tu^s)^y \underset{l}{\mid} t_1)$.

Theorem 4. *Semidirect product of semigroups $U = T_\theta \times S$ is left Archimedean semigroup if and only if next conditions are fulfield:*

(i) *S is a left Archimedean semigroup, and*
(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists u \in T)(\exists y, v \in S)(\exists k \in \mathbf{Z}^+)(s_1^k = ys \ \Rightarrow t^y \underset{l}{\mid} \prod_{i=0}^{k-1} t_1^{s_1^i})$.

Proof. By multiplying in U we can simply prove that U is a left Archimedean iff for all $t, t_1 \in T$, $s, s_1 \in s$ exist $x \in U$, $y \in S$ and $k \in \mathbf{Z}^+$ such that:

$$(t_1, s_1)^k = (t, t_1^{s_1} \ldots t_1^{s_1^{k-1}}, s_1^k) = (xt^y, ys) = (x, y)(t, s) \tag{9}$$

holds. Suppose that U is a left Archimedean semigroup. Then from (9) we can obtain that conditions (i) and (ii) holds.

The converse of the theorem can be obtained immediately from (9). □

From Theorem 3, putting $k = 1$, we obtain next corollary.

Corollary 2. *Semidirect product of semigroups $U_\theta \times S$ is a left simple semigroup if and only if next conditions are fulfield:*

(i) *S is a left simple semigroup, and*
(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists y \in S)(s_1 = ys \ \Rightarrow \ t^y \underset{l}{\mid} t_1)$.

The proof for next result is similar as the proof of Theorem 4.

Theorem 5. *Semidirect product of semigroups $U = T_\theta \times S$ is a right Archimedean semigroup if and only if next conditions are fulfield:*

(i) *S is a right Archimedean semigroup, and*
(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists x \in T)(\exists k \in \mathbf{Z}^+)(t, t_1^{s_1} \ldots t_1^{s_1^{k-1}} = tx^s)$.

Corollary 3. *Semidirect product of semigroups $U_\theta \times S$ is a right simple semigroup if and only if next conditions are fulfield:*

(i) *S is a right simple semigroup, and*
(ii) $(\forall t, t_1 \in T)(\forall s, s_1 \in S)(\exists x \in S)(t_1 = xt^s)$.

G. B. Preston in [5] gave the proof for Wilkinson's theorem, we shall give here one more proof for this theorem.

Theorem 6. *Semidirect product of semigroups $U_\theta \times S$ is a group if and only if T and S are groups and $S\theta \subseteq AutT$.*

Proof. Let U be group. Then by Corollaries 2 and 3 we obtain that S is a group. Let e be the identity element in group S. Then from (ii) of Corollary 2 we obtain:

$$(\forall t, t_1 \in T)(\exists x \in S)(t_1 = tx^e) \tag{10}$$

and

$$(\forall t, t_1 \in T)(\exists x \in S)(t_1 = xt^e). \tag{11}$$

Hence, from (9) we conclude that for every $u \in T$ there exists $x \in t$ such that $u = u^l x^l$ whence $u^l = (u^l x^l)^l = u^l x^l = u$. Consequently, $e\theta$ is an identity mapping. Now by (10) and (11) we obtain that equations $t_1 = tx$, $t_1 = xt$ have solutions in T, so T is a group.

Since for every $t \in T$ and $s \in S$

$$t = t^l = t^{ss^{-1}} = (t^{s^{-1}})^{s^{-1}},$$

we conclude that every mapping $s\theta$ has it's inverse mapping $s^{-1}\theta$, so $S\theta \subseteq AutT$.

Conversely, let S and T be groups and let e be identity in S, f identity in T. By streightforward verification we obtain that (f, e) is an identity of semigroup$T$ and that every $(t, s) \in U$ has it's inverse element $(t, s)^{-1} = (t^{-1}, s^{-1})$, so U is a group. □

References

[1] S. Bogdanović and M. Ćirić, *Polugrupe*, Prosveta, Niš, 1993.

[2] M. Ćirić and S. Bogdanović, *Decompositions of semigroups induced by identities*, Semigroup Forum **46** (1993), 329-346.

[3] F. Catino and M. Miccoli, *On semidirect product of semigroups*, Note di matematica **IX-n2** (1989), 189-194.

[4] B. H. Neumann, *Embeding theorems for semigroups*, Jour. London Math. Soc. **35** (1960), 184-192.

[5] G. B. Preston, *Semidirect product of semigroups*, Proc. Royal Soc. Edinburgh (1986), 91-102.

[6] T. Saito, *Orthodox semidirect products of monoids*, Semigroup Forum **38** (1989), 347-354.

Faculty of Civil Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 931–940
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

DETERMINANTAL REPRSENTATION OF GENERALIZED INVERSES OVER INTEGRAL DOMAINS

Predrag Stanimirović and Miomir Stanković

Abstract. In this paper we introduce a general form of determinantal representation of generalized inverses, for matrices which admit rank factorizations over an integral domain. We investage necessary and sufficient conditions for existence of generalized inverses. Finally, we examine correlations between the minors of generalized inverses and minors of the source matrix.

1. Introduction and preliminaries

We consider an integral domain $\mathbb{I}$ with an involution $\lambda : a \mapsto \overline{a}$. For an $m \times n$ matrix A let $\alpha = \{\alpha_1, \dots, \alpha_r\}$ and $\beta = \{\beta_1, \dots, \beta_r\}$ be the subsets of $\{1, \dots, m\}$ and $\{1, \dots, n\}$, respectively. Then $A\begin{pmatrix}\alpha_1 & \dots & \alpha_r \\ \beta_1 & \dots & \beta_r\end{pmatrix} = | A^\alpha_\beta |$ denotes the minor of A determined by the rows indexed by α and the columns indexed by β. If $\alpha = \{1, \dots, m\}$, then $| A^\alpha_\beta |$ can be simply denoted $| A_\beta |$, and similarly if $\beta = \{1, \dots, n\}$, then $| A^\alpha_\beta |$ can be denoted by $| A^\alpha |$. Also, the algebraic complement of $| A^\alpha_\beta |$ is defined by

$$\frac{\partial}{\partial a_{ij}} | A^\alpha_\beta | = A_{ij}\begin{pmatrix}\alpha_1 & \dots & \alpha_{p-1} & i & \alpha_{p+1} & \dots & \alpha_r \\ \beta_1 & \dots & \beta_{q-1} & j & \beta_{q+1} & \dots & \beta_r\end{pmatrix} = (-1)^{p+q} A\begin{pmatrix}\alpha_1 & \dots & \alpha_{p-1} & \alpha_{p+1} & \dots & \alpha_r \\ \beta_1 & \dots & \beta_{q-1} & \beta_{q+1} & \dots & \beta_r\end{pmatrix}.$$

The r-th compound matrix $C_r(A)$ of A is a matrix of order $\binom{m}{r} \times \binom{n}{r}$ defined on the following way. The rows of $C_r(A)$ are indexed by the r-element subsets of $\{1, \dots, m\}$; the columns are indexed by the r-element subsets of $\{1, \dots, n\}$, and the (α, β) entry of $C_r(A)$ is defined as $| A^\alpha_\beta |$.

For $A \in \mathbb{C}^{m\times n}$ consider the following Penrose [13] equations in X (where $A^* = (\overline{A})^T$):

$$(1) \quad AXA = A \quad (2) \quad XAX = X \quad (3) \quad (AX)^* = AX \quad (4) \quad (XA)^* = XA.$$

If $m = n$ we add

$$(5) \qquad AX = XA\,.$$

For a subset $\mathcal{S}$ of $\{1,2,3,4,5\}$ the set of matrices G obeying the conditions represented in $\mathcal{S}$ will be denoted by $A\{\mathcal{S}\}$. A matrix $G \in A\{\mathcal{S}\}$ is called an $\mathcal{S}$-inverse of A and is denoted by $A^{(\mathcal{S})}$.

The starting point of the investigations of this paper is the determinantal representation of generalized inverses of complex matrices [1, 4, 7, 15, 16]. Also, we use the determinantal representation of the *Moore-Penrose inverse*, the *weighted Moore-Penrose inverse* and the *group inverse* over integral domain ([2], [11] and [12]). Let us recall

Theorem 1.1. [2] *Let A be an $m \times n$ matrix of rank r over $\mathbb{I}$, and let $A = PQ$ be a rank factorization of A. Then the following conditions are equivalent;*

(i) *A has a Moore-Penrose inverse.*
(ii) *P^*P and QQ^* are invertible over $\mathbb{I}$.*
(iii) *$C_r(A)$ has a Moore-Penrose inverse.*
(iv) *$\sum\limits_{\alpha,\beta} |\overline{A}^\alpha_\beta| \cdot |A^\alpha_\beta|$ is invertible in $\mathbb{I}$, where α, β run over r-element subsets of $\{1, \dots, m\}$ and $\{1, \dots, n\}$ respectively.*

Furthermore, the Moore-Penrose inverse $G = (g_{ij})$, if it exists is given by $G = A^\dagger = Q^*(QQ^*)^{-1}(P^*P)^{-1}P^*$ and

$$g_{ij} = \left(\sum_{\gamma,\delta} |\overline{A}^\gamma_\delta|\,|A^\gamma_\delta|\right)^{-1} \cdot \sum_{\alpha:j\in\alpha;\,\beta:i\in\beta} |\overline{A}^\alpha_\beta|\, \frac{\partial}{\partial a_{ji}}\, |A^\alpha_\beta|\,.$$

Similar results are obtained for the weighted Moore-Penrose inverse $A^\dagger_{M,N}$ [12], which satisfies equations (1), (2) and

$$(6) \qquad (MAX)^* = MAX \qquad (7) \qquad (NXA)^* = NXA.$$

Theorem 1.2. [12] *Let A be an $m \times n$ matrix of rank r over $\mathbb{I}$, and let $A = PQ$ be a rank factorization of A. Then the following conditions are equivalent:*

(i) *A has a weighted Moore-Penrose inverse with respect to M and N.*
(ii) *P^*MP and $QN^{-1}Q^*$ are invertible over $\mathbb{I}$.*
(iii) *$C_r(A)$ has a weighted Moore-Penrose inverse with respect to $C_r(M)$ and $C_r(N)$.*

(iv) $\sum_{\alpha,\beta} |(N^{-1}A^*M)^{\beta}_{\alpha}||A^{\alpha}_{\beta}| = \sum_{\alpha,\beta} |(\overline{MAN^{-1}})^{\alpha}_{\beta}||A^{\alpha}_{\beta}|$ *is invertible in* $\mathbb{I}$.

Determinantal representation of the group inverse over an integral domain $\mathbb{I}$ is introduced in [11]:

Theorem 1.3. [11] *Let A be an $m \times n$ matrix of rank r over $\mathbb{I}$, and let $A = PQ$ be a rank factorization of A. Then the following conditions are equivalent:*

(i) *A has a group inverse.*

(ii) *$C_r(A)$ has a group inverse.*

(iii) *$\sum_{\gamma} |A^{\gamma}_{\gamma}|$ is invertible in* $\mathbb{I}$.

Furthermore, the group inverse $G = (g_{ij})$, if it exists, is given by

$$g_{ij} = \left(\sum_{\gamma} |A^{\gamma}_{\gamma}|\right)^{-2} \cdot \sum_{\alpha:j\in\alpha\,;\,\beta:i\in\beta} |(A^T)^{\alpha}_{\beta}| \frac{\partial}{\partial a_{ji}} |A^{\alpha}_{\beta}|\,.$$

The main results of this paper are:

(1) Generalization of the *algebraic complement* and determinant, and obtain general form of the *determinantal representation* for different classes of generalized inverses: the Moore-Penrose inverse, the weighted Moore-Penrose inverse, the group inverse and the left(right) inverses. In this way, we generalize the results obtained in [16].

(2) Necessary and sufficient conditions for existence of the *general determinantal representation*, and partially, existence of the *Moore-Penrose* inverse, the *weighted Moore-Penrose* inverse and the group inverse.

(3) Correlations between the minors of diferent classes of generalized inverses and minors of the given matrix.

2. Determinantal representations of generalized inverses

First we generalize the concepts of determinant and *algebraic complement* (see [1, 2, 4, 7, 8, 11, 12, 15, 16, 17]).

Definition 2.1. The generalized determinant of an $m \times n$ matrix A of rank r, denoted by $N_{(R,r)}(A)$, is defined by

$$N_{(R,r)}(A) = \sum_{\alpha,\beta} |\overline{R}^{\alpha}_{\beta}||A^{\alpha}_{\beta}|\,, \tag{2.1}$$

where R is an $m \times n$ matrix satisfying condition

$$rank(AR^*) = rank(R^*A) = rank(A)\,. \tag{2.2}$$

Note that (2.2) is satisfied if and only if $rank(R) \geq rank(A) = r$.

Definition 2.2. Let A, R be $m\times n$ matrices over $\mathbb{I}$ and let R satisfies (2.2). The *generalized algebraic complement* of A corresponding to a_{ij} is defined by

$$A_{ij}^{(\dagger,R)} = \sum_{\alpha:j\in\alpha;\beta:i\in\beta} |\,\overline{R}_{\beta}^{\alpha}\,|\,\frac{\partial}{\partial a_{ji}}\,|\,A_{\beta}^{\alpha}\,|\,. \tag{2.3}$$

In a similar way can be generalized the notion of *adjoint matrix*.

Definition 2.3. Matrix whose elements are equal to $A_{ij}^{(\dagger,R)}$ we denote by $adj^{(\dagger,R)}(A)$, and we write it as the generalized adjoint matrix of A, corresponding to R.

Finally, in the following definition we introduce the *general determinantal representation* of generalized inverses over an integral domain.

Definition 2.4. Let given an $m\times n$ matrix A of rank r over $\mathbb{I}$ and $m\times n$ matrix R which satisfies condition (2.2). General determinantal representation of generalized inverses of A is defined by

$$A^{(\dagger,R)} = \left(N_{(R,r)}(A)\right)^{-1}\cdot adj^{(\dagger,R)}(A). \tag{2.4}$$

For two full-rank matrices A and R we have:

Lemma 2.1. *If A is an $m\times n$ matrix of full-rank and matrix R has the same dimensions and rank, then:*

(i) $N_{(R,r)}(A) = \begin{cases} |\,AR^*\,|, & r=m \\ |\,R^*A\,|, & r=n. \end{cases}$

(ii) $A_{ij}^{(\dagger,R)} = \begin{cases} (R^*adj(AR^*))_{ij}\,, & r=m \\ (adj(R^*A)R^*)_{ij}\,, & r=n. \end{cases}$

(iii) $A^{(\dagger,R)} = \begin{cases} R^*(AR^*)^{-1}, & r=m \\ (R^*A)^{-1}R^*, & r=n. \end{cases}$

Proof. (i) Follows from the Cauchy-Binet Theorem.

(ii) The relation $(A^*adj(AA^*))_{ij} = \sum_{\beta:i\in\beta} |\,\overline{A}_{\beta}\,|\,\frac{\partial}{\partial a_{ji}}\,|\,A_{\beta}\,|$ is obtained in [1], [6]. The result (*ii*) can be obtained in a similar way, substituting the matrix A^* by the matrix R^*.

(*iii*) It is implied by (*i*) and (*ii*). □

Now we investage main properties of the *generalized adjoint matrix*, *generalized algebraic complement* and *generalized determinant*.

Lemma 2.2. *Let $A = PQ$ be a full-rank factorization of an $m \times n$ matrix A of rank r, R_1 be an $r \times n$ matrix of rank r and R_2 be an $m \times r$ matrix of rank r. Generalized adjoint matrix satisfies the following:*

$$adj^{(\dagger,R_1)}(Q) \cdot adj^{(\dagger,R_2)}(P) = adj^{(\dagger,R_2R_1)}(A).$$

Proof. For $1 \le i \le n$, $1 \le j \le m$ we obtain the following representation for (i,j)-th element of the matrix product $adj^{(\dagger,R_1)}(Q) \cdot adj^{(\dagger,R_2)}(P)$:

$$\left(adj^{(\dagger,R_1)}(Q) \cdot adj^{(\dagger,R_2)}(P)\right)_{ij} = \sum_{k=1}^{r} Q_{ik}^{(\dagger,R_1)} \cdot P_{kj}^{(\dagger,R_2)} =$$

$$= \sum_{k=1}^{r} \sum_{\beta:i\in\beta} |(\overline{R_1})_\beta| \frac{\partial}{\partial q_{ki}} |Q_\beta| \cdot \sum_{\alpha:j\in\alpha} |(\overline{R_2})^\alpha| \frac{\partial}{\partial p_{jk}} |P^\alpha| =$$

$$= \sum_{\alpha:j\in\alpha;\beta:i\in\beta} |(\overline{R_2R_1})_\beta^\alpha| \cdot \sum_{k=1}^{r} \frac{\partial}{\partial p_{jk}} |P^\alpha| \frac{\partial}{\partial q_{ki}} |Q_\beta| \,.$$

Using the Cauchy-Binet formula we get

$$\sum_{k=1}^{r} \frac{\partial}{\partial p_{jk}} |P^\alpha| \frac{\partial}{\partial q_{ki}} |Q_\beta| = \frac{\partial}{\partial a_{ji}} |A_\beta^\alpha|,$$

which implies

$$\left(adj^{(\dagger,R_1)}(Q) \cdot adj^{(\dagger,R_2)}(P)\right)_{ij} = \left(adj^{(\dagger,R_2R_1)}(A)\right)_{ij}. \quad \square$$

Similarly, the following lemma can be proved:

Lemma 2.3. *If $A = PQ$ is a full-rank factorization of an $m\times n$ matrix A of rank r, R_1 and R_2 are matrices of apropriate sizes, and satisfy $rank(QR_1) = rank(R_2P) = r$, then the generalized determinant satisfies*

$$N_{(R_1,r)}(Q) \cdot N_{(R_2,r)}(P) = N_{(R_2R_1,r)}(A).$$

Proof. From Lemma 2.1 and the Cauchy-Binet theorem we obtain:

$$\begin{aligned} N_{(R_1,r)}(Q) \cdot N_{(R_2,r)}(P) &= |QR_1^*| \cdot |R_2^*P| = \\ &= \sum_{\beta} |(\overline{R_1})_\beta| |Q_\beta| \cdot \sum_{\alpha} |(\overline{R_2})^\alpha| |P^\alpha| = \\ &= \sum_{\alpha,\beta} |(\overline{R_2R_1})_\beta^\alpha| |A_\beta^\alpha| = N_{(R_2R_1,r)}(A). \quad \square \end{aligned}$$

From Lemma 2.2 and Lemma 2.3 we have.

Corollary 2.1. *If $A = PQ$ is a full-rank factorization of an $m \times n$ matrix A of rank r, R_1 and R_2 satisfy conditions from Lemma 2.2 and Lemma 2.3, then the generalized determinantal representation satisfies*

$$Q^{(\dagger,R_1)} \cdot P^{(\dagger,R_2)} = A^{(\dagger,R_2R_1)}.$$

The following theorem showes the properties of determinantal representation of generalized inverses.

Theorem 2.1. *Let A, R be $m \times n$ matrices of rank r and $A = PQ$ be a full-rank factorization of A. Then:*

(i) $A^\dagger = Q^{(\dagger,Q)}P^{(\dagger,P)} = A^{(\dagger,A)}$;

(ii) $A^\dagger_{M,N} = Q^{(\dagger,QN^{-1})}P^{(\dagger,MP)} = A^{(\dagger,MAN^{-1})}$;

(iii) $A^\# = Q^{(\dagger,Q^*)}P^{(\dagger,P^*)} = A^{(\dagger,A^*)}$;

(iv) $A^{-1} = A^{(\dagger,R)}$, *for arbitrary regular R and regular A;*

(v) $A^{(\dagger,R)}$ *represents the left(right) inverses, for a full-rank matrix A.*

Proof. (i) Follows from $A^\dagger = Q^\dagger P^\dagger$ and Lemma 2.1.

(ii) It is implied by $A^\dagger_{M,N} = (QN^{-1})^*(Q(QN^{-1})^*)^{-1}((MP)^*P)^{-1}(MP)^*$ [12], [17] and Lemma 2.1. Furthermore, from Definition 2.4, we obtain the following determinantal representation for $A^\dagger_{M,N}$, (see [17]) and [12, Theorem 8]:

$$\left(A^\dagger_{M,N}\right)_{ij} = \left(\sum_{\gamma,\delta} |(\overline{MAN^{-1}})^\gamma_\delta|\,|A^\gamma_\delta|\right)^{-1} \cdot \sum_{\alpha,\beta} |(\overline{MAN^{-1}})^\alpha_\beta|\, \frac{\partial}{\partial a_{ji}} |A^\alpha_\beta|.$$

(iii) Follows from Theorem 1.3 and

$$\left(\sum_\gamma |A^\gamma_\gamma|\right)^{-2} \cdot \sum_{\alpha:j\in\alpha;\,\beta:i\in\beta} |(A^T)^\alpha_\beta|\, \frac{\partial}{\partial a_{ji}} |A^\alpha_\beta| =$$

$$= \left(\sum_{\gamma,\delta} |(A^T)^\gamma_\delta|\,|A^\gamma_\delta|\right)^{-1} \cdot \sum_{\alpha:j\in\alpha;\,\beta:i\in\beta} |(A^T)^\alpha_\beta|\, \frac{\partial}{\partial a_{ji}} |A^\alpha_\beta|.$$

(v) For example, suppose $r = m$. Using the Laplace's development for the square minors A_β we get

$$N_{(R,m)}(A) = \sum_\beta \overline{R}_\beta \left[\sum_{k=1}^{r} a_{ij_k} \frac{\partial}{\partial a_{ij_k}} |A_\beta|\right] =$$

$$= \sum_{l=1}^{n} a_{il} \left[\sum_{\beta,l\in\beta} \overline{R}_\beta \frac{\partial}{\partial a_{il}} |A_\beta|\right] = \sum_{l=1}^{n} a_{il} A^{(R,m)}_{li}.$$

For $p \neq q$, $1 \leq p, q \leq m$, substituting in the minors of A, the q-th row by the p-th row, and using $N_{(R,m)}(A) = \sum\limits_{\beta} \overline{R}_\beta A_\beta = 0$, in the same way we prove $\sum\limits_{l=1}^{n} a_{pl} A_{lq}^{(R,m)} = 0$. Hence, $g_{ij} = \delta_{ij} N_{(R,m)}(A)$, and consequently $A \cdot A^{(\dagger,R)} = I_m$, for arbitrary R. This means that $A^{(\dagger,R)}$ represents the class of *right inverses* of the full-rank matrix A in the case $r = m \le n$.

On the other hand, it can be proved that $A^{(\dagger,R)}$ represents the class of *left inverses* of A, in the case $r = n \le m$. □

In the following theorem we examine existence of the general determinantal representation.

Theorem 2.2. *Let A, R be $m \times n$ matrices of rank r over $\mathbb{I}$, $A = PQ$ be a full-rank factorization of A and $R = ST$ be a full-rank factorization of R. Then the following conditions are equivalent:*

(i) *$A^{(\dagger,R)}$ exists.*

(ii) *QT^* and S^*P are invertible matrices in $\mathbb{I}$.*

(iii) *$N_{(T,r)}(Q)$ and $N_{(S,r)}(P)$ are invertible in $\mathbb{I}$.*

(iv) *$N_{(R,r)}(A)$ is invertible in $\mathbb{I}$.*

Proof. (i)⇒(ii): If $A^{(\dagger,R)}$ exists, from Corollary 2.1 and Lemma 2.1, we get

$$A^{(\dagger,R)} = A^{(\dagger,ST)} = Q^{(\dagger,T)} \cdot P^{(\dagger,S)} = T^*(QT^*)^{-1}(S^*P)^{-1}S^*.$$

From $AA^{(\dagger,R)}A = A$ follows

$$QT^*(QT^*)^{-1}(S^*P)^{-1}S^*P = I,$$

which implies (ii).

(ii)⇒(i): If QS^* and T^*P are invertible, from Lemma 2.1 and Corollary 2.1, we conclude

$$T^*(QT^*)^{-1}(S^*P)^{-1}S^* = Q^{(\dagger,T)} \cdot P^{(\dagger,S)} = A^{(\dagger,ST)} = A^{(\dagger,R)}.$$

(ii)⇔(iii) A square matrix over a ring $\mathbb{I}$ is invertible if and only if its determinant is invertible in $\mathbb{I}$ [9], [10]. Hence, QT^* and S^*P are invertible matrices if and only if $|QT^*|$ and $|S^*P|$ are invertible in $\mathbb{I}$. Finally, from Lemma 2.1 we obtain

$$|QT^*| = N_{(T,r)}(Q), \qquad |S^*P| = N_{(S,r)}(P).$$

(iii)⇔(iv) An application of Lemma 2.3 implies

$$N_{(T,r)}(Q) \cdot N_{(S,r)}(P) = N_{(ST,r)}(A) = N_{(R,r)}(A).$$

Therefore, $N_{(R,r)}(A)$ is invertible if and only if both $N_{(S,r)}(P)$ and $N_{(T,r)}(Q)$ are invertible. □

Corollary 2.2. *Let A be an $m \times n$ matrix of rank r, and $A = PQ$ be its full-rank factorization. The following conditions are equivalent:*

(i) $A^\dagger$ *exists.*
(ii) QQ^* *and* P^*P *are invertible matrices in* $\mathbb{I}$.
(iii) $N_{(Q,r)}(Q)$ *and* $N_{(P,r)}(P)$ *are invertible in* $\mathbb{I}$.
(iv) $N_{(A,r)}(A)$ *is invertible in* $\mathbb{I}$.

Corollary 2.3. *For an $m \times n$ matrix A of rank r the following conditions are equivalent:*

(i) $A^\dagger_{M,N}$ *exists.*
(ii) P^*MP *and* $QN^{-1}Q^*P$ *are invertible matrices in* $\mathbb{I}$.
(iii) $N_{(QN^{-1},r)}(Q)$ *and* $N_{(PM,r)}(P)$ *are invertible in* $\mathbb{I}$.
(iv) $N_{(MAN^{-1},r)}(A)$ *is invertible in* $\mathbb{I}$.

Corollary 2.4. *Let A be a square matrix of order n, rank r and $rank(QP) = r$. Then the following conditions are equivalent:*

(i) $A^\#$ *exists.*
(ii) QP *is invertible matrix in* $\mathbb{I}$.
(iii) $N_{(P^*,r)}(Q)$ *and* $N_{(Q^*,r)}(P)$ *are invertible in* $\mathbb{I}$.
(iv) $N_{(A^*,r)}(A)$ *is invertible in* $\mathbb{I}$.
(v) $\sum\limits_{\gamma} |\,A_\gamma^\gamma\,|$ *is invertible in* $\mathbb{I}$.

Proof. Note that $(iv) \Leftrightarrow (v)$ follows from

$$\left(N_{(A^*,r)}(A)\right)^{-1} = \left(\sum_{\gamma} |\,A_\gamma^\gamma\,|\right)^{-2}. \quad \square$$

In the following part of this section we represent minors of generalized of A, in terms of minors of A and arbitrary matrix R, which satisfies the condition (2.2).

Theorem 2.3. *Let A, R be matrices of type $m \times n$ whose rank is r and let $A = PQ$ be a full-rank factorization of A. Then for all α, β we have*

$$|\,(A^{(\dagger,R)})_\beta^\alpha\,| = \left(\sum_{\gamma,\delta} |\,A_\delta^\gamma\,|\,|\,(\overline{R})_\delta^\gamma\,|\right)^{-1} \cdot |\,(\overline{R})_\beta^\alpha\,| = \tag{2.5}$$

$$= \left(N_{(R,r)}(A)\right)^{-1} |\,(\overline{R})_\beta^\alpha\,|.$$

Proof. In ([2], Theorem 3.) is proved the following relation for the reflexive generalized inverses $G = (g_{ij})$ of A:

$$g_{ij} = \sum_{\alpha,\beta} |\, G^{\alpha}_{\beta} \,| \frac{\partial}{\partial a_{ji}} |\, A^{\alpha}_{\beta} \,| .$$

The proof can be completed using that $A^{(\dagger,R)}$ is a reflexive generalized inverse and

$$\left(A^{(\dagger,R)}\right)_{ij} = \left(\sum_{\gamma,\delta} |\, A^{\gamma}_{\delta} \,|\, |\, (\overline{R})^{\gamma}_{\delta} \,| \right)^{-1} \cdot \sum_{\alpha, j\in\alpha \;,\delta, i\in\delta} |\, (\overline{R})^{\alpha}_{\beta} \,| \frac{\partial}{\partial a_{ji}} |\, A^{\alpha}_{\beta} \,| . \quad \square$$

In particular, the last theorem and Theorem 2.1 imply:

Corollary 2.5. *Let A, R be matrices of type $m \times n$ and rank r over $\mathbb{I}$. Then for all α, β is valid:*

$$|\, A^{\#}{}^{\alpha}_{\beta} \,| = \left(\sum_{\gamma} |\, A^{\gamma}_{\gamma} \,| \right)^{-2} |\, A^{\beta}_{\alpha} \,| = \left(N_{(A^{*},r)}(A)\right)^{-1} |\, (A^{T})^{\alpha}_{\beta} \,| ;$$

$$|\, A^{\dagger}{}^{\alpha}_{\beta} \,| = \left(\sum_{\gamma,\delta} |\, A^{\gamma}_{\delta} \,|\, |\, \overline{A}^{\gamma}_{\delta} \,| \right)^{-1} |\, \overline{A}^{\alpha}_{\beta} \,| = \left(N_{(A,r)}(A)\right)^{-1} |\, A^{\alpha}_{\beta} \,| ;$$

$$\begin{aligned} |\, (A^{\dagger}_{M,N})^{\alpha}_{\beta} \,| &= \left(\sum_{\gamma,\delta} |\, A^{\gamma}_{\delta} \,|\, |\, (\overline{MAN^{-1}})^{\gamma}_{\delta} \,| \right)^{-1} |\, (\overline{MAN^{-1}})^{\alpha}_{\beta} \,| = \\ &= \left(N_{(MAN^{-1},r)}(A)\right)^{-1} |\, (\overline{MAN^{-1}}^{\alpha}_{\beta} \,| . \end{aligned}$$

Proof. If $m = n$, and $R = A^{*}$, in (2.5) we obtain $G = A^{\#}$. Similarly, for $R = A$ we obtain $G = A^{\dagger}$, and $G = A^{\dagger}_{M,N}$ is induced by $R = MAN^{-1}$. □

Note that correlations between minors of A and corresponding minors of the Moore-Penrose and group inverse are proved (in another way) in [2] and [11], respectively.

References

[1] Arghiriade, E. e Dragomir, A., *Une nouvelle definition de l'inverse generalisee d'une matrice*, Rendiconti dei Lincei, serir XIII Sc. fis. mat. e nat. **35** (1963), 158–165.

[2] Bapat, R.B.; Bhaskara, K.P.S. and Manjunatha Prasad, K., *Generalized inverses over integral domains*, Linear Algebra Appl. **140** (1990), 181–196.

[3] Ben-Israel, A. ; Grevile, T.N.E., *Generalized Inverses: Theory and Applications*, Wiley-Interscience, New York, 1974.

[4] Ben-Israel, A., *Generalized inverses of matrices: a perspective of the work of Penrose*, Math. Proc. Camb. Phil. Soc. **100** (1986), 407–425.

[5] Bhaskara Rao, K.P.S., *On generalized inverses of matrices over integral domains*, Linear Algebra Appl. **49** (1983), 179–189.

[6] Gabriel, R., *Das verallgemeinerte inverse einer matrix ü ber einem beliebigen Körper - analytish betrachtet*, J. Rewie Ansew Math. **244(V)** (1968), 83–93.

[7] Gabriel, R., *Extinderea complementilor algebrici generalizati la matrici oarecare*, Studii si cercetari matematice **17 -Nr. 10** (1965), 1566–1581.

[8] Gabriel, R., *Das allgemeine element des verallgemeinerte inverse von Moore-Penrose*, Rev. Roum. Math. Pures et appl. **37 No 6** (1982), 689–702.

[9] Kesler, S.S., Krupnik N.J., *On invertibility matrices with elements from the ring*, Uchennie Zapiski, Kishinjev **91** (1967), 51–54. (in Russian)

[10] Krupnik N.J., *On normal solvability problem and singular and integral equations index*, Uchennie Zapiski, Kishinjev **82** (1965), 3–7. (in Russian)

[11] Prasad, K.M.; Bhaskara Rao, K.P.S. and Bapat, R.B., *Generalized inverses over integral domains. II. Group inverses and Drazin inverses*, Linear Algebra Appl. **146** (1991), 31–47.

[12] Prasad, K.M. and Bapat, R.B., *The Generalized Moore-Penrose inverse*, Linear Algebra Appl. **165** (1992), 59–69.

[13] Penrose, R., *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc. **51** (1955), 406–413.

[14] Radić, M., *Some contributions to the inversions of rectangular matrices*, Glasnik Matematički **1 (21) -No. 1** (1966), 23–37.

[15] Stanimirović, P. and Stanković, M., *Generalized algebraic complement and Moore-Penrose inverse*, FILOMAT **8** (1994), 57–64.

[16] Stanimirović, P., *Computing pseudoinverses using minors of an arbitrary matrix*, FILOMAT (1994).

[17] Stanimirović, P. and Stanković, M., *Determinantal representation of weighted Moore -Penrose inverse*, Matematički Vesnik **46** (1994), 41–50.

University of Niš, Philosophical Faculty, Department of Mathematics, Ćirila i Metodija 2, 18000 Niš, Yugoslavia

University of Niš, Faculty of Occupational Safety, Čarnojevića 10a, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 941–946
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

SEMI-FREDHOLM ELEMENTS IN BANACH ALGEBRAS

Nebojša Stojković

ABSTRACT. In this paper we define the set of semi-Fredholm elements in a semisimple Banach algebra and we prove that the perturbation class of this set is a closed twosided ideal of this algebra.

1. Introduction

Let X be Banach space and let $B(X)$ be Banach space of all bounded linear transformations of X into X. For $T \in B(X)$ we let $N(T)$ denote the kernel of T, $N(T) = \{x \in X | T(x) = 0\}$ and we let $R(T)$ denote the range of T, $R(T) = \{y \in X | T(x) = y$ for some $x \in X\}$. If $T \in B(X)$ and $R(T)$ is closed, we say that T is a semi-Fredholm operator if either $dim(N(T)) < \infty$ or $dim(X/R(T)) < \infty$. We have two classes of semi-Fredholm operators,

$$\Phi_+(X) = \{T \in X | R(T) \text{ is closed and } dimN(T) < \infty\} \text{ and}$$

$$\Phi_-(X) = \{T \in B(X) | R(T) \text{ is closed and } dimX/R(T) < \infty\}.$$

We also set $\Phi(X) = \Phi_+(X) \cap \Phi_-(X)$ and call this set of Fredholm operators on X. It is known that $T \in B(X)$ is Fredholm if and only if $\pi(T)$ is invertible in the Calkin algebra $C(X)(C(X) = B(X)/K(X)$ where $K(X)$ is set of compact operators and π denote the natural homomorphism of $B(X)$ onto $C(X), \pi(T) = T + K(X)$. Index of T is defined by $i(T) = dim(N(T)) - dim(X/R(T))$. Set of finite dimensional operators is denoted by $F(X)$ and let $G_l(A)(G_r(A))$ be set left (right) invertible elements of algebra A.

If A is a semisimple Banach algebra, x is defined to be a Fredholm element of A if there exists a $y \in A$ such that $xy-1, yx-1 \in soc(A)$. Set of Fredholm elements is denoted by $\Phi(A)$.

2. Preliminaries

Let X be Banach space and let $T \in B(X)$. Next two theorems are proved by Yood in [8]:

Theorem 2.1. *$T \in \Phi_+(X)$ and $i(T) \leq 0$ if and only if there exists $T_0 \in B(X)$ and $K \in K(X)(K \in F(X))$ such that $T = T_0 + K$ where T_0 is bounded below.*

Theorem 2.2. *$T \in \Phi_-(X)$ and $i(T) \geq 0$ if and only if there exists $T_0 \in B(X)$ and $K \in K(X)(K \in F(X))$ such that $T = T_0 + K$ and $R(T_0) = X$.*

By [2, Theorem 57.19] $T \in B(X)$ is bounded below if and only if T is not left topological divisor of zero and operator $T \in B(X)$ is onto if and only if T is not right topological divisor of zero [2, Corollary 57.17]. Let A be a semisimple Banach algebra and let H_l and H_r be a sets defined by

$$H_l = \{x \in A | x \text{ is not left topological divisor of zero }\},$$

$$H_r = \{x \in A | x \text{ is not right topological divisor of zero }\}.$$

Sets H_l and H_r are open semigroups [5, p.21].

In [6] Rowell defined set of left-Fredholm elements

$$\Phi_l(A) = \{x \in A | \exists y \in A \text{ such that } yx - 1 \in soc(A)\},$$

$$\Phi_l^{\leq 0} = \{x \in \Phi_l(A) | ind(x) \leq 0\},$$

and proved [6, Theorem 5.3] that $x \in \Phi_l^{\leq 0}(A)$ if and only if there exists $u \in soc(A)$ and $a \in G_l(A)$ such that $x = a + u$.

3. Results

In this paper we define sets

$$\Phi_+^-(A) = \{x \in A | \exists a \in H_l \exists k \in soc(A) \text{ such that } x = a + k\},$$

$$\Phi_-^+(A) = \{x \in A | \exists a \in H_r \exists k \in soc(A) \text{ such that } x = a + k\}.$$

Remark. If we put $A = B(H), H$ is Hilbert space, then $\Phi_+^-(A) = \Phi_l^{\leq 0}$ because T is not left topological divisor of zero if and only if T is left invertible in $B(H)$ [2, Theorem 57.19].

In general case is $G_l(A) \subset H_l$ and $G_r(A) \subset H_r$ [5, p.20]. From this fact we get $\Phi_l^{\leq 0}(A) \subset \Phi_+^-(A)$. If X is Banach space then $\Phi_l(X) \subset \Phi_+(X)$.

Definition 3.1. Let A be semisimple Banach Algebra. We defined sets of semi-Fredholm elements Φ_+ and Φ_- in A by

$$\Phi_+(A) = \Phi_+^-(A) \cup \Phi(A) \text{ and } \Phi_-(A) = \Phi_-^+(A) \cup \Phi(A).$$

Lemma 3.2. (1) *If $x, y \in \Phi_+^-(A)$ then $xy \in \Phi_+^-(A)$.*

(2) *If $x \in \Phi_+^-(A)$ and $k \in soc(A)$ then $x + k \in \Phi_+^-(A)$.*

(3) *If $x \in \Phi_+^-(A)$ and $\lambda \in \mathbb{C}, \lambda \neq 0$, then $\lambda x \in \Phi_+^-(A)$.*

Proof. (1) Let $x = a_1 + k_1$ and $y = a_2 + k_2$ such that $a_1, a_2 \in H_l$ and $k_1, k_2 \in soc(A)$. Then we have

$$xy = (a_1 + k_1)(a_2 + k_2) = a_1a_2 + k_1a_2 + a_1k_2 + k_1k_2 = a_1a_2 + k,$$

where we put $k_1a_2 + a_1k_2 + k_1k_2 = k \in soc(A)$. From fact that H_l is semigroup we have $a_1a_2 \in H_l$ and from this $xy \in \Phi_+^-(A)$.

(2) and (3) is obvious. □

Let $Min(A)$ be a set of minimal idempotents of A and let e be a fixed minimal idempotent of A. We shall write $\widehat{x}$ to denote left regular representation of Banach algebra A on the Banach space Ae, that is $\widehat{x}(y) = xy$ for $y \in Ae$.

Theorem 3.3. *If $x \in \Phi_+^-(A)$ then $\widehat{x}$ is semi-Fredholm operator on Ae and $i(\widehat{x}) \leq 0$.*

Proof. Let $x \in \Phi_+^-(A)$. Then there exist $a \in H_l$ and $k \in soc(A)$ such that $x = a + k$. Let $ye \in Ae$ be arbitrary element. Then we have

$$\widehat{x}(ye) = xye = (a + k)ye = aye + kye = \widehat{a}(ye) + \widehat{k}(ye) = (\widehat{a} + \widehat{k})ye.$$

From this we get $\widehat{x} = \widehat{a} + \widehat{k}$ and $\widehat{k}$ is compact on Ae because $dim(\widehat{k}) < \infty$. As $\widehat{a}$ is not left topological divisor of zero [2, Theorem 57.4], we get from Theorem 2.1 that $\widehat{x}$ is semi-Fredholm operator on Ae and $i(\widehat{x}) \leq 0$. □

If $y \in \Phi(A)$ then $\widehat{y}$ is Fredholm operator on Ae [1, Theorem F.2.6]. From this fact and Theorem 3.3 we have that if $x \in \Phi_+(A)$ then $\widehat{x}$ is semi-Fredholm operator on Ae. From [1, Example F.4.2] and fact that algebra $A/K(X)$ is commutative because it is generated by $T + K(X)$ we get that the converse of Theorem 3.3 is false.

Lemma 3.4. *Set $\Phi_+^-(A)$ is open.*

Proof. It is known that H_l is open. Let $x \in \Phi_+^-(A)$. Then there exist $a \in H_l$ and $k \in soc(A)$ such that $x = a + k$, and there exists $\epsilon > 0$ such that for $u \in A, \|u\| < \epsilon$ implies $a - u \in H_l$. Let $y \in A$ and let $\|x - y\| < \epsilon$. Then we have

$$y = x - (x - y) = a - (x - y) + k,$$

and $a - (x - y) \in H_l$. That means that $y \in \Phi_+^-(A)$. □

Let S be a subset of Banach space A. Perturbation class of set S is

$$P(S) = \{a \in A | a + s \in S \text{ for all } s \in S\}.$$

Next two lemmas are valid with assumption $\lambda S \subset S$ for every $\lambda \neq 0$.

Lemma 3.5. [3, Lemma 5.5.3] *$P(S)$ is linear subspace of A. If S is open subset of A, then $P(S)$ is closed.*

Lemma 3.6. [3, Lemma 5.5.5] *Let A be a Banach algebra with unit and let G be a group invertible elements in A. If $GS \subset S$, then $P(S)$ is a left ideal; if $SG \subset S$, then $P(S)$ is a right ideal.*

Theorem 3.7. *$P(\Phi_+^-(A))$ is a closed two sided ideal of A.*

Proof. In Lemma 3.4 is shown that $\Phi_+^-(A)$ is a open set and from Lemma 3.5 it follows that $P(\Phi_+^-(A))$ is a closed set.

Let $b \in G(A)$ and $x = a + k, a \in H_l, k \in soc(A)$. Then we have

$$bx = ba + bk = ba + k_1,$$

where we put $bk = k_1 \in soc(A)$. Suppose that $ba \notin H_l$. Then there exists a sequence $\{y_n\}_{n=1}^\infty$ in A such that $bay_n \to 0, n \to \infty$. But in this case $b^{-1}bay_n = ay_n \to 0$,
$n \to \infty$, which is impossible. Thus $ba \in H_l$. From this it follows that $bx \in \Phi_+^-(A)$, so we have

$$G(A)\Phi_+^-(A) \subset \Phi_+^-(A). \tag{1}$$

From the other side

$$xb = ab + kb = ab + k_2,$$

where $k_2 \in soc(A)$ and $ab \in H_l$. (If $ab \notin H_l$ then there exists sequence $\{z_n\}_{n=1}^\infty$ such that $abz_n = a(bz_n) \to 0, n \to \infty$ and $a \notin H_l$ what is contradiction.) It follows that

$$\Phi_+^-(A)G(A) \subset \Phi_+^-(A). \tag{2}$$

From (1), (2) and Lemma 3.6 it follows that $P(\Phi_+^-(A))$ is two sided ideal of A. □

Let now suppose that A and B are Banach algebras with identity 1 and $T : A \to B$ is a homomorphism of Banach algebras. Suppose that T is bounded and $T(1) = 1$. In [4] Harte defined $a \in A$ as Fredholm element with respect to T if and only if $T(a) \in G(B)$. Analogously we define left and right Fredholm elements with respect to T as

$$\Phi_l(A) = \{a \in A | T(a) \in G_l(B)\},$$

$$\Phi_r(A) = \{a \in A | T(a) \in G_r(B)\}.$$

Next Lemma follows immediately:

Lemma 3.8. *(1) If $x, y \in \Phi_l(A)$ then $xy \in \Phi_l(A)$.*
(2) If $xy \in \Phi_l(A)$ then $y \in \Phi_l(A)$.
(3) If $x \in \Phi_l(A)$ and $u \in N(T)$ then $x + u \in \Phi_l(A)$.
(4) If $x \in \Phi_l(A)$ and $\lambda \in \mathbb{C}, \lambda \neq 0$ then $\lambda x \in \Phi_l(A)$.
(5) $\Phi_l(A)$ is open set.

Theorem 3.9. *$P(\Phi_l(A))$ is a closed two sided ideal of A.*

Proof. $\Phi_l(A)$ is open set, so $P(\Phi_l(A))$ is closed set of A.

Let $z \in G(A), y \in \Phi_l(A)$ and let $b \in B$ be a left inverse for $T(y)$. Now we have

$$bT(z^{-1})T(zy) = bT(z^{-1})T(z)T(y) = bT(y) = 1,$$

so $zy \in \Phi_l(A)$ and $G(A)\Phi_l(A) \subset \Phi_l(A)$.

Similarly $\Phi_l(A)G(A) \subset \Phi_l(A)$, and by Lemma 3.6 it follows that $P(\Phi_l(A))$ is a two sided ideal of A. □

References

[1] Barnes, B. A., Murphy, G. J., Smyth, M. R. F. and West, T. T., *Riesz and Fredholm theory in Banach algebras*, Pitman Research Notes in Math. 67 Boston, 1982..

[2] Berberian, S. K., *Lectures in Functional Analysis and Operator Theory*, Springer-Verlag, New York, Heidelberg Berlin, 1974..

[3] Caradus, S. R., Pfaffenberger, W. and Yood, B., *Calkin algebras and algebras of operators on Banach spaces*, Marcel Dekker Inc. New York, 1974..

[4] Harte, R., *Fredholm theory relative to a Banach algebra homomorphism*, Math.Z. **179** (1982), 431–436.

[5] Rickart, C. E., *General theory of Banach algebras*, Princeton, Van Nostrand 1960..

[6] Rowell, J. W., *Unilateral Fredholm theory and unilateral spectra*, Proc.R.Ir.Acad. Vol. **84a,No.1** (1984), 69–85.

[7] Stojković, N., *Fredholmova teorija sa algebarskog aspekta*, Magistarski rad, Filozofski fakultet u Nišu, 1993..

[8] Yood, B., *Properties of linear transformations preserved under addition of a completely continuous transformation*, Duke Math.J. Vol **18** (1951), 599–612.

University of Niš, Faculty of Economics, Trg JNA 11, 18000 Niš, Yugoslavia

FILOMAT (Niš) 9:3 (1995), 947–959

Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

ON THE ANTIINVERSE AND COREGULAR SEMIGROUPS AND SOME THEIR APPLICATIONS

Kalčo Todorov

ABSTRACT. In the present survey are considered some basic results, conserning the two – sided identical semigroups, their subclasses of antiinverse and coregular semigroups, as well as some application of this investigations towards problem of the Hadamard matrices.

1. Introduction and preliminaries

An element a of a semigroup S we call *two – sided identical* if there exists an element $b \in S$ (which we call neutral to a) such that

$$a = bab \tag{1}$$

A semigroup S we call *two – sided identical* if every element of S is two-sided identical.

For elements a and b of a semigroup S, we say that they are *mutually antiinverse* if the following conditions hold

$$aba = b \quad \text{and} \quad bab = a. \tag{2}$$

A semigroup S is *antiinverse* if for every $a \in S$ there exists its antiinverse element $b \in S$.

An element a of a semigroup S we call *coregular* and $b \in S$ its *coinverse*, if

$$a = aba = bab. \tag{3}$$

A semigroup S we call *coregular*, if every element of S is coregular.

It is evident that every coregular semigroup is two – sided identical. Moreover, every coregular semigroup is simultaneously regular and antiinverse, but the converse is not true.

The first systematic investigation of the above shown classes semigroups are the ones of the antiinverse semigroups and are included in a sequence of papers of S. Bogdanović and other as well as in the paper [21] of Sharp. Almost simultaneously with the investigations of antiinverse semigroups have appeared also the ones of the coregular semigroups (see Bijev, Todorov [2]).

Somewhat later have been introduced and investigated by the author the two – sided identical semigroups (see K. Todorov [22]).

The class of the two – sided identical semigroups is a quite large class of semigroups. To it belong all the semigroups, possesing two – sided identity element (in particular the groups). A lot of examples indicate however the presence of two – sided identical semigroups without two – sided identity element. To the class of two – sided identical semigroups belong some of the regular semigroups – such as for example any antiinverse semigroup. This two classes does not cover each other, as shown in the coming two examples.

A more complete idea about semigroups, belonging to the intersection of the classes of two – sided identical, regular and coregular semigroups can be obtain from the paper [3] by Bijev, Todorov , where has been delivered a complete classification of the abstract semigroups, included in the symmetric semigroup T_3 of degree 3, whereby for each one of them is remarked, whether it is reguilar, two – sided identical or coregular.

From the results known for the two – sided identical semigroups I'll cite the following statements.

Theorem 1. *Let* $\mathbf{E}(S)$ *be the set of the idempotents of the semigroup* S, *let* $N(+)$ *be the additive semigroup of natural numbers and let for the elements* $a,\ b \in S,\ e \in \mathbf{E}(S)$ *and* $m \in N(+)$ *we have*

$$a = bab \quad \text{and} \quad b^m = e. \tag{4}$$

Then

a) $a^2 = (ab^i)^2 = (b^ia)^2, \qquad i = 1, 2, ..., m;$

b) $a^{2k} = b^pa^{2k}b^{m-p};\ a^{2k-1} = b^pa^{2k-1}b^p,\ \ p = 1, 2, ..., m,\ k = 1, 2, ...$

c) Any element of the semigroup $< a, b >$ *can be represented as the type* a^ib^j, *where* $i = 0, 1, ...;\ j = 0, 1, ..., m-1$ *and* $i + j > 0$.

Both the antiinverse and coregular semigroups may be considered as a subclasses of the class of the two – sided identical.

Coregular and antiinverse form subclasses of the class of regular semigroups. Although coregular semigroups form a subclass of the class of antiinverse semigroups, strictly containing the class of commutative antiinverse semigroups.

From the theory of the antiinverse semigroups I schall cite the following statements

Theorem 2. *Let S be a semigroup. Then the semigroup S is a antiinverse iff*

$$(\forall a \in S)(\exists b \in S)(a^2 = b^2,\ ba = a^3 b,\ a^5 = a).$$

Theorem 3. *Let G be a group. Then G is antiinverse iff G is a union of subgroups which belong to the class of the trivial group, of the cyclic group of order 2 and of the quaternion group.*

Theorem 4. (*Bogdanović, [7] Theorem* 4.1) *Let S be a semigroup. Then all proper subsemigroups of S are antiinverse iff one of the following conditions hold:*

1. $(\forall a \in S)(a = a^3)$,
2. *S is a cyclic group of prime-power order* (> 1),
3. *S is the cyclic group of order* 4,
4. *S is $M(2,1)$ semigroup, i.e.* $S =< a >: a^{2+1} = a^2$,
5. *S is $M(2,2)$ semigroup.*

Define a relation ρ on a semigroup S as follows: $a\rho b \Leftrightarrow a$ and b are antiinverse in S. Let

$$\mathrm{S}[a] = \{x \in S \mid x\rho a\},$$

for all $a \in S$.

Theorem 5. *If S is a commutative semigroup, the following are equivalent:*

(i) *S is antiinverse.*
(ii) *ρ is a congruence on S.*
(iii) **S**$[a]$ *is a subsemigroup of S , for all a in S.*
(iv) $a^3 = a$ *, for all a in S.*

Coregular in the multiplicative matrix semigroup $\mathbf{M}_2(\mathbf{R})$ of real matrix is the matrix

$$\begin{pmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{pmatrix}, \quad 0 \le \beta < 2\pi,\ \beta \ne \pi,$$

defining the axial symmetries of the figures of a given plane.

In the class of coregular semigroups are valid the following statements:

Theorem 6. *Let the elements a and b of a semigroup S satisfy condition* (e). *Then:*

a) $a^3 = a$.
b) $a^2b^2a^2 = a^2 = b^2a^2b^2$.

c) $a^2b = ab^2 = a$.
d) $b^2 = b$ *implies* $a^2 = a$.
e) $a^2 = b^2$ *implies* $a = b^3$.

Theorem 7. *For a semigroup* S *the following conditions are equivalent:*
a) S *is coregular.*
b) $a^3 = a$ *for every element* a *of* S.
c) S *is a union of disjoint groups, the elements of which are of order* ≤ 2.

Some interesting continuations of the already shown classes of semigroups are containted further in the papers of Bijev [1] and Chvalina and Matouškova [10].

Accordingh to Bijev [1] any representation of the element x of an arbitrary semigroup S in the form

$$x = ab, \quad a, b \in S \text{ for } a^3 = a \text{ and } b^3 = b$$

is called coregular. A good motivation for the examination of this representations is the fact, that in the multiplicative semigroup of all ortogonal matrices of rank 2

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \quad \begin{pmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{pmatrix}, \quad 0 \leq \beta < 2\pi,$$

the matrix of the second type are coregular, and the matrix of the first type are coregular representative.

Theorem 8. *Let* $(\mathbf{A}, f)$ *be a unar. Then the semigroup End* $(\mathbf{A}, f)$ *is coregular iff it is antiinverse.*

Theorem 9. *Let* T_X *be the symmetric semigroup on the set* $\mathbf{X}$. *The following conditions are equivalent:*
1) T_X *is coregular.* 2) T_X *is antiinverse.* 3) *card* $\mathbf{X} \leq 2$.

Theorem 10. *Let* $\mathbf{X}$ *be an infinite set. There exists a coregular commutative subsemigroup* H_X *of the symmetric semigroup* T_X *(not generated by idempotent elements only) such that card*$\mathbf{H}_X$ = *card* $\mathbf{X}$.

A quite striking application of the two – sided identical semigroups turned out to be the one bound to the problem of the Hadamard matrices.

The quaternion group Q is an origin in the study of the antiinverse semigroup, which as well the coregular semigroups, present subclasses of the class of two-sided identical semigroups. In this respect the quaternion group Q as a two – sided identical semigroup admits further generalizations (see Magnus and al.[18] and Neuman [19]) about its genetic code $Q = \langle i, j : \ i = jij, \ j = iji \rangle$

The following statements hold:

Theorem 11. *Let S be the semigroup generated by the elements of a countable set $\mathbf{M} = \{a_1, a_2, ...\}$ subject only to the relations*

$$a_1 = a_2 a_1 a_2, \; a_i = a_j a_i a_j = a_{i+1} a_i a_{i+1} \tag{5}$$

for each two elements $a_i, a_j \in \mathbf{M}$ with $1 \leq j < i$. Then:

1) $a_i^2 = a_j^2 = (a_k a_l)^2$ *for* $a_i, a_j, a_l, a_l \in \mathbf{M}$ *with* $k \neq 1$. *Further is putting* $a_i^2 = -1$.

2) $a_i a_j = -a_j a_i$ *for* $a_i, a_j \in \mathbf{M}$ *with* $i \neq j$.

3) $a_i a_j a_i = a_j$ *for every two elements* $a_i, a_j \in \mathbf{M}$ *with* $i \neq j$.

4) S *is a group.*

5) *Let* $a \in \mathbf{M}$ *and let* l *denote the number of the factors of the element* $c = c_1 c_2 ... c_k$, $(c_i \in \mathbf{M}, i = 1, 2, ..., k)$ *equal to the element* a. *Then* $ca = (-1)^{k-l} ac$.

6) $a^4 = 1$ *for every element* $a \in S$.

7) $|\mathbf{C}(S)| = 2$.

8) *The group* S *contains only two conjugate singleton classes and each one of its other conjugate classes is a two – element one.*

9) *Any subgroup of* S *is a normal subgroup of* S *iff it contains at least one conjugate class distinct from the class of* 1.

10) *The centralizer* $\mathbf{C}_S(c)$ *of every element* $c \in S$ *is infinite.*

11) $\langle a_i, a_j \rangle \cong Q$ *for* $a_i, a_j \in \mathbf{M}$ *with* $i \neq j$.

12) S *is a locally finite group.*

13) *The subsemigroups of* S *generated by the subsets* $\mathbf{U}$ *of* $\mathbf{M}$ *with* $n = |\mathbf{U}| > 1$ *are groups having properties* 1)–6), 9), 11).

14) *Let* $\mathbf{C}(S_n)$ *denote the center of the semigroup* $S_n = \langle a_1, ..., a_n \rangle$ *then*

$$\mathbf{C}(S_n) = \begin{cases} \{\pm 1, \pm a_1 a_2 ... a_n\} & \textit{is odd;} \\ \{\pm 1\} & \textit{is even;} \end{cases}.$$

15) *Each element* $c \in S_n$ *may be written in the form*

$$c = (-1)^{u_o} a_1^{u_1} a_2^{u_2} ... a_n^{u_n},$$

where $u_i \in \{0; 1\}, i = 0, 1, ..., n$.

16) *Let* $a \in \mathbf{M}_n = \{a_1, ..., a_n\}$ *and let* l *denote the number of the factors of the element* $c = c_1 c_2 ... c_k, (c_i \in \mathbf{M}_n, i = 1, 2, ..., k)$, *coinciding with the element* a. *Then* $ca = (-1)^{k-l} ac$.

17) *The conjugate class* $K_a = \{bab^{-1},\ b \in S_n\}$ *for every element* $a \in \mathbf{S}_n \backslash \{1, -1\}$ *is two-element and coincides with the set* $\{\pm a\}$.

18) $|\mathbf{S}_n| = 2^{n+1}$.

Corollary 1. *Every element c of the semigroup S of order 4 (defined in theorem 8) generates a normal subgroup of S.*

The properties 1)–18) of the semigroup S in Theorem 11 strongly depend on its genetic code (5). Adding a new relation give a new semigroup. The following Theorem 12 is a special case of Theorem 11.

Theorem 12. *Let S be a semigroup generated by the elements of the countable set $\mathrm{M} = \{a_1, a_2, ...\}$ subject only to the relations (5) and $a_k = a_k^2$ for some $a_k \in \mathrm{M}$. Then S is a commutative group with identity element $1 = a_k$ in which each non-identity element is an element of order 2.*

The groups $S_n = \langle a_1, a_2, ..., a_n\rangle$ in Theorem 11 we shall call n – *generated q. groups.*

There exists a close connection between the so constructed n–generated q. groups and the Hadamard matrices.

As is well known, in 1893 Hadamard proved that if $X = (x_{ij})$ is a square matrix of order n then holds the inequality

$$|detX|^2 \leq \prod_{i}^{n} \sum_{j=1}^{n} |x_{ij}|^2,$$

where the equality

$$|detX|^2 = \prod_{i}^{n} \sum_{j=1}^{n} |x_{ij}|^2$$

holds iff

$$\sum_{k=1}^{k=n} x_{ik}x_{jk} = 0,\ i \neq j,\ i,j = 1, ..., n \ \ or \ \ x_{ij} = 0 \ \ for\ some\ i.$$

By definition, a square matrix H of order n whose entries are $+1$ and -1 is called a Hadamard matrix of order n provided that its rows are pairwise orthogonal, in other words

$$\mathbf{HH'} = \mathbf{H'H} = n\mathbf{E},$$

where $\mathbf{H'}$ is the transposed matrix of $\mathbf{H}$.

It is known (see Hedayat and Wallis [15]) that there exist Hadamard matrices of orders 1 and 2, but it can be shown that every other Hadamard matrix has order $4t$ for some positive integer t. The question: "How many different Hadamard matrices of a given order might exist?" is a very difficult question to answer and researchers' interest in it varies for almost a whole century.

Hadamard matrices of infinitely many orders have been constructed, and it has been conjectured that one exists for every t, but no general proof is available, and the number of unsettled orders is infinite.

From the above described basic properties of the n–generated quaternion group one can construct Hadamard matrices by fixing a representative from every conjugate class K_a, $a \notin \{1,-1\}$ and from the set $\{1,-1\}$. Taken in a given order the so fixed elements of the n–generated quaternion group S_n, together with the rows and columns corresponding to them comprise a subtable of the multiplicative table of the semigroup S_n. The signs of the elements of this subtable, taken in the same order, form the Hadamard matrix. In this way we obtain from the q. group Q the following Hadamard matrix:

$$\begin{array}{cccc} {}_1\backslash^1 & i & j & k \\ i & -1 & k & -j \\ j & -k & -1 & i \\ k & j & -i & -1 \end{array} \qquad \Rightarrow \qquad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

These relations are described by the author in details in [24].

What are the further possibilities for obtaining of the Hadamard matrices on the basis of the considered n–generated q. groups? As the calculations show with the multiplication table of the integral quaternions

$$\pm 1,\ \pm i,\ \pm j,\ \pm k, \quad \frac{\pm 1\ \pm i\ \pm j\ \pm k}{2}$$

is associated Hadamard matrices of order 12 of the type

$$\begin{pmatrix} A_1 & A_2 & A_3 \\ -B_1 & B_2 & B_3 \\ C_1 & -C_2 & -B_2 \end{pmatrix}$$

where $\mathbf{A}_i$, $\mathbf{B}_j$, $\mathbf{C}_k$ are Hadamard matrices of order 4.

All this is a good motivation for the further investigations of the corresponding algebras of the n-generated q. group. For the present because of the principal difficulties as compared to the already known results a more thorough investigation is made on algebra of the 3–generated q. group. The received basic results can be formulated as follows:

Lemma 1. *The (group) algebra H_3 (of the quaternion group S_3 in Theorem* 11) *over the field* $\mathbf{R}$ *of the real number is a* 8*–dimensional (associative) algebra, containing the quaternion algebra H.*

Every element $a=\sum_{i=0}^{i=7}a_ie_i\in H_3$ *is a divisor of zero iff*

$$a_7=\varepsilon a_0,\ a_6=-\varepsilon a_1,\ a_5=\varepsilon a_2,\ a_4=-\varepsilon a_3,$$

where $\varepsilon=\pm1$. *If*

$$a=a_0(1+\varepsilon e_7)+a_1(e_1-\varepsilon e_6)+a_2(e_2+\varepsilon e_5)+a_3(e_3-\varepsilon e_4)$$

and

$$a^*=b_0(1-\varepsilon e_7)+b_1(e_1+\varepsilon e_6)+b_2(e_2-\varepsilon e_5)+b_3(e_3+\varepsilon e_4).$$

then

$$aa^*=a^*a=0.$$

Theorem 13. *Let us put*

$$I^-=\{q^-=a_0(e_0-e_7)+a_1(e_1+e_6)+a_2(e_2-e_5)+a_3(e_3+e_4)\},$$

$$I^+=\{q^+=b_0(e_0+e_7)+b_1(e_1-e_6)+b_2(e_2+e_5)+b_3(e_3-e_4)\},$$

where $a_i,b_j\in\mathbf{R}$. $i,j=0,1,..,7$. *Then:*

1) $I^-\cap I^+=\{0\}$,

2) I^- *and* I^+ *are the uniquely non-trivial* minimal *ideals of the algebra* H_3,

3) $I^-\cong H\cong I^+$, *where* H *is the quaternion algebra,*

4) $H_3=I^-+I^+$.

Let us put

$$2f_i=\begin{cases} e_i-e_{7-i}, & \text{if } i=0,2;\\ e_i+e_{7-i}, & \text{if } i=1,3,5,7;\\ e_{7-i}-e_i, & \text{if } i=4,6.\end{cases}$$

Here, to every element $a\in H_3$, the norm $N(a)\in\mathbf{R}$ and minimal polynomial are associated. When determining these characteristics is used the regular representation of the quaternions of the algebra H_3 by means of finite matrices.

Further, for every element $a\in H_3$ are determined the essential properties of its conjugate element $\overline{a}\in H_3$, as well as the integral elements of the algebra H_3.

By formulating the propositions and their proofs equally both the basis f and the basis e of the algebra H_3 is used.

2. Norm and minimal polynomial

Let us recall that if

$$q = a_0 + a_1 i + a_2 j + a_3 k \tag{6}$$

is an arbitrary element of the quaternion algebra H, then:

1) the positive number $N(q) = \sum_{i=0}^{i=3} a_i^2$ is the norm of the quaternion q;
2) $h(x) = x^2 - 2a_0 x + N(q)$ is the minimal polynomial, satisfied by the quaternion q;
3) $\bar{q} = a_0 - a_1 i - a_2 j - a_3 k$ is the conjugate to a quaternion, as $q\bar{q} = \bar{q}q = N(q)$;
4) by the regular representation of the algebra H (over the field of real numbers $\mathbf{R}$) by Skornjakov [28] and Deuring [13] to the quaternion q in (6) corresponds the matrix

$$\mathbf{A} = \begin{pmatrix} a_0 & -a_1 & -a_2 & -a_3 \\ a_1 & a_0 & -a_3 & a_2 \\ a_2 & a_3 & a_0 & -a_1 \\ a_3 & -a_2 & a_1 & a_0 \end{pmatrix},$$

where $N(q) = \sqrt{|A|}$ and $|A|$ is the determinant of the matrix A.

Remark 1. Here and on, $(a)^{1/2k}$ for $a \geq 0$ and k - a natural number means the arithmetical root of the number a, i. e. the module of the complex number $(a)^{1/2k}$.

Let $q = \sum_{i=0}^{i=7} c_i f_i = q_1 + q_2$ be an arbitrary element of the quaternion algebra H_3, where

$$q_1 = \sum_{i=0}^{i=3} c_i f_i \in I^-, \qquad q_2 = \sum_{i=4}^{i=7} c_i f_i \in I^+,$$

i.e.

$$q_1 = (c_0, c_1, c_2, c_3, 0, 0, 0, 0), \qquad q_2 = (0, 0, 0, 0, c_4, c_5, c_6, c_7) \in H_3.$$

The positive number $N(q)$:

$$N(q) = N(q_i) \in \mathbf{R}, \qquad if\, q = q_i,\ i = 1, 2$$

or

$$N(q) = (N(q_1)N(q_2))^{1/2} \in \mathbf{R}, \qquad if\ q \neq q_i, i = 1, 2$$

we shall call the *norm* of the quaternion q.

For the elements of the quaternion algebra H_3 a describtion of the above characteristics is given in the following.

Theorem 14. *Let $q = \sum_{i=0}^{i=7} c_i f_i$ be an arbitrary element of the quaternion algebra H_3 and let*

$$q_1 = \sum_{i=0}^{i=3} c_i f_i \in I^-,\ q_2 = \sum_{i=4}^{i=7} c_i f_i \in I^+,\ q \neq q_i,\ i = 1, 2.$$

Then:

1) *of the basis f by a regular representation of H_3, the matrix*

$$A_f = \begin{pmatrix} B_f & 0 \\ 0 & C_f \end{pmatrix},$$

corresponds to the quaternion q, where

$$B_f = \begin{pmatrix} c_0 & -c_1 & -c_2 & -c_3 \\ c_1 & c_0 & -c_3 & c_2 \\ c_2 & c_3 & c_0 & -c_1 \\ c_3 & -c_2 & c_1 & c_0 \end{pmatrix}, \quad C_f = \begin{pmatrix} c_7 & -c_6 & c_5 & c_4 \\ c_6 & c_7 & -c_4 & c_5 \\ -c_5 & c_4 & c_7 & c_6 \\ -c_4 & -c_5 & -c_6 & c_7 \end{pmatrix}.$$

2) $N(q) = \sqrt[4]{|A_f|}1 = ((\sum_{i=0}^{i=3} c_i^2)(\sum_{i=4}^{i=7} c_i^2))^{1/2} = (N(q_1)N(q_2))^{1/2}$ *is the norm of the quaternion q (see Remark 1).*

3) *the quaternion $q = \sum_{i=0}^{i=7} c_i f_i$ satisfies the minimal polynomial*

$$h(x) = x^4 - 2(c_0 + c_7)x^3 + (N(q_1) + N(q_2) + 4c_0c_7)x^2 - \\ -2(c_7 N(q_1) + c_0 N(q_2))x + N^2(q).$$

Corollary 2. *Let $q = \sum_{i=0}^{i=7} c_i f_i$ be an arbitrary element of the quaternion algebra H_3 and let*

$$q_1 = \sum_{i=0}^{i=3} c_i f_i \in I^-,\ q_2 = \sum_{i=4}^{i=7} c_i f_i \in I^+.$$

The norm $N(q)$ of every element $q \in H_3$ has the following basic properties:

1. $N(q) \geq 0$, *and* $N(q) = 0$ *iff* $q = 0$.

2. $N(\alpha q) = \alpha^2 N(q)$ *for* $\alpha \in \mathbf{R}$ *and* $q \in H_3$.

3. $N(xy) = N(x)N(y) = N(yx),\ x, y \in H_3$.

4. Let $q = \sum_{i=0}^{i=7} a_i e_i$,

$$s = \sum_{i=0}^{i=7} a^2 \quad \text{and} \quad t = 2(-a_0a_7 + a_1a_6 - a_2a_5 + a_3a_4), \tag{7}$$

then

i) $N^2(q) = s^2 - t^2 = (s+t)(s-t) =$
$= [(a_0 - a_7)^2 + (a_1 + a_6)^2 + (a_2 - a_5)^2 + (a_3 + a_4)^2]x$
$[(a_0 + a_7)^2 + (a_1 - a_6)^2 + (a_2 + a_5)^2 + (a_3 - a_4)^2]$.

ii) $h(x) = x^4 - 4a_0x^3 + 2[2a_0^2 + s - 2a_7^2]x^2 - 4(a_0 s + a_7 t)x + s^2 - t^2$.

3. Conjugate elements

Let

$$q = \sum_{i=0}^{i=7} c_i f_i = q_1 + q_2,$$

where

$$q_1 = \sum_{i=0}^{i=3} c_i f_i \in I^-, \quad q_2 = \sum_{i=4}^{i=7} c_i f_i.$$

The quaternion $\bar{q} \in H_3$:

$$\bar{q} = \bar{q}_i, \text{ if } q = q_i, \ i = 1, 2$$

or

$$\bar{q} = (c_0 - \sum_{i=0}^{i=3} c_i f_i)(N^{-1}(q_1)N(q_2))^{1/2} +$$

$$+(c_7 f_7 - \sum_{i=4}^{i=6} c_i f_i)(N(q_1)N^{-1}(q_2))^{1/2},$$

if $q \neq 0$ and $q \neq q_i$, $i = 1, 2$, we shall call *conjugate* of the quaternion $q \in H_3$.

Theorem 15. *For every quaternion q of H_3 and its conjugate $\bar{q}$ there hold:*

1) $q\bar{q} = \bar{q}q = N(q)$.
2) $N(q) = N(\bar{q})$.
3) $\bar{\alpha}\bar{q} = \alpha\bar{q}$ *for every* $\alpha \in \mathbf{R}$.
4) $\bar{\bar{q}} = q$.
5) $\bar{x}\bar{y} = \bar{y}\bar{x}$ *for* $x, y \in H_3$.

4. Integral elements

The integral quaternions of the algebra H have been determined for a first time by Hurwitz in [16] (see Hurvitz [17]). The method which we use in the present paragraph is more sophisticated than the one given by Dickson in [14].

The integral quaternions of the algebra H (see Dickson p. 148, Theorem 1) are given by

$$q = a_0\rho + a_1 i + a_2 j + a_3 k, \ \rho = \frac{1 + i + j_k}{2} \tag{8}$$

for integral values of a_0, a_1, a_2, a_3. Expressed otherwise, they are the quaternions whose four coordinates are either all integers or all halves of odd integers.

Proposition 1. *(see Dickson, p. 157) The first components of the elements of any (maximal) set of integral elements, with properties* **R**, **C**, **U** *of a direct sum $B+C$ constitute a (maximal) set of integral elements of the first component algebra B, and similarly for the second components. Conversely, given a (maximal) set* **b** *of integral elements b of a rational algebra B and a (maximal) set* **c** *of integral elements c of another rational algebra C, such that B and C have module β and γ and have a direct sum, then if we add every b to every c we obtain sums forming a (maximal) set of integral elements of the direct sum $B+C$.*

The next theorem is also immediate from the foregoing Proposition and (8).

Theorem 16. *The integral quaternions of the algebra H_3 are given by*

$$q = a_0\rho + a_1 f_1 + a_2 f_2 + a_3 f_3 + a_4 f_4 + a_5 f_5 + a_6 f_6 + a_7\tau,$$

where

$$\rho = \frac{f_0 + f_1 + f_2 + f_3}{2}, \qquad \tau = \frac{f_4 + f_5 + f_6 + f_7}{2},$$

for integral values of a_0, a_1, ..., a_7.

References

[1] Бијев Г., *Корегулярни представяния на полугрупи*, vol. XXV, Год. ВУЗ –приложна математика, 1980.

[2] Bijev G. and Todorov K., *Coregular semigroups*, Notes on Semigroups VI **4** (Budapest 1980), 1-11.

[3] Bijev G. and Todorov K., *On the representation of abstract semigroups by transformation semigroups: computer investigations*, Semigroup Forum **43** (1991), 253 - 256.

[4] Bogdanović S., Milic S. and Pavlovic V., *Antiinverse semigroups*, Publ. Inst. Math. **24 (38)** (1978), 19-28.

[5] Bogdanović S., *Deux caracterisations des semigroupes anti–inverses*, Bull.des travaux de *la* Faculté des Sciences - Université de Novi Sad **8** (1978), 79 - 81.

[6] Bogdanović S., *On anti – inverse semigroups*, Publ. Inst. Math. **25 (39)** (1979), 25 - 31.

[7] Bogdanović S., *Sur les demi – groupes dans lesquels tous les sous – demi – groupes properes sont idempotents*, Mathematics Seminar Notes **9** (1981), 17 - 24.

[8] Bogdanović S., *Semigroups with a system of subsemigroups*, Inst. of Math., Novi Sad, 1985.

[9] Bogdanović S. and Ćirić M., *Polugrupe*, Prosveta, Niš, 1993.

[10] Chvalina J. and Matouškova K., *Coregularity of endomorphism monoid of unars*, Archivum Mathematicum **20** (1984), 43 - 48.

[11] Clifford H. A. and Preston B. G., *The algebraic theory of semigroups I*, Amer. Math. Soc., 1961.

[12] Clifford H. A. and Preston B. G., *The algebraic theory of semigroups I*, Amer. Math. Soc., 1961.

[13] Deuring M., *Algebren*, Springer - Verlag, Berlin - Heidelberg - New York, 1968.

[14] Dickson L. E., *Algebras and their arithmetics*, New York, 1960.

[15] Hedayat A. and Wallis D. W., *Hadamard Matrices and their applications*, Annals of Statistics 6 **6** (1978), 1184-1238.

[16] Hurvitz A., *Vorlesungeüber die Zahlentheorie der Quaternionen*, Springer Verlag, Berlin, 1919.

[17] Hurvitz A., *Mathematische Werke* Bd - 2, Basel etc. Birkhäuser, 1963.

[18] Magnus W., Karras A. and Solitar D., *Combinatorial Group Theory*, Interscience, New York, 1966.

[19] Neuman B. H., *Some remarks on semigroup presentations*, Canad. J. Math. **19** (1967), 1018-1026.

[20] Sharp J. C., *Anti-regular Semigroups*, Not. Amer. Math. Soc. **24:2** (1977), A -206.

[21] Sharp J. C., *Anti-regular Semigroups*, Publ. Inst. Math. **24 (38)** (1978), 147 - 150.

[22] Todorov K., *Two - sided identical (antiregular) semi- groups*, Conference on Theory and Applications of Semigroups, Greifswald, Nov 12-16, 1984, pp. 123.

[23] Todorov K., *Generalized quaternion group*, MTA SZTAKI Kozlemwnyek **38** (1988), 53-65.

[24] Todorov K., *On the generalized quaternion group and the Hadamard matrices*, MTA SZTAKI Kozlemwnyek **38** (1988), 67-79.

[25] Todorov K., *On a generalization of the quaternion group and of the quaternion algebra*, Potsdamer Forschungen. Reihe **B**. Heft **62** (1989), 101 -107.

[26] Todorov K., *On one generalization of the quaternion algebra*, Math. Balcanica (to appear).

[27] Дрозд, Ю, А. В. Кириченко, *Конечномерные алгебры*, Киев, "Виша школа", 1980.

[28] Скорняков, Л. А, *Элементы алгебры*, "Наука", Москва, 1966.

Bulgarian Academy of Sciences, Institute of Mathematics, POBox 373, 1090 Sofia, Bulgaria

E-mail address: `kjt@ bgearn.acad.bg`

FILOMAT (Niš) 9:3 (1995), 961–965
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

GENERATION OF PRIMITIVE BINARY POLYNOMIALS

Miodrag Živković

ABSTRACT. Binary linear recurrent sequences with the primitive characteristic polynomial are used as a good approximation of the random sequence.

A known algorithm is implemented in a program for generation of primitive binary polynomials of degree < 5000 with the given number of terms.

An account is given of problems solved during the program developement. Practically hardest among them is the problem of obtaining the factorizations of the numbers $2^n - 1$.

Let F_q denote the finite field of order $q = p^n$, where p is prime and $n \geq 1$. The multiplicative group F_q^* of nonzero elements of F_q is cyclic and a generator of F_q^* is called a primitive element. A monic irreducible polynomial whose roots are primitive elements is called a primitive polynomial.

It is known [2] that the binary (over $GF(2)$) sequence $\{a_n\}_{n\geq 0}$ satisfying the linear recurrent relation $a_{k+n} = \sum_{i=0}^{n-1} a_{k+i} f_i$ possesses good statistical properties if its characteristic polynomial $f = \sum_{i=0}^{n} f_i x^i$ is primitive. For example, the period length of such sequence (so called m–sequence) is $N = 2^n - 1$, the difference between the number of ones and zeros across the period is exactly one, and each n–tuple from $\{0,1\}^n$ except $\mathbf{0}$ appears exactly once in one period. The two sequences $\{a_{n+r}\}$ and $\{a_{n+s}\}$, $\quad 0 \leq r < s \leq N$, are mutually orthogonal, i.e., the equality

$$\sum_{n=0}^{N} (-1)^{a_{n+r}} (-1)^{a_{n+s}} = \begin{cases} N+1, & \text{for } 0 \leq r = s \leq N \\ -1, & \text{for } 0 \leq r < s \leq N. \end{cases}$$

holds. m–sequences are used for obtaining uniformly distributed random numbers [9]. Another field where m–sequences are widely used is cryptology [6]. Quality of m–sequences grows with n, and therefore there is a need to obtain primitive polynomials of degree as large as possible.

Supported by Grant 0401B of RFNS through Math. Inst. SANU.

There are several published tables of primitive binary polynomials. Watson [11] gives for $n < 100$ one primitive of degree n, and Stahnke [8] lists for each $n \leq 168$ a primitive with a minimum number of nonzero coefficients (trinomial or pentanomial). Zierler and Brillhart [13,14] extended this work by listing all primitive and irreducible trinomials of degree $n \leq 1000$, with the period for some for which the factorization of $2^n - 1$ is known. Rodemich and Rumsey [7] have listed all primitive trinomials of degree M_j, $12 \leq j \leq 17$ (here M_j denotes the jth Mersenne exponent, the prime for which $2^{M_j} - 1$ is also prime). The list has been extended by Zierler [12], Kurita and Matsumoto [4] and Heringa, Blöte and Compagner [3] up to $M_{23} = 11213$, $M_{28} = 86243$ and $M_{31} = 216091$ correspondingly. One primitive pentanomial of each degree M_j, $8 \leq j \leq 27$ is also listed in [4]. For those $n < 5000$, for which the factorization of $2^n - 1$ is known, in [1̆3,1̆4] the first primitive trinomial (if such exists) and a randomly generated primitive 5– and 7–nomial of degree n in $GF(2)$ are given.

In this paper we give some characteristics of the algorithm for generation of primitive binary polynomials which is used to assemble the tables in [1̆3,1̆4].

Generation of primitive polynomials is performed by testing primitivity of the sequence of trial polynomials from the given set. Here we deal with the set of polynomials f of degree n with t terms, for given n and odd t, with the constraint $f(0) = 1$. The number t is usually small, to enable simple calculation of the corresponding linear recurrent sequence. The sequence of trial polynomials is formed using the linear recurrent sequence of order 127 as a source of random numbers.

The primitivity test of a given polynomial f is effectively performed using the following set of conditions [5, Th 3.18]

$$f(0) = f(1) = 1, \tag{1}$$

$$\min\{k \mid \ f \mid x^{2^k} - x\} = n, \tag{2}$$

$$\text{for all prime } p \mid 2^n - 1 \qquad f \nmid x^{(2^n-1)/p} - 1. \tag{3}$$

The condition (1) eliminates polynomials divisible by x and $x + 1$. As t is odd for trial polynomials, this condition is automatically fulfilled.

The polynomial $x^{2^k} - x$ is equal to the product of all irreducible polynomials of degree dividing k [5, Th 3.20], and therefore the irreducible polynomials satisfy the conditions (1) and (2). The inverse is not true: a polynomial equal to the product of different irreducible polynomials of degrees dividing n satisfy (1) and (2) (and it is not irreducible). The number of irreducible polynomials of degree n equals to [5, Th 3.25]

$$\frac{1}{n}\sum_{d|n}\mu(d)2^{n/d},$$

where μ is the Moebius function, defined by

$$\mu(n)=\begin{cases}1 & \text{if } n=1\\ (-1)^r & \text{if } n \text{ is the product } r \text{ distinct primes}\\ 0 & \text{if } n \text{ is divisible by the square of a prime.}\end{cases}$$

To check if f satisfy (2), it is necessary to calculate n residues $x^{2^k}-x\,(\mathrm{m}od\,f)$. Squaring in $\mathrm{G}F(2)$ is simple, because $(a+b)^2=a^2+b^2$. After each squaring, a residue is calculated from the division of a polynomial of degree at most $2n-2$ by f. The fact that f is sparse is used to perform division more efficiently. The total number of elementary (in$\mathrm{G}F(2)$) operations needed to test the condition (2) is bounded by $O(tn^2)$, which is not small, having in mind the number of polynomials that need to be checked. The problem is solved in the usual way (see for example [3]): the condition (2) is modified by previously checking the conditions

$$(f,x^{2^k}-x)=1,\quad 2\le k\le 12, \tag{2a}$$

$$(f,f')=1, \tag{2b}$$

where (f,g) is the greatest common divisor of f and g, which is computed by the Euclidean algorithm. This makes the complete test (2) more complicated, but for the large part of trial polynomials (85% according to the estimate from [3]) it is ended by (2a). The condition (2b) eliminates the polynomials divisible by the square of a polynomial.

The numerical complexity of finding the factorization of 2^n-1 is very large. This makes it unreasonable to include the factorization as a part of the primitivity check. Even more, it is unreasonable to compute these factorizations at all, because all those of them that are known can be found in [1,10] (which is an output of the famous Cunningham project). Therefore the factorizations from [1,10] are input in a data base using a special program. The process is not straightforward, because in cited references there

are actually four tables, containing (not always complete) factorizations of the numbers
(A) $2^{2k-1}-1,\ \ k\le 600,$
(B) $2^{2k-1}+1,\ \ k\le 600,$
(C) $2^{4k-2}+1=LM, L=2^{2k-1}-2^k+1,\ \ M=2^{2k-1}+2^k+1,\ \ k\le 600,$
(D) $2^{4k+1}+1,\ \ k\le 300.$
The first step is to list all $n<4800$ for which all the prime factors of 2^n-1 can be found. For some values of n, the number 2^n-1 has simple algebraic factors. These algebraic factors are then further split in the algebraic factors or their prime factors are taken from one of the tables A, B, C or D. The program for updating the factorization data base tests automatically the factors during the input. It uses algebraic factors and the factors that are already in the data base. This is useful for example if the factorization of $2^{2n}-1$ is to be input when the factorization of 2^n-1 is already input.

The efficiency of the primitivity check depends also on the order in which the prime factors of 2^n-1 are used in (3): the check is carried out for the small factors p first, because according to [5, Th 3.5] the probability that (3) is not satisfied is greater for small than for large prime factors.

The number of primitive polynomials of degree n is $\phi(N)/n$ [5, Th 3.5] (here $\phi(n)$ is Euler's function, showing the number of integers i with $1\le i\le n$ that are relatively prime to n). Therefore a randomly chosen binary polynomial is primitive with the probability α/n where

$$\alpha=(1-2^{-n})\prod_{p|N}(1-\frac{1}{p}).$$

The complexity of this primitivity check is $O(ktn^2)$, where k is the number if different prime factors of 2^n-1. Taking into account the "density" of primitive polynomials, an upper bound for the complexity of generation of one primitive polynomial is roughly estimated by $O(ktn^3)$. The program, when running on a PC with the 80486 microprocessor on 66MHz, gives one primitive polynomial of degree 500 (1000) after about 2 min (20 min).

References

[1] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, S. S. Wagstaf, Jr, *Factorization of $b^n\pm 1$, $b=2,3,5,6,7,10,11,12$ up to high powers*, 2nd ed., Contemp. Math., vol. 22, Amer. Math. Soc., Providence, RI, 1988.

[2] S. W. Golomb, *Shift Register Sequences*, Holden–Day, San Francisco, 1967.

[3] J. R. Heringa, H. W. Blöte and A. Compagner, *New Primitive trinomials of Mersenne–exponent degrees for random number generation*, International Journal of Modern Physics C **3** (1992), 561–564.

[4] Y. Kurita, M. Matsumoto, *Primitive t–nomials ($t = 3, 5$) over $GF(2)$ whose degree is a Mersenne exponent* ≤ 44497, Math. Comp. **56** (1991), 817–821.

[5] R. Lidl, H. Niederreiter, *Finite Fields*, Encyclopedia Math. Appl., Vol.20, Addison–Wesley, Reading, Mass., 1983.

[6] W. Meier and O. Staffelbach, *Fast correlation attacks on certain stream ciphers*, J. Cryptology **1** (1989), 159–176.

[7] E. R. Rodemich, H. Rumsey, Jr., *Primitive trinomials of high degree*, Math. Comp. **22** (1968), 863–865.

[8] W. Stahnke, *Primitive binary polynomials*, Math. Comp. **27** (1973), 977–980.

[9] R. C. Tausworthe, *Random numbers generated by linear recurrence modulo two*, Math. Comp. **19** (1965), 201–209.

[10] S. S. Wagstaf, Jr., *Update 2.6 to the Second Edition of Factorization of* $b^n \pm 1$,, 1993.

[11] E. J. Watson, *Primitive polynomials (mod 2)*, Math. Comp. **16** (1962), 368–369.

[12] N. Zierler, *Primitive trinomials whose degree is a Mersenne exponent*, Inform. Control **15** (1969), 67–69.

[13] N. Zierler, J. Brillhart, *On Primitive trinomials* (mod 2), Inform. Control **13** (1968), 541–554.

[14] N. Zierler, J. Brillhart, *On primitive trinomials* (mod 2), *II*, Inform. Control **14** (1969), 566–569.

[15] M. Živković, *A Table of Primitive Binary Polynomials*, Math. Comp. **92** (1993), 1368–1369.

[16] M. Živković, *Table of Primitive Binary Polynomials,* II, Math. Comp. **93** (1993), 368–372.

Matematički fakultet, Studentski trg 11, 11000 Belgrade, Yugoslavia

FILOMAT

FILOMAT is a continuation of *Zbornik radova Filozofskog fakulteta u Nišu, Serija Matematika (vol. 1(1987)– vol. 6(1992)* and is published yearly. It publishes original papers in all fields of pure and applied mathematics.

Instructions for authors

All manuscripts (the original and a copy) must be written in English. The extent of the papers is limited to ten pages in length; manuscripts over 10 pages are accepted only exceptionally. Manuscripts should not require many language corrections.

FILOMAT is produced using $\mathcal{AMS}$-TEX. Authors are encouraged to prepare their manuscripts using $\mathcal{AMS}$-TEX. Only a hard copy should be submitted for assessment, but if the paper is accepted the author will be asked to send the text on an IBM PC compatible diskette.

The author(s) should write their names, addresses and the title of the paper on a separate sheet. All manuscripts should start with a short Abstract and include the footnote 1991 Mathematics Subject Classification on the first page. Definitions, theorems, lemmas, remarks, proofs etc. should be written using only one of two alternative styles: **Definition** 2.1. or 3.2. **Theorem** consistently throughout the paper.

Figures are included in the text and must be numbered (by arabic numbers) and mentioned in the text. Equations (or formulas) must be numbered (for future references) in parentheses ( ) at the left margin.

References should be listed alphabetically in the following form:

[3] E.Hewitt and K.A.Ross, *Abstract Harmonic Analysis*, Vol. I, Springer-Verlag, Berlin, 1963.

[11] Đ.Kurepa, *On regressive functions*, Z. Math. Logik 4 (1958), 148-156.

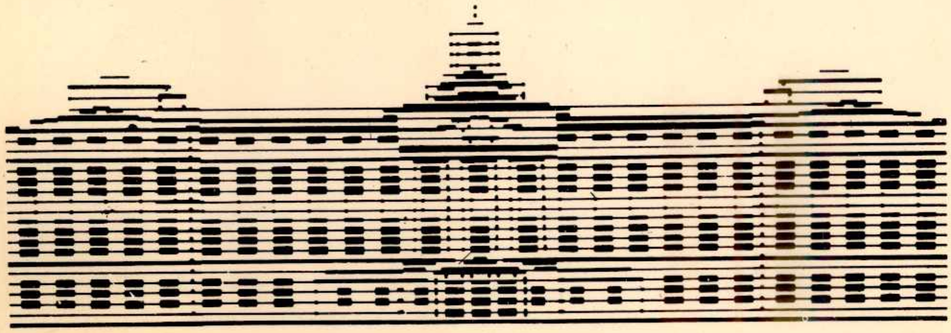
[15] P.Petrović, *Neka svojstva ...*, Doctoral dissertation, University of Belgrade, 1980.

A total of 30 reprints of each paper will be available free of charge; additional reprints can be ordered.

All correspondence concerning both manuscripts and exchange should be addressed to:

Filozofski fakultet (FILOMAT), ul. Ćirila i Metodija 2, 18000 Niš, Yugoslavia.
The subscription price is 20 dollars USA per volume, post free.
The price for this volume is 60 dollars USA, post free.
For subscriptions write to the same address.

rović



ISSN 0354–5180