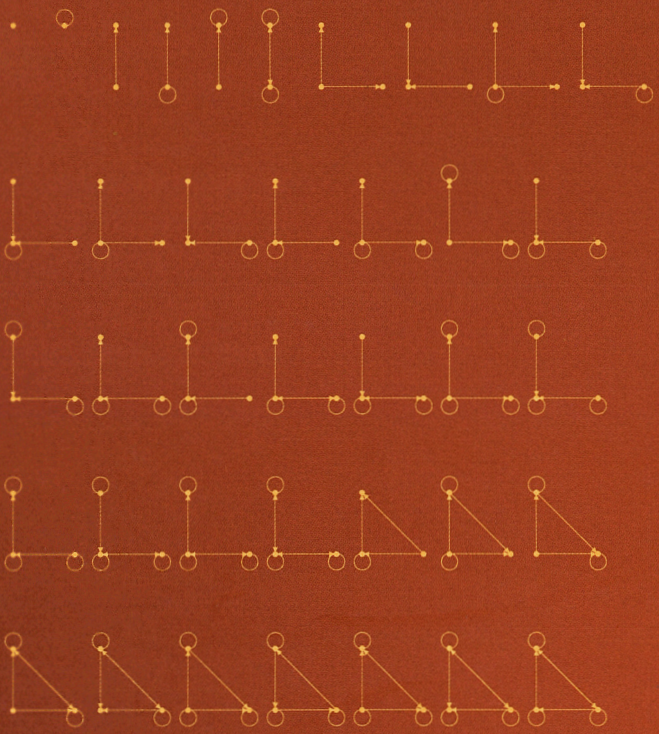
Advances in Algebra and Combinatorics



K P Shum • E Zelmanov
Jiping Zhang • Li Shangzhi
editors

World Scientific

Advances in Algebra and Combinatorics

This page intentionally left blank

Advances in Algebra and Combinatorics

Proceedings of the Second International Congress in Algebra and Combinatorics

Guangzhou, China	2 – 4 July 2007
Beijing, China	6 – 11 July 2007
Xian, China	12 – 15 July 2007

editors

K P Shum
The University of Hong Kong, Hong Kong

E Zelmanov
University of California, San Diego, USA

Jiping Zhang
Peking University, China

Li Shangzhi
Beihang University, China

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ADVANCES IN ALGEBRA AND COMBINATORICS

Proceedings of the Second International Congress in Algebra and Combinatorics

Copyright © 2008 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-279-000-2

ISBN-10 981-279-000-4

Printed in Singapore by B & JO Enterprise

Preface

The first international congress of algebras and combinatorics (in short, ICAC) was held in Hong Kong dated 17-23 August 1997. The aim of the congress was to celebrate the re-unification of Hong Kong to China and to celebrate the 25th anniversary of the Southeast Asian Mathematical Society. Also 1997 was the 100th anniversary of the International Mathematical Union (IMU). On this very special occasion, the congress was initiated by Professor Earl Taft, the former chief editor of Communications in Algebras and Professor Zhe-xian Wan, the chief editor of Algebra Colloquium. There were over 350 participants attended this congress from over 24 countries.

The experts who were invited to give main lectures in the congress include B. H. Neumann, L. A. Bokut, E. Zelmanov, A. V. Mikhalev, P. Solé, J. M. Howie, Z. Arad, H. J. Hoehnke, K. Saito, M. M. Deza, V. Dlab, L. J. Carbone, A. M. Hinz, C. P. Milies, H. J. Vogel, R. Wisbauer, P. Vamos, O. Grösek, A. Cherubini, K. Denecke, B. Piochi, M. Ito, S. Cohen, M. Harada, Nguyen Van Sanh, R. Gonchigdorzh, K. H. Kim, A. C. Kim, B. Schein, Y. H. Xu, M. Cohen, M. K. Sen, M. Tokizawa, J. C. Meakin, J. P. Zhang, N. Q. Ding, etc. Professor Stephen Hill, Regional Director of UNESCO, specially came to Hong Kong address the opening ceremony of ICAC97.

The ICAC97 in Hong Kong was regarded as a great success. Professor B. H. Neumann who was 88 at 1997 said in the closing ceremony that he hoped the participants in ICAC to keep up their momentum for more activities and he expected that the second ICAC to be held again every 10 years with more participants. In fact, since August 1997, there were many international conferences in algebra and combinatorics being held in Manila, Bangkok, Yujarkata, Hanoi, Chongqing, Moscow, Dennison, Kyoto, Islamabad, Tirupathi, Calcutta, Taipei, Pusan, Tainan, Kashan, $\cdots$. Fruitful results in this area were produced by many young mathematicians.

The second ICAC2007 was first initiated by President Li Wei of Beihang University, that is, Beijing University of Aeronautics and Astronautics and President Xu Delong of Xi'an University of Architecture and Technology.

In particular, Professor Li Wei was very keen to host ICAC2007 because he realizes that algebras and combinations have extensive applications in coding theory, in the design of computers, in electronic communications and even in space technology. Moreover, Beihang University has helped China to develop the project of launching the moon in 2007. He thinks that this would be an excellent occasion to hold the ICAC, 2007 at their university. On the other hand, Professor L. A. Bokut has been appointed by the South China Normal University as the research chair and to help the establishment of a center of combinatorial algebra in the university. On the occasion of his 70th anniversary, President Wang Guo-jiang was more than happy to host a conference for his birthday. This idea was strongly applaud by Professor E. Zelmanov, one of the former students of L. A. Bokut. Hence, with such a background, the ICAC2007 was composed by three chapters. The first chapter was held in South China Normal University at Guangzhou, 2-4 July; the second chapter was held in Beijing, 6-11 July; and the third chapter was held in Xi'an, 12-15 July.

There were over 150 participants coming to Guangzhou. The main speakers included Professors E. Zelmanov (Chairman, Scientific Committee), Lance Small, L. A. Bokut, A. V. Mikhalev, B. Schein, K. Kalorkoti, S. H. Ng, Y. M. Wang, W. J. Shi, W. B. Guo, X. Y. Guo, N. Q. Ding, Z. F. Hao, Q. H. Zhang, Michel Jumbu, K. Denecke, etc. A boat tour along the Pearl River at night was arranged. For the second chapter in Beijing, there were around 350 participants. Professor M. Bróue, the chief editor of Journal of Algebra has made a special effort to come to the congress for giving a special plenary talk. The other main speakers included Professors Zvi Arad, Susan Montgomery, A. V. Mikhalev, A. A. Mikhalev, Pavel Kolesnikov, Li Wei, Agata Smoktunowicz, Ivan Shestakov, Ts. Dajdorz, A. P. Pojidaev, A. Kemer, A. Yu. Olshanskii, V. A. Artamonov, N. A. Vavilov, Kenji Ueno, V. K. Kharchenko, G. O. H. Katona, Jan Okninski, Yong-Chuen Chen, Jiping Zhang, Tasuro Ito, Ilias Kotsireas, Leslie Hogben, Simone Rinaldi, M. R. R. Moghaddom, Das Dorz, Shangzhi Li, P. H. Lee, T. K. Lee, Y. Q. Zhou, W. S. Cheung, Ling Long, George Szeto, M. R. R. Moghaddam, A. R. Moghaddamfar, Wanida Hemakul, L. Sabinina, L. Sbitneva, etc. A tour to Great Wall was arranged. For the chapter in Xi'an, there were over 400 participants, including many graduate students. The main speakers in Xi'an included Professors Claus Ringel, Victoria Gould, Stephen Pride, Masami Ito, Lószló Márki, J. C. Meakin, S. K. Jain, Pal Domosi, V. Bovdi, K. C. Chowdhury, Nguyen Van Sanh, Masami Ito, Y. Kobayashi, K. B. Nam, Le Anh Vu, Y. Q. Guo, G. C. M. Gomes, H. France-Jackson, K. P.

Shum, etc. A tour to visit Terre Cotta was specially arranged. There were over 600 participants in total in the ICAC2007, coming from 27 countries. This congress is in fact one of the largest scale international conference held in China in recent years. We all hope that the quality of research in algebras and combinatorics will be further improved and promoted after the ICAC2007.

To organize an international conference is not an easy task. Apart from inviting the eminent main speakers, there are many arranging difficulties to be overcome. I would like to take this opportunity to thank the staffs and students, especially, the main organizers Professors Yuqun Chen, Shangzhi Li and Xueming Ren, in South China Normal University, Beihang University and Xi'an University of Architecture and Technology for their help and effort. Without them, the congress will not be so successful.

The congress has been made possible by the following sponsors for whose generosity the organizing committee is much indebted.
-NNSF, China
-IMU(CDC) Committee
-South China Normal University
-Beihang University
-Xi'an University of Architecture Technology
-Southeast Asian Mathematical Society
-Southeast Asian Bulletin of Mathematics
-World Scientific INC at Singapore

K. P. Shum
Chairman, Organizing Committee

March 10, 2008

This page intentionally left blank

Contents

Preface v

Hyperidentities in the Class $(x(yz)) \approx ((x(yy))z)$ Graph Algebras of Type $(2,0)$
A. Anantpinitwatna and T. Poomsa-Ard 1

Quantum Polynomials
V. A. Artamonov 19

Gröbner–Shirshov Bases: Some New Results
L. A. Bokut and Y. Q. Chen 35

Isomorphisms and Derivations of Algebras
S. H. Choi and K.-B. Nam 57

Semigroup Properties of Cooperations on Finite Sets
K. Denecke and K. Saengsura 69

Algebras Derived by Surjective Hypersubstitutions
K. Denecke and R. Srithus 83

Continuous Coalgebra Endomorphisms of Some Complete Ultrametric Hopf Algebras
B. Diarra 95

θ^*-Pairs and the Structure of Finite Groups
H. H. Feng and X. Y. Guo 119

Stability of the Theory of Existentially Closed S-Acts over a Right Coherent Monoid S
J. Fountain and V. Gould 129

Paper-Folding, Polygons, Complete Symbols, and the Euler Totient Function: An Ongoing Saga Connecting Geometry, Algebra, and Number Theory
P. Hilton, J. Pedersen and B. Walden 157

Koszul Algebras and Hyperplane Arrangements
M. Jambu 179

Some Problems in *PI*-Theory
A. Kemer and A. Ilya 189

On Irreducible Subalgebras of Matrix Weyl Algebras
P. S. Kolesnikov 205

On the Length of Conjugacy Classes and P-Nilpotence of Finite Groups
Q. J. Kong and X. Y. Guo 219

Computations with Finite Index Subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ Using Farey Symbols
C. A. Kurth and L. Long 225

Gröbner–Shirshov Bases and Normal Forms for the Coxeter Groups E_6 and E_7
D. Lee 243

On Overgroups in $\mathrm{GL}(n, F)$ over a Subfield of F
S. Li 257

A Symbolic Calculus on Defect Revisions of Axiomatic Systems
W. Li 275

Some Remarks on the Burnside Problem for Loops
P. Plaumann and L. Sabinina 293

Rpp Semigroups, Its Generalizations and Special Subclasses
K. P. Shum 303

Conformal Field Theory and Modular Functor
K. Ueno 335

Classification of 5-Dimensional MD-Algebras Having Commutative Derived Ideals
Le Anh Vu and K. P. Shum 353

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 1–18)

HYPERIDENTITIES IN THE CLASS $(x(yz)) \approx ((x(yy))z)$ GRAPH ALGEBRAS OF TYPE (2,0) *

AMPORN ANANTPINITWATNA and TIANG POOMSA-ARD

Department of Mathematics, Faculty of Science, Mahasarakham University
Mahasarakham 44150, Thailand
E-mail: tiang@kku.ac.th

Graph algebras establish a connection between directed graphs without multiple edges and special universal algebras of type (2,0). We say that a graph G satisfies an identity $s \approx t$ if the corresponding graph algebra $\underline{A}(G)$ satisfies $s \approx t$. A graph $G = (V, E)$ is called an $(x(yz)) \approx ((x(yy))z)$ graph if the graph algebra $\underline{A}(G)$ satisfies the equation $(x(yz)) \approx ((x(yy))z)$. An identity $s \approx t$ of terms s and t of any type τ is called a hyperidentity of an algebra $\underline{A}$ if whenever the operation symbols occurring in s and t are replaced by any term operations of $\underline{A}$ of the appropriate arity, the resulting identities hold in $\underline{A}$.

In this paper, we characterize the class of $(x(yz)) \approx ((x(yy))z)$ graph algebras, $(x(yz)) \approx ((x(yy))z)$ class, identities and hyperidentities in the class of $(x(yz)) \approx ((x(yy))z)$ graph algebras.

Keywords: Identities; Hyperidentities; Term; Normal form term; Binary algebra; Graph algebra; $(x(yz)) \approx ((x(yy))z)$ graph algebra.

1. Introduction

An identity $s \approx t$ of terms s, t of any type τ is called a *hyperidentity* of an algebra $\underline{A}$ if whenever the operation symbols occurring in s and t are replaced by any term operations of $\underline{A}$ of the appropriate arity, the resulting identity holds in $\underline{A}$. Hyperidentities can be defined more precisely using the concept of a hypersubstitution, which was introduced by K. Denecke, D. Lau, R. Pöschel and D. Schweigert in [1].

We fix a type $\tau = (n_i)_{i \in I}$, $n_i > 0$ for all $i \in I$, and operation symbols $(f_i)_{i \in I}$, where f_i is $n_i - ary$. Let $W_\tau(X)$ be the set of all terms of type τ over some fixed alphabet X, and let $Alg(\tau)$ be the class of all algebras of

*2000 Mathematics Subject Classification. 05C25, 08B15.

type τ. Then a mapping

$$\sigma : \{f_i | i \in I\} \longrightarrow W_\tau(X)$$

which assigns to every $n_i - ary$ operation symbol f_i an $n_i - ary$ term will be called a *hypersubstitution* of type τ (for short, a hypersubstitution). By $\hat{\sigma}$, we denote the extension of the hypersubstitution σ to a mapping

$$\hat{\sigma} : W_\tau(X) \longrightarrow W_\tau(X).$$

The term $\hat{\sigma}[t]$ is defined inductively by

(i) $\hat{\sigma}[x] = x$ for any variable x in the alphabet X, and
(ii) $\hat{\sigma}[f_i(t_1, ..., t_{n_i})] = \sigma(f_i)^{W_\tau(X)}(\hat{\sigma}[t_1], ..., \hat{\sigma}[t_{n_i}])$.

Here $\sigma(f_i)^{W_\tau(X)}$ on the right hand side of (ii) is the operation induced by $\sigma(f_i)$ on the term algebra with the universe $W_\tau(X)$.

Graph algebras have been initiated in [13] to obtain examples of non-finitely based finite algebras. To recall this concept, let $G = (V, E)$ be a (directed) graph with the vertex set V and the set of arcs $E \subseteq V \times V$. Define the *graph algebra* $\underline{A(G)}$ corresponding to G with the underlying set $V \cup \{\infty\}$, where ∞ is a symbol outside V, and with two basic operations, namely a nullary operation pointing to ∞ and a binary one denoted by juxtaposition, given for $u, v \in V \cup \{\infty\}$ by

$$uv = \begin{cases} u, & \text{if} \quad (u, v) \in E, \\ \infty, & \text{otherwise.} \end{cases}$$

The graph identities were characterized in [5] by using the rooted graph of a term t, where the vertices correspond to the variables occurring in t. Since on a graph algebra we have one nullary and one binary operation, $\sigma(f)$ in this case is a binary term in $T(X_2)$, i.e., a term built up from variables of a two-element alphabet and a binary operation symbol f corresponding to the binary operation of the graph algebra. In [12], R. Pöschel showed that any term over the class of all graph algebras can be uniquely represented by a normal form term and that there is an algorithm to construct the normal form term to every given term t. K. Denecke and T. Poomsa-ard [2], T. Poomsa-ard [8], T. Poomsa-ard, J. Wetweerapong and C. Samartkoon [9,10] characterized graph hyperidentities, associative graph hyperidentities, idempotent graph hyperidentities and transitive graph hyperidentities respectively by using normal form graph hypersubstitutions.

A graph $G = (V, E)$ is called an $(x(yz)) \approx ((x(yy))z)$ graph if the corresponding graph $\underline{A(G)}$ satisfied the identity $(x(yz)) \approx ((x(yy))z)$. In this paper, we characterize the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras,

$(x(yz)) \approx ((x(yy))z)$ class, identities and hyperidentities in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras.

2. $(x(yz)) \approx ((x(yy))z)$ graph algebras

R. Pöschel introduced terms for graph algebras in [11], the underlying formal language has to contain a binary operation symbol (juxtaposition) and a symbol for the constant ∞ (denoted by ∞, too).

Definition 2.1. The set $T(X)$ of all terms over the alphabet

$$X = \{x_1, x_2, x_3, ...\}$$

is defined inductively as follows:

(i) every variable $x_i, i = 1, 2, 3, ...$, and ∞ are terms;
(ii) if t_1 and t_2 are terms, then $(t_1 t_2)$ is a term;
(iii) $T(X)$ is the set of all terms which can be obtained from (i) and (ii) in finitely many steps.

The leftmost variable of a term t is denoted by $L(t)$ and rightmost variable of a term t is denoted by $R(t)$. A term, in which the symbol ∞ occurs, is called a *trivial term.*

Definition 2.2. To each non-trivial term t, one can define a directed graph $G(t) = (V(t), E(t))$, where the vertex set $V(t)$ is the set $Var(t)$ of all variables occurring in t, and where the edge set $E(t)$ is defined inductively by $E(t) = \phi$ if t is a variable and $E((t_1 t_2)) = E(t_1) \cup E(t_2) \cup \{(L(t_1), L(t_2))\}$. where $t = (t_1 t_2)$ is a compound term and $L(t_1)$, $L(t_2)$ are the leftmost variables in t_1 and t_2, respectively.

$L(t)$ is called the *root* of the graph $G(t)$, and the pair $(G(t), L(t))$ is the *rooted graph* corresponding to t. We recall, a root of a graph G is a vertex v, such that every vertex $w \in V(G)$ is accessible from v via a directed path. A *rooted graph* is a pair (G, v), where G is a nonempty graph and v is root. Formally, to every trivial term t we assign the empty graph ϕ.

Definition 2.3. A graph $G = (V, E)$ *is said to satisfy an identity* $s \approx t$ if the corresponding graph algebra $\underline{A(G)}$ satisfies $s \approx t$ (i.e., we have $s = t$ for every assignment $V(s) \cup V(t) \to \overline{V \cup \{\infty\}}$). In this case, we write $G \models s \approx t$.

Definition 2.4. Let $G = (V, E)$ and $G' = (V', E')$ be graphs. A *homomorphism* h from G into G' is a mapping $h : V \to V'$ carrying edges to edges, that is, for which $(u, v) \in E$ implies $(h(u), h(v)) \in E'$.

The following result was proved in [5].

Proposition 2.1. *Let s and t be non-trivial terms from $W_\tau(X)$ with variables $V(s) = V(t) = \{x_0, x_1, ..., x_n\}$ and $L(s) = L(t)$. Then a graph $G = (V, E)$ satisfies $s \approx t$ if and only if the graph algebra $\underline{A(G)}$ has the following property:*

a mapping $h : V(s) \longrightarrow V$ is a homomorphism from $G(s)$ into G if and only if it is a homomorphism from $G(t)$ into G.

Proposition 2.1 gives a method to check whether a graph $G = (V, E)$ satisfies the equation $s \approx t$. Hence, we can check whether a graph $G = (V, E)$ has an $(x(yz)) \approx ((x(yy))z)$ graph algebra.

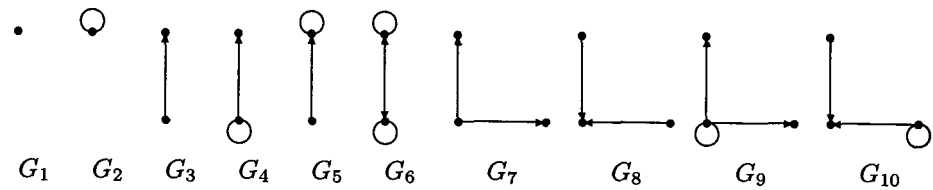
Proposition 2.2. *Let $G = (V, E)$ be a graph. Then the following statements are equivalent:*

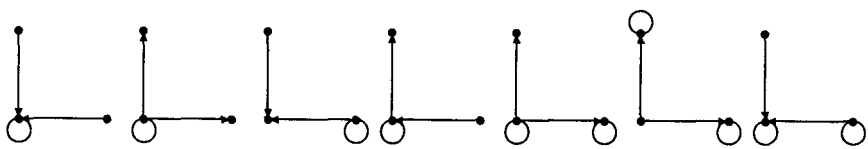
(i) G has an $(x(yz)) \approx ((x(yy))z)$ graph algebra,
(ii) if $(a, b) \in E$ and $c \in V$, then $(b, c) \in E$ if and only if $(a, c), (b, b) \in E$.

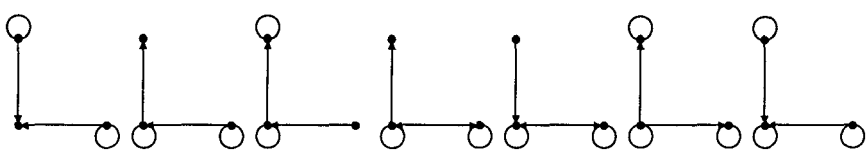
Proof. $(i) \Rightarrow (ii)$: Suppose that $G = (V, E)$ has an $(x(yz)) \approx ((x(yy))z)$ graph algebra. Let s and t be terms such that $s = (x(yz))$, $t = ((x(yy))z)$. Let $(a, b), (b, c) \in E$ and $h : V(s) \to V$ be a function such that $h(x) = a$, $h(y) = b$ and $h(z) = c$. We see that h is a homomorphism from $G(s)$ into G. By Proposition 2.1, we have h is a homomorphism from $G(t)$ into G. Since $(x, z), (y, y) \in E(t)$, $(h(x), h(z)) = (a, c) \in E$ and $(h(y), h(y)) = (b, b) \in E$. Let $(a, b), (a, c), (b, b) \in E$ and $h : V(t) \to V$ be a function such that $h(x) = a$, $h(y) = b$ and $h(z) = c$. We see that h is a homomorphism from $G(t)$ into G. By Proposition 2.1, h is a homomorphism from $G(s)$ into G. Since $(y, z) \in E(s)$, $(h(y), h(z)) = (b, c) \in E$.

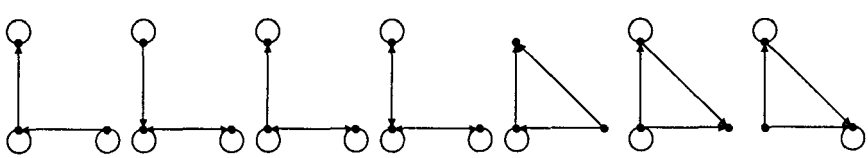
$(ii) \Rightarrow (i)$: Suppose that $G = (V, E)$ is a graph satisfing (ii). Let s and t be terms such that $s = (x(yz))$, $t = ((x(yy))z)$. Suppose that $h : V(s) \to V$ is a homomorphism from $G(s)$ into G. Since $(x, y), (y, z) \in E(s)$, $(h(x), h(y)), (h(y), h(z)) \in E$. By assumption, we get $(h(x), h(z)), (h(y), h(y)) \in E$. Therefore, h is a homomorphism from $G(t)$ into G. By the same way, if h is a homomorphism from $G(t)$ into G, then we can prove that it is a homomorphism from $G(s)$ into G. Hence, by Proposition 2.1, $\underline{A(G)}$ satisfies $s \approx t$. □

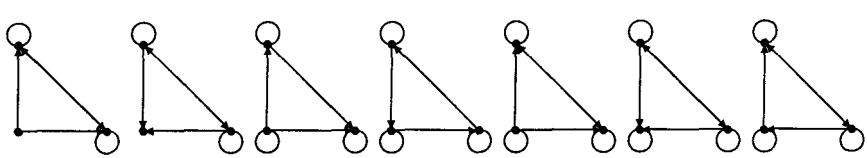
By Proposition 2.2, the following graphs have $(x(yz)) \approx ((x(yy))z)$ graph algebras;

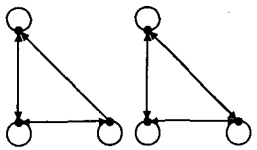

G_1 G_2 G_3 G_4 G_5 G_6 G_7 G_8 G_9 G_{10}


G_{11} G_{12} G_{13} G_{14} G_{15} G_{16} G_{17}


G_{18} G_{19} G_{20} G_{21} G_{22} G_{23} G_{24}


G_{25} G_{26} G_{27} G_{28} G_{29} G_{30} G_{31}


G_{32} G_{33} G_{34} G_{35} G_{36} G_{37} G_{38}



G_{39} G_{40}

and all graphs such that each component of every subgraph induced by at most three vertices is one of these graphs.

3. Identities in $(x(yz)) \approx ((x(yy))z)$ graph algebras

Graph identities were characterized in [5] by the following Proposition:

Proposition 3.1. *A non-trivial identity $s \approx t$ is an identity in the class of all graph algebras if and only if either both terms s and t are trivial or none of them is trivial, $G(s) = G(t)$ and $L(s) = L(t)$.*

Further it was proved.

Proposition 3.2. *Let $G = (V, E)$ be a graph and let $h : X \cup \{\infty\} \longrightarrow V \cup \{\infty\}$ be an evaluation of the variables such that $h(\infty) = \infty$. Consider the canonical extension of h to the set of all terms. Then the following holds: if t is a trivial term then $h(t) = \infty$. Otherwise, if $h : G(t) \longrightarrow G$ is a homomorphism of graphs, then $h(t) = h(L(t))$, and if h is not a homomorphism of graphs, then $h(t) = \infty$.*

The following lemma was proved in [8].

Lemma 3.1. *Let $G = (V, E)$ be a graph, let t be a term and let*

$$h : X \longrightarrow V \cup \{\infty\}$$

be an evaluation of the variables. Then:

(i) If h is a homomorphism from $G(t)$ into G with the property that the subgraph of G induced by $h(V(t))$ is complete, then $h(t) = h(L(t))$.

(ii) If h is a homomorphism from $G(t)$ into G with the property that the subgraph of G induced by $h(V(t))$ is disconnected, then $h(t) = \infty$.

Now we apply our results to characterize all identities in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras. Clearly, if s and t are trivial, then $s \approx t$ is an identity in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras

and $x \approx x$ $(x \in X)$ is an identity in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras, too. Further, if s is a trivial term and t is a non-trivial term, then $s \approx t$ is not an identity in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras, since for a complete graph G, we have an evaluation of the variables h such that $h(s) = \infty$ and $h(t) \neq \infty$. Hence, we only need to consider the case that s and t are non-trivial and are different from variables. Before we do this, we first introduce some notation. For any term $t \in T(X)$ and the graph $G(t) = (V(t), E(t))$, let $N_o^t(x) = \{y \in V(t) \mid (x,y) \in E(t)\}$ be the set of all out-neighbors of the vertex $x \in V(t)$ and $N_i^t(x') = \{y' \in V(t) \mid (x',y') \in E(t)\}$ be the set of all in-neighbors of the vertex $x' \in V(t)$. Then all identities in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras are characterized by the following theorem:

Theorem 3.1. *Let s and t be non-trivial terms. Then $s \approx t$ is an identity in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras if and only if the following conditions are satisfied:*

(i) $L(s) = L(t)$,
(ii) $V(s) = V(t)$,
(iii) $N_i^s(L(s)) \neq \phi$ *if and only if* $N_i^t(L(t)) \neq \phi$,
(iv) for any $x \in V(s)$, $N_o^s(x) \neq \phi$ *if and only if* $N_o^t(x) \neq \phi$.

Proof. Suppose that $s \approx t$ is an identity in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras.

Suppose that $V(s) \neq V(t)$ and let $x \in V(s)$ but $x \notin V(t)$. Consider the $(x(yz)) \approx ((x(yy))z)$ graph $G = (V, E)$ with $V = \{0\}$, $E = \{(0,0)\}$ and an evaluation of variables $h : V(s) \cup V(t) \to V \cup \{\infty\}$ such that $h(x) = \infty$ and $h(y) = 0$ for all other $y \in V(s) \cup V(t)$. We have $h(s) = \infty$ and $h(t) = 0$. Hence $\underline{A(G)}$ does not satisfy $s \approx t$. Now let $G = (V, E)$ be a complete graph with $V = V(s) = V(t)$ and let $h : V(s) \to V$ be an identity evaluation of variables. By Lemma 3.1, we have $L(s) = h(L(s)) = h(s) = h(t) = h(L(t)) = L(t)$.

Suppose that $N_i^s(L(s)) \neq \phi$ but $N_i^t(L(t)) = \phi$. Consider the graph $G = (V, E)$ such that $V = \{0,1\}$, $E = \{(0,1),(1,1)\}$. By Proposition 2.2, $\underline{A(G)}$ has an $(x(yz)) \approx ((x(yy))z)$ graph algebra. Let $h : V(s) \to V$ be the restriction of an evaluation of the variables such that $h(L(s)) = 0$ and $h(y) = 1$ for all other $y \in V(s)$. We see that $h(s) = \infty$ and $h(t) = 0$. Hence, $\underline{A(G)}$ does not satisfy $s \approx t$. Similarly, we can prove the converse.

Suppose that there exists $x \in V(s)$ such that $N_o^s(x) \neq \phi$ but $N_o^t(x) = \phi$. Consider the graph $G = (V, E)$ such that $V = \{0,1\}$, $E = \{(0,0),(0,1)\}$.

Then by Proposition 2.2, $\underline{A(G)}$ has an $(x(yz)) \approx ((x(yy))z)$ graph algebra. Let $h : V(s) \to V$ be the restriction of an evaluation of the variables such that $h(x) = 1$ and $h(y) = 0$ for all other $y \in V(s)$. We see that $h(s) = \infty$ and $h(t) = 0$. Hence, $\underline{A(G)}$ does not satisfy $s \approx t$. Similarly, we can prove the converse.

Conversely, suppose that s and t are non-trivial terms satisfying (i), (ii), (iii) and (iv). Let $G = (V, E)$ be an $(x(yz)) \approx ((x(yy))z)$ graph. Suppose that a function $h : V(s) \to V$ is a homomorphism from $G(s)$ into G and let $(x, y) \in E(t)$. If $x = y = L(s)$, (i.e., $(L(s), L(s)) \in E(t)$), then by (iii) there exists $u \in V(s)$ such that $(u, L(s)) \in E(s)$. Hence there exists $v \in V(s)$ such that $(u, L(s)), (L(s), v) \in E(s)$. We have $(h(u), h(L(s)), (h(L(s)), h(v)) \in E$. By Proposition 2.2, we have $(h(L(s)), h(L(s))) \in E$. If $x = y \neq L(s)$, then by (iv) and $G(s)$ is a rooted graph with root $L(s)$, there exist $u, v \in V(s)$ such that $(u, x), (x, v) \in E(s)$. Hence $(h(u), h(x)), (h(x), h(v)) \in E$. By Proposition 2.2, we have $(h(x), h(x)) \in E$. If $x \neq y$ and $x = L(s)$, then because $G(s)$ is a rooted graph with root $L(s)$, there exists a dipath from $L(s)$ into y. By the homomorphism of h and Proposition 2.2, we get $(h(x), h(y)) \in E$. If $x \neq y$ and $x \neq L(s)$, then there exist $u, v \in V(s)$ such that $(u, x), (x, v) \in E(s)$. Hence we get $(h(x), h(x)) \in E$. Since $G(s)$ is a rooted graph with rooted $L(s)$ again, there exist dipaths from $L(s)$ into x and y. Thus $(h(L(s)), h(x)), (h(L(s)), h(y)) \in E$. Since $(h(L(s)), h(x)), (h(L(s)), h(y)), (h(x), h(x)) \in E$. By Proposition 2.2, we get $(h(x), h(y)) \in E$. This shows that h is a homomorphism from $G(t)$ into G. By the same way, if h is a homomorphism from $G(t)$ into G, then we can prove that it is a homomorphism from $G(s)$ into G. By Proposition 2.1, we prove that $\underline{A(G)}$ satisfies $s \approx t$. □

4. The $(x(yz)) \approx ((x(yy))z)$ class

Let $\mathcal{G}'$ be the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras and $Id(\mathcal{G}')$ the set of all identities satisfied in $\mathcal{G}'$. In this Section, we characterize the equation of terms $s \approx t$ such that the class of all $s \approx t$ graph algebras is the class $\mathcal{G}'$. We begin with a definition of a relation between the equation of terms.

Definition 4.1. For any term equation $s \approx t$ and $s' \approx t'$, we call $s \approx t$ relate to $s' \approx t'$ if and only if the class of all $s \approx t$ graph algebras and the class of all $s' \approx t'$ graph algebras are the same. In this case, we write $s \approx t \sim s' \approx t'$.

We see that this relation is an equivalence relation. If $s \approx t \sim (x(yz)) \approx ((x(yy))z)$, then we call $s \approx t$ belongs to the class $(x(yz)) \approx ((x(yy))z)$. Clearly, if s and t are trivial terms, then $s \approx t$ does not belong to the class $(x(yz)) \approx ((x(yy))z)$ and $x \approx x$, $x \in X$ does not belong to the class $(x(yz)) \approx ((x(yy))z)$, too. Hence we consider the case that s and t are non-trivial and are different from variables. Before we do this, we introduce some notation. For any term t, let $L_i(t) = \{y \in V(t) \mid \vec{d}(L(t), y) = i \text{ in } G(t)\}$. Then all equations $s \approx t$ in the class $(x(yz)) \approx ((x(yy))z)$ are characterized by the following theorem:

Theorem 4.1. *Let s and t be non-trivial terms and different from variables. Then $s \approx t$ belongs to the $(x(yz)) \approx ((x(yy))z)$ class if and only if the following conditions are satisfied:*

(i) $s \approx t \in Id(\mathcal{G}')$,

(ii) one of $G(s)$ and $G(t)$, $G(t)$ say, has the properties that $L_2(t) \neq \phi$, $L_3(t) = \phi$ and for any $x \in V(s)$ if $x \in L_i(t)$, then there exists no $x' \in L_j(t)$ which $j \geq i$ such that $(x', x) \in E(t)$,

(iii) if $G(t)$ has the properties in (ii), then (a) there exists $z \in L_2(t)$ such that $N_i^s(z) = \{L(s)\}$ or (b) there exists $w \in L_1(t)$ which $N_o^t(w) = \phi$ such that $(L(s), w) \notin E(s)$ or (c) there exists $z \in L_2(t)$ and there exist $y, w \in L_1(t)$ which $N_o^t(w) = \phi$ such that $(L(s), z), (y, w) \in E(s)$.

Proof. Suppose that $s \approx t$ belongs to the class $(x(yz)) \approx ((x(yy))z)$. Clearly, $s \approx t \in Id(\mathcal{G}')$.

Suppose that both of $G(s)$ and $G(t)$ do not have the properties in (ii). Consider the graph $G = (V, E)$ such that $V = \{0, 1, 2\}$, $E = \{(0, 1), (1, 2)\}$. We see that $G \notin \mathcal{G}'$ and $h(s) = h(t) = \infty$ for all assignment $h : V(s) \to V$ such that the range of h is V. Hence, $h(s) = h(t)$ for all assignment $h : V(s) \to V$. Thus, G is an $s \approx t$ graph but $G \notin \mathcal{G}'$. Consequently $s \approx t$ does not belong to the class $(x(yz)) \approx ((x(yy))z)$.

Suppose that $G(t)$ has the properties in (ii). then, by Theorem 3.1(iii), we have $N_i^s(L(s)) = \phi$. If $G(s)$ does not have the properties (a), (b) and there exists no $z \in L_2(t)$ such that $(L(s), z) \in E(s)$, then consider the graph $G = (V, E)$ such that $V = \{0, 1, 2\}$, $E = \{(0, 1), (1, 1), (1, 2)\}$. We see that $G \notin \mathcal{G}'$. Consider the assignment $h : V(s) \to V$ such that the range of h is V. Since $s \approx t \in Id(\mathcal{G}')$ and by our assumption, we see that if $h(L(t)) = 0$, $h(z) = 2$ for some $z \in L_2(t)$, and $h(y) = 1$ for all other $y \in V(t)$, then $h(t) = h(s) = 0$. Otherwise $h(t) = h(s) = \infty$. Thus, we have $h(s) = h(t)$ for all assignment $h : V(s) \to V$. Hence G is an $s \approx t$ graph but $G \notin \mathcal{G}'$. Therefore $s \approx t$ does not belong to the class $(x(yz)) \approx ((x(yy))z)$.

Suppose that $G(s)$ does not have the properties (a), (b) and there exist no $y', w' \in L_1(t)$ which $N_o^t(w') = \phi$ such that $(y', w') \in E(s)$. Consider the graph $G = (V, E)$ such that $V = \{0, 1, 2\}$, $E = \{(0, 1), (1, 1), (0, 2)\}$. We see that $G \notin \mathcal{G}'$. Consider the assignment $h : V(s) \to V$ such that the range of h is V. Since $s \approx t \in Id(\mathcal{G}')$ and by our assumption, we see that if $h(L(t)) = 0$, $h(w') = 2$ for some $w' \in L_1(t)$, $N_o^t(w') = \phi$ and $h(z) = 1$ for all other $z \in V(s)$, then $h(t) = h(s) = 0$. Otherwise $h(t) = h(s) = \infty$. Thus, we have $h(s) = h(t)$ for all assignment $h : V(s) \to V$. Hence G is an $s \approx t$ graph but $G \notin \mathcal{G}'$. Therefore, $s \approx t$ does not belong to the class $(x(yz)) \approx ((x(yy))z)$.

Conversely, suppose that s and t are non-trivial terms and are different from variables satisfying (i), (ii) and (iii). Since $s \approx t \in Id(\mathcal{G}')$, $\mathcal{G}'$ is a subset of the class of all $s \approx t$ graphs. Hence, we only prove that the class of all $s \approx t$ graphs is a subset of $\mathcal{G}'$. Further, we suppose that $G(t)$ has the properties in (ii). Then by Theorem 3.1(iii), $N_i^s(L(s)) = \phi$. Let $G = (V, E)$ be an $s \approx t$ graph. Suppose that there exists $z' \in L_2(t)$ such that $N_i^s(z') = \{L(s)\}$. If $(a, b), (b, c) \in E$, then let $h : V(s) \to V$ be the function such that $h(L(t)) = a$, $h(y) = b$ for all $y \in L_1(t)$ and $h(z) = c$ for all $z \in L_2(t)$. We see that h is a homomorphism from $G(t)$ into G. By Proposition 2.1, h is a homomorphism from $G(s)$ into G. We have $(h(L(s)), h(z)) = (a, c) \in E$. If there are $y, y' \in L_1(t)$ such that $(y, y') \in E(s)$, then $(h(y), h(y')) = (b, b) \in E$. If there are no $y, y' \in L_1(t)$ such that $(y, y') \in E(s)$, then the function $h' : V(s) \to V$ such that $h'(L(s)) = a$, $h'(y) = b$, for all $y \in N_o^s(L(s))$ and $h'(z) = c$, for all other $z \in V(s)$ is a homomorphism from $G(s)$ into G. We have h is a homomorphism from $G(t)$ into G. Since there exists $y \in N_o^t(L(t))$ such that $(y, z') \in E(t)$, $(h'(y), h'(z')) = (b, b) \in E$. If $(a, b), (b, b), (a, c) \in E$, then let $h : V(s) \to V$ be a function such that $h(L(s)) = a$, $h(z') = c$ and $h(y) = b$ for all other $y \in V(s)$. We see that h is a homomorphism from $G(s)$ into G. By Proposition 2.1, h is a homomorphism from $G(t)$ into G. Since there exists $y' \in L_1(t)$ such that $(y', z') \in E(t)$, we have $(h(y'), h(z')) = (b, c) \in E$. Hence G is an $(x(yz)) \approx ((x(yy))z)$ graph. Suppose that there exists $w \in L_1(t)$ which $N_o^t(w) = \phi$ such that $(L(s), w) \notin E(s)$. If $(a, b), (b, c) \in E$, then let $h : V(s) \to V$ be a function such that $h(L(t)) = a$, $h(y) = b$ for all $y \in L_1(t)$ and $h(z) = c$ for all $z \in L_2(t)$. We see that h is a homomorphism from $G(t)$ into G. By Proposition 2.1, h is a homomorphism from $G(s)$ into G. Since $(L(s), w) \notin E(s)$, there exists $y \in L_1(t)$ such that $(y, w) \in E(s)$ and so $(h(y), h(w)) = (b, b) \in E$. Now let $h' : V(s) \to V$ be the function such that $h'(L(s)) = a$, $h'(w) = c$ and $h'(y) = b$ for all other $y \in V(s)$. We see that h' is a homomorphism from

G(s) into G. By Proposition 2.1, *h'* is a homomorphism from *G(t)* into *G.* Since *(L(t),.I) E E(t), (h'(L(t)), h'(w)) = (a, C) E E.* **If** (a, *b), (b, b),* (a, *c)* E *E,* then let *h : V(s)* –+ *V* be the function such that *h(L(t)) = a, h(w) = c* and *h(y) = b* for all other *y E V(t).* We see that *h* is a homomorphism from G(t) into G. By Proposition 2.1, *h* is a homomorphism from *G(s)* into G. Since there exists *y' E L*$_l$*(t)* such that *(y',w)* E *E(s), (h(y'),h(w))* = (b, c) E E. Hence G is an *(z(yz))* x *((z(yy))z)* graph. Suppose that there exists z' *E &(t)* and there exist *y',w' E Ll(t)* with N:(w') = q5 such that (L(s),z'),(y',w') E E(s). If (a,b),(b,c) E *E,* then let h : *V(s)* -, *V* be a function such that *h(L(t))* = a, h(y) = b for all *y* E Ll(t) and *h(z) = c* for all *z* E L$_2$(t). We see that *h* is a homomorphism from *G(t)* into G. By Proposition 2.1, *h* is a homomorphism from *G(s)* into G. Since *(L(s),4,(Y',4* E E(s), (h(L(s)),h(z')) = E E S and *(h(Y')>h(w'))* = (b, b) E *E.* Hence G is an *(~(yz))* x ((z(yy))z) graph. If (a, b), (b, b), (a, c) E E, then let *h* : V(s) -+ V be the function such that *h(L(t))* = a, *h(w')* = c and h(y) = b for all other *y E V(t).* Hence, *h* is a homomorphism from *G(t)* into G. By Proposition 2.1, *h* is also a homomorphism from G(s) into G. Since (y',w') E *E(s), (h(y'),h(w')) = (b,c)* E *E.* Hence G is an *(~(yz))* x *((z(yy))z)* graph. Therefore, the class of all s x t graph algebras is a subset of *0'.* 0

By Theorem 4.1, we see that *((x(yz))w)* « *((x(yw))z), ((x((yz)u))w)* w (((z(i/z))uM, (((x(yz))u)ti;) w ((x(((j/z)u))w), *(x{((yz)w)u))* » ((x((yz)u))^{i}u;), ((o;((yz)iy))ii) w (((x((yy)z))iu)u) are the examples of the equations of terms which belong to the *(x(yz))* w *((x(yy))z)* class.

5. Hyperidentities in the class (~(yz)) M ((z(yy))r) graph algebras

Now we want to formulate precise a concept of a hypersubstitution for graph algebras.

Definition 5.1. A mapping *u : {f, ~} -, T(X$_2$),* where X;$_!$ = (21,~~) and f is the operation symbol corresponding to the binary operation of a graph algebra is called *graph hypersubstitution* if U(M) = 00 and *a(f)* = *s* E *77x$_2$).* The graph hypersubstitution with *g(f) = s* is denoted by u$_8$.

Definition 5.2. An identity s M t is an (~(yz)) x ((z(yy))z) ^rapft %-*peridentity* if and only if for all graph hypersubstitutions *u,* the equations *S[s]* x &[t] are identities in Q'.

Let $s \approx t$ be in the class $(x(yz)) \approx ((x(yy))z)$. Then, by Theorem 4.1, we see that if we want to find all hyperidentities in the class of all $s \approx t$ graph algebras, then it is enough to find all hyperidentities in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras.

Further if we want to check that $s \approx t$ is a hyperidentity in the class of $\mathcal{G}'$, we can restrict our considerations to a (small) subset of $Hyp(\mathcal{G})$ - the set of all graph hypersubstitutions.

The following relation between hypersubstitutions was defined in [6].

Definition 5.3. Two graph hypersubstitutions σ_1, σ_2 are called $\mathcal{G}'$-*equivalent* if and only if $\sigma_1(f) \approx \sigma_2(f)$ is an identity in $\mathcal{G}'$. In this case, we write $\sigma_1 \sim_{\mathcal{G}'} \sigma_2$.

The following lemma was proved in [3,6].

Lemma 5.1. *If* $\hat{\sigma}_1[s] \approx \hat{\sigma}_1[t] \in Id(\mathcal{G}')$ *and* $\sigma_1 \sim_{\mathcal{G}'} \sigma_2$ *then* $\hat{\sigma}_2[s] \approx \hat{\sigma}_2[t] \in Id(\mathcal{G}')$.

Therefore, it suffices to consider the quotient set $Hyp(\mathcal{G})/\sim_{\mathcal{G}'}$. It was shown that any non-trivial term t over the class of graph algebras has a uniquely determined normal form term $NF(t)$ and there is an algorithm to construct the normal form term to a given term t. Now, we are going to describe how to construct the normal form term[11]. Let t be a non-trivial term. Then, the *normal form term of* t is the term $NF(t)$ constructed by the following algorithm:

(i) Construct $G(t) = (V(t), E(t))$.

(ii) Construct for every $x \in V(t)$ the list $l_x = (x_{i_1}, ..., x_{i_{k(x)}})$ of all out-neighbors (i.e., $(x, x_{i_j}) \in E(t), 1 \leq j \leq k(x)$) ordered by increasing indices $i_1 \leq ... \leq i_{k(x)}$ and let s_x be the term $(...((xx_{i_1})x_{i_2})...x_{i_{k(x)}})$.

(iii) Starting with $x := L(t), Z := V(t), s := L(t)$, choose a variable $x_i \in Z \cap V(s)$ with the least index i, substitute the first occurrence of x_i by the term s_{x_i}, denote the resulting term again by s and put $Z := Z \setminus \{x_i\}$. While $Z \neq \phi$ continue this procedure. The resulting term is the normal form $NF(t)$.

The algorithm stops after a finite number of steps, since $G(t)$ is a rooted graph. Without difficulties, one shows that $G(NF(t)) = G(t), L(NF(t)) = L(t)$.

The following definition was given in [2].

Definition 5.4. The graph hypersubstitution $\sigma_{NF(t)}$ is called the *normal form graph hypersubstitution.* Here $NF(t)$ is the normal form of the binary

term t.

Since for any term t the rooted graphs of t and $NF(t)$ are the same, we have $t \approx NF(t) \in Id(\mathcal{G}')$. Then for any graph hypersubstitution σ_t with $\sigma_t(f) = t \in T(X_2)$, one obtains $\sigma_t \sim_{\mathcal{G}'} \sigma_{NF(t)}$.

All rooted graphs with at most two vertices were considered in [2]. Now we formed the corresponding terms and use the algorithm to construct normal form terms. The results are given in the following table.

normal form term	graph hypers.	normal form term	graph hypers.
(x_1x_2)	σ_0	x_1	σ_1
x_2	σ_2	(x_1x_1)	σ_3
(x_2x_2)	σ_4	(x_2x_1)	σ_5
$((x_1x_1)x_2)$	σ_6	$((x_2x_1)x_2)$	σ_7
$(x_1(x_2x_2))$	σ_8	$(x_2(x_1x_1))$	σ_9
$(x_1x_1)(x_2x_2)$	σ_{10}	$((x_2(x_1x_1))x_2)$	σ_{11}
$(x_1(x_2x_1))$	σ_{12}	$(x_2(x_1x_2))$	σ_{13}
$((x_1x_1)(x_2x_1))$	σ_{14}	$(x_2((x_1x_1)x_2))$	σ_{15}
$(x_1((x_2x_1)x_2))$	σ_{16}	$((x_2(x_1x_2))x_2)$	σ_{17}
$((x_1x_1)((x_2x_1)x_2))$	σ_{18}	$((x_2((x_1x_1)x_2))x_2)$	σ_{19}

By Theorem 3.1, we have the following relations:

(i) $\sigma_{10} \sim_{\mathcal{G}'} \sigma_{12} \sim_{\mathcal{G}'} \sigma_{14} \sim_{\mathcal{G}'} \sigma_{16} \sim_{\mathcal{G}'} \sigma_{18}$,

(ii) $\sigma_{11} \sim_{\mathcal{G}'} \sigma_{13} \sim_{\mathcal{G}'} \sigma_{15} \sim_{\mathcal{G}'} \sigma_{17} \sim_{\mathcal{G}'} \sigma_{19}$.

For $\sigma_{10} \sim_{\mathcal{G}'} \sigma_{12}$ consider by the following way: since $L(\sigma_{10}(f)) = L(\sigma_{12}(f)) = x_1$, $V(\sigma_{10}(f)) = V(\sigma_{12}(f))$, $N_i^{\sigma_{10}(f)}(x_1) \neq \phi$ and

$NT^{2(f)}(x_1) \neq \phi$, $N_o^{\sigma_{10}(f)}(x_1) \neq j>$ and $7V_O^{CTI2(f)}(x_1) \neq \phi$, $N_o^{\sigma_{10}(f)}(x_2) \neq \phi$ and $No^{\sigma_{2(-f\backslash x2)}} \wedge 0$. By Theorem 3.1 we have *<rio~g><Ti2-* For the other cases, we consider by the similar way. Let *Mg>* be the set of all normal form graph hypersubstitutions in *Q'.* Then we get

$$Mgi = \{<To,(Ti,(T2,<T3,<74,cr5,cr6,0'7,a8,a'9,(7lOjO'll\}-$$

We defined the product of two normal form graph hypersubstitutions in Mgi as follows.

Definition 5.5. The *product O-\N °N <*2N of two normal form graph hypersubstitutions* is defined by (CTI/V O_N <72./V)(/) = *NF{&iN\o'2N{f})}-*

The following table gives the multiplication of elements in Mg'.

O_N	00	O^r I	02	^"3	(74	05	06	(77	08	09	010	011
00	00	O^r I	02	^*3	(74	05	06	(77	08	09	010	811
0-1	(71	0^r I	02	O\	0^2	0^2	0-1	0^2	σ_1	0^2	σ_1	02
02	(72	0^r I	0^2	01	(Jo,	0-1	0-2	0-2	σ_2	0-1	σ_2	0^2
0-3	O3	0^r I	0"2	03	O~4	04	03	04	03	(74	^3	0-4
04	04	0^r I	0-2	03	(74	0^3	(74	(74	0-4	0^3	(74	04
Oh	Oh	0^r I	0^2	03	(74	0^0	0-9	0-11	07	06	0-11	(T\\
O_6	06	C^r I	0"2	0^3	(74	(77	0-6	(77	010	O-n	010	(T\\
σ_7	(77	0^r I	0"2	0-3	(74	0"6	O11	011	07	0-6	0-11	O11
08	0-8	0^r I	02	0^3	(74	0-9	OlO	0-11	08	0^9	0-10	(J\\
09	09	0^r I	02	03	(74	08	O9	0-11	0-11	0^10	O i l	(J\\
010	O'iO	C^{r1}	O_2	0^3	O_{-4}	0^11	0-10	O"II	0-10	011	010	(J\\
Oil	01i	c_r I	02	0-3	(74	o'IO	011	011	0-11	010	0-11	(J\\

The concept of a leftmost normal form graph hypersubstitution was defined in [2].

Definition 5.6. A graph hypersubstitution *0* is called *leftmost* hypersubstitution if *L(o(f)) = 21.*

The set $M_L(\sim)$ of all leftmost normal form graph hypersubstitutions in *MGI* is

$$ML(Q') = \{go,01,03,06708,010).$$

The concept of a proper hypersubstitution of a class of algebras was introduced in [6].

Definition 5.7. A hypersubstitution *0* is called *proper with respect to a class K of algebras* if *&[s]* M^*8[t]* E *Id(K)* for all *s* x *t* E $Id(K)$.

A graph hypersubstitution with the property that $\sigma(f)$ contains both variables x_1 and x_2 is called *regular*, which was introduced in [4]. It is easy to check that the set of all regular graph hypersubstitutions forms a groupoid M_{reg}.

We want to prove that $\{\sigma_0, \sigma_6, \sigma_8, \sigma_{10}\}$ is the set of all proper graph hypersubstitutions with respect to $\mathcal{G}'$.

The following lemma was proved in [2].

Lemma 5.2. *For each non-trivial term s, $(s \neq x \in X)$ and for all $u, v \in X$, we have*

$$E(\hat{\sigma_6}[s]) = E(s) \cup \{(u,u) | (u,v) \in E(s)\}$$

and

$$E(\hat{\sigma_8}[s]) = E(s) \cup \{(v,v) | (u,v) \in E(s)\}.$$

Then we obtain the following theorem.

Theorem 5.1. *$\{\sigma_0, \sigma_6, \sigma_8, \sigma_{10}\}$ is the set of all proper graph hypersubstitutions with respect to the class $\mathcal{G}'$ of $(x(yz)) \approx ((x(yy))z)$ graph algebras.*

Proof. If $s \approx t \in Id(\mathcal{G}')$ and s, t are trivial terms, then $\hat{\sigma_6}[s]$, $\hat{\sigma_8}[s]$, $\hat{\sigma_6}[t]$ and $\hat{\sigma_8}[t]$ are also trivial terms and thus $\hat{\sigma_6}[s] \approx \hat{\sigma_6}[t] \in Id(\mathcal{G}')$ and $\hat{\sigma_8}[s] \approx \hat{\sigma_8}[t] \in Id(\mathcal{G}') \in Id(\mathcal{G}')$. In the same manner, we see that $\hat{\sigma_6}[s] \approx \hat{\sigma_6}[t] \in Id(\mathcal{G}')$ and $\hat{\sigma_8}[s] \approx \hat{\sigma_8}[t] \in Id(\mathcal{G}')$ if $s = t = x$.

Now, assume that s and t are non-trivial terms, different from variables, and $s \approx t \in Id(\mathcal{G}')$. Then (i), (ii), (iii) and (iv) of Theorem 3.1 hold.

For σ_6, σ_8, we obtain:

$$L(\hat{\sigma_6}[s]) = L(s) = L(t) = L(\hat{\sigma_6}[t]),$$

and

$$L(\hat{\sigma_8}[s]) = L(s) = L(t) = L(\hat{\sigma_8}[t]).$$

Since σ_6 and σ_8 are regular, we have:

$$V(\hat{\sigma_6}[s]) = V(s) = V(t) = V(\hat{\sigma_6}[t])$$

and

$$V(\hat{\sigma_6}[s]) = V(s) = V(t) = V(\hat{\sigma_6}[t]).$$

By Lemma 4.2, we get:

$$E(\hat{\sigma_6}[s]) = E(s) \cup \{(u,u) \mid (u,v) \in E(s)\},$$

$$E(\hat{\sigma}_6[t]) = E(t) \cup \{(u,u) \mid (u,v) \in E(t)\},$$

$$E(\hat{\sigma}_8[s]) = E(s) \cup \{(v,v) \mid (u,v) \in E(s)\}$$

and

$$E(\hat{\sigma}_8[t]) = E(t) \cup \{(v,v) \mid (u,v) \in E(t)\}.$$

For σ_6, by Lemma 4.2, we see that $(L(s), L(s)) \in E(\hat{\sigma}_6[s])$ and $(L(t), L(t)) \in E(\hat{\sigma}_6[t])$. Further for any $x \in V(s)$, we get $N_o^{\hat{\sigma}_6[s]}(x) \neq \phi$ if and only if $N_o^s(x) \neq \phi$ and $N_o^{\hat{\sigma}_6[t]}(x) \neq \phi$ if and only if $N_o^t(x) \neq \phi$. By Theorem 3.1 (iii), we have $N_o^{\hat{\sigma}_6[s]}(x) \neq \phi$ if and only if $N_o^{\hat{\sigma}_6[t]}(x) \neq \phi$. By Theorem 3.1, we get $\hat{\sigma}_6[s] \approx \hat{\sigma}_6[t] \in Id(\mathcal{G}')$.

For σ_8, by Lemma 4.2 we see that for any $x \in V(s)$ with $x \neq L(s)$, $(x,x) \in E(\hat{\sigma}_8[s])$ and $(x,x) \in E(\hat{\sigma}_8[t])$, hence $N_o^{\hat{\sigma}_8[s]}(x) \neq \phi$ and $N_o^{\hat{\sigma}_8[t]}(x) \neq \phi$. Further we get $N_i^{\hat{\sigma}_8[s]}(L(s)) \neq \phi$ if and only if $N_i^s(L(s) \neq \phi$ and $N_i^{\hat{\sigma}_8[t]}(L(t)) \neq \phi$ if and only if $N_i^t(L(t) \neq \phi$. By Theorem 3.1(iii) we get $N_i^{\hat{\sigma}_8[s]}(L(s)) \neq \phi$ if and only if $N_i^{\hat{\sigma}_8[t]}(L(t)) \neq \phi$. By Theorem 3.1, we get $\hat{\sigma}_8[s] \approx \hat{\sigma}_8[t] \in Id(\mathcal{G}')$.

For σ_{10}, since $\sigma_6 \circ_N \sigma_8 = \sigma_{10}$ and σ_6, σ_8 are proper, we see that σ_{10} is a proper graph hypersubstitution.

For any $\sigma \notin \{\sigma_0, \sigma_6, \sigma_8, \sigma_{10}\}$, we give an identity $s \approx t$ in $\mathcal{G}'$ such that $\hat{\sigma}[s] \approx \hat{\sigma}[t] \notin Id(\mathcal{G}')$. Clearly, if s and t are trivial terms with different leftmost and different rightmost, then $\hat{\sigma_1}[s] \approx \hat{\sigma_1}[t] \notin Id(\mathcal{G}')$, $\hat{\sigma_3}[s] \approx \hat{\sigma_3}[t] \notin Id(\mathcal{G}')$, $\hat{\sigma_2}[s] \approx \hat{\sigma_2}[t] \notin Id(\mathcal{G}')$ and $\hat{\sigma_4}[s] \approx \hat{\sigma_4}[t] \notin Id(\mathcal{G}')$.

Now, let $s = (x_1(x_2x_1))$ and $t = (x_1((x_2x_1)x_2))$. By Theorem 3.1, we get $s \approx t \in Id(\mathcal{G}')$. If $\sigma \in \{\sigma_5, \sigma_7, \sigma_9, \sigma_{11}\}$, then $L(\sigma(f)) = x_2$. We see that $L(\hat{\sigma}[s]) = x_1$ and $L(\hat{\sigma}[t]) = x_2$ for $\sigma \in \{\sigma_5, \sigma_7, \sigma_9, \sigma_{11}\}$. Thus $\hat{\sigma}[s] \approx \hat{\sigma}[t] \notin Id(\mathcal{G}')$. □

Now, we apply our results to characterize all hyperidentities in the class of all $(x(yz)) \approx ((x(yy))z)$ graph algebras. Clearly, if s and t are trivial terms, then $s \approx t$ is a hyperidentity in $\mathcal{G}'$ if and only if they have the same leftmost variable and the same rightmost variable and $x \approx x$, $x \in X$ is a hyperidentity in $\mathcal{G}'$, too. Now, we consider the case that s and t are non-trivial and different from variables.

The concept of a *dual term* s^d of a non-trivial term s can be defined in the following way in [2].

If $s = x \in X$, then $x^d = x$, if $s = (t_1t_2)$, then $s^d = (t_2^d t_1^d)$. The dual term s^d can be obtained by application of the graph hypersubstitution σ_5, namely $\hat{\sigma}_5[s] = s^d$. Then, we prove the following Theorem:

Theorem 5.2. *An identity $s \approx t$ in $\mathcal{G}'$, where s, t are non-trivial and $s \neq x$, $t \neq x$, is a hyperidentity in $\mathcal{G}'$ if and only if the dual identity $s^d \approx t^d$ is also an identity in $\mathcal{G}'$.*

Proof. If $s \approx t$ is a hyperidentity in $\mathcal{G}'$, then $\hat{\sigma}_5[s] \approx \hat{\sigma}_5[t]$ is an identity in $\mathcal{G}'$, i.e., $s^d \approx s^d$ is an identity in $\mathcal{G}'$. Conversely, assume that $s \approx t$ is an identity in $\mathcal{G}'$ and that $s^d \approx t^d$ is an identity in $\mathcal{G}'$, too. We have to prove that $s \approx t$ is closed under all graph hypersubstitutions from $M_{\mathcal{G}'}$.

If $\sigma \in \{\sigma_0, \sigma_6, \sigma_8, \sigma_{10}\}$, then σ is a proper and we get that $\hat{\sigma}[s] \approx \hat{\sigma}[t] \in Id(\mathcal{G}')$. By assumption, $\hat{\sigma}_5[s] = s^d \approx t^d = \hat{\sigma}_5[t]$ is an identity in $\mathcal{G}'$.

For $\sigma_1, \sigma_2, \sigma_3$ and σ_4, we have $\hat{\sigma}_1[s] = L(s) = L(t) = \hat{\sigma}_1[t]$, $\hat{\sigma}_2[s] = L(s^d) = L(t^d) = \hat{\sigma}_2[t]$, $\hat{\sigma}_3[s] = L(s)L(s) = L(t)L(t) = \hat{\sigma}_3[t]$ and $\hat{\sigma}_4[s] = L(s^d)L(s^d) = L(t^d)L(t^d) = \hat{\sigma}_4[t]$.

Because of $\sigma_6 \circ_N \sigma_5 = \sigma_7$, $\sigma_8 \circ_N \sigma_5 = \sigma_9$, $\sigma_{10} \circ_N \sigma_5 = \sigma_{11}$ and $\hat{\sigma}_6[\hat{\sigma}_5[t']] = \hat{\sigma}_6[t'^d]$, $\hat{\sigma}_8[\hat{\sigma}_5[t']] = \hat{\sigma}_8[t'^d]$, $\hat{\sigma}_{10}[\hat{\sigma}_5[t']] = \hat{\sigma}_{10}[t'^d]$ for all $t' \in T(X)$, we have $\hat{\sigma}_7[s] \approx \hat{\sigma}_7[t]$, $\hat{\sigma}_9[s] \approx \hat{\sigma}_9[t]$ and $\hat{\sigma}_{11}[s] \approx \hat{\sigma}_{11}[t]$ are identities in $\mathcal{G}'$. □

References

1. K. Denecke, D. Lau, R. Pöschel and D. Schweigert: *Hyperidentities,hyperequational classes and clone congruences*, Contributions to General Algebra 7, Verlag Hölder-Pichler-Tempsky, Wien 1991, 97-118.
2. K. Denecke and T. Poomsa-ard: *Hyperidentities in graph algebras*, Contributions to General Algebra and Applications in Discrete Mathematics, Potsdam 1997, 59-68.
3. K. Denecke and M. Reichel: *Monoids of Hypersubstitutions and M-solid varieties*, Contributions to General Algebra, Wien 1995, 117-125.
4. K. Denecke and S. L. Wissmath: *Hyperidentities and Clones*, Gordon and Breach Science Publishers 2000.
5. E. W. Kiss, R. Pöschel and P. Pröhle: *Subvarieties of varieties generated by graph algebras*, Acta Sci. Math., 54, 1990, 57-75.
6. J. Płonka: *Hyperidentities in some of vareties*, in: General Algebra and discrete Mathematics ed. by K. Denecke and O. Lüders, Lemgo 1995, 195-213.
7. J. Płonka: *Proper and inner hypersubstitutions of varieties*, in: Proceedings of the International Conference: Summer School on General Algebra and Ordered Sets 1994, Palacký University Olomouce 1994, 106-115.
8. T. Poomsa-ard: *Hyperidentities in associative graph algebras*, Discussiones Mathematicae General Algebra and Applications 20(2000), 169-182.
9. T. Poomsa-ard, J. Wetweerapong and C. Samartkoon: *Hyperidentities in idempotent graph algebras*, Thai journal of Mathematics 2(2004), 171-181.
10. T. Poomsa-ard, J. Wetweerapong and C. Samartkoon: *Hyperidentities in transitive graph algebras*, Discussiones Mathematicae General Algebra and Applications 25(2005), 23-37.

11. R. Pöschel: *The equational logic for graph algebras*, Zeitschr. f. math. Logik und Grundlagen d. Math. Bd. 35, S. 1989, 273-282.
12. R. Pöschel: *Graph algebras and graph varieties*, Algebra Universalis, 27, 1990, 559-577.
13. C. R. Shallon: *Nonfinitely based finite algebras derived from lattices*, Ph. D. Dissertation, Uni. of California, Los Angeles, 1979.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 19–34)

QUANTUM POLYNOMIALS*

VYACHESLAV A.ARTAMONOV†

Department of Algebra, Faculty of Mechanics and Mathematics Moscow State University,
Leninsky Gory, 119992, GSP-2, Moscow, RUSSIA
E-mail: artamon@mech.math.msu.su

The paper contains a survey of some recent results on ring-theoretical properties of quantum affine space and actions on them of finite dimensional pointed Hopf algebras

1. Introduction

Let k be a field with a fixed matrix

$$\mathbf{q} = (q_{ij}) \in Mat(n, k), \quad n \geqslant 2,$$

whose entries $q_{ij} \in k^*$ satisfy the relations $q_{ii} = q_{ij}q_{ji} = 1$ for all $1 \leqslant i, j \leqslant n$. Let also r be an integer and $0 \leqslant r \leqslant n$. Denote by

$$\mathcal{O}_q = k_{\mathbf{q}}[X_1^{\pm 1}, \ldots, X_r^{\pm 1}, X_{r+1}, \ldots, \ldots, X_n] \tag{1}$$

the associative k-algebra with a unit element generated by elements

$$X_1, X_1^{-1}, \ldots, X_r, X_r^{-1}, X_{r+1}, \ldots, X_n$$

subject to defining relations

$$\begin{aligned} X_i X_i^{-1} = X_i^{-1} X_i = 1, \quad & 1 \leqslant i \leqslant r; \\ X_i X_j = q_{ij} X_j X_i, \qquad & 1 \leqslant i, j \leqslant n. \end{aligned} \tag{2}$$

The algebra (2) is an algebra of *quantum polynomials.* The elements q_{ij} are *multiparameters.* The algebra $\mathcal{O}_{\mathbf{q}}$ is a *generic* algebra of quantum polynomials if all multiparameters q_{ij} with $1 \leqslant i < j \leqslant n$, are independent in the multiplicative group k^* of the field k.

*Research partially supported by grant RFBR 06-01-00037
†To Prof. L.A. Bokut on the occasion of his 70th anniversary.

The algebra $\mathcal{O}_{\mathbf{q}}$ can be viewed as a coordinate algebra of a quantum affine space $\mathbb{A}^n_{\mathbf{q}}$ if $r = 0$ [13] and a coordinate algebra of a quantum torus $\mathbb{T}^n_{\mathbf{q}}$ if $r = n$ [13]. Unifying both cases we can consider a case of an arbitrary r where $0 \leqslant r \leqslant n$. Then the algebra $\mathcal{O}_{\mathbf{q}}$ can be considered as a coordinate algebra of $\mathbb{T}^r_{\mathbf{q}} \times \mathbb{A}^{n-r}_{\mathbf{q}}$ [13]. In the case $n = 2$, $r = 0$, the algebra $\mathcal{O}_q$ is also called a *quantum plane*, where $q = q_{12}$.

It is an easy exercise to check that every element of $\mathcal{O}_{\mathbf{q}}$ has a unique representation as a linear combination of monomials

$$X^i = X_1^{i_1} \cdots X_r^{i_r} X_{r+1}^{i_{r+1}} \cdots X_n^{i_n},$$

where $i = (i_1, \ldots, i_n) \in \mathbb{Z}^n$ and $i_{r+1}, \ldots, i_n \geqslant 0$. The ring $\mathcal{O}_{\mathbf{q}}$ a left and right Noetherian domain, it satisfies Ore condition and it has a division ring of fractions $F = Fract\mathcal{O}_{\mathbf{q}} = k_{\mathbf{q}}(X_1, \ldots, X_n)$.

2. Valuations

Let Γ be a linearly ordered additive (not necessarily commutative) group. A Γ-*valuation* is a surjective group homomorphism ν from the multiplicative group F^* of the quantum field F onto the group Γ with the following properties:

(1) if $a, b, a + b \in F^*$, then $\nu(a + b) \geqslant \min[\nu(a), \nu(b)]$;
(2) if $a, b, a + b \in F^*$ and $\nu(a) \neq \nu(b)$, then $\nu(a + b) = \min[\nu(a), \nu(b)]$.

We shall also assume that $\nu(k^*) = 0$.

Theorem 2.1 (Sabitov A. Yu.). *A valuation of a quantum division ring F, not necessarily a generic one, is Abelian in the sense that the group Γ is Abelian.*

Let $\nu_1 : F \to \Gamma_1$ and $\nu_2 : F \to \Gamma_2$ be two valuations. Set $\nu_1 \geqslant \nu_2$ if there exists an epimorphism of ordered groups $\tau : \Gamma_1 \to \Gamma_2$ such that $\tau\nu_1 = \nu_2$. It means that the diagram

$$\begin{array}{ccc} F & \xrightarrow{\gamma_1} & \Gamma_1 \\ \| & & \downarrow \tau \\ F & \xrightarrow{\gamma_2} & \Gamma_2 \end{array} \qquad (3)$$

is commutative. A valuation $\nu_1 : F \to \Gamma_1$ has a *maximal rank* if for any valuation $\nu_2 : F \to \Gamma_2$ such that $\nu_1 \geqslant \nu_2$ the morphism τ in (3) is an isomorphism.

Theorem 2.2. *A valuation $\nu : F \to \Gamma$ of a general quantum division ring F is has maximal rank if and only if $\Gamma \simeq \mathbb{Z}^n$.*

Observe that the set of all maximal valuations on F can be identified with the set of all linear orderings of the additive group $\mathbb{Z}^n$.

Let $\mathbb{R}^n$ be a vector space of all rows $(r_1, \ldots, r_n)$, $r_i \in \mathbb{R}$, of a length n. We shall assume that $\mathbb{R}^n$ is equipped with the lexicographic order. The following result is presented in ([16], Chapter 6]).

Theorem 2.3. *Let $\leqslant$ be a linear order in the additive group $\mathbb{Z}^n$. Then there exists order preserving group embedding $\mathbb{Z}^n \to \mathbb{R}^n$.*

The general linear group $GL(n, \mathbb{R})$ acts on the set of embeddings $\mathbb{Z}^n \to \mathbb{R}^n$. A linear order *essentially lexicographic* if it belongs to the orbit under the action of the group $GL(n, \mathbb{Z})$ of the standard embedding of $\mathbb{Z}^n$ into $\mathbb{R}^n$. If we apply an element $A = (a_{ij}) \in GL(n, \mathbb{Z})$ to the standard embedding of $\mathbb{Z}^n$ into $\mathbb{R}^n$ then we choose in F a new set of variables

$$Y_i = X_1^{a_{i1}} \cdots X_n^{a_{in}}, \quad i = 1, \ldots, n. \tag{4}$$

In other words we choose a new quantum polynomial algebra

$$D_{\mathbf{q}', \alpha'}[Y_1^{\pm 1}, \ldots, Y_n^{\pm 1}]$$

which is again a generic one and it has the same division ring F.

Thus if we consider an essentially lexicographic order then making a change of variables of the form (4) we obtain a lexicographic order with respect to the new set of variables $Y_1, \ldots, Y_n$.

According to ([25], Chapter 6) points of affine quantum space $\mathbb{A}_{\mathbf{q}}^n$ are related to valuations of F. The coordinate algebra of a quantum affine space $\mathbb{A}_{\mathbf{q}}^n$ is the subalgebra $k_{\mathbf{q}}[X_1, \ldots, X_n] \subseteq \mathcal{O}_{\mathbf{q}}$ generated by $X_1, \ldots, X_n$. Since the algebra $\mathcal{O}_{\mathbf{q}}$ is simple [18] the only prime ideals of $k_{\mathbf{q}}[X_1, \ldots, X_n]$ are just the ideals generated by some sets of variables. So a choice of a set of variables belonging to $\ker \nu$ can be identified with an element of $Speck_{\mathbf{q}}[X_1, \ldots, X_n]$. Recall that prime ideals of an algebra $k_{\mathbf{q}}[X_1, \ldots, X_n]$, which is not supposed to be a generic one, are classified in [13].

3. Completions of quantum fields

Throughout the rest of the paper we fix a maximal $\mathbb{Z}^n$-valuation ν of F and Theorem 2.2 where $\mathbb{Z}^n$ is equipped with (essentially) lexicographic linear order.

Definition 3.1. Following [2] denote by $\mathcal{F}$ the set of all maps $f : \mathbb{Z}^n \to k$ and the zero element such that $suppf = \{m \in \mathbb{Z}^n \mid f(m) \neq 0\}$ is Artinian with respect to the lecxicographic order on $\mathbb{Z}^n$. An element $f \in \mathcal{F}$ can be identified with a power series $f = \sum_{m \in \mathbb{Z}^n} f(m)X^m$. We say that a monomial $X^u = X_1^{u_1} \cdots X_n^{u_n}$, $u \in \mathbb{Z}^n$, *occurs* in an element $f \in \mathcal{F}$ if $f(u) \neq 0$.

Expand the valuation ν to $f \in \mathcal{F}$ in the following way. If $f \in \mathcal{F}$ then $\nu(f)$ – the least element from $suppf$. Put

$$\mathcal{O} = \{f \in \mathcal{F} \mid \nu(f) \geqslant 0\}, \quad \mathfrak{m} = \{f \in \mathcal{F} \mid \nu(f) > 0\}.$$

As in [3] it can be shown the set $\mathcal{F}$ is a division ring containing F. The skew field $\mathcal{F}$ is called a *completion* of F with respect to ν. Moreover $\mathcal{O}$ is a subring in $\mathcal{F}$ and $\mathfrak{m}$ is an ideal in $\mathcal{O}$. It is generated as an ideal in $\mathcal{O}$ by the elements $X_1^{\varepsilon_1}, \ldots, X_n^{\varepsilon_n}$ provided $\nu(X_1^{\varepsilon_1}), \ldots, \nu(X_n^{\varepsilon_n}) > 0$, where $\varepsilon_1, \ldots, \varepsilon = \pm 1$. Furthermore $\mathcal{O}/\mathfrak{m} \simeq k$.

Conjecture 3.1. *A valuation ν is associated to an essentially lexicographic order on $\mathbb{Z}^n$ if and only if $\cap_{n \geqslant 1} \mathfrak{m}^i = 0$.*

4. Projective module, elementary matrices and Morita-equivalence

A survey of results on projective modules over quantum polynomials can be found in [1]. We shall remind some of the basic results on the subject.

Theorem 4.1. *([1]) Let P be a finitely generated projective modules over a generic quantum polynomial algebra $\mathcal{O}_\mathbf{q}$. Suppose that the rank of P is at least 2. Then P is free. If $n \geqslant 2$ then there exists a non-free projective $\mathcal{O}_\mathbf{q}$-modules of rank 1.*

Theorem 4.2. *([1]) Assume that each multiparameter q_{ij} is a root of one. Let P be a finitely generated projective modules over a quantum polynomial algebra $\mathcal{O}_\mathbf{q}$. Suppose that the rank of P is at least 2. Then P is free. If the ring $\mathcal{O}_\mathbf{q}$ is non-commutative then there exists a non-free projective $\mathcal{O}_\mathbf{q}$-modules of rank 1. Moreover the group of elementary matrices $\mathbb{E}(m, \mathcal{O}_\mathbf{q})$ acts transitively on the set of unimodular rows of length $m \geqslant 3$. If $t \geqslant 4$ then $GL(t, \mathcal{O}_\mathbf{q}) = \mathbb{E}(t, \mathcal{O}_\mathbf{q})D$, where D is the group of invertible diagonal matrices.*

In the paper [2] we study Morita-equivalent generic quantum polynomials and prove Zarisky-type theorem for these algebras.

Theorem 4.3 (2). *Let $\mathcal{O}_{\mathbf{q}}$ and $\mathcal{O}_{\mathbf{q}'}$ be two non-commutative quantum polynomial algebras, one of which a generic one. If $\mathcal{O}_{\mathbf{q}}$ and $\mathcal{O}_{\mathbf{q}'}$ are Morita-equivalent then they are isomorphic. Moreover the Picard group of a generic quantum polynomial algebra is trivial.*

This result was generalized in [23].

Theorem 4.4 (L. Richard). *Let $\mathcal{O}_{\mathbf{q}}$, $\mathcal{O}_{\mathbf{q}'}$ be two quantum polynomial algebras with $r = n$ and the algebra $\mathcal{O}_{\mathbf{q}}$ is simple. The following are equivalent:*

(1) $\mathcal{O}_{\mathbf{q}} \simeq \mathcal{O}_{\mathbf{q}'}$;
(2) the division rings of fractions of $\mathcal{O}_{\mathbf{q}}$ and of $\mathcal{O}_{\mathbf{q}'}$ are isomorphic;
(3) the algebras of differential operators on $\mathcal{O}_{\mathbf{q}}$ and on $\mathcal{O}_{\mathbf{q}'}$ are isomorphic;
(4) the division rings of fractions of algebras of differential operators on $\mathcal{O}_{\mathbf{q}}$ and on $\mathcal{O}_{\mathbf{q}'}$ are isomorphic.

The next theorem is a solution of the Zarisky problem for generic quantum polynomials.

Theorem 4.5. *([2]) Let $\mathcal{O}_{\mathbf{q}}$ be a generic quantum polynomial algebra which is a skew polynomial extension either of the form $\mathcal{O}_{\mathbf{q}} = B[X, \alpha]$ or $\mathcal{O}_{\mathbf{q}} = B[X^{\pm 1}, \alpha]$ for some subalgebra B in $\mathcal{O}_{\mathbf{q}}$ and for some automorphism α of B. Then the subalgebra B is itself a generic quantum polynomial algebra.*

5. Automorphisms of generic quantum polynomials

Each quantum polynomial algebra is provided with a subgroup of *toric* automorphisms γ such that

$$\gamma(X_i) = \gamma_i X_i, \qquad \gamma_i \in k^*, \quad 1 \leqslant i \leqslant n.$$

Each toric automorphism can be expanded to automorphisms of division rings F, $\mathcal{F}$. If $r = n$ then the algebra $\mathcal{O}_q$ has also *mirror* automorphisms τ, where

$$\tau(X_i) = \tau_i X_i^{-1}, \qquad \tau_i \in k^*, \quad 1 \leqslant i \leqslant n.$$

Each mirror automorphism can be expanded to the division ring F, but it cannot be expanded to the division ring $\mathcal{F}$ from Section 3. A mirror automorphism has always order 2.

Now we shall consider the case $n = 2$. Then the algebra $\mathcal{O}_q$, $q = q_{12}$, is generated by two elements X, Y (possibly also by, X^{-1}, Y^{-1} if $r > 0$) subject to the defining relation $XY = qYX$. This algebra is a generic one

if and only if q is not a root of 1. If h is a rational function in one variable, then the maps

$$\begin{array}{ll} X \mapsto h(Y)X, & X \mapsto X, \\ Y \mapsto Y; & Y \mapsto h(X)Y; \end{array} \tag{5}$$

determine automorphisms of the division ring F. Moreover F has also toric and mirror automorphisms.

Conjecture 5.1 (J. Alev). *The automorphism group of F is generated by toric, mirror automorphisms, by automorphisms of the form* (5) *and by conjugations.*

The next theorem refines one of results from [22] in the case of quantum torus $r = n = 2$

Theorem 5.1. *Let γ be an endomorphism of a generic algebra $\mathcal{O}_q$ where $n = 2$. If $r = n = 2$ then there exist integers l, s, t, v and elements $\beta, \xi \in k$ such that*

$$\gamma(X) = \beta X^l Y^s, \quad \gamma(Y) = \xi X^t Y^v,$$

and $lv - st = 1$. In particular γ is an automorphism.

If $r = 1$ and $\gamma(Y) \neq 0$ then there exists elements $h \in k[X^{\pm 1}]$ and $\beta \in k^*$ such that

$$\gamma(X) = \beta X, \quad \gamma(Y) = h(X)Y. \tag{6}$$

If $r = 0$ then any automorphism of $\mathcal{O}_q$ is toric.

Assign the matrix

$$\pi(\gamma) = \begin{pmatrix} l & s \\ t & v \end{pmatrix} \in SL(2, \mathbb{Z})$$

to an automorphism γ from Theorem 5.1. The map $\pi : Aut\mathcal{O}_q \to SL(2, \mathbb{Z})$ is a group homomorphism and $\ker \pi$ consists of toric automorphisms. If $r = 2$ then the map π is surjective.

Theorem 5.2. *Let G be a finite automorphism group of a generic quantum plane. If $r < 2$ then G consists of toric automorphisms. If $r = n = 2$ then the group G is a semidirect product of a normal subgroup of toric automorphisms in G and a cyclic group of order 1 — 4, 6.*

An automorphism group of a generic quantum polynomial algebra $\mathcal{O}_q$ in the case $n \geqslant 3$ is determined in [11]. If $r = n \geqslant 3$ the following theorem can also be deduced from [22].

Theorem 5.3. *([11,22]) Let γ be an injective endomorphism of a generic quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$ where $n \geqslant 3$. Then γ is either toric or in the case $r = n$ either toric or a mirror automorphism. In particular the group $Aut\mathcal{O}_{\mathbf{q}}$ is a semidirect product of a normal subgroup of toric automorphisms and a cyclic group of order 2. If $r < n$, then $Aut\mathcal{O}_{\mathbf{q}}$ consists of toric automorphisms and therefore is Abelian.*

Theorem 5.4. *([11]) Let G be a finite automorphism group of a generic quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$ and $n \geqslant 3$. Then the subalgebra of invariants $\mathcal{O}_{\mathbf{q}}^G$ is left and right Noetherian and $\mathcal{O}_{\mathbf{q}}$ is a left and right finitely generate $\mathcal{O}_{\mathbf{q}}^G$-module.*

This theorem was proved in [22] for any quantum polynomial algebra provided $r = n$.

Theorem 5.5. *([11])) Let G be a finite automorphism group of a generic quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$ and $n \geqslant 3$. Let F be the division ring of fraction of $\mathcal{O}_{\mathbf{q}}$. Then the subalgebra of invariants F^G is the division ring of fractions of $\mathcal{O}_{\mathbf{q}}^G$.*

Conjecture 5.2. *Let $n \geqslant 3$ and F from Theorem 5.5. Prove that the automorphism group of the division ring F is generated by toric, mirror automorphisms and by all conjugations.*

Theorem 5.6. *([23]) Let $\mathcal{O}_{\mathbf{q}}$ with $r = n$ be a simple algebra (not necessarily a generic one). Then any endomorphism of $\mathcal{O}_{\mathbf{q}}$ is an automorphism.*

Conjecture 5.3. *Let $\mathcal{O}_{\mathbf{q}}$ be a generic quantum polynomial algebra and ϕ a nonzero endomorphism of the division ring F. Prove that ϕ is an automorphism of F.*

A partial solution of this conjecture was presented by J. Alev and F. Dumas in [10].

An endomorphism γ of the division ring of Laurent power series $\mathcal{F}$ is *continuous*, if it is determined by its images $\gamma(X_1), \ldots, \gamma(X_n)$.

Another partial solution of Conjecture 5.2 is presented in

Theorem 5.7. *Suppose that $\mathcal{O}_{\mathbf{q}}$ is a generic quantum polynomial algebra and γ is a continuous automorphism of $\mathcal{F}$. Assume γ has a finite order if $n = 2$. Then there exists an element $z \in \mathcal{F}$ and a toric automorphism γ' such that γ is a product $(Ad\, z)\gamma'$ where $(Ad\, z)x = zxz^{-1}$.*

Theorem 5.8. *Let $\mathcal{O}_{\mathbf{q}}$ be a generic quantum polynomial algebra and G a finite automorphism group of continuous automorphism of $\mathcal{F}$. Then there exists an element $w \in \mathcal{F}$ such that $(Adw)G(Adw)^{-1}$ consists of toric automorphisms.*

6. Commutative subalgebras

In this section we are exposing some results on commutative subalgebra in quantum polynomial algebras $\mathcal{O}_{\mathbf{q}}$ (not necessarily generic ones) and in division ring of fraction F.

Theorem 6.1. *([8]) Let $f \in k_q(X,Y) \setminus k$ and q is not a root of 1. Then the centralizer $C(f)$ is a commutative subalgebra (in fact a maximal subfield) in $k_q(X,Y)$.*

Theorem 6.2. *([24]) Let $\mathcal{F}$ be obtained from a generic quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$ and $f \in \mathcal{F} \setminus k$. Then the centralizer $C(f)$ of f in $\mathcal{F}$ is commutative and is a maximal subfield in $\mathcal{F}$.*

Theorem 6.3 (K. Goodearl, private communication). *Let K be a subfield in F (it is not necessarily assumed that $\mathcal{O}_{\mathbf{q}}$ is a generic quantum polynomial algebra). Then the transcendent degree of K over k does no exceed Krull dimension of $k_{\Pi}[X_1^{\pm 1}, \ldots, X_n^{\pm 1}]$.*

If $r = n$ the Krull and global dimensions of $\mathcal{O}_{\mathbf{q}}$ are equal to the maximal number of commuting monomials whose multi-indices are independent in $\mathcal{Z}^n$ [12].

Theorem 6.4 (Zelenova S.). *Let K be a commutative subalgebra of the algebra $\mathcal{O}_{\mathbf{q}}$ which is not supposed to be generic. Then the maximal number of algebraically independent elements in K does not exceed Krull and global dimensions of $\mathcal{O}_{\mathbf{q}}$.*

Conjecture 6.1. *Any maximal commutative subalgebra in $\mathcal{O}_{\mathbf{q}}$ is affine (finitely generated).*

Conjecture 6.2. *Let K be a maximal subfield in $k_q(X,Y)$ (q is not a root of 1). Then the extension K/k is a purely transcendental of degree 1.*

Observe that the division ring $k_q(X,Y)$ coincides with the division ring of quantized Weyl algebra $A_q(1) = k\langle a, b \mid ab - qba = 1\rangle$. In the meantime as it was shown by J. Dixmier the conjecture similar to Conjecture 6.2 did not hold for the division ring of an ordinary Weyl algebra A_1.

7. Actions of Hopf algebras

We start this section with explication of some basic ideas related to noncommutative algebraic geometry and the theory of Hopf algebras.

Let V, W be affine algebraic varieties with coordinate k-algebras $k[V], k[W]$ respectively. Then

$$k[V \times W] \simeq k[V] \otimes k[W]. \tag{7}$$

In particular if G is an affine algebraic group then morphisms of multiplication, of inverse, and of unit

$$\begin{gathered} G \times G \to G,\ G \to G,\ \{e\} \to G, \\ (x, y) \mapsto xy,\ x \mapsto x^{-1},\ e \to e. \end{gathered}$$

induce by (7) k-algebra morphisms

$$\begin{array}{lllll} \Delta : k[G] \to k[G] \otimes k[G], & (\Delta f)(x, y) & = & f(xy), \\ S : k[G] \to k[G], & (Sf)(x) & = & f(x^{-1}), \\ \varepsilon : k[G] \to k, & \varepsilon f & = & f(e). \end{array}$$

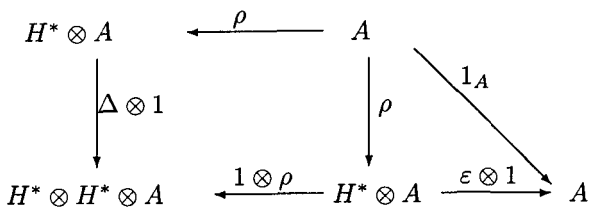
These morphisms are called *comultiplication, antipode* and a *counit*, respectively. It means that $k[G]$ is a commutative Hopf algebra. Recall that a Hopf algebra H with a comultiplication Δ is *cocommutative* if $\Delta = \tau\Delta$, where $\tau(u \otimes v) = v \otimes u$ for all $u, v \in H$.

Proposition 7.1. *A Hopf algebra $k[G]$ is cocommutative if and only if G is an Abelian group.*

In algebraic geometry we consider actions $G \times V \to V$ of algebraic group G on affine algebraic variety V. According to (7) it means that there is given a k-algebra (*structure*) morphism

$$\rho : k[V] \to k[G \times V] \simeq k[G] \otimes k[V]$$

such that $(\Delta \otimes 1)\rho = (1 \otimes \rho)\rho$, $(\varepsilon \otimes 1)\rho = 1$. It means that the following diagram with $A = k[V]$ and $H = k[G]$ is commutative



(8)

The main idea of noncommutative algebraic geometry is to consider a coordinate algebra A (not necessarily commutative) which is an *H-comodule algebra* for some Hopf algebra H. It means that there is given a k-algebra morphism $\rho : A \to H \otimes A$ with the properties (8).

A dual notion of H-comodule algebra is a notion of *H-module* algebra. We say that an algebra A is an *H-module algebra* if A is an H-module and $h(ab) = \sum_h (h_{(1)}a)(h_{(2)}b)$, where $\Delta(h) = \sum_h h_{(1)} \otimes h_{(2)}$. If H is a finite dimensional Hopf algebra then $H^* = Hom_k(H,k)$ is again a Hopf algebra and $H^{**} \simeq H$. An algebra A is an H-module algebra if and only if A is an H^*-comodule algebra.

In noncommutative geometry [14] a quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$ with $r = 0$ is a coordinate algebra of a quantum affine space $A^n_{\mathbf{q}}$ is considered together with quantum Grassman algebra Γ which is generated by elements $\xi_1, \ldots, \xi_n$ subject to the defining relations

$$\xi_i^2 = 0, \quad \xi_i \xi_j = p_{ij} \xi_j \xi_i \tag{9}$$

where $p_{ij} \in k^*$, and $p_{ii} = p_{ij}p_{ji} = 1$ for any $i, j = 1, \ldots, n$. The algebra of functions on matrices of size n is introduced in the book [14] as a universal bialgebra $M_{\mathbf{p},\mathbf{q}}(n)$ such that $\mathcal{O}_{\mathbf{q}}$, Γ are $M_{\mathbf{p},\mathbf{q}}(n)$-comodule algebras.

Starting from now we shall assume in this section that $\mathcal{O}_{\mathbf{q}}$ is a generic quantum polynomial algebra. We shall consider cocommutative and pointed Hopf algebras H such that $\mathcal{O}_{\mathbf{q}}$ is an H-(co)module algebra. From geometrical point of view (see Proposition 7.1) we consider commutative quantum groups acting on the affine space $A^n_{\mathbf{q}}$. We shall now study the Lie algebra $Der\mathcal{O}_{\mathbf{q}}$ of derivations of $\mathcal{O}_{\mathbf{q}}$ and the Lie algebra $Der\mathcal{F}$ of continuous derivations of $\mathcal{F}$. A derivation of $\mathcal{F}$ is *continuous* if it is determined by the images of $X_1, \ldots, X_n$. The subalgebra $Derint\mathcal{O}_{\mathbf{q}}$ of inner derivations is always an ideal in $Der\mathcal{O}_{\mathbf{q}}$. Similarly the subalgebra $Derint\mathcal{F}_\nu$ of inner derivations is an ideal in $Der\mathcal{F}$. We also have derivation $\partial_1, \ldots, \partial_n$ such that $\partial_j(X_i) = \delta_{ij}X_i$. Observe that $[\partial_i, \partial_j] = 0, \quad [\partial_i, ad_u] = ad_{\partial_i u}$. Moreover if $chark = p > 0$ then $\partial_i^p = \partial_i$. Thus the span L of $\partial_1, \ldots, \partial_n$ is an Abelian Lie algebra of dimension n. The following theorem can be deduced from [9] in the case $r = 0$ and from [22] in the case $r = n$.

Theorem 7.1. *([4-6]) Let $\mathcal{O}_{\mathbf{q}}$ be a generic quantum polynomial algebra. There is a direct decomposition of vector spaces $Der\mathcal{O}_{\mathbf{q}} = Derint\mathcal{O}_{\mathbf{q}} \oplus L$. Similarly $Der\mathcal{F} = Derint\mathcal{F} \oplus L$. Any finite dimensional Lie subalgebra in $Der\mathcal{O}_{\mathbf{q}}$ and in $Der\mathcal{F}$ is Abelian.*

Let γ be an automorphism of $\mathcal{O}_{\mathbf{q}}$. A linear operator D on $\mathcal{O}_{\mathbf{q}}$ is a *γ-derivation* if $D(xy) = D(x)y + \gamma(x)D(y)$ for all elements $x, y \in \mathcal{O}_{\mathbf{q}}$. D is

an *inner* γ-derivation if there exists an element $w \in \mathcal{O}_{\mathbf{q}}$ such that

$$D(x) = (ad_{\gamma} w)\, x = wx - \gamma(x)w$$

for all $x \in \mathcal{O}_{\mathbf{q}}$. The next theorem expands Corollaries 1.10 and 1.14 from [5].

Theorem 7.2. *([6]) Suppose that $n \geqslant 3$ and that γ has a finite order. Let D be a γ-derivation of a generic algebra $\mathcal{O}_{\mathbf{q}}$. Then either D is an inner γ-derivation or γ is a toric automorphism and there exists an element $w \in \mathcal{O}_{\mathbf{q}}$ such that*

$$D(X) = (ad_{\gamma} w)X + \theta X, \quad D(Y) = (ad_{\gamma} w)Y + \tau Y$$

where $(\beta - 1)\tau = (\xi - 1)\theta = 0$.

A similar result holds for a continuous γ-derivation of $\mathcal{F}$ with $n \geqslant 3$. Suppose that $n = 2$ and γ from Theorem 5.1.

Theorem 7.3. *([6]) If either $l \neq 1$, or $v \neq 1$, then D is an inner γ-derivation of a generic algebra $\mathcal{O}_q$.*

Theorem 7.4. *([6]) Suppose that D is a γ-derivation of a generic algebra $\mathcal{O}_q$ such that $\gamma(X) = \beta X$, $\gamma(Y) = \xi X^t Y$. If $t \neq 0$, then there exists an element $w \in \mathcal{O}_q$ such that*

$$D(X) = (ad_{\gamma} w)X, \quad D(Y) = (ad_{\gamma} w)Y + g(X)Y^{b+1},$$

where $g(X) \in k[X^{\pm 1}]$ and $\beta = q^{-b}$. If $t = 0$ then there exists an element $w \in \mathcal{O}_q$ such that

$$\begin{aligned} D(X) &= (ad_{\gamma} w)X + \theta X^{d+1}, \quad \theta \in k, \quad \xi = q^d, \\ D(Y) &= (ad_{\gamma} w)Y + \tau X^d Y^{b+1}, \quad \tau \in k, \quad \beta = q^{-b}. \end{aligned} \tag{10}$$

Theorem 7.5. *([6]) Let D be a algebraic γ-derivation of a generic algebra $\mathcal{O}_{\mathbf{q}}$ with a non-toric automorphism γ. Then $D = 0$.*

Theorem 7.6. *([6]) Let k have characteristic zero, D an algebraic γ-derivation of a generic algebra $\mathcal{O}_{\mathbf{q}}$ and γ a toric automorphism. Then $D = 0$.*

Similar results can be proved for continuous algebraic automorphism of $\mathcal{F}$.

Observe that if $r = n$ then the algebra $\mathcal{O}_{\mathbf{q}}$ is simple [18]. Hence by [15] $Derint\mathcal{O}_{\mathbf{q}} \simeq \mathcal{O}_{\mathbf{q}}/k$ is a simple Lie algebra. Similarly [15] the special Jordan algebra $\mathcal{O}_{\mathbf{q}}^{+}$ with respect to the new multiplication $a \circ b = \frac{1}{2}[ab + ba]$ is a simple Jordan algebra.

Proposition 7.2. *Let char $k = 0$ and ∂ either a derivation of a generic $\mathcal{O}_\mathbf{q}$ or a continuous derivation of $\mathcal{F}$. Suppose that there exists a nonzero polynomial $f(T) \in k[T]$ such that $f(\partial) = 0$. Then $\partial = 0$.*

There is a natural cocommutative Hopf algebra $H_{\mathcal{O}_\mathbf{q}}$ for which $\mathcal{O}_\mathbf{q}$ is an $H_{\mathcal{O}_\mathbf{q}}$-module algebra. Let U be the (restricted) universal enveloping algebra of Der. Then the group $G = Aut\mathcal{O}_\mathbf{q}$ acts on U by conjugations. In fact by Theorem 5.3

$$\alpha\partial_j\alpha^{-1} = \partial_j, \quad \alpha ad_u\alpha^{-1} = ad_{\alpha(u)}.$$

So we can form smash product $H_{\mathcal{O}_\mathbf{q}} = U\sharp kG$. Then both $\mathcal{O}_\mathbf{q}$ and Γ (from above) are $H_{\mathcal{O}_\mathbf{q}}$-module algebras.

Theorem 7.7. *([6]) Suppose that k is an algebraically closed field of characteristic zero and H is a cocommutative Hopf algebra such that $\mathcal{O}_\mathbf{q}$ is a generic quantum polynomial algebra which is an H-module algebra. Then there exists a Hopf algebra homomorphism $\xi : H \to H_{\mathcal{O}_\mathbf{q}}$ such that the action of H on $\mathcal{O}_\mathbf{q}$ is induced by ξ and the action of $H_{\mathcal{O}_\mathbf{q}}$ on $\mathcal{O}_\mathbf{q}$.*

Definition 7.1. ([19], p. 45) Let H be Hopf algebra acting in $\mathcal{O}_\mathbf{q}$. An extension of algebras $\mathcal{O}_\mathbf{q}/\mathcal{O}_\mathbf{q}^H$ is *Schelter-integral* if for any element $f \in \mathcal{O}_\mathbf{q}$ there exists a positive integer m such that $f^m + q(f) = 0$, where q is a sum of elements of the form

$$a_1 f^{n_1} \cdots a_d f^{n_d} a_{d+1}, \quad n_1 + \cdots + n_d < m$$

for some $a_1, \ldots, a_{d+1} \in \mathcal{O}_q^H$.

Proposition 7.3. *Let H be a Hopf subalgebra in $H_{\mathcal{O}_\mathbf{q}}$, the field k algebraically closed of characteristic zero and the extension $\mathcal{O}_\mathbf{q}/\mathcal{O}_\mathbf{q}^H$ Schelter-integral. If $P(H) \neq 0$ then $\mathcal{O}_\mathbf{q}^H$ and $P(H)$ are commutative.*

Proposition 7.4. *The action of $H = H_{\mathcal{O}_\mathbf{q}}, n \geqslant 3$, can be expanded to $F = k_\mathbf{q}(X_1, \ldots, X_n)$ and to $\mathcal{F}$. Let F^H be the set of all invariants for each element of H. Then F^H is a skew subfield in F.*

Theorem 7.8. *([6]) Let H be a Hopf subalgebra in $H_{\mathcal{O}_\mathbf{q}}$ such that F has a finite left (right) finite dimension over F^H. Suppose k has characteristic zero and some multiparameter q_{ij}, $1 \leqslant i < j \leqslant n$, is algebraically independent over the subfield of rationales $\mathbb{Q}$ in k. Then $P(H) = 0$ and $H = kG$, where G is a subgroup of $Aut\mathcal{O}_\mathbf{q}$.*

8. Actions of pointed finite dimensional Hopf algebras

A Hopf algebra H over a field k is *pointed* if all simple co-subalgebras in H have dimension 1. For example group algebra and universal (restricted) algebra are pointed. A Hopf algebra is pointed if and only if any maximal ideal of the dual algebra H^* has codimension 1. In particular any cocommutative Hopf algebra over an algebraically closed field is pointed. Consider *standard* pointed Hopf algebras acting of quantum polynomial algebras $\mathcal{O}_{\mathbf{q}}$. Let $C = \langle \tau \rangle$ is a cyclic group of order $1 - 4, 6$. Fix an injective group homomorphism $\Phi : C \to SL(n, \mathbb{Z})$ such that if $n > 2$ then the order of C is either 1 or 2. By Theorem 5.1, Theorem 5.2 Theorem 5.3 the required homomorphism Φ exists.

Let U be a $\mathbb{Z}C$-submodule in $\mathbb{Z}^n$ of a finite index. Denote by Λ the finite abelian group $\mathbb{Z}^n/U$ written multiplicatively.

The homomorphism $\Phi : C \to SL(n, \mathbb{Z})$ induces an action of kC on the dual algebra $(k\Lambda)^*$. Namely if $f \in (k\Lambda)^*$, $c \in C$, and $x \in \Lambda$ then

$$f^c(x) = f\left(\Phi(c)^{-1}(x)\right). \tag{11}$$

The smash product $(k\Lambda)^* \sharp kC$ is a Hopf algebra. We shall call it a *standard* Hopf algebra. Note that if $n > r$, then a standard Hopf algebra has the form $(k\Lambda)^*$.

An algebra $\mathcal{O}_{\mathbf{q}}$ with $r = n$ admits the natural Λ-grading, induced by $\mathbb{Z}^n$-grading with respect to $X_1, \ldots, X_n$. According to[19] there exists a left coaction

$$\rho : \mathcal{O}_q \to k\Lambda \otimes \mathcal{O}_q, \quad \rho(X_i) = (e_i U) \otimes X_i, \quad i = 1, \ldots, n, \tag{12}$$

where $e_i = (0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0) \in \mathbb{Z}^n$. In other words $\mathcal{O}_q$ admits a left action of the dual Hopf algebra $(k\Lambda)^*$ [19], namely

$$f(X^v) = f(v + U)X^v, \quad v \in \mathbb{Z}^n. \tag{13}$$

We can define an action of kC on $\mathcal{O}_q$ with $r = n$ as follows. If the order of C is equal to 2, then

$$\tau \circ X_i = \xi_i X_i^{-1}, \quad \xi_i \in k^*, \ i = 1, \ldots, n. \tag{14}$$

Suppose now that $n = r = 2$ and

$$\Phi(\tau) = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in SL(2, \mathbb{Z})$$

has one of the orders $3, 4, 6$. Then we put

$$\tau \circ (X) = \xi X^a Y^b, \quad \tau \circ (Y) = \eta X^c Y^b, \tag{15}$$

where $\xi, \eta \in k^*$.

Proposition 8.1. *There exists an action of* $(k\Lambda)^* \sharp kC$ *on* $\mathcal{O}_{\mathbf{q}}$ *extending actions of* $(k\Lambda)^*$ *and the action* (14), (15) *of* kC.

An element $a \in \mathcal{O}_{\mathbf{q}}$ is an *invariant* under an action of H if $ha = \varepsilon(h)a$ for all $h \in H$. All invariants form a subalgebra $\mathcal{O}_q^H$ in $\mathcal{O}_q$.

Theorem 8.1. *([7]) Suppose that* $\mathcal{O}_{\mathbf{q}}$ *is a generic quantum polynomial algebra and* H *is a finite dimensional Hopf algebra. Then an action of* H *in* $\mathcal{O}_q$ *is a composition of* Ψ *and a "standard" action of a "standard" Hopf algebra in* $\mathcal{O}_{\mathbf{q}}$. *If* $\mathcal{O}_q^H$ *is the subalgebra of invariants then it is left and right Noetherian ring and* $\mathcal{O}_{\mathbf{q}}$ *is a finitely generated left and right* $\mathcal{O}_{\mathbf{q}}^H$*-module.*

Theorem 8.2. *([6]) Let a finite dimensional Hopf algebra* H *act continuously on* $\mathcal{F}$. *Then there exists a nonzero element* $z \in \mathcal{F}$ *such that* $h(X^v) = z\chi_v(h)X^v z^{-1}$ *for any monomial* $X^v \in \mathcal{F}$, $v \in \mathbb{Z}^n$, *and any element* $h \in H$. *Here* $\chi_v \in H^*$. *The division ring* $\mathcal{F}$ *has a finite left and right dimension over the subdivision ring on invariants* $\mathcal{F}^H$.

9. Poisson structures

A *Poisson structure* on a k-algebra A is a k-bilinear multiplication *Poisson bracket* $\{\ ,\ \} : A \otimes A \to A$ such that

(1) A is a Lie algebra with respect to the multiplication $\{x, y\}$;
(2) $\{xy, z\} = \{x, z\}y + x\{y, z\}$ for all $x, y, z \in A$.

An algebra A with a Poisson bracket is called *Poisson algebra.*

Poisson algebras are considered in [17]. It is shown in [20] that under some assumptions on a set of multiparameters in an algebra $\mathcal{O}_{\mathbf{q}}$ with $r = 0, n$ there exists a Poisson algebra $\mathcal{O}_{\mathbf{q}'}$ such that the topological spaces of primitive (prime) ideals in $\mathcal{O}_{\mathbf{q}}$ and of sympectic (prime Poisson) ideals in $\mathcal{O}_{\mathbf{q}'}$ are homeomorphic. A study of Poisson brackets is related to a study of the Lie algebra $Der\mathcal{O}_{\mathbf{q}}$ of derivation of the algebra $\mathcal{O}_{\mathbf{q}}$, because that map $a \mapsto \{b, a\}$ is a derivation of $\mathcal{O}_{\mathbf{q}}$ for any $b \in \mathcal{O}_{\mathbf{q}}$.

Theorem 9.1. *([6]) Let a Poisson bracket be given in a generic quantum polynomial algebra* $\mathcal{O}_{\mathbf{q}}$. *Then there exists an element* $\xi \in k$ *such that* $\{a, b\} = \xi[a, b]$.

A Poisson bracket on $\mathcal{F}$ is *continuous* if it is uniquely determined by the set of values $\{X_i, X_j\}$, $1 \leqslant i < j \leqslant n$. As above we can prove

Theorem 9.2. *([6]) Let a continuous Poisson bracket be given on Laurent quantum power series $\mathcal{F}$ which is associated with a generic quantum polynomial algebra $\mathcal{O}_{\mathbf{q}}$. Then there exists an element $\xi \in k$ such that $\{a, b\} = \xi[a, b]$.*

References

1. Artamonov V. A., Quantum Serre's conjecture, Uspehi mat. nauk. 53(1998), N 4, 3-76.
2. Artamonov V. A., General quantum polynomials: irreducible modules and Morita-equivalence, Izv. RAN, ser. math. 63(1999), N 5, 3-36.
3. Artamonov V. A., Automorphisms of division ring of rational quantum functions, Math. sbornik, 191(2000), N 12, 3-26.
4. Artamonov V. A., Valuations on quantum fields, Commun. Algebra, 29(2001), N 9,
5. Artamonov V. A., Pointed Hopf algebras acting on quantum polynomials, J. Algebra 259(2003), N 2, 323-352.
6. Artamonov V. A., Actions of Hopf algebras on general quantum Mal'tsev power series and quantum planes.// J. Math. Sci. – 2006. – 134, N 1. – p. 1773 - 1798.
7. Artamonov V.A., Actions of pointed Hopf algebras on quantum torus, in Proceedings of the "Ferrara Algebra Workshop" jointly with the "Workshop on Hopf Algebras, Swansea", a special issue of the "Annali dell'Universita' di Ferrara, sez. VII, Scienze Matematiche, Vol. LI, 29-60, 2005.
8. Artamonov V. A., Cohn P. M., The skew field of rational functions on the quantum plane. J. Math. Sci. 93(1999), N 6, 824-829.
9. Alev J., Chamarie M., Dérivations et automorphismes de quelques algèbres quantiques, Comm. Algebra 20(1992), N 6, 1787-1802.
10. Alev, J., Dumas F., Field generators for the quantum plane. in Algebra, arithmetic a nd geometry with applications (West Lafayette, IN, 2000), 127-141, Springer, Berlin, 2004
11. Artamonov V. A., Wisbauer R., Homological properties of quantum polynomials, Algebras and representation theory, 4(2001), N 3, 219 - 247.
12. Brookes C. J. B., Crossed products and finitely presented groups, J. Group Theory, 3(2000), 433-444.
13. Brown K. A., Goodearl K. R., Lecture on algebraic quantum groups. Birkhäuser, Basel, Boston, 2002.
14. Demidov E. E., Quantum group, Moscow: Factorial, 1998, 127P.
15. Herstein I. N., On the Lie and Jordan rings of a simple associative ring, Amer. J. Math. 77(1955), 279-285.
16. Kokorin A. I., Kopytov V. M., Linearly ordered groups. Moscow: Nauka, 1972.
17. Korogodski Leonid I., Soibelman Yan I., Algebra of functions on quantum groups, Part I, Providence RI:Amer. Math. Soc., 1998.
18. McConnell J.C., Pettit J.J., Crossed products and multiplicative analogues of Weyl algebras, J. London Math. Soc. 38(1988), N 1, 47-55.

19. Montgomery S., Hopf algebras and their actions on rings. Regional Conference Series in Mathematics, v. 82 – American Math. Soc.: Providence R. I., 1993.
20. Oh Sei-Qwon, Park Chun-Gil, Shin Youg-Yeon, Quantum n-space and Poisson n-space, Commun. Algebra, 30(2002), N 9, 4197-4209.
21. Freddy van Oystaeyen, Algebraic geometry for associative algebras. Marcel Dekker Inc. NY, 2000, 302 pp.
22. Osborn J. P., Passman D., Derivations of skew polynomial rings, J. Algebra, 176(1995), N 2, 417-448.
23. Richard L., Sur les endomorphismes des tores quantiques. Commun. Algebra, 30(2002), N 11, 5282-5306.
24. Zelenova S., Commutative subalgebras in a ring of quantum polynomials and a skew field of quantum Laurent series, Math. Sbornik, 192(2001), N 3, 53-64.
25. Zariski O., Samuel P., Commutative algebra, vol. II, D. van Nostrand Company Inc.: Princeton, New Jersy, 1960.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 35–56)

GRÖBNER-SHIRSHOV BASES: SOME NEW RESULTS*

L. A. Bokut†

School of Mathematical Sciences, South China Normal University
Guangzhou 510631, P. R. China
Sobolev Institute of Mathematics, Russian Academy of Sciences
Siberian Branch, Novosibirsk 630090, Russia
E-mail: bokut@math.nsc.ru

Yuqun Chen‡

School of Mathematical Sciences, South China Normal University
Guangzhou 510631, P. R. China
Email: yqchen@scnu.edu.cn

In this survey article, we report some new results of Gröbner-Shirshov bases, including new Composition-Diamond lemmas, applications of some known Composition-Diamond lemmas and content of some expository papers.

Keywords: Composition-Diamond lemma; Group; HNN-extension; Schreier extension; Dialgebra; Lie algebra; Module; Chinese monoid.

1. Introduction

In this survey, we report the activities of the first author who has been staying in the South China Normal University, at Guangzhou, in Spring 2006, Spring 2007 and November-December 2007. With the participation of the second author, we are running an algebra seminar with 12 students (most of them are master degree students) and some young teachers at 5 times per week, two hours each session. The subjects of this seminar include from Combinatorial group theory, Free Lie algebras, Semi-simple Lie algebras to Non-associative algebras, Conformal algebras, Quantum groups, Semigroups and Dialgebras with emphasizing in Gröbner-Shirshov bases.

*Supported by the NNSF of China (No.10771077) and the NSF of Guangdong Province (No.06025062).
†Supported by the RFBR and the Integration Grant of the SB RAS (No. 1.9).
‡Corresponding author.

The second author visited Sobolev Institute of Mathematics at Novosibirsk as a visiting professor in July-October, 2006.

As the result of all these activities, more than 10 papers have been prepared. We now give a brief survey for some of the papers. We also mention some papers which were done during the first author's visiting to Prof. K. P. Shum, the Chinese University of Hong Kong, 2003-2005.

We divide these papers into three blocks:

(I) *New Composition-Diamond (CD-) lemmas,*
(II) *Applications of known CD-lemmas,*
(III) *Expository papers.*

We first explain what it means of "CD-Lemma" for a class (variety or category) $\mathcal{M}$ of linear Ω-algebras over a field k (here Ω is a set of linear operations on $\mathcal{M}$) with free objects.

$\mathcal{M}$-CD-Lemma Let $\mathcal{M}$ be a class of (in general, non-associative) Ω-algebras, $Free_{\mathcal{M}}(X)$ a free Ω-algebra in $\mathcal{M}$ generated by X with a linear base consisting of "normal (non-associative Ω-) words" $[u]$, $S \subset Free_{\mathcal{M}}(X)$ a subset and $<$ a monomial well order on normal words. Let S be a Gröbner-Shirshov basis (this means that any "composition" of elements in S is "trivial"). Then

(a) If $f \in Ideal(S)$, then $[\bar{f}] = [a\bar{s}b]$, where $[\bar{f}]$ is the "leading monomial" of f and $[asb]$ is a "normal s-word", $s \in S$.
(b) $Irr(S) = \{[u] | [u] \neq [a\bar{s}b], s \in S, [asb]$ *is a normal* $s-word\}$ is a linear basis of the algebra $\mathcal{M}(X|S)$ with defining relations S.

In many cases, each of conditions (a) and (b) is equivalent to the condition that S is a Gröbner-Shirshov basis in $Free_{\mathcal{M}}(X)$. But in some of our "new CD-Lemmas", this is not the case. Here, $\mathcal{M}$ may be as follows:

–Associative algebras ([56], [7], [3]).

A free associative algebra is $k\langle X\rangle$, the algebra of non-commutative polynomials; normal words are words on X; the A. I. Shirshov's composition $(f,g)_w$ is equal to $fb - ag$, if $w = \bar{f}b = a\bar{g}$, $\deg(\bar{f})+\deg(\bar{g}) > \deg(w)$, or $f - agb$, if $w = \bar{f} = a\bar{g}b$, where $a, b \in X^*$ and X^* the free monoid generated by X; a polynomial h is called trivial $mod(S)$ if it goes to 0 by using the Eliminations of Leading Words (ELW) of S (see below an equivalent definition).

–Lie algebras ([56]).

A free Lie algebra is $Lie(X)$, the algebra of Lie polynomials in $k\langle X\rangle$

(this theorem was proved by W. Magnus and E. Witt); by the normal words we mean the non-associative Lyndon-Shirshov words $[u]$ on X; the leading word $\bar{f}$ of a Lie polynomial f is the same as the associative polynomial; A. I. Shirshov's composition $[f,g]_w$ of two Lie polynomials is its associative composition with some extra bracketing defined in [53]; a normal s-word for $s \in Lie(X)$ has the form $[asb]$ with extra bracketing as before.

–Commutative algebras ([17], [18]).

A free commutative algebra is $k[X]$, the algebra of polynomials on X over a field k; normal words are monomials; the composition $S(-,-)$ is the operation of taking the B. Buchberger's S-polynomial: $S(f,g) = fb - ag$ for any polynomials f,g, where $w = \bar{f}b = a\bar{g} = l.c.m(\bar{f},\bar{g})$ and $\deg(\bar{f})+\deg(\bar{g}) > \deg(w)$.

–(Commutative, anti-commutative) non-associative algebras ([55]).

There is only composition of inclusion in the cases.

–Lie superalgebras ([45]).

The Composition-Diamond lemma for Lie superalgebras is known and proved.

–Grassmann algebras ([57]).

There is new composition of multiplication by a monomial.

–Supercommutative associative superalgebras ([45], [46]).

There is new composition of multiplication by a monomial.

–Conformal associative algebras (C, (n), $n \geq 0$, D) ([12]).

There are 6 types of compositions including inclusion, intersection, D-inclusion, D-intersection, left (right) multiplication by a generator. The condition (a) ((b)) in the CD-lemma is not equivalent to the condition that S is a Gröbner-Shirshov basis.

–Modules ([36], [29]).

2. CD-lemma for associative algebras

In this section, we cite some concepts and results from the literature which are related to the Gröbner-Shirshov bases for the associative algebras.

Definition 2.1. ([56], see also [6], [7]) Let f and g be two monic poly-

nomials in k$\langle X\rangle$ and $<$ a well order on X^*. Then, there are two kinds of compositions:

(i) If w is a word such that $w = \bar{f}b = a\bar{g}$ for some $a, b \in X^*$ with $\deg(\bar{f})+\deg(\bar{g}) > \deg(w)$, then the polynomial $(f,g)_w = fb - ag$ is called the intersection composition of f and g with respect to w.

(ii) If $w = \bar{f} = a\bar{g}b$ for some $a, b \in X^*$, then the polynomial $(f,g)_w = f - agb$ is called the inclusion composition of f and g with respect to w.

Definition 2.2. ([6], [7], [56]) Let $S \subset k\langle X\rangle$ with each $s \in S$ monic. Then the composition $(f,g)_w$ is called trivial modulo (S,w) if $(f,g)_w = \sum \alpha_i a_i s_i b_i$, where each $\alpha_i \in k$, $a_i, b_i \in X^*$, $s_i \in S$ and $\overline{a_i s_i b_i} < w$. If this is the case, then we write $(f,g)_w \equiv 0 \mod(S,w)$

Definition 2.3. ([6], [7], [56]) We call the set S with respect to the well order " $<$ " a Gröbner-Shirshov set (basis) in $k\langle X\rangle$ if any composition of polynomials in S is trivial modulo S.

If a subset S of $k\langle X\rangle$ is not a Gröbner-Shirshov basis, then we can add to S all nontrivial compositions of polynomials of S, and by continuing this process (maybe infinitely) many times, we eventually obtain a Gröbner-Shirshov basis S^{comp}. Such a process is called the Shirshov algorithm. It is an infinite algorithm as well as Kruth-Bendix algorithm (see [38]).

A well order $>$ on X^* is monomial if it is compatible with the multiplication of words, that is, for $u, v \in X^*$, we have

$$u > v \Rightarrow w_1 u w_2 > w_1 v w_2, \ \textit{for all} \ w_1, \ w_2 \in X^*.$$

A standard example of monomial order on X^* is the deg-lex order to compare two words first by degree and then lexicographically, where X is a linearly ordered set.

The following lemma was proved by Shirshov [56] for the free Lie algebras (with deg-lex ordering) in 1962 (see also Bokut [6]). In 1976, Bokut [7] specialized the approach of Shirshov to associative algebras, see also Bergman [3]. For commutative polynomials, this lemma is known as the Buchberger's Theorem in [17] and [18].

Lemma 2.1. *(Composition-Diamond Lemma) Let k be a field, $A = k\langle X|S\rangle = k\langle X\rangle/Id(S)$ and $<$ a monomial order on X^*, where $Id(S)$ is the ideal of $k\langle X\rangle$ generated by S. Then the following statements are equivalent:*

(i) S is a Gröbner-Shirshov basis in $k\langle X\rangle$.

(ii) $f \in Id(S) \Rightarrow \bar{f} = a\bar{s}b$ *for some* $s \in S$ *and* $a, b \in X^*$.
(iii) $Irr(S) = \{u \in X^* | u \neq a\bar{s}b, s \in S, a, b \in X^*\}$ *is a basis of the algebra* $A = k\langle X|S\rangle$.

3. New CD-lemmas

3.1. *CD-Lemma and HNN-extensions*

Y. Q. Chen and C. Y. Zhong in [27] give a version of CD-lemma in which the order may not be monomial. A Gröbner-Shirshov basis for HNN extensions of groups is obtained by using the new CD-lemma. This is the first paper to give a Gröbner-Shirshov basis by using a non-monomial order.

Theorem 3.1. *([27]) Let* $S \subseteq k\langle X\rangle$ *and* " $<$ " *a well order on* X^* *such that*

(I) $\overline{asb} = a\bar{s}b$ *for any* $a, b \in X^*$, $s \in S$;
(II) for each composition $(s_1, s_2)_w$ *in* S, *there exists a presentation*

$$(s_1, s_2)_w = \sum_i \alpha_i a_i t_i b_i,\ a_i \bar{t}_i b_i < w, \quad \text{where } t_i \in S,\ a_i, b_i \in X^*,\ \alpha_i \in k$$

such that for any $c, d \in X^*$, *we have* $ca_i\bar{t}_i b_i d < cwd$.

Then, the following statements hold.

(i) S *is a Gröbner-Shirshov basis in* $k\langle X\rangle$.
(ii) For any $f \in k\langle X\rangle$, $f \in Id(S) \Rightarrow \bar{f} = a\bar{s}b$ *for some* $s \in S$, $a, b \in X^*$.
(iii) The set

$$Irr(S) = \{u \in X^* | u \neq a\bar{s}b, s \in S, a, b \in X^*\}$$

is a linear basis of the algebra $k\langle X|S\rangle$.

We call the order satisfying the conditions in Theorem 3.1 an S-weak monomial order.

Let $G = gp\langle H, t | t^{-1}at = \varphi(a), a \in A\rangle$ be an HNN-extension of a group H, where A is a subgroup of H and φ a group isomorphism. By using Theorem 3.1, it is proved in [27] that there exists an explicit Gröbner-Shirshov basis S of G relative to some explicit S-weak monomial order such that the set $Irr(S)$ of S-irreducible words coincides with the set of normal forms in the Normal Form Theorem for HNN-extensions (see [4] and [43]).

3.2. *Dialgebreas*

In this section, we report some recent results of L. A. Bokut, Y. Q. Chen and C. H. Liu [10].

Let $D(X)$ be a free dialgebra (J.-L. Loday, 1995, [40]), where multiplications " $\vdash$ ", " $\dashv$ " are both associative and for any $a, b, c \in D(X)$,

$$a \dashv (b \vdash c) = a \dashv b \dashv c,$$
$$(a \dashv b) \vdash c = a \vdash b \vdash c,$$
$$a \vdash (b \dashv c) = (a \vdash b) \dashv c.$$

A linear basis of $D(X)$ consists of normal diwords

$$[u] = x_{-m} \vdash \cdots \vdash x_0 \dashv \cdots \dashv x_k = x_{-m} \cdots \dot{x_0} \cdots x_k,$$

where $x_i \in X, m, k \geq 0, x_0$ is the center of $[u]$ (see J.-L. Loday, 1995, [40]). We define the deg-lex order $[u] < [v]$, by using the lex-order of the weight $wt[u] = (k + m + 1, m, x_{-m}, \ldots, x_k)$.

Now, for $f, g \in S$, we define the compositions of inclusion, intersection and left(right) multiplication by a letter.

We call the set S a Gröbner-Shirshov set (basis) in $D(X)$ if any composition of polynomials in S is trivial modulo S (and $[w]$).

Theorem 3.2. *([10] CD-Lemma for dialgebras) Let $S \subset D(X)$ be a monic set and the order $<$ as before. Then $(i) \Rightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv)$, where*

(i) S is a Gröbner-Shirshov basis in $D(X)$.
(ii) For any $f \in D(X)$, $f \in Id(S) \Rightarrow [\overline{f}] = [a[\overline{s}]b]$ for some $s \in S$, $a, b \in [X^*]$ and $[asb]$ a normal S-diword.
(iii) The set $Irr(S) = \{u \in [X^*] | u \neq [a[\overline{s}]b], s \in S, a, b \in [X^*], [asb]$ is normal S-diword$\}$ is a linear basis of the dialgebra $D(X|S)$.
(iv) Each composition is trivial modulo S.

As an application of the above theorem, we obtain a Gröbner-Shirshov basis for the universal enveloping algebra of a Leibniz algebra. It is the PBW-Theorem for the Leibniz algebras. This is the third proof of the theorem after M. Aymon and P. P. Grival (2003) in [1], and P. Kolesnikov (2007) in [39].

Recall that a Leibniz algebra L is a non-associative algebra with a multiplication $[xy] \in L$ such that $[[xy]z] - [[xz]y] - [[yz]x] = 0$ (see [40]). For any dialgebra $(D, \dashv, \vdash)$, the linear space D with the multiplication $[xy] = x \dashv y - y \vdash x$ is a Leibniz algebra. For any Leibniz algebra

$L = Lei(\{e_i\}_I | [e_i e_j] = \sum_k \alpha_{ij}^k e_k,\ i,j \in I)$, one can define the universal enveloping D-algebra $U(L) = D(\{e_i\}_I | e_i \dashv e_j - e_j \vdash e_i = \sum_k \alpha_{ij}^k e_k,\ i,j \in I)$, where $\{e_i\}_I$ is a basis of L.

Theorem 3.3. *([10]) Let $\mathcal{L}$ be a Leibniz algebra over a field k with the product $\{,\}$. Let $\mathcal{L}_0$ be the subspace of $\mathcal{L}$ generated by the set $\{\{a,a\}, \{a,b\}+\{b,a\} \mid a,b \in \mathcal{L}\}$. Let $\{x_i | i \in I_0\}$ be a basis of $\mathcal{L}_0$ and $X = \{x_i | i \in I\}$ a linearly ordered basis of $\mathcal{L}$ such that $I_0 \subseteq I$. Let $(D(X), \dashv, \vdash)$ be the free dialgebra and the order $<$ on $[X^*]$ as before. Let S be the set which consists of the following polynomials:*

1. $f_{ji} = x_j \vdash x_i - x_i \dashv x_j + \{x_i, x_j\}$ $\quad (i,j \in I)$
2. $f_{ji\vdash t} = x_j \vdash x_i \vdash x_t - x_i \vdash x_j \vdash x_t + \{x_i, x_j\} \vdash x_t$ $\quad (i,j,t \in I,\ j > i)$
3. $h_{i_0 \vdash t} = x_{i_0} \vdash x_t$ $\quad (i_0 \in I_0,\ t \in I)$
4. $f_{t \dashv ji} = x_t \dashv x_j \dashv x_i - x_t \dashv x_i \dashv x_j + x_t \dashv \{x_i, x_j\}$ $\quad (i,j,t \in I,\ j > i)$
5. $h_{t \dashv i_0} = x_t \dashv x_{i_0}$ $\quad (i_0 \in I_0,\ t \in I)$.

Then

(i) S is a Gröbner-Shirshov basis in $(D(X)$.

(ii) The set

$$\{x_j \dashv x_{i_1} \dashv \cdots \dashv x_{i_k} \mid j \in I, i_p \in I - I_0,\ 1 \le p \le k,\ i_1 \le \cdots \le i_k,\ k \ge 0\}$$

is a linear basis of the universal enveloping algebra $U(\mathcal{L}) = D(X|S)$. In particular, $\mathcal{L}$ can be embedded into $U(\mathcal{L})$.

3.3. *Free Γ-algebras $k\langle X; \Gamma\rangle$*

In this section, we summary the results given by L. A. Bokut and K. P. Shum [14].

Let X be a set, Γ a group, $\Gamma(x)$, $\Gamma'(x)$ isomorphic subgroups, $x \in X$. Then the algebra $k\langle X; \Gamma\rangle$ with defining relations

$$\gamma x = x\gamma' \ (\gamma \in \Gamma(x),\ \gamma' \in \Gamma'(x), x \in X), \quad \gamma\delta = \mu \ (\gamma, \delta, \mu \in \Gamma)$$

is called *free Γ-algebra*.

A linear basis of $k\langle X; \Gamma\rangle$ consists of Γ-*words*

$$u = \gamma_0 x_{i_1} \gamma_1 \cdots x_{i_k} \gamma_k,\ x_i \in X, \gamma_i \in \Gamma, k \ge 0,$$

which are equivalent under transformations $\gamma x \to x\gamma'$ above.

We input a quasi-order on Γ-words:

$$u \le v \Leftrightarrow [u] \le [v],$$

where $[u] = x_{i1} \cdots x_{ik}$ is the projection of u, and $[u] \leq [v]$ a monomial order on X^*.

A Γ-polynomial f may have several leading monomials of $\bar{f}$. We call f a *strong* polynomial if $\bar{f}$ is unique. We define compositions of inclusion and intersection of two strong Γ-polynomials, and a strong Γ-Gröbner-Shirshov basis. The later is a set of strong Γ-polynomials that is closed under compositions.

Theorem 3.4. *Let $k\langle X;\Gamma\rangle$ be a free strong Γ-algebra, $S \subset k\langle X;\Gamma\rangle$ a strong Γ-Gröbner-Shirshov basis. Then*

(a) If $f \in Id(S)$, then $\bar{f} = a\bar{s}b$, where $\bar{f}$ is a leading monomial of f, $s \in S$, a, b Γ-words.

(b) $Irr(S) = \{u \neq a\bar{s}b | s \in S, a, b$ are Γ-words$\}$ is a linear basis of $k\langle X;\Gamma|S\rangle$.

There are many examples of Γ-algebras with strong Γ-Gröbner-Shirshov bases.

(a) Group algebras of universal groups $G(R^*)$ of multiplicative semigroups R^* of some rings R.

Let $R = \overline{k(S)}$, where $S = sgp\langle X|w_ih_i = u_if_i, w_i, h_i, u_i, f_i \in X\rangle$, k a field, $k(S)$ the semigroup algebra, $\overline{k(S)}$ the algebra of formal series over S. In particular, if S is a free semigroup, then $\overline{k(S)} = \overline{k\langle X\rangle}$ is the Magnus algebra of formal series over X.

These examples are from Bokut's solution to the Malcev embedding problem: There exists a semigroup S such that $k(S)^* \subset G$ (the multiplicative semigroup of $k(S)$ is embeddable into a group), but $k(S) \not\subseteq D$ ($k(S)$ is not embeddable into any division ring) (see [5]).

(b) Group algebras $k(G)$ for Tits systems (G, B, N, S) (see [16]). Here G has strong Γ-Gröbner-Shirshov basis, where $\Gamma = B$ and Γ-normal form is the Bruhat normal form.

3.4. *Tensor product of free algebras*

In A. A. Mikhalev and A. A. Zolotykh [47], a CD-lemma for the algebra $k[X] \otimes k\langle Y\rangle$ was found, where $k[X]$ is a polynomial algebra generated by X and $k\langle X\rangle$ is a free algebra.

In this section, we introduce the CD-lemma for tensor product $k\langle X\rangle \otimes k\langle Y\rangle$ of free algebras, which is from L. A. Bokut, Y. Q. Chen and Y. S. Chen [23].

Let X and Y be linearly ordered sets, $S = \{yx = xy | x \in X,\ y \in Y\}$. Then, with the deg-lex order ($y > x$ for any $x \in X,\ y \in Y$) on $(X \cup Y)^*$, S is a Gröbner-Shirshov basis in $k\langle X \cup Y\rangle$. Then, the set

$$N = X^*Y^* = Irr(S) = \{u = u^X u^Y | u^X \in X^* \text{ and } u^Y \in Y^*\}$$

is the normal words of the tensor product of the free algebras

$$k\langle X\rangle \otimes k\langle Y\rangle = k\langle X \cup Y \mid S\rangle.$$

Let kN be a k-space spanned by N. For any $u = u^X u^Y, v = v^X v^Y \in N$, we define the multiplication of the normal words as follows

$$uv = u^X v^X u^Y v^Y \in N.$$

Then, kN clearly coincides with the tensor product $k\langle X\rangle \otimes k\langle Y\rangle$.

Now, we order the set N. For any $u = u^X u^Y, v = v^X v^Y \in N$,

$$u > v \Leftrightarrow |u| > |v| \text{ or } (|u| = |v| \text{ and } (u^X > v^X \text{ or } (u^X = v^X \text{ and } u^Y > v^Y))),$$

where $|u| = |u^X| + |u^Y|$ is the length of u. It is obvious that $>$ is a monomial order on N. Such an order is also called the deg-lex order on $N = X^*Y^*$.

Let f and g be monic polynomials of kN and $w = w^X w^Y \in N$. Then we have found 16 types of compositions of inclusion and intersection.

S is called a *Gröbner-Shirshov basis* in $kN = k\langle X\rangle \otimes k\langle Y\rangle$ if all compositions of elements in S are trivial modulo S.

Theorem 3.5. *Let $S \subseteq k\langle X\rangle \otimes k\langle Y\rangle$ with each $s \in S$ monic and "$<$" the deg-lex order on $N = X^*Y^*$ as before. Then the following statements are equivalent:*

(1) S is a Gröbner-Shirshov basis in $k\langle X\rangle \otimes k\langle Y\rangle$.
(2) $f \in Id(S) \Rightarrow \overline{f} = a\bar{s}b$ for some $a, b \in N,\ s \in S$.
(3) $Irr(S) = \{w \in N | w \neq a\bar{s}b,\ a, b \in N,\ s \in S\}$ is a k-linear basis for the factor $k\langle X \cup Y | yx = xy, x \in X, y \in Y\rangle / Id(S)$.

4. Application of known CD-Lemmas

4.1. *Schreier extensions of groups*

Consider a Schreier extension of group

$$1 \to A \to G \to B \to 1.$$

Then we have Schreier's theorem (see [50]): A group G is a Schreier extension of A by B if and only if there exist a factor set $\{(b, b') | b, b' \in B\}$

of B in A and $\{b: A \to A,\ a \mapsto a^b$ is an automorphism$\}$ such that for any $b, b', b'' \in B, a \in A$,

$$(b,b')a^{bb'} = a^{[bb']}(b,b') \quad \text{and} \quad (b,b'b'')(b',b'') = (bb',b'')(b,b')^{b''},$$

where $[bb']$ is the product of elements b, b' in G.

M. Hall in his book [33] wrote down the following statement: "It is difficult to determine the identities [in A] leading to conditions for an extension", where the group B is presented by generators and relations.

In a recent paper, Y. Q. Chen [21], by using Gröbner-Shirshov bases, the structure of Schreier extensions of groups is completely characterized and an algorithm is given to find conditions for any Schreier extension of a group A by B, where B is presented by a presentation. Therefore, the above problem of M. Hall is solved.

Let A, B be groups. By a factor set of B in A, we mean a subset of A which is related to the presentation of B, see below.

Let the group B be presented as semigroup by generators and relations: $B = sgp\langle Y|R\rangle$, where R is a Gröbner-Shirshov basis for B with the deg-lex order $<_B$ on Y^*. For the sake of convenience, we can assume that R is a minimal Gröbner-Shirshov basis in a sense that the leading monomials are not contained each other as subwords, in particular, they are pairwise different. Let G be as in the following Theorem 4.1, where $A_1 = A\backslash\{1\}$, $S = \{aa' = [aa'], v = h_v \cdot (v),\ ay = ya^y | v \in \Omega,\ a, a' \in A_1,\ y \in Y\}$, $\{(v)|v \in \Omega\} \subseteq A$ a factor set of B in A, $\psi_y: A \to A,\ a \mapsto a^y$ an automorphism.

We define a tower order on $(A_1 \cup Y)^*$ which extends the order $<_B$ on Y^*.

For $w_1 = w = v_1 c = dv_2,\ v_1, v_2 \in \Omega,\ c, d \in Y^*,\ deg(v_1) + deg(v_2) > deg(w)$, we have,

$$f_{v_1}c - df_{v_2} = dh_{v_2} - h_{v_1}c \equiv 0 \ \ mod(R, w)$$

It means that there exists a $z \in Y^*$ such that

$$h_{v_1}c \equiv dh_{v_2} \equiv z \ \ mod(R, w)$$

and thus, there exist $\xi_{(v_1,v_2)_w}(v),\ \zeta_{(v_1,v_2)_w}(v) \in A$ such that

$$g \equiv z(\xi_{(v_1,v_2)_w}(v) - \zeta_{(v_1,v_2)_w}(v)) \ mod(S, w_1) \tag{1}$$

where $\xi_{(v_1,v_2)_w}(v)$ and $\zeta_{(v_1,v_2)_w}(v)$ are functions of $\{(v)|v \in \Omega\}$, and $g = (v_1 - h_{v_1} \cdot (v_1), v_2 - h_{v_2} \cdot (v_2))_{w_1}$.

In fact, by the previous formulas, we have an algorithm to find the functions $\xi_{(v_1,v_2)_w}(v)$ and $\zeta_{(v_1,v_2)_w}(v)$.

Theorem 4.1. *([21]) Let A, B be groups, $B = sgp\langle Y|R\rangle$, where $R = \{v - h_v | v \in \Omega\}$ is a minimal Gröbner-Shirshov basis for B and v the leading term of the polynomial $f_v = v - h_v \in R$. Let*

$$G = E(A, Y, a^y, (v)) = sgp\langle A_1 \cup Y|S\rangle$$

where $A_1 = A\backslash\{1\}$, $S = \{aa' = [aa'],\ v = h_v \cdot (v),\ ay = ya^y | v \in \Omega,\ a, a' \in A_1,\ y \in Y\}$, $\psi_y : A \to A,\ a \mapsto a^y$ an automorphism, $\{(v)|v \in \Omega\} \subseteq A$ a factor set of B in A with $(v) = 1$ if $f_v = y^\epsilon y^{-\epsilon} - 1,\ y \in Y,\ \epsilon = \pm 1$.

(i) For the tower order, S is a Gröbner-Shirshov basis for G if and only if for any $v \in \Omega,\ a \in A$ and any composition $(f_{v_1}, f_{v_2})_w$ of R in $k\langle Y\rangle$,

$$(v)a^v = a^{h_v}(v) \quad \text{and} \quad \xi_{(v_1,v_2)_w}(v) = \zeta_{(v_1,v_2)_w}(v) \tag{2}$$

hold in A, where $\xi_{(v_1,v_2)_w}(v)$ and $\zeta_{(v_1,v_2)_w}(v)$ are defined by (1). Moreover, if this is the case, G is a Schreier extension of A by B in a natural way.

(ii) A group C is a Schreier extension of A by B if and only if there exist $\{a^y | y \in Y,\ A \to A,\ a \mapsto a^y$ is an isomorphism$\}$ and a factor set $\{(v)|v \in \Omega\}$ of B in A with $(v) = 1$ if $f_v = y^\epsilon y^{-\epsilon} - 1,\ y \in Y,\ \epsilon = \pm 1$ such that (2) holds. Moreover, if this is the case, $C \cong G = E(A, Y, a^y, (v)) = sgp\langle A_1 \cup Y|S\rangle$.

Remark. In the above theorem, let the group $A = gp\langle X|R_A\rangle$ be also presented by generators and relations, where $R_A = \{u = f_u | u \in \Lambda\}$ is a Gröbner-Shirshov basis for A with the deg-lex order $<_A$ on X^*. Then, by replacing A_1 with X, $aa' = [aa']$ with R_A and x with a, the results hold.

As a corollary of the above theorem, in [21], by using the result in Y. Q. Chen and C. Y. Zhong in [27], a criteria in the case that B is HNN-extension is formulated.

Another solution to the M. Hall's problem can be referred in [2].

4.2. *Extensions of algebras*

In the paper of Y. Q. Chen [20], he gave the same kind of answer to an analogy of the M. Hall's problem in the above section for Schreier extensions of algebras.

Definition 4.1. Let k be a field, $M, B, \mathcal{R}$ k-algebras (not necessarily with 1). Then $\mathcal{R}$ is called an extension of M by B if $M^2 = 0$, where M is an ideal of $\mathcal{R}$ and $\mathcal{R}/M \cong B$ as algebras. Such an extension is called a singular extension in [34].

The following classical result is known. Let $M, B, \mathcal{R}$ be k-algebras with $M^2 = 0$. Then $\mathcal{R}$ is an extension of M by B if and only if M is a B-bimodule and there exists a factor set $\{(b, b') | b, b' \in B\}$ of B in M such that for any $b, b', b'' \in B$,

$$b(b', b'') - (bb', b'') + (b, b'b'') - (b, b')b'' = 0.$$

In [20], by using Gröbner-Shirshov bases, the structure of extensions of algebras is completely characterized and an algorithm is given to find conditions for any extension of an algebra M by B, where B is presented by a presentation.

As results, by using this theorem, in [20], a characterization theorem of the extension of M by B is given, when B is a cyclic algebra, free commutative algebra, universal envelope of a Lie algebra, and Grassmann algebra, respectively.

4.3. *Anti-commutative algebras*

In the paper of L. A. Bokut, Y. Q. Chen and Y. Li [11], an application of Shirshov's CD-lemma for anti-commutative algebras (see [55]) is given. This application gives an anti-commutative Gröbner-Shirshov basis of a free Lie algebra.

Let $X = \{x_i | i \in I\}$ be a linear ordered set. Let X^{**} be the set of all non-associative words (u) in X. We assume that (u) is a bracketing of u. Then we define normal words $N = \{[u]\}$ and order them by using induction on the length $n = |[u]|$ of $[u]$:

(i) If $n = 1$, then $[u] = x_i$ is a normal word. Define $x_i > x_j$ if $i > j$.
Let $N_{n-1} = \{[u] | [u]$ is a normal word and $|[u]| \leq n - 1\}$, $n > 1$ and suppose that " $<$ " is a well order on N_{n-1}. Then

(ii) If $n > 1$ and $(u) = ((v)(w))$ is a word of length n, then (u) is a normal word, if and only if

(a) both (v) and (w) are normal words, that is, $(v) = [v]$ and $(w) = [w]$, and

(b) $[v] > [w]$.

Let $[u]$, $[v]$ be normal words of length $\leq n$. Then $[u] < [v]$, if and only if one of the following three cases holds:

(a) $|[u]| < n$, $|[v]| < n$ and $[u] < [v]$.

(b) $|[u]| < n$ and $|[v]| = n$.

(c) If $|[u]| = |[v]| = n$, $[u] = [[u_1][u_2]]$ and $[v] = [[v_1][v_2]]$, then $[u_1] < [v_1]$ or $([u_1] = [v_1]$ *and* $[u_2] < [v_2])$.

It is clear that the order "$<$" on N is a well order.

Let $AC(X)$ be a k-space spanned by normal words. Now, we define the product of normal words by the following way:

$$\widetilde{[u][v]} = \begin{cases} [[u][v]] & : \ [u] > [v] \\ -[[v][u]] & : \ [u] < [v] \\ 0 & : \ [u] = [v] \end{cases}$$

Then $AC(X)$ is the free anti-commutative algebra generated by X.

Let $S \subset AC(X)$ be a set of monic polynomials, $s \in S$ and $(u) \in X^{**}$. We define S-word $(u)_s$ by induction as a non-associative word in $X \cup S$ with only one occurrence of $s \in S$. An S-word $(u)_s$ is called a normal S-word if $(u)_{\bar{s}} = (a[\bar{s}]b)$ is a normal word.

There is only one kind of compositions that is inclusion one.

Theorem 4.2. *([55], [11]) Let $S \subset AC(X)$ be a nonempty set of monic polynomials and the order " $<$ " as before. Then the following statements are equivalent:*

(i) S is a Gröbner-Shirshov basis in $AC(X)$.

(ii) $f \in Id(S) \Rightarrow [\bar{f}] = [a[\bar{s}]b]$ for some $s \in S$ and $a, b \in X^*$, where $[asb]$ is normal S-word.

(iii) $Irr(S) = \{[u] \in N | [u] \neq [a[\bar{s}]b] \ a, b \in X^*, \ s \in S$ and $[asb]$ is a normal S-word$\}$ is a basis of the algebra $AC(X|S)$.

By using this theorem, a Gröbner-Shirshov basis S in $AC(X)$ is given in [11] which shows that the Hall words in X forms a basis for the free Lie algebra $Lie(X)$, where $S = \{([u][v])[w] - ([u][w])[v]] - [u]([v][w]) \ |[u] > [v] > [w]$ *and* $[u], [v], [w]$ *are Hall words*$\}$.

4.4. *Akivis algebras*

This section is from the paper of Y. Q. Chen and Y. Li [25].

An Akivis algebra is a vector space A over a field k endowed with a skew-symmetric bilinear product $[x, y]$ and a trilinear product (x,y,z) that satisfy the identity $[[x, y], z] + [[y, z], x] + [[z, x], y] = (x, y, z) + (z, x, y) + (y, z, x) - (x, z, y) - (y, x, z) - (z, y, x)$. For any (non-associative) algebra B, one may obtain an Akivis algebra $Ak(B)$ by considering in B the usual commutator $[x, y] = xy - yx$ and associator $(x, y, z) = (xy)z - x(yz)$.

The CD-lemma for non-associative algebras is invented by Shirshov in [55]. By applying this lemma, in [25], a Gröbner-Shirshov basis in $AC(X)$ is given for the universal enveloping algebra of an Akivis algebra which gives an another proof of I. P. Shestakov's result (see [51)] that any Akivis algebra is linear. An Akivis algebra A is linear if A can be embedded in some non-associative algebra B with above operations.

Theorem 4.3. *([25]) Let $(A, +, [,], (,,))$ be an Akivis algebra with a linearly ordered basis $\{e_i | \ i \in I\}$. Let*

$$[e_i, e_j] = \sum_m \alpha_{ij}^m e_m, \ (e_i, e_j, e_k) = \sum_n \beta_{ijk}^n e_n,$$

where $\alpha_{ij}^m, \beta_{ijk}^n \in k$. We denote $\sum_m \alpha_{ij}^m e_m$ and $\sum_n \beta_{ijk}^n e_n$ by $\{e_i e_j\}$ and $\{e_i e_j e_k\}$, respectively. Let

$$U(A) = M(\{e_i\}_I | \ e_i e_j - e_j e_i = \{e_i e_j\}, \ (e_i e_j)e_k - e_i(e_j e_k) = \{e_i e_j e_k\}, \ i, j, k \in .$$

be the universal enveloping algebra of A. Let

$$S = \{e_i e_j - e_j e_i - \{e_i e_j\} \ (i > j), \ (e_i e_j)e_k - e_i(e_j e_k) - \{e_i e_j e_k\} \ (i, j, k \in I),$$
$$e_i(e_j e_k) - e_j(e_i e_k) - \{e_i e_j\}e_k - \{e_j e_i e_k\} + \{e_i e_j e_k\} \ (i > j, \ k \geq j)\}.$$

Then

(i) S is a Gröbner-Shirshov basis for $U(A)$,
(ii) A can be embedded into the universal enveloping algebra $U(A)$.

4.5. *Some one-relator groups*

This section contains the results of Y. Q. Chen and C. Y. Zhong in [28].

It is well known the Magnus algorithm for a solution of the word problem for any one-relator group.

Using the Magnus rewriting procedure (see R. C. Lyndon and P. E. Schupp [43)], one may embed any one-relator group into a tower of HNN-extensions. For towers of HNN-extensions of groups, L. Bokut (see [13]) developed a method of groups with the standard normal forms in 1965. Actually, for these groups, Gröbner-Shirshov bases are also "standard" in a sense, so we may speak about "groups with the standard Gröbner-Shirshov bases".

By the way, we give here a story about applications of groups with the standard normal forms. At our seminar, we have studied the following interesting paper

K. Kalorkoti, Turing degree and the word and conjugacy problem for finitely presented groups, Internet.

Actually, it is a part of his Thesis in London University, 1979.

C. Y. Zhong was the speaker for almost two months at the seminar and she pointed out that K. Kalorkoti used successfully the method of L. A. Bokut on groups with the standard normal forms.

As a result of this study, we suggested K. Kalorkoti to publish his paper in the SEA Bull. Math. (see [35]).

Now, we back to one-relator groups. There is a chance that any tower of HNN-extensions produced by the Magnus method has the standard Gröbner-Shirshov basis and the standard normal form. In particular, it would give another algorithm for the word problem for any one-relator group.

The problem is that the Magnus embedding is not easy to observe and write down explicitly. Hence, we need to go step by step. Any one-relator group can be effectively embedded into one-relator group with two generators

$$gp\langle X|r\rangle \hookrightarrow G = gp\langle x,\ y|x^{n_1}y^{m_1}\cdots x^{n_k}y^{m_k} = 1\rangle,$$

where $x_i \mapsto x^{-q}yx^l$ (W. Magnus, see also [43]), $n_i,\ m_i \neq 0,\ k \geq 0$. We call k the depth of G.

Theorem 4.4. *([28]) Any two-generator one-relator group of the depth 3 is effectively Magnus embeddable into a tower of HNN-extensions, which is a group with the effective standard Gröbner-Shirshov basis and effective standard normal form.*

4.6. *The Chinese monoid*

The Chinese monoid $CH(X)$ on a well ordered set X has the following defining relations:

$$cba = bca = cab,\quad c \geq b \geq a,\quad a,b,c \in X.$$

A fundamental paper on the Chinese monoid has been published in 2001 [19].

In the paper of Y. Q. Chen and J. J. Qiu [26], a Gröbner-Shirshov basis for the Chinese monoid is found.

Theorem 4.5. *([26]) Let* $S = \{x_ix_jx_k - x_jx_ix_k,\ x_ix_kx_j - x_jx_ix_k,\ x_ix_jx_j - x_jx_ix_j,\ x_ix_ix_j - x_ix_jx_i,\ x_ix_jx_ix_k - x_ix_kx_ix_j,\ x_i,x_j,x_k \in X,\ i > j > k\}$. *Then*

(i) $sgp\langle X|T\rangle = sgp\langle X|S\rangle$ *and with deg-lex order,* S *is a Gröbner-Shirshov basis of the Chinese monoid* $CH(X)$.

(ii) *The* $Irr(S)$ *is the set which consists of words on* X *of the form* $u_n = w_1 w_2 \cdots w_n, n \geq 0$, *where*

$$\begin{aligned} w_1 &= x_1^{t_{11}} \\ w_2 &= (x_2 x_1)^{t_{21}} x_2^{t_{22}} \\ w_3 &= (x_3 x_1)^{t_{31}} (x_3 x_2)^{t_{32}} x_3^{t_{33}} \\ &\cdots \\ w_n &= (x_n x_1)^{t_{n1}} (x_n x_2)^{t_{n2}} \cdots (x_n x_{n-1})^{t_{n(n-1)}} x_n^{t_{nn}} \end{aligned}$$

with $x_i \in X$, $x_1 < x_2 < \cdots < x_n$ *and all exponents are non-negative integers.*

Then $Irr(S)$ coincides with the set of staircase words of the paper by J. Cassaigne et al. [19]. Also, the insertion algorithm of J. Cassaigne et al. [19] coincides with the elimination of leading words algorithm.

4.7. *Markov and Artin normal form theorem for braid groups*

In the paper of L. A. Bokut, V. V. Chaynikov and K. P. Shum [8], the authors present the classical results of Artin-Markov on braid groups by using the Gröbner-Shirshov bases. As an application, one can reobtain the normal form of Artin-Makov-Ivanovskiy as an easy corollary.

5. Expository papers

5.1. *Gröbner and Gröbner-Shirshov bases: an elementary approach*

There is an elementary approach in L. A. Bokut and K.P. Shum [15] to Gröbner-Shirshov bases theory with quite a few examples, including the example for Lie algebras.

5.2. *Shirshov's CD-lemma for Lie algebras*

What is now called the Gröbner-Shirshov method for Lie algebras originally invented by A. I. Shirshov in 1962 ([56]). Actually, that paper based on his paper [53] where Shirshov invented a new linear basis for a free Lie algebra now called the Lyndon-Shirshov basis (it was defined independently in

the paper [24] in the same year). Remarkably, Lyndon–Shirshov basis is a particular case of the series of bases of a free Lie algebra invented by A. I. Shirshov in his Candidate of Doctor of Science Thesis (Moscow State University, 1953, advisor was A. G. Kurosh) and published in 1962 ([54]) (cf. [49] where these bases are called Hall Bases). We now cite the Zbl review by P. M. Cohn [31] of the paper [53]: "The author varies the usual construction of basis commutators in Lie rings by ordering words lexicographically and not by length. This is used to give a very short proof of the theorem (Magnus [44], Witt [58]) that the Lie algebra obtained from a free associative algebra is free. Secondly he derives Friedrich's criterion (this Zbl 52,45) for Lie elements. As the third application he proves that every Lie algebra L can be embedded in a Lie algebra M such that in M any subalgebra of countable dimension is contained in a 2-generated subalgebra." We would like to add that it was a beginning of Gröbner-Shirshov bases theory for Lie and associative algebras. Lemma 4 of the paper, on special bracketing of a regular (Lyndon-Shirshov) associative word with a fix regular subword, leads to the algorithm of elimination of the leading word of one Lie polynomial in other Lie polynomial , i.e., to the reduction procedure that is very familiar in the cases of associative and associative-commutative polynomials. Also this Lemma 4 leads to the crucial notion of composition of two Lie polynomials that will be defined lately in the paper [56]. Last but not least, Shirshov [52] proved the following result for connections of some ideals of free Lie and free associative algebras.

Let $Lie(a,b)$ be the Lie algebra of Lie polynomials of $k\langle a,b\rangle$ (it is the free Lie algebra over a set $\{a,b\}$ and a field k). Let $J = J([a^2b^kab] = [[a[[ab]\cdots b]]][ab],\ k \geq 1)$ be the Lie ideal of $Lie(a,b)$ generated by $\{[a^2b^kab],\ k \geq 1\}$ and I the associative ideal of $k\langle a,b\rangle$ generated by J. Then, $I \cap Lie(a,b) = J$. The proof is dealing with leading monomials of Lie and associative polynomials. This result and its proof are the real beginning of Gröbner-Shirshov bases theory for Lie and associative algebras.

As for the paper [56] itself, it is a fully pioneer paper in the subject. He defines a notion of the composition $(f,g)_w$ of two Lie (associative) polynomials relative to an associative word w (it was called lately by S-polynomial for commutative polynomials by B. Buchberger [17] and [18]). It leads to the algorithm for construction of a Gröbner-Shirshov basis ($GSB(S)$) of Lie (associative) ideal generated by some set S: to joint to S all nontrivial compositions and to eliminate leading monomials of one polynomial of S in others. Shirshov proves the lemma, now known as the Composition, or Composition-Diamond lemma, that if $f \in Id_{Lie}(S)$, then $\overline{f}$, the leading

associative word of f, has a form $\overline{f} = u\bar{s}v$, where $s \in GSB(S)$, $u, v \in X^*$. Several years later the first author formulated this lemma in the modern form (see [6]). Let S be a set of Lie polynomials that is complete under composition (i.e., any composition of polynomials of S is trivial; on the other word, S is a Gröbner-Shirshov basis). Then if $f \in Id_{Lie}(S)$, then $\overline{f} = u\bar{s}v$, where $s \in S$, $u, v \in X^*$. Of course, from Shirshov's Composition-Diamond lemma it follows that the set $Irr(S)$ of $S-$irreducible Lyndon-Shirshov words constitutes a linear basis of the quotient algebra $Lie(X)/Id(S)$. The converse is also true.

Lately explicitly Shirshov's Composition-Diamond lemma for associative algebra was formulated by L. A. Bokut [7] in 1976 and G. Bergman [3] in 1978.

The paper by L. A. Bokut and Y. Q. Chen [9] contains a comprehensive proof of the Shirshov's Composition-Diamond lemma for Lie algebras. We follow Shirshov's ideas of his papers [52], [53], [54] and [56]. In particular, we prove all necessary properties of both associative (see also [42]) and non-associative Lyndon-Shirshov words by using the Shirshov's elimination in [52] (that is, the so called Lazard elimination in [41] and [49]).

5.3. *CD-lemma for modules*

Composition-Diamond lemma for modules was first formulated and proved by S.-J. Kang and K.-H. Lee in [36] and [37]. According to their approach, a Gröbner-Shirshov basis of a cyclic module M over an algebra A is a pair (S, T), where S is the set of the defining relations of A, $A = k\langle X|S\rangle$, and T is the defining relations for the A-module ${}_AM = {}_AM(e|T)$. Then Kang-Lee's Lemma says that (S, T) is a Gröbner-Shirshov pair for the A-module ${}_AM = {}_AM(e|T)$ if S is a Gröbner-Shirshov basis of A and T is closed under the right-justified composition with respect to S, and for $f \in S$, $g \in T$, such that $(f, g)_w$ is defined and $(f, g)_w \equiv 0 \ mod(S, T, w)$. A composition $(f, g)_w$ is called right-justified if $w = \overline{f} = a\overline{g}$ for some $a \in X^*$.

Some years later, E. S. Chibrikov [29] suggested a new Composition-Diamond lemma for modules that treat any module as a factor module of "double-free" module.

Let X, Y be sets and $mod_{k\langle X\rangle}\langle Y\rangle$ a free left $k\langle X\rangle$-module with the basis Y. Then $mod_{k\langle X\rangle}\langle Y\rangle = \oplus_{y\in Y} k\langle X\rangle y$ is called a "double-free" module.

Suppose that $<$ is a monomial order on X^*, $<$ a well order on Y and $X^*Y = \{uy|u \in X^*, \ y \in Y\}$. We define an order " $\prec$ " on X^*Y: for any

$w_1 = u_1y_1, w_2 = u_2y_2 \in X^*Y$,

$$w_1 \prec w_2 \Leftrightarrow u_1 < u_2 \quad \text{or } u_1 = u_2,\ y_1 < y_2 \tag{3}$$

Let $S \subset mod_{k\langle X\rangle}\langle Y\rangle$ with each $s \in S$ monic. We define a composition $(f,g)_w = f - ag$, where $w = \bar{f} = a\bar{g}$, $a \in X^*$, $f, g \in mod_{k\langle X\rangle}\langle Y\rangle$ are monic.

If $(f,g)_w = f - ag = \sum \alpha_i a_i s_i$, where $\alpha_i \in k$, $a_i \in X^*$, $s_i \in S$ and $a_i\bar{s}_i \prec w$, then this composition is called trivial modulo (S,w) and is denoted by $(f,g)_w \equiv 0\ mod(S,w)$.

Definition 5.1. ([29]) Let $S \subset mod_{k\langle X\rangle}\langle Y\rangle$ be a non-empty set with each $s \in S$ monic. Let the order "$\prec$" be as before. Then we call S a Gröbner-Shirshov basis in the module $mod_{k\langle X\rangle}\langle Y\rangle$ if all the compositions of polynomials in S are trivial modulo S.

Lemma 5.1. *([29], Composition-Diamond lemma for "double-free" modules) Let $S \subset mod_{k\langle X\rangle}\langle Y\rangle$ be a non-empty set with each $s \in S$ monic and "$\prec$" the order on X^*Y as before. Then the following statements are equivalent:*

(i) *S is a Gröbner-Shirshov basis in $mod_{k\langle X\rangle}\langle Y\rangle$.*
(ii) *$f \in k\langle X\rangle S \Rightarrow \bar{f} = a\bar{s}$ for some $a \in X^*$, $s \in S$.*
(iii) *$Irr(S) = \{w \in X^*Y | w \neq a\bar{s},\ a \in X^*,\ s \in S\}$ is a k-linear basis for the factor $mod_{k\langle X\rangle}\langle Y|S\rangle = mod_{k\langle X\rangle}\langle Y\rangle / k\langle X\rangle S$.*

As applications of Lemma 5.1, Y. Q. Chen, Y. S. Chen and C. Y. Zhong [22], found the Gröbner-Shirshov bases for highest weight modules over Lie algebra sl_2, Verma modules over Kac-Moody algebras, Verma modules over Lie algebras of coefficients of free conformal algebras and the universal enveloping modules for Sabinin algebras. (The last modules are defined in [48]).

References

1. M. Aymon and P.-P. Grivel, Un theoreme de Poincare-Birkhoff-Witt pour les algebres de Leibniz, *Comm. Algebra*, **31**, 527-544(2003).
2. Y. G. Baik, J. Harlander and S. J. Pride, The geometry of group extensions, *J. Group Theory*, **1**, 4, 395-416(1998).
3. G. M. Bergman, The diamond lemma for ring theory, *Adv. in Math.*, **29**, 178-218(1978).
4. L. A. Bokut, On one property of the Boone groups, *Algebra i Logika*, **5**, 5, 5-23(1966), II, **6**, 1, 25-38(1967).

5. L. A. Bokut, Groups of functions of multiplication semigroups of certain rings I, II, III, IV, *Sibir. Math. J.*, **10**, 2, 4, 4, 5, 246-286, 744-199, 800-819, 965-1005(1969).
6. L. A. Bokut, Unsolvability of the word problem, and subalgebras of finitely presented Lie algebras, *Izv. Akad. Nauk. SSSR Ser. Mat.*, **36**, 1173-1219(1972).
7. L. A. Bokut, Imbeddings into simple associative algebras, *Algebra i Logika*, **15**, 117-142(1976).
8. L. A. Bokut, V. V. Chaynikov and K. P. Shum, Markov and Artin normal form theorem for braid groups, *Comm. Algebra*, **35**, 2105-2115(2007).
9. L. A. Bokut and Yuqun Chen, Gröbner-Shirshov bases for Lie algebras: after A.I. Shirshov, *SEA Bull Math.*, **31**, 1057-1076(2007).
10. L. A. Bokut, Yuqun Chen and Cihua Liu, Gröbner-Shirshov bases for dialgebras, submitted.
11. L. A. Bokut, Yuqun Chen and Yu Li, Anti-commutative Gröbner-Shirshov bases of a free Lie algebra, submitted.
12. L. A. Bokut, Y. Fong and W. F. Ke, Composition Diamond Lemma for associative conformal algebras, *J. Algebra*, **272**, 739-774(2004).
13. L. A. Bokut and G. Kukin, Undecidable algrithmic problems for semigroup, *Algebra, Topology and Geometry*, **25**, 3-66. VINITI, Moscow, 1987.
14. L. A. Bokut, and K. P. Shum, Relative Gröbner-Shirshov bases for algebras and groups, *Algebra and Analisis*, to appear.
15. L. A. Bokut and K. P. Shum, Gröbner and Gröbner-Shirshov bases in algebra: an elementary approach, *SEA Bull. Math.*, **29**, 227-252(2005).
16. N. Bourbaki, Lie Groups and Lie Algebras, Chapter 4-6, Elements of Mathematics, Springer-Verlag, 2002.
17. B. Buchberger, An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal [in German], Ph.D. thesis, University of Innsbruck, Austria, 1965.
18. B. Buchberger, An algorithmical criteria for the solvability of algebraic systems of equations [in German], *Aequationes Math.*, **4**, 374-383(1970).
19. J. Cassaigne, M. Espie, D. Krob, J. C. Novelli and F. Hivert, The Chinese Monoid, *International Journal of Algebra and Computation,* **11**, 3, 301-334(2001).
20. Yuqun Chen, Gröbner-Shirshov basis for Schreier extensions of groups, *Comm. Algebra*, to appear. DOI: 10.1080/00927870701869899
21. Yuqun Chen, Gröbner-Shirshov basis for extensions of algebras, *Algebra Colloq.*, to appear.
22. Yuqun Chen, Yongshan Chen and Chanyan Zhong, Composition-Diamond Lemma for Modules, submitted.
23. L. A. Bokut, Yuqun Chen and Yongshan Chen, Composition-Diamond lemma for tensor product of free algebras, submitted.
24. K. T. Chen, R. H. Fox, and R. C. Lyndon, Free differential calculus, IV: the quotient groups of the lower central series. *Annals of Mathematics*, **68**, 81-95(1958).
25. Yuqun Chen and Yu Li, An application of non-associative Composition-Diamond lemma, submitted.

26. Yuqun Chen and Jianjun Qiu, Gröbner-Shirshov Basis for the Chinese Monoid, submitted.
27. Yuqun Chen and Chanyan Zhong, Gröbner-Shirshov basis for HNN extensions of groups and for the alternative group, *Comm. Algebra*, **36**, 1, 94-103(2008).
28. Yuqun Chen and Chanyan Zhong, Gröbner-Shirshov basis for some one-relator groups, submitted.
29. E. S. Chibrikov, On free Lie conformal algebras, *Vestnik Novosibirsk State University*, **4**, 1, 65-83(2004).
30. E. S. Chibrikov, A right normed basis of free Lie algebras and Lyndon–Shirshov words. *J. Algebra*, **302**, 593-612(2006).
31. P. M. Cohn, review Zbl 0080.25503 Shirshov, A. I. Über freie Liesche Ringe. (Russian) Mat. Sb., N. Ser., **45**(87), 113-122(1958).
32. G. Duchamp and D. Krob, *Plactic-growth-like monoids*, in Words, Languages and Combinatorics II, Kyoto, Japan, 25-28 August 1992, eds. M Jürgensen, World Scientific, 1994, 124-142.
33. Marshall Hall, Jr., The Theory of Groups, The Macmillan Company, 1959.
34. G. Hochschild, On the cohomology groups of an associative algebra, *Ann. Math.*, **46**, 58-67(1945).
35. K. Kalorkoti, Turing degrees and the word conjugacy problems for finitely presented groups, *SEA. Bull. Math.*, **30**, 855-888(2006).
36. S.-J. Kang, K.-H. Lee, Gröbner-Shirshov bases for representation theory, *J. Korean Math. Soc.*, **37**, 55-72(2000).
37. S.-J. Kang, K.-H. Lee, Gröbner-Shirshov bases for irreducible sl_{n+1}-modules, *J. Algebra*, **232**, 1-20(2000).
38. D. E. Knuth and P. B. Bendix, Simple word problems in universal algebras, *Comput. Probl. abstract Algebra, Proc. Conf. Oxford* 1967, 263-297 (1970).
39. P. Kolesnikov, Conformal representations of Leibniz algebras, arXiv:math/0611501.
40. J.-L. Loday, Algebras with two associative operations (dialgebras), *C. R. Acad. Sci. Paris*, **321**, 141-146(1995).
41. M. Lothaire, "Combinatorics on Words", Encyclopedia of Mathematics, Vol. 17, Addison-Wesley , 1983; reprented M. Lothaire, "Combinatorics on Words", Cambridge Mathematical Library, Cambridge University Press, 1997.
42. R. C. Lyndon, On the Burnside problem, *Trans. Amer. Math. Soc.*, **77**, 202-215(1954).
43. R. C. Lyndon and P. E. Schupp, Combinatorial Group Theory, Springer-Verlag, 1977.
44. W. Magnus, Über Beziehungen zwischern hören Kommutatoren, *J. Reine Angew. Math*, **177**, 105-115(1937).
45. A. A. Mikhalev, The junction lemma andthe equality problem for color Lie superalgebras, *Vestnik Moskov. Univ. Ser. I Mat. Mekh.*, **5**, 88-91(1989). English translation: *Moscow Unv. Math. Bull.*, **44**, 87-90(1989).
46. A. A. Mikhalev and E. A. Vasilieva, Standard bases of ideals of free supercommutative polynomial algebras (ε-Gröbner bases). Proc. Second International Taiwan-Moscow Algebra Workshop, Springer-Verlag, in press.

47. A. A. Mikhalev and A. A. Zolotykh, Standard Gröbner-Shirshov bases of free algebras over rings, I. Free associative algebras, preprint.
48. J. M. Perez-Izquierdo, Algebras, hyperalgebras, nonassociative bialgebras and loops, *Advances in Mathematics*, **208**, 834-876(2007).
49. C. Reutenauer, Free Lie algebras. Oxford Science Publications, 1993.
50. Otto Schreier, Uber die Erweiterung von Gruppen I, *Monats. fur Math. u. Phys.*, **34**, 165-180(1926).
51. I. P. Shestakov, Every Akivis algebra is linear, *Geometriae Dedicata*, **77**, 215-223(1999).
52. A. I. Shirshov, Subalgebras of free Lie algebras, *Mat. Sb.*, **33**, 441-452(1953) (in Russian).
53. A. I. Shirshov, On free Lie rings, *Mat. Sbornik*, **45**(87), 113-122(1958).
54. A.I. Shirshov, Bases for free Lie algebras, *Algebra Logic*, **1**, 1, 14-19(1962).
55. A. I. Shirshov, Some algorithmic problems for ε-algebras, *Sibirsk. Mat. Z.* **3**, 1, 132-137(1962).
56. A. I. Shirshov, Some algorithmic problem for Lie algebras, *Sibirsk. Mat. Z.*, **3**, 292-296(1962) (in Russian); English translation in SIGSAM Bull., **33**(2), 3-6(1999).
57. T. Stokes, Gröbner bases in exterior algebra, *J. Automated Reasoning*, **6**, 233-250(1990).
58. E. Witt, Treue Darstellungen Lieschen Ringe, *J. Reine Angew. Math.*, **177**, 152-160(1937).

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 57–68)

ISOMORPHISMS AND DERIVATIONS OF ALGEBRAS

SEUL HEE CHOI

Dept. of Mathematics, Jeonju University,
Chon-ju 560-759, Korea
E-mail: chois@jj.ac.kr

KI-BONG NAM

Dept. of Mathematic and Computer science, University of Wisconsin-Whitewater
Whitewater, WI 53190, USA
E-mail: namk@uww.edu

Using a given ring, we define the Weyl-type non-associative rings in this work. If the ring is simple, then we show that the Weyl-type non-associative ring is simple in this work. All the derivations of a Weyl-type non-associative algebra are found in the papers (see [4, 6]). In this work, we find all the derivations of a Weyl-type algebra which includes the algebras in the papers (see [4, 6]). We show that the associators of the algebras $\overline{WN_{n,0,n_1}}$ and $\overline{WN_{n,n,0_1}}$ are zeroes.

Keywords: Simple algebra; Graded; Antisymmetrized ring; Isomorphism; Derivation invariant.

1. Introduction

For a given ring, we define a non-associative ring and an antisymmetrized ring. We show that if a ring R is simple, then we can define a simple non-associative ring and a simple antisymmetrized ring such that these rings contain R and R^- respectively. In Herstein's book (see [9]), the well known result is that if A is a finite dimensional simple algebra over its center, then any derivation of A is inner. For a Lie algebra case, if L is a finite dimensional semi-simple Lie algebra, then every derivation of L is inner (see [10]). There is an infinite dimensional simple and Lie ring (algebra) which have an outer derivation. If an algebra is not (semi-simple) simple, then the algebra may have an outer derivation (see [14, 18]). Because of the additive group $Der(\overline{WN_{n,0,n_1}})$ of all the derivations of the algebra $\overline{WN_{n,0,n_1}}$, we can find that any two algebras in the class $\{\overline{WN_{n,0,n_1}} | n \in \mathbb{N}\}$ are not isomorphic (see Section 2).

2. Preliminaries

Let $\mathbb{F}$ be a field of characteristic zero (not necessarily algebraically closed). Throughout this paper, $\mathbb{N}$ (resp. $\mathbb{N}^+$) and $\mathbb{Z}$ will denote the non-negative (resp. positive) integers and the integers respectively. Let R be an associative ring and $M = \{\delta | \delta$ is a mapping from R to itself$\}$. The evaluation ring $R_M = \{a\delta | a \in A, \delta \in R\}$ with the obvious addition and the multiplication $*$ is defined as follows:

$$a_1\delta_1 * a_2\delta_2 = a_1\delta_1(a_2)\delta_2$$

for any $a_1\delta_1, a_2\delta_2 \in R_M$ (see [17]). For R_M, if $M = \{id\}$, then the ring $R_M = R$ where id is the identity map of R. Note that $R_M =< R_M, +, * >$ is generally not an associative ring. Using the commutator $[,]$ of R_M, we can define the antisymmetrized ring $R_M{}^- =< R_M, +, [,] >$. If the Jacobi identity holds in $R_M{}^-$. Then $R_M{}^-$ is a Lie ring. Let $p_n = \{g_1, \cdots, g_n\}$ be the set of given polynomials in the polynomial ring $\mathbb{F}[x_1, \cdots, x_{m+s}]$. For p_n and $m, s \in \mathbb{N}$, we define the $\mathbb{F}$-algebra $\mathbb{F}_{p_n,m,s} = \mathbb{F}[e^{\pm g_1}, \cdots, e^{\pm g_n}, x_1^{\pm 1}, \cdots, x_m^{\pm 1}, x_{m+1}, \cdots, x_{m+s}]$ in the formal power series ring $\mathbb{F}[[x_1, \cdots, x_{m+s}]]$ such that $\mathbb{F}_{p_n,m,s}$ is called a stable algebra(see [12]) with the standard basis

$$\mathbf{B} = \{e^{a_1 g_1} \cdots e^{a_n g_n} x_1^{i_1} \cdots x_{m+s}^{i_{m+s}} | a_1, \cdots, a_n, i_1, \cdots, i_m \in \mathbb{Z}, \\ i_{m+1}, \cdots, i_{m+s} \in \mathbb{N}\} \qquad (1)$$

and under obvious addition and usual multiplication (see [1, 2, 4, 5, 6]) where we take $g_1, \cdots, g_n$ so that $\mathbf{B}$ is the standard basis of $\mathbb{F}_{p_n,m,s}$. Let ∂_w, $1 \leq w \leq m+s$. Denote the usual partial derivative with respect to x_w on $\mathbb{F}_{p_n,m,s}$. For partial derivatives $\partial_u, \cdots, \partial_v$ of $\mathbb{F}_{p_n,m,s}$, the composition $\partial_u^{j_u} \circ \cdots \circ \partial_v^{j_v}$ of them is denoted by $\partial_u^{j_u} \cdots \partial_v^{j_v}$, where $j_1, \cdots, j_{m+s} \in \mathbb{N}$. We put $D = \{\partial_1^{j_1} \cdots \partial_{m+s}^{j_{m+s}} | j_1, \cdots, j_{m+s} \in \mathbb{N}\}$ where $\partial_{u_1}^0 \cdots \partial_{u_r}^0 = id$ is the identity map of $\mathbb{F}_{p_n,m,s}$. For a given ring R, we can define an infinite group ring $R_{p_n,m,s}$ as the algebra $\mathbb{F}_{p_n,m,s}$. For a subset D_1 of D, we can define the evaluation rings $\mathbb{F}_{p_n,m,s_{D_1}}$ and R_{p_n,m,s_D} respectively. Thus we can also define their antisymmetrized rings $\mathbb{F}_{p_n,m,s_{D_1}}{}^-$ and $R_{p_n,m,s_D}{}^-$. If D_1 is D, then the algebra $\mathbb{F}_{p_n,m,s_{D_1}}$ is the algebra $\overline{WN_{p_n,m,s}}$. If $D_1 = \{\partial_1^{j_1} \cdots \partial_{m+s}^{j_{m+s}} | 0 \leq j_1 + \cdots + j_{m+s} \leq r\}$, then we put $\mathbb{F}_{p_n,m,s_{D_1}} = \overline{WN_{p_n,m,s}}_r$. If $D_1 = \{\partial_1^{j_1} \cdots \partial_{m+s}^{j_{m+s}} | 0 < j_1 + \cdots + j_{m+s} \leq r\}$, then we put $\mathbb{F}_{p_n,m,s_{D_1}} = \overline{WN_{p_n,m,s}}_{r+}$. If $D_1 = \{\partial_u^v | v = r, 1 \leq u \leq m+s\}$, then we put $\mathbb{F}_{p_n,m,s_{D_1}} = \overline{WN_{p_n,m,s}}_{[r]}$. Using $R_{p_n,m,s_{D_1}}$ in-

stead of $\mathbb{F}_{p_n,m,s}$, we can define the rings $\overline{WN_{p_n,m,s}}(R)$, $\overline{WN_{p_n,m,s_r}}(R)$, $\overline{WN_{p_n,m,s_{r^+}}}(R)$, and $\overline{WN_{p_n,m,s_{[r]}}}(R)$ similarly. Thus we can define their corresponding antisymmetrized rings or algebras naturally. For any basis element $e^{a_1g_1}\cdots e^{a_ng_n}\, x_1^{i_1}\cdots x_{m+s}^{i_{m+s}}\partial_u^{j_u}\cdots\partial_v^{j_v}$ of the algebra $\overline{WN_{p_n,m,s}}$, we define the degree $deg_r(e^{a_1g_1}\cdots e^{a_ng_n}\, x_1^{i_1}\cdots x_{m+s}^{i_{m+s}}\partial_u^{j_u}\cdots\partial_v^{j_v})$ of x_r as i_r and the total degree $deg_{tot}(e^{a_1g_1}\cdots e^{a_ng_n}x_1^{i_1}\cdots x_{m+s}^{i_{m+s}}\,\partial_u^{j_u}\cdots\partial_v^{j_v})$ of the element $e^{a_1g_1}\cdots e^{a_ng_n}x_1^{i_1}\cdots x_{m+s}^{i_{m+s}}\partial_u^{j_u}\cdots\partial_v^{j_v}$ as $|i_1|+\cdots+|i_m|+i_{m+1}+\cdots+i_{m+s}$. Thus for any element l of the algebra $\overline{WN_{p_n,m,s}}$, we define $deg_r(l)$ of x_r as the highest degree of x_r in the basis terms of l and $deg_{tot}(l)$ as the highest total degree of the basis terms of l. It is well known that the algebra $\overline{WN_{p_n,m,s}}$ is simple, even though it has the right annihilator (see [7, 8]). For any element l of the algebra $\overline{WN_{p_n,m,s}}$, an element l_1 (resp. l_2) is a right (resp. left) identity of l, if $l*l_1=l$ (resp. $l_2*l=l$). Note that the algebra $\overline{WN_{p_n,m,s}}$ contains the algebra $\mathbb{F}_{p_n,m,s}$. The ring $\overline{WN_{p_n,m,s}}$ is $\mathbb{Z}^n$-graded as follows:

$$\overline{WN_{p_n,m,s}} = \bigoplus_{(a_1,\cdots,a_n)\in\mathbb{Z}^n} N_{(a_1,\cdots,a_n)} \tag{2}$$

where $(a_1,\cdots,a_n)$-homogeneous component $N_{(a_1,\cdots,a_n)}$ is the vector space spanned by $\{e^{a_1g_1}\cdots e^{a_ng_n}x_1^{i_1}\cdots\ x_{m+s}^{i_{m+s}}\partial_1^{j_1}\cdots\partial_{m+s}^{j_{m+s}}|i_1,\cdots,i_m\ \in\ \mathbb{Z},\ i_{m+1},\cdots,i_{m+s},j_1,\cdots,j_{m+s}\ \in\ \mathbb{N}\}$. Thus $\overline{WN_{p_n,m,s_r}}$, $\overline{WN_{p_n,m,s_{[r]}}}$, $\overline{WN_{p_n,m,s}}(R)$, $\overline{WN_{p_n,m,s_r}}(R)$, $\overline{WN_{p_n,m,s_{[r]}}}(R)$, and their appropriate corresponding antisymmetrized algebras or rings are also $\mathbb{Z}^n$-graded. If p_n is $\{x_u|1\le u\le n\}$ and D_1 is $\{\partial_u^{j_u}|1\le u\le n\}$, then we define the nonassociative algebra $\overline{WN_{p_n,0,0}}_{D_1} := \overline{WN_{n,0,0}}_{D_1}$ and its antisymmetrized algebra $\overline{WN_{p_n,0,0}}_{D_1}^{-} := \overline{WN_{n,0,0}}_{D_1}^{-}$ appropriately. We can also define the degree of each of the W-type rings or algebras respectively in this paper. The algebra $\overline{WN_{n,0,0}}_{D_1}$ has neither a right nor left identity. Throughout the paper, for a positive integer t, let us define the finite dimensional subalgebra $M_n^t(\mathbb{F})$ of the algebra $\overline{WN_{n,0,n}}$ such that the algebra $M_n^t(\mathbb{F})$ is spanned by

$$\{x_1^{i_1}\cdots i_n^{i_n}\partial_1^{j_1}\cdots\partial_n^{j_n}|\sum_{u=1}^{n} i_u=\sum_{u=1}^{n} j_u=t, i_1,\cdots,i_n,j_1,\cdots,j_n\in\mathbb{N}\}$$

For a given ring R, we can define the ring $M_n^t(R)$ as the ring $\overline{WN_{p_n,m,s}}(R)$ (see [15]). Note that $M_n^1(R)$ is the matrix ring over R.

3. Simplicities

Note that a ring R is simple if and only if every non-zero element l of R is full, i.e., the ideal generated by l is R. If I is a two-sided ideal of the ring R, then $\overline{WN_{p_n,m,s}}(I)$ (resp. $\overline{WN_{p_n,m,s_r}}(I)$ or $\overline{WN_{p_n,m,s_{r^+}}}(I)$) is a two-sided ideal of the ring $\overline{WN_{p_n,m,s}}(R)$ (resp. $\overline{WN_{p_n,m,s_r}}(R)$ or $\overline{WN_{p_n,m,s_{r^+}}}(I)$).

Theorem 3.1. *If R is a simple ring, then the non-associative ring $\overline{WN_{p_n,m,s_r}}(R)$ is simple. If R is a simple ring and its antisymmetrized ring R^- is simple, then the antisymmetrized ring $\overline{WN_{p_n,m,s_r}}(R)^-$ is simple.*

Proof. Since $\overline{WN_{p_n,m,s_r}}(R)$ (resp. $\overline{WN_{p_n,m,s_{[r]}}}(R)^-$) is graded, it is easy to prove that every non-zero element of the algebra $\overline{WN_{p_n,m,s_r}}(R)$ (resp. $\overline{WN_{p_n,m,s_r}}(R)^-$) is full. □

Corollary 3.1. *If R is the field $\mathbb{F}$, then the non-associative algebra $\overline{WN_{p_n,m,s}}(\mathbb{F}) = \overline{WN_{p_n,m,s}}$ (resp. antisymmetrized algebra $\overline{WN_{p_n,m,s}}(\mathbb{F})^- = \overline{WN_{p_n,m,s}}^-$) is simple. The matrix algebra $M_n(\mathbb{F})$ (resp. $(sl_n(\mathbb{F}))$ is a simple (resp. Lie) subalgebra of the algebra $\overline{WN_{p_n,m,s}}(\mathbb{F})$ (resp. $\overline{WN_{p_n,m,s}}^-$).*

Proof. The proof of the corollary is straightforward by Theorem 3.2. □

Corollary 3.2. *If R is simple, then the non-associative ring $\overline{WN_{p_n,m,s_{[1]}}}(R)$ is simple. If R and its antisymmetrized ring R^- are simple, then the Lie ring $\overline{WN_{p_n,m,s_{[1]}}}(R)^-$ is simple.*

Proof. Since it is easy to check the Jacobian identity of the ring $\overline{WN_{p_n,m,s_{[1]}}}(R)$, the remaining proofs of the corollary are straightforward by Theorem 3.1 and Theorem 3.2. Hence, the proof of the corollary is omitted. □

Proposition 3.1. *If p_n is $\{x_u | 1 \leq u \leq n\}$ and D_1 is $\{\partial_u^{\alpha_u} | 1 \leq u \leq n, 0 \neq \alpha_u \in \mathbb{N}\}$, then the algebra $\overline{WN_{n,0,0}}_{D_1}$ and the Lie algebra $\overline{WN_{n,0,0}}_{D_1}^-$ are simple. If n is greater than or equal to 2, then the matrix algebra is not a subalgebra of the algebra $\overline{WN_{n,0,0}}_{D_1}$.*

Proof. The proofs of the Proposition are also easy, so they are omitted. □

Theorem 3.2. *The $C(n+t-1,t)^2$-dimensional algebra $M_n^t(\mathbb{F})$ is simple. If R is a simple ring, then $M_n^t(R)$ is also simple.*

Proof. Every non-zero element of the algebra $M_n^t(\mathbb{F})$ and $M_n^t(R)$ is full, so the algebras $M_n^t(\mathbb{F})$ and $M_n^t(R)$ are simple. This completes the proofs of the theorem. □

4. Derivations of the algebras $\overline{WN_{n,0,n_1}}$ and $\overline{WN_{n,n,0_1}}$

The algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) can be decomposed as follows:

$$\overline{WN_{n,0,n_1}} = \mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n] \bigoplus N_1^\partial$$
$$(\text{ resp. } \overline{WN_{n,0,n_1}} = \mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}] \bigoplus N_2^\partial) \qquad (3)$$

where N_1^∂ (resp. N_2^∂) is the appropriate complimentary vector subspace of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,0,n_1}}$).

Note. 1 For any basis elements $e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}$ and $e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k$ of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$), $k \in \{1, \cdots, n\}$, if we define $\mathbb{F}$-maps $D_{v,w}$, D'_u, and S_v, $1 \leq u, v, w \leq n$, from the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) to itself as follows:

$$D_{v,v}(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}) = a_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v+1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n},$$
$$D_{v,v}(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k)$$
$$= a_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v+1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n} \partial_k$$
$$+(i_v - \delta_{k,v}) e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k, \text{ for } 1 \leq v \leq n$$
$$D_{v,w}(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}) = a_w e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v+1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n},$$
$$D_{v,w}(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k)$$
$$= a_w e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v+1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n} \partial_k,$$
$$+i_w e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v+1} \cdots x_w^{i_w-1} \cdots x_n^{i_n} \partial_k$$
$$-\delta_{k,v} e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_w, \text{ for } 1 \leq v \neq w \leq n$$
$$D'_v(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}) = a_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n},$$
$$D'_v(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k) = a_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k,$$
$$S_v(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}) = i_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v-1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n}$$
$$S_v(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \partial_k)$$
$$= i_v e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_v^{i_v-1} x_{v+1}^{i_{v+1}} \cdots x_n^{i_n} \partial_k$$

then the maps $\mathbb{F}$-maps $D_{v,w}$, D'_v, and S_v, $1 \leq u, v, w \leq n$, of the algebra $\overline{WN_{n,0,n_{[1]}}}$ (resp. $\overline{WN_{n,n,0_{[1]}}}$) can be linearly extended to non-associative

algebra derivations of $\overline{WN_{n,0,n_{[1]}}}$ (resp. $\overline{WN_{n,n,0_{[1]}}}$) (see [2, 6]).

Note. 2 We know that the derivations $D_{v,w}$ and S_v, $1 \leq u, v, w \leq n$, of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) are inner. We call the derivations D'_v, $1 \leq v \leq n$, of the algebra $\overline{WN_{n,0,n_1}}$ scalar derivations.

Remark 4.1. The algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is Lie admissible, but only the elements $x_u\partial_v$ and ∂_u, $1 \leq u, v \leq n$, of the algebra induce inner derivations of the algebra. The algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is simple, but it has an outer derivation. By Note 2, we know that the inner derivation $D'_u + S'_u$ induced by ∂_u, $1 \leq u \leq n$, is the sum of the scalar derivation D'_u and the degree reducing derivation S_v.

Proposition 4.1. *Every derivation of* $\mathbb{F}[x_1, \cdots, x_n]$ *is of the form* $\sum_{u=1}^{r} f_u$, *for* $f_1, \cdots, f_r \in \mathbb{F}[x_1, \cdots, x_n]$.

Proof. The reader is referred Proposition 1.3 in the text (see [7]). □

Corollary 4.1. *Every derivation of* $\mathbb{F}[x_1^{\pm 1}, \cdots, x_m^{\pm 1}, x_{m+1}, \cdots, x_{m+s}]$ *is of the form* $\sum_{u=1}^{r} f_u$, *for* $f_1, \cdots, f_r \in \mathbb{F}[x_1^{\pm 1}, \cdots, x_m^{\pm 1}, x_{m+1}, \cdots, x_{m+s}]$.

Proof. The proof of the Corollary is straightforward by Proposition 4.1, so it is omitted. □

Note. 3 The restrictions $D''_v := D'_v|_{\mathbb{F}[x_1^{\pm 1}, \cdots, x_n^{\pm 1}]}$ and $S'_v := S_v|_{\mathbb{F}[x_1^{\pm 1}, \cdots, x_n^{\pm 1}]}$ of the derivations D'_v and S_v in Note 1 to the subalgebra $\mathbb{F}[x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$ (resp. $D''_v := D'_v|_{\mathbb{F}[x_1, \cdots, x_n]}$ and $S'_v := S_v|_{\mathbb{F}[x_1, \cdots, x_n]}$) can be a derivation of the $\mathbb{F}$-algebra $\mathbb{F}[x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$ (resp. $\mathbb{F}[x_1, \cdots, x_n]$).

Proposition 4.2. *Every derivation of* $\mathbb{F}[e^{\pm x_1}, \cdots, e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$ *is of the form* $\sum_{p,u}^{r} f_{1,p,u}\partial_u + \sum_{v=1}^{s} f_{2,v}\ D''_v + \sum_{w=1}^{t} f_{3,w}S'_v$, *where* D''_v *and* S'_w *are the derivations in Note 3 for* $f_{1,u,p}, f_{2,1}, \cdots, f_{2,v}, f_{3,1}, \cdots, f_{3,w} \in \mathbb{F}[e^{\pm x_1}, \cdots, e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$.

Proof. The proof of the Proposition is easy, thus it is omitted. □

Corollary 4.2. *If* m *is not equal to* m, *then the* $\mathbb{F}$*-algebras* $\mathbb{F}[e^{\pm x_1}, \cdots, e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$ *and* $\mathbb{F}[e^{\pm x_1}, \cdots, e^{\pm x_m}, x_1^{\pm 1}, \cdots, x_m^{\pm 1}]$ *are not isomorphic.*

Proof. By Proposition 4.2, the algebras in the Proposition have different dimensions of all outer derivations of them, so they are not isomorphic. □

Lemma 4.1. *Only the elements ∂_u, $1 \leq u \leq n$, and $x_v\partial_w$, $1 \leq u, w \leq n$, of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) induce inner derivations of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$).*

Proof. Let D be the derivation in the lemma. Since $\mathbb{F}$ is a field of characteristic zero, we can assume that either $D(1)$ is zero or $D(1)$ is not in the $\mathbb{F}$-algebra $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n]$ (resp. $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$). First let us assume that $D(1)$ is not in the $\mathbb{F}$-algebra $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n]$ (resp. $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$). Note that 1 is the identity of itself. We can put $D(1)$ as follows:

$$D(1) = \sum_{a_1,\cdots,a_n,i_1,\cdots,i_n,1\leq u\leq n} c_{a_1,\cdots,a_n,i_1,\cdots,i_n,1\leq u\leq n} e^{a_1x_1}\cdots e^{a_nx_n}x_1^{i_1}\cdots x_n^{i_n}\partial_u$$

with appropriate coefficients. Without loss of generality, we can assume that a_1 is not zero. Since 1 is in the right annihilator of ∂_1, we have

$$D(\partial_1) * 1 = -\partial_1 * \Big(\sum_{a_1,\cdots,a_n,i_1,\cdots,i_n,1\leq u\leq n} c_{a_1,\cdots,a_n,i_1,\cdots,i_n,1\leq u\leq n} e^{a_1x_1}\cdots e^{a_nx_n}x_1^{i_1}\cdots x_n^{i_n}\partial_u \tag{4}$$

Since a_1 is a non-zero integer, $D(1)$ has no term in the $\mathbb{F}$-algebra $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n]$ (resp. $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$). This implies that the equality (4) does not hold. Thus we can assume that the element $D(1)$ is zero. Since 1 is the right identity of x_u, $1 \leq u \leq n$, $D(x_u)$ is in the $\mathbb{F}$-algebra $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n]$ (resp. $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$). Similarly, we can prove that the $\mathbb{F}$-algebra $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1, \cdots, x_n]$ (resp. $\mathbb{F}[e^{\pm x_1}, \cdots e^{\pm x_n}, x_1^{\pm 1}, \cdots, x_n^{\pm 1}]$) is derivation invariant. We can also prove that $D(N_1^\partial) \subset N_1^\partial$ (resp. $D(N_2^\partial) \subset N_2^\partial$) holds where N_1^∂ (resp. N_2^∂) is in (3). By Proposition 4.2, we know that only the elements ∂_u, $1 \leq u \leq n$, and $x_v\partial_w$, $1 \leq u, w \leq n$, induce inner derivations of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$). Therefore we the Lemma is prove. □

Lemma 4.2. *For any basis elements $x_1^{i_1}\cdots x_n^{i_n}$ and $x_1^{i_1}\cdots x_n^{i_n}\partial_u$, $1 \leq u \leq n$, of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) and for any derivation D of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$), $D(x_1^{i_1}\cdots x_n^{i_n}\partial_u)$ holds in Note 1 or $D(x_1^{i_1}\cdots x_n^{i_n}) = 0$ and $D(x_1^{i_1}\cdots x_n^{i_n}\partial_u) = 0$ i.e., the homogeneous component $N_{(0,\cdots,0)}$ of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is derivation invariant.*

Proof. Let D be the derivation in the Lemma. Let $x_1^{i_1} \cdots x_n^{i_n}$ be a basis element of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$). Let us prove the lemma by induction on the degree of the element $l = x_1^{i_1} \cdots x_n^{i_n}$. If $deg(l)$ is zero, then by Lemma 4.1, it is easy to prove that $D(l)$ and $deg(D(l\partial_u))$, $1 \leq u \leq n$, are zeroes. If $deg(l)$ is one, then it is also easy to prove that either $D(l)$ and $D(l\partial_u)$, $1 \leq u \leq n$, hold in Note 1 such that $deg(D(l)), deg(D(l\partial_u)) \leq 1$, $1 \leq u \leq n$, or $D(l)$ and $D(l\partial_u)$ are zeroes. By induction, we can assume that if $deg(l)$ is p, then we can assume that either $D(l)$ and $D(l\partial_u)$, $1 \leq u \leq n$, hold in Note 1 such that $p-1 \leq deg(D(l)), deg(D(l\partial_u)) \leq p$, or $D(l)$ and $D(l\partial_u)$ are zeroes. Let us assume that the degree of the element l is $p+1$. Since $\sum_{1 \leq u \leq n} x_u \partial_u$ is a right identity of the subalgebra $\overline{WN_{n,0,n}}_{[1]}$ (resp. $\overline{WN_{n,n,0}}_{[1]}$) of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$), we are able to prove that either $D(l)$ and $D(l\partial_u)$, $1 \leq u \leq n$, hold Note 1 such that $p \leq deg(D(l)), deg(D(l\partial_u)) \leq p+1$, or $D(l)$ and $D(l\partial_u)$ are zeroes. Therefore by induction, we can prove the lemma routinely. □

Remark 4.2. By Lemma 4.1 and Theorem 2 (resp. Theorem 2) of the paper [6] (resp. [2]), the restrictions $D_{v,w}|_{\overline{WN_{0,0,n_1}}}$, (resp. $D_{v,w}|_{\overline{WN_{0,n,0_1}}}$), $D'_u|_{\overline{WN_{0,0,n_1}}}$ (resp. $D'_u|_{\overline{WN_{0,n,0_1}}}$), and $S_v|_{\overline{WN_{0,0,n_1}}}$ (resp. $S_v|_{\overline{WN_{0,n,0_1}}}$), $1 \leq u, v, w \leq n$, of $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) of the derivations to the subalgebra $\overline{WN_{0,0,n_1}}$ (resp. $\overline{WN_{0,n,0_1}}$) are derivations of the algebra $\overline{WN_{0,0,n_1}}$ (resp. $\overline{WN_{0,n,0_1}}$). The reader is referred to Lemma 1–Lemma 4 of the paper ([6]) for more details of similar proofs of Lemma 4.1 and Lemma 4.2.

Theorem 4.1. *For any derivation D and of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$), D is the linear sum of the derivations $D_{v,w}$, D'_v, and S_v, $1 \leq u, v, w \leq n$, which are defined in Note 1.*

Proof. Let D be a derivation of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$). We first show that every homogeneous component of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is a derivation invariant. By Lemma 4.1 and Lemma 4.2, the homogeneous component $N_{(0,\cdots,0)}$ is derivation invariant. Since is $\mathbb{Z}^n$-graded and by Remark 4.2, for any element $e^{a_1x_1} \cdots e^{a_nx_n}$ of the algebra, we can prove that either $D(e^{a_1x_1} \cdots e^{a_nx_n})$ and $D(e^{a_1x_1} \cdots e^{a_nx_n}\partial_u)$, $1 \leq u \leq n$, are in the homogeneous component $N_{(a_1,\cdots,a_n)}$ or $D(e^{a_1x_1} \cdots e^{a_nx_n}) = D(e^{a_1x_1} \cdots e^{a_nx_n}\partial_u) = 0$. By Remark 4.2 and Lemma 4.1, we can easily prove that either $D(l)$ holds Note 1 or $D(l) = 0$. This implies that $D(N_{(a_1,\cdots,a_n)}) \subset N_{(a_1,\cdots,a_n)}$. For any element $l = e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n} \in N_{(a_1,\cdots,a_n)}$, if $deg(l)$ is zero, then we can also prove that either $D(l) =$

$0 = D(l\partial_u)$, $1 \leq u\partial_u$, or $D(l)$ and $D(l\partial_u)$ hold in Note 1 such that $deg(D(l)), deg(D(l\partial_u)) \leq 1$. By induction on the degree $deg(l)$ of the element l, we can assume that if $deg(l)$ is p, then either $D(l)$ and $D(l\partial_u)$, $1 \leq u \leq n$, hold Note 1 such that $p \leq deg(D(l)), deg(D(l\partial_u)) \leq p+1$, or $D(l) = D(l\partial_u) = 0$. If we assume that $deg(l)$ is $p+1$, then we are able to prove that either $D(l)$ and $D(l\partial_u)$, $1 \leq u \leq n$, hold Note 1 such that $p+1 \leq deg(D(l)), deg(D(l\partial_u)) \leq p+2$, or $D(l) = 0 = D(l\partial_u)$ holds easily. Therefore, the Theorem is proved. □

Corollary 4.3. *For any derivation D and of the algebra $\overline{WN_{n,n,0_1}}$, D is the linear sum of the derivations $D_{v,w}$, D'_v, and S_v, $1 \leq v, w \leq n$, in Note 1.*

Proof. Let D be the derivation in the corollary. For any basis element $e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}\partial_k$ of the algebra $\overline{WN_{n,0,n_1}}$, if $deg(e^{a_1x_1} \cdots e^{a_nx_n} x_1^{i_1} \cdots x_n^{i_n}\partial_k) = p$, then by Theorem 4.1 either $p \leq deg(D(e^{a_1x_1} \cdots e^{a_nx_n}\ x_1^{i_1} \cdots x_n^{i_n}\partial_k)) \leq p+1$ holds or $D(l) = 0$ holds. Either of these two cases, we can prove that D is the linear sum of the derivations $D_{v,w}$, D'_u, and S_v, $1 \leq u, v, w \leq n$, in Note 1 holds routinely (see [8]). This completes the proof the corollary. □

Corollary 4.4. *The associator of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is the matrix ring $M_n(\mathbb{F})$.*

Proof. The proof of the corollary is straightforward by Note 1 and Theorem 4.1. This completes the proof of the Corollary. □

Corollary 4.5. *The dimension of $Der(\overline{WN_{n,0,n_1}})$ (resp. $Der(\overline{WN_{n,n,0_1}})$) of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) is $n^2 + 2n$.*

Proof. The proof of the corollary is straightforward by Note 1 and Theorem 4.1. This completes the proof of the Corollary. □

Corollary 4.6. *The matrix ring $M_n(\mathbb{F})$ is a subalgebra of $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$). For a derivation D of $M_n(\mathbb{F})$, D can be extended to a derivation of the algebra $\overline{WN_{n,0,n_1}}$ (resp. $\overline{WN_{n,n,0_1}}$) in Note 1 (see [1, 6]).*

Proof. The proof of the corollary is straightforward by Note 1 and Theorem 4.1. This completes the proof of the corollary. □

Proposition 4.3. *The algebras* $\overline{WN_{1,0,1_1}}$ *(resp.* $\overline{WN_{1,1,0_1}}$*) and* $\overline{WN_{0,0,1_1}}$ *(resp.* $\overline{WN_{0,1,0_1}}$*) are not isomorphic.*

Proof. Let θ be isomorphism from the algebra $\overline{WN_{1,0,1_1}}$ (resp. $\overline{WN_{1,1,0_1}}$) to the algebra $\overline{WN_{0,0,1_1}}$ (resp. $\overline{WN_{0,1,0_1}}$. Since $x\partial$ is a right identity of $\overline{WN_{1,0,1_1}}$ (resp. $\overline{WN_{1,1,0_1}}$), we have that $\theta(x\partial) = x\partial + c\partial$ and $\theta(\partial) = c_0\partial$ where $c \in \mathbb{F}$ and $c_0 \in \mathbb{F}^\bullet$. Since ∂ is a left identity of $e^x\partial$, we have that

$$c_0\partial * \theta(e^x\partial) = \theta(e^x\partial) \tag{5}$$

There is no element $\theta(e^x\partial)$ in $\overline{WN_{0,0,1_1}}$ (resp. $\overline{WN_{0,1,0_1}}$). This contradiction shows that there is no isomorphism from the algebra $\overline{WN_{1,0,1_1}}$ (resp. $\overline{WN_{1,1,0_1}}$) to the algebra $\overline{WN_{0,0,1_1}}$ (resp. $\overline{WN_{0,1,0_1}}$). Therefore the Proposition is proved. □

Theorem 4.2. *The algebras* $\overline{WN_{n,0,n_1}}$ *and* $\overline{WN_{0,0,n_1}}$ *(resp.* $\overline{WN_{0,n,0_1}}$*) are not isomorphic.*

Proof. Since $Dim(Der(\overline{WN_{n,0,n_1}}))$ is $n^2 + 2n$ and $Dim(Der(\overline{WN_{0,0,n_1}}))$ (resp. $\overline{WN_{0,n,0_1}}$) is $n^2 + n$, the algebras of the theorem are not isomorphic (see [6]). Therefore the Theorem is proved. □

Corollary 4.7. *If* m *is greater than* n*, then* $Hom(\overline{WN_{0,0,m_1}}, \overline{WN_{0,n,0_1}}) = \{0\}$ *where* 0 *is the zero homomorphism.*

Proof. Since the algebras $\overline{WN_{0,0,m_1}}$ and $\overline{WN_{0,n,0_1}}$ are simple and they have right identities, if there is a non-zero homomorphism from the algebra $\overline{WN_{0,0,m_1}}$ to the algebra $\overline{WN_{0,n,0_1}}$, then we can derive a contradiction easily. This completes the proof of the corollary. □

Theorem 4.3. *If* m *is not equal to* n*, then the algebras* $\overline{WN_{m,0,m_1}}$ *and* $\overline{WN_{n,0,n_1}}$ *(resp.* $\overline{WN_{m,0,m_1}}$ *and* $\overline{WN_{n,n,0_1}}$*) are not isomorphic.*

Proof. Since m is not equal to n, we have $Dim(Der(\overline{WN_{m,0,m_1}})) \neq Dim(Der(\overline{WN_{n,0,n_1}}))$ (resp.$Dim(Der\ (\overline{WN_{m,0,m_1}})) \neq Dim\ (Der\ (\overline{WN_{n,n,0_1}}))$) (see [6]). This implies that the appropriate algebras in the theorem are not isomorphic. Therefore the theorem is proved. □

Corollary 4.8. *If* m *is not equal to* n*, then the algebras* $\overline{WN_{m,0,m_1}}$ *and* $\overline{WN_{n,n,0_1}}$ *are not isomorphic.*

Proof. *By the assumptions of the Corollary and Theorem 4.3, the proof of the corollary is obvious, so it is omitted.* □

Corollary 4.9. *If m is greater than n, then $Hom(\overline{WN_{m,0,m_1}}, \overline{WN_{n,n,0_1}}) = \{0\}$ where 0 is the zero homomorphism.*

Proof. *Since the algebras $\overline{WN_{m,0,m_1}}$ and $\overline{WN_{n,n,0_1}}$ are simple, the results of the corollary is straightforward by Theorem 4.3 and Corollary 4.6.* □

Note that if m is not equal to n, then it is an interesting problem to find $Hom(\overline{WN_{0,0,m_1}}, \overline{WN_{0,n,0_1}})$ and $Hom(\overline{WN_{m,0,m_1}}, \overline{WN_{n,n,0_1}})$.

References

1. Mohammad H. Ahmadi, Ki-Bong Nam, and Jonathan Pakianathan, *Lie admissible non-associative algebras*, Algebra Colloquium,Vol. 12, No. 1, World Scientific, (March) 2005, 113-120.
2. Choi, Seul Hee, *Derivations of a restricted Weyl type algebra on a Laurent extension*, Comm. Korean Math. Soc. 21 (2006), No. 2, 227–236.
3. Choi, Seul Hee, *Derivations of a Weyl type non-associative algebra on Laurent extension*, Bull. Korean Math. Soc. 43 (2006), 627–634.
4. Seul Hee Choi and Ki-Bong Nam, *The Derivation of a Restricted Weyl Type Non-Associative Algebra*, Vol. 28, No. 3, Hadronic Journal, 2005, 287-295.
5. Seul Hee Choi and Ki-Bong Nam, *Derivation of Symmetric Non-associative Algebra I*, Algebras, Groups and Geometries, Volume 22, Number 3, Sep. 2005, 341-352.
6. Seul Hee Choi and Ki-Bong Nam, *Derivations of a restricted Weyl Type Algebra I*, Rocky Mountain Journal of Mathematics, Volume 37, Number 6, 2007, 1813-1830.
7. S. C. Coutinho, *A Primer of Algebraic D-modules*, Cambridge University Press, 1995.
8. J. Dixmier, *ENVELOPING ALGEBRAS*, North-Holland Pub. Com., New York, 1977.
9. I. N. Herstein, *Noncommutative Rings*, Carus Mathematical Monographs, Mathematical association of America.
10. J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1987.
11. T. Ikeda, N. Kawamoto and Ki-Bong Nam, *A class of simple subalgebras of Generalized W algebras*, Proceedings of the International Conference in 1998 at Pusan (Eds. A. C. Kim), Walter de Gruyter Gmbh Co. KG, 2000, 189-202.
12. V. G. Kac, *Description of Filtered Lie Algebra with which Graded Lie algebras of Cartan type are Associated*, Izv. Akad. Nauk SSSR, Ser. Mat. Tom, **38**, 1974, 832-834.
13. Naoki Kawamoto, Atsushi Mitsukawa, Ki-Bong Nam, and Moon-Ok Wang, *The automorphisms of generalized Witt type Lie algebras*, Journal of Lie Theory, 13 Vol. (2), Heldermann Verlag, 2003, 571-576.
14. Ki-Bong Nam, *Generalized W and H Type Lie Algebras*, Algebra Colloquium **6**:3, (1999), 329-340.

15. Ki-Bong Nam, *On Some Non-Associative Algebras Using Additive Groups*, Southeast Asian Bulletin of Mathematics, Vol. 27, Springer Verlag, 2003, 493-500.
16. Ki-Bong Nam, Yanggon Kim and Moon-Ok Wang, *Weyl-type Non-Associative Algebras I*, IMCC Proceedings, 2004, SAS Publishers, 147-155.
17. Ki-Bong Nam and Seul Hee Choi, *On the Derivations of Non-Associative Weyl-type Algebras*, Southeast Asian Bull. Math. (2007), 31: 341-348, 2007.
18. J. Marshall Osborn, *Derivations and isomorphisms of Lie algebras of characteristic* 0, Modular interfaces (Riverside, CA, 1995), 95–108, AMS/IP Stud. Adv. Math., 4, Amer. Math. Soc., Providence, RI, 1997.
19. A. N. Rudakov, *Groups of Automorphisms of Infinite-Dimensional Simple Lie Algebras*, Math. USSR-Izvestija, **3**, 1969, 707-722.
20. R. D. Schafer, *Introduction to nonassociative algebras*, Dover, 1995, 128-138.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 69–82)

SEMIGROUP PROPERTIES OF COOPERATIONS ON FINITE SETS*

K. DENECKE and K. SAENGSURA

Institute of Mathematics, Universität Potsdam,
Am Neuen Palais, 14415 Potsdam, Germany
E-mail: kdenecke@rz.uni-potsdam.de

Let $A^{\sqcup n} := \{1, \ldots, n\} \times A$ be the n-th copower of the set A. An n-ary cooperation is a mapping $f : A \to A^{\sqcup n}$. If f is an n-ary cooperation and if $g_1, \ldots, g_n$ are k-ary cooperations on A, we may define a composition $comp_k^n(f, g_1, \ldots, g_n)$ and obtain a new k-ary cooperation defined on A. Together with the injection cooperations $i_i^{n,A} : A \to A^{\sqcup n}$ defined by $a \mapsto (i, a)$ for $1 \leq i \leq n$, all cooperations defined on A form a multibased algebra. This algebra is called the clone of all cooperations defined on A. If we define by $f + g := comp_n^n(f, g, \ldots, g)$ a binary operation on the set of all n-ary cooperations, the set of all n-ary cooperations defined on A forms a semigroup with respect to this operation. We determine the order of the elements of this semigroup, characterize all idempotent and regular elements, ask for bands of n-ary cooperations and characterize Green's relations $\mathcal{L}$ and $\mathcal{R}$.

Keywords: Cooperations; Regular elements; Green's relations.

1. Introduction

Let A be a non-empty set. For each $n \geq 1$ we denote the n-th copower of A, that is the union of n disjoint copies of A by $A^{\sqcup n}$. Specifically, $A^{\sqcup n} := \{1, \ldots, n\} \times A$, and an element (i, a) corresponds to the element a in the i-th copy of A. N-ary cooperations on A are mappings $f : A \to A^{\sqcup n}$ and are uniquely determined by a pair (f_1, f_2) of mappings $f_1 : A \to \{1, \ldots, n\}$ and $f_2 : A \to A$.

Let cO_A^n be the set of all n-ary cooperations defined on A and let $cO_A := \bigcup_{n \geq 1} cO_A^n$ be the set of all cooperations defined on A. If $f \in cO_A^n$ and $g_1, \ldots, g_n \in cO_A^k$, then we define a k-ary cooperation $f[g_1, \ldots, g_n]$:

*December 26,2007
AMS Subject Classification: 68Q55, 08A46.

$A \to A^{\sqcup k}$ by $a \mapsto (g_{f_1(a)})(f_2(a))$ for all $a \in A$ (see [1]). The cooperation $f[g_1, \ldots, g_n]$ is called the composition of f and $g_1, \ldots, g_n$. Instead of $f[g_1, \ldots, g_n]$ we will also write $comp_k^n(f, g_1, \ldots, g_n)$. The injections $i_i^{n,A}$ are special cooperations which are defined by $i_i^{n,A} : A \to A^{\sqcup n}$ with $a \mapsto (i, a)$ for $1 \leq i \leq n$. Then we get a multi-based algebra $((cO_A^n)_{n\geq 1}; (comp_k^n)_{k,n\geq 1}, (i_i^{n,A})_{1\leq i\leq n})$. This multi-based algebra is an abstract clone, i.e., satisfies the clone axioms $(C1)$, $(C2)$, $(C3)$ (see [3]) with

$(C1)$ $S_m^p(z, S_m^n(y_1, x_1, \ldots, x_n), \ldots, S_m^n(y_p, x_1, \ldots, x_n))$
$\approx S_n^n(S_n^p(z, y_1, \ldots, y_p), x_1, \ldots, x_n)$, for $m, n, p = 1, 2, 3, \ldots$,
$(C2)$ $S_m^n(\lambda_j, x_1, \ldots, x_n) \approx x_j$, for $1 \leq j \leq n$ and $m, n = 1, 2, 3, \ldots$,
$(C3)$ $S_m^m(x_j, \lambda_1, \ldots, \lambda_m) \approx x_j$, for $1 \leq j \leq m$ and $m = 1, 2, 3, \ldots$,

where $S_m^p, S_m^n, S_n^p, S_m^m$ are operation symbols corresponding to the operations $comp_m^p, comp_m^n, comp_n^p$ and $comp_m^m$, respectively; $\lambda_1, \ldots, \lambda_m$ are nullary operation symbols, and $z, y_1, \ldots, y_p, x_1, \ldots, x_n$ are variables.
Using the operation $comp_n^n$ we define a binary operation $+$ on $cO_A^n, n \geq 1$ and by $(C1)$ we obtain the semigroup $(cO_A^n; +)$ of all n-ary cooperations on A. We will study this semigroup and its subsemigroups.

We recall the following notions from semigroup theory. Let $\underline{S} = (S; \circ)$ be a semigroup. Then an element $a \in S$ is called *regular* if there is an element $b \in S$ such that $aba = a$. Clearly every idempotent element is regular and more general, every element $a \in S$ satisfying $a^n = a$, $n \geq 2$ is regular. If every element of a semigroup is regular, then the semigroup is called *regular*. A regular semigroup $\underline{S}$ is called *orthodox* if its set $E(S)$ of idempotents forms a subsemigroup. *Bands* are idempotent semigroups. A semigroup $(S; \circ)$ is called a *rectangular band* if it is idempotent and if $xyx = x$ for all $x, y \in S$, a *normal band*, if it is idempotent and $xyuv = xuyv$ for all $x, y, u, v \in S$. Semigroups satisfying $xy = x$ for all $x \in S$ are called *left-zero semigroups*. If $xy = y$ for all $x, y \in S$ we speak of a *right-zero semigroup* and if $xy = uv$ for all $x, y, u, v \in S$, the semigroup S is called a *constant semigroup*.

Green's relations $\mathcal{L}$ and $\mathcal{R}$ are defined by

$$a\mathcal{L}b :\Leftrightarrow a = b \text{ or } \exists\, c, d \in S(ca = b \wedge db = a),$$

$$a\mathcal{R}b :\Leftrightarrow a = b \text{ or } \exists\, e, f \in S(ae = b \wedge bf = a).$$

Let $a \in S$ be an element of the universe of the semigroup $\underline{S} = (S; \circ)$. The order of a is the cardinality of the cyclic subsemigroup $\langle a \rangle$ of $\underline{S}$: $o(a) = |\langle a \rangle|$. If S is finite, then the set $\{i \in \mathbb{N} \setminus \{0\} | \exists j \in \mathbb{N}(a^i = a^j, i \neq j)\}$ is non-empty and has a least element m which is called the *index* of a. Then the set

$\{x \in \mathbb{N} \setminus \{0\} | a^{m+x} = a^m\}$ is non-empty, and so it also has a least element r which is called the *period* of a. By the minimality of m and r we deduce that $\langle a \rangle = \{a, a^2, \ldots, a^{m+r-1}\}$ and thus $o(a) = m + r - 1$. Further, for all $u, v \in \mathbb{N} \setminus \{0\}$ we have $a^{m+u} = a^{m+v}$ iff $u \equiv v \bmod r$.

If $g : A \to A$ is a transformation on the finite set A with $|A| > 1$, then g is an element of the semigroup $(H_A; \circ)$ of all transformations defined on A. If $Im(g) = \{b | \exists a \in A(b = g(a))\}$ is the image of g, then the index of g is the least non-negative integer $\lambda(g)$ such that $Im(g^{\lambda(g)}) = Im(g^{\lambda(g)+1})$ and the period of g is the order of the restriction function $h := g|Im(g^{\lambda(g)})$ (which is a permutation).

The kernel of a mapping $f : A \to A$ is the equivalence relation $Kerf := \{(a, b) \mid f(a) = f(b), a, b \in A\}$. For more background on semigroups see [4].

2. Idempotent and Regular Elements

We consider the semigroup $(cO_A^n; +)$ of all n-ary cooperations defined on the finite set A and its subsemigroups. First of all we are interested in idempotent and regular elements of $(cO_A^n; +)$.

Theorem 2.1.

(i) The n-ary cooperation $f \in cO_A^n$ is an idempotent element of $(cO_A^n; +)$ iff $a \in f^{-1}((i, a))$ for all $(i, a) \in Im(f)$.

(ii) f is a regular element of $(cO_A^n; +)$ iff for any two elements $(i, a), (j, b)$ of $Im(f)$ there follows $a = b \Rightarrow i = j$.

Proof.

(i) Let $(i, a) \in Im(f)$. Then there is an element $x \in A$ such that $f(x) = (i, a)$. Since $(f + f)(x) = f(x)$ and $(f + f)(x) = f(f_2(x)) = f(a)$ we get $f(x) = f(a) = (i, a)$ and $a \in f^{-1}((i, a))$.
Conversely, we assume that $a \in f^{-1}((i, a))$ for all $(i, a) \in Im(f)$. Let $x \in A$. Then $f(x) = (f_1(x), f_2(x)) \in Im(f)$ and by assumption $f_2(x) \in f^{-1}(f_1(x), f_2(x))$ and then $(f + f)(x) = f(f_2(x)) = (f_1(x), f_2(x)) = f(x)$ and f is idempotent.

(ii) If f is a regular element, then there exists a cooperation $g \in cO_A^n$ such that $f + g + f = f$. Let $(i, a), (j, b) \in Im(f)$ such that $a = b$. Since $(i, a), (j, b) \in Im(f)$, there are elements $x, y \in A$ such that $f(x) = (i, a)$ and $f(y) = (j, b)$ (i.e., $f_1(x) = i$, $f_2(x) = a$, $f_1(y) = j$, $f_2(y) = b$) and $f_2(x) = f_2(y)$. Moreover, we use that $(f+g)(x) = g(f_2(x))$ by definition

of the composition. Then we have

$$\begin{aligned} f_1(x) &= (f+g+f)_1(x) \\ &= f_1(g_2(f_2(x))) \\ &= f_1(g_2(f_2(y))) \\ &= (f+g+f)_1(y) \\ &= f_1(y). \end{aligned}$$

Conversely, assume that for any $(i,a),(j,b) \in Im(f)$ we have $a = b \Rightarrow i = j$. If $(i,a) \in Im(f)$, then there is an element $d_a \in f^{-1}((i,a))$. We define a cooperation $g : A \to A^{\sqcup n}$ as follows

$$g(x) = \begin{cases} (i,d_x) & if \quad x \in \{a|(i,a) \in Im(f)\} \\ (j,x) & if \quad x \notin \{a|(i,a) \in Im(f)\} \ for \ some \ j \in \{1,\ldots,n\}. \end{cases}$$

Let now $x \in A$, then $x \in f^{-1}((l,b))$ for some $(l,b) \in Im(f)$, i.e.

$$f(x) = (l,b), f_1(x) = l, f_2(x) = b.$$

Then we have

$$\begin{aligned} (f+g+f)(x) &= ((f+g)+f)(x) \\ &= f((f+g)_2(x)) \\ &= f(g_2(f_2(x))) \\ &= f(g_2(b)) \\ &= f(d_b) \\ &= (l,b) \\ &= f(x). \end{aligned}$$

This shows that f is regular. □

Let $E(cO_A^n)$ and $Reg(cO_A^n)$ be the sets of idempotent and regular elements, respectively. In [2] was proved that for $A = \{0,1\}$ the regular elements of $(cO_{\{0,1\}}^n; +)$ form a subsemigroup and the idempotent elements form a subsemigroup of the semigroup of all regular elements, i.e., the set of all regular elements forms an orthodox semigroup. In the general situation, i.e., if $|A| > 2$, this is not the case as the following example shows:
Let $A = \{a,b,c\}$ and let $f,g \in cO_A^n$ be defined by the following tables

	f	g
a	(i,a)	(k,a)
b	(i,a)	(l,b)
c	(j,c)	(l,b)

for some $i,j,k,l \in \{1,\ldots,n\}$.

Since $a \in f^{-1}((i,a))$, $c \in f^{-1}((j,c))$, $a \in g^{-1}((k,a))$, $b \in g^{-1}((l,b))$, the cooperations f and g are idempotent. Since

$$(f+g)(c) = g(f_2(c)) = g(c) = (l,b)$$

and

$$(f+g)(b) = g(f_2(b)) = g(a) = (k,a)$$

we have $b \notin (f+g)^{-1}(l,b)$ and therefore $f+g$ is not idempotent. But the sum of two regular elements is regular since if f and g are regular and if $(i,a),(j,b) \in Im(f+g)$, then because of $Im(f+g) \subseteq Im(g)$ from $a=b$ we obtain $i=j$. This shows that $f+g$ is regular.

Any cooperation f on the set A is uniquely determined by the pair (f_1,f_2) consisting of the mapping $f_1 : A \to \{1,\ldots,n\}$ and the transformation $f_2 : A \to A$. Because of $(if)(a) = f((f_2)^{i-1}(a))$ for any $i \geq 2$, $i \in \mathbb{N}$, any $a \in A$ and any cooperation $f \in cO_A^n$, the order of f is determined by the order of f_2. Let $\lambda(f_2)$ be the index and let r be the period of f_2. Then

$$f_2^{\lambda(f_2)}(a) = f_2^{\lambda(f_2)+r}(a)$$

for all $a \in A$ and then

$$o(f_2) = \lambda(f_2) + r - 1$$

and we have

$$\begin{aligned}((\lambda(f_2)+r+1)f)(a) &= f(f_2^{\lambda(f_2)+r}(a)) \\ &= f(f_2^{\lambda(f_2)}(a)) \\ &= ((\lambda(f_2)+1)f)(a).\end{aligned}$$

This means, $o(f) \leq \lambda(f_2)+r$ and $o(f) \leq o(f_2)+1$. For the index m of the cooperation f, we have $m \leq \lambda(f_2)+1$. If f is regular, then for any $s,t \in \mathbb{N}^+$ we have $f_2^s = f_2^t$ iff $sf = tf$. Since for each $a \in A$,

$$\begin{aligned}f_2^s(a) = f_2^t(a) &\Leftrightarrow (f_1(f_2^{s-1}(a)), f_2^s(a)) = (f_1(f_2^{t-1}(a)), f_2^t(a)) \\ &\Leftrightarrow (sf)(a) = (tf)(a).\end{aligned}$$

Therefore in this case $o(f) = o(f_2)$ and $m = \lambda(f_2)$. To characterize the order of $f \in cO_A^n$ such that f is not regular we need the following lemma.

Lemma 2.1. *Let $|A| = m$. Then for every $f \in cO_A^n$ the cooperation mf is regular.*

Proof. Suppose that mf is not regular.
Claim: If kf is regular, then $(k+1)f$ is regular.
Assume that kf is regular. Since $(k+1)f = f+kf$, we have $Im(f+kf) \subseteq Im(kf)$. By the characterization of regular elements, $(k+1)f$ must be regular. Since mf is not regular and by using the contraposition of our claim we get that $(m-1)f$ is not regular. By using the same argument we get that

$(m-2)f, \ldots, f$ are not regular. Then $mf, (m-1)f, \ldots, f$are not regular. Since $(m-1)f$ is not regular, then there are $(i,c), (j,c) \in Im((m-1)f)$ such that $i \neq j \in \{1, \ldots, n\}$. Let $a, b \in A$ such that $((m-1)f)(a) = (i,c)$ and $((m-1)f)(b) = (j,c)$. Then

$$\begin{aligned}(mf)(a) &= f(((m-1)f)_2(a)) \\ &= f(c) \\ &= f(((m-1)f)_2(b)) \\ &= (mf)(b).\end{aligned}$$

We consider the mapping $\varphi : Im((m-1)f) \to Im(mf)$ defined by $\varphi((m-1)f)(x)) = (mf)(x)$. As we have seen, the surjective mapping φ maps the different elements $(i,c), (j,c) \in Im((m-1)f)$ to $f(c) \in Im(mf)$. Therefore together with $Im(mf) \subseteq Im((m-1)f)$ we get $Im(mf) \subset Im((m-1)f)$. Similarly, one can show that $Im((m-1)f) \subset \ldots \subset Im(f)$. Then there follows that $|Im(mf)| < \ldots < |Im(f)| = m$. This implies that $|Im(mf)| = 1$. Therefore mf is regular, a contradiction. □

From Lemma 2.2, we have $\{k \in \mathbb{N}^+ | kf \textit{ is regular } \}$ is non-empty.

Definition 2.1. Let $f \in cO_A^n$ such that f is not regular. Let $\beta(f)$ be the least positive integer such that $\beta(f)f$ is regular.

For every cooperation $f \in cO_A^n$ we have $\beta(f) \leq \lambda(f_2) + 1$.

Corollary 2.1. *Let $f \in cO_A^n$ such that f is not regular. Then*

(i) If $\beta(f) \leq \lambda(f_2)$, then $o(f) = o(f_2)$.
(ii) If $\beta(f) = \lambda(f_2) + 1$, then $o(f) = o(f_2) + 1$.

Proof.

(i) Assume that $\beta(f) \leq \lambda(f_2)$. This implies that $\lambda(f_2)f$ is regular. Then it follows that for any $s, t \in \mathbb{N}$, we have $f_2^{\lambda(f_2)+s} = f_2^{\lambda(f_2)+t}$ iff $(\lambda(f_2)+s)f = (\lambda(f_2)+t)f$. Therefore $o(f) = o(f_2)$.

(ii) Assume that $\beta(f) = \lambda(f_2) + 1$. Then $(\lambda(f_2)+1)f$ is regular and by the minimality of $\beta(f)$ we get that $(\lambda(f_2)f)$ is not regular. Let r be the period of f_2. This implies that $f_2^{\lambda(f_2)+r} = f_2^{\lambda(f_2)}$, but $(\lambda(f_2)+r)f \neq \lambda(f_2)f$. Therefore $o(f) = \lambda(f_2) + r = o(f_2) + 1$. □

For unary cooperations, we have $(cO_A^1; +) \cong (H_A; \circ)$ and therefore by the theorem of Cayley every abstract semigroup is isomorphic to a semigroup of cooperations.

3. Bands of n-ary Cooperations

After having determined all idempotent elements of $(cO_A^n; +)$ we ask for subsemigroups which consist only of idempotent elements, i.e., for bands of n-ary cooperations.

First of all, we are interested in right-zero semigroups and in left-zero semigroups. If $f + g = g$, then $(f + g)(x) = g(f_2(x)) = g(x)$ for all $x \in A$. From this it follows that $Ker(f) = Ker(g)$. Indeed, we have

$$\begin{aligned}(x, y) \in Ker(f) &\Rightarrow f(x) = f(y)\\ &\Rightarrow f_2(x) = f_2(y)\\ &\Rightarrow g(f_2(x)) = g(f_2(y))\end{aligned}$$

and therefore $g(x) = g(y)$ and then $(x, y) \in Ker(g)$ and $Ker(f) \subseteq Ker(g)$. Similarly, from $g + f = f$ there follows $Ker(g) \subseteq Ker(f)$. If $f + g = f$, then clearly we have $Im(f) \subseteq Im(g)$ and $g + f = g$ implies $Im(g) \subseteq Im(f)$. Altogether we have:

Proposition 3.1. *Let $S \subseteq E(cO_A^n)$. Then*

(i) $\underline{S}$ is a right-zero semigroup iff for all $f, g \in S$ we have $Ker(f) = Ker(g)$,
(ii) $\underline{S}$ is a left-zero semigroup iff for all $f, g \in S$ we have $Im(f) = Im(g)$.

Proof.

(i) By the previous remark, if $\underline{S}$ is a right-zero semigroup, then $Ker(f) = Ker(g)$. Let conversely $f, g \in S$ such that $Ker(f) = Ker(g)$. Since f is idempotent, we have $f(a) = (f + f)(a) = f(f_2(a))$ and thus $(a, f_2(a)) \in Ker(f) = Ker(g)$. Therefore $g(a) = g(f_2(a)) = (f + g)(a)$ for all $a \in A$ and thus $f + g = g$. In a similar way we prove $g + f = f$. Clearly, S is closed with respect to $+$ and therefore it is a right-zero semigroup.

(ii) If $\underline{S}$ is a left-zero semigroup, then for all $f, g \in S$ we have $Im(f) = Im(g)$. Conversely, assume that $Im(f) = Im(g)$ for all $f, g \in S$. Let $a \in A$. Since $f(a) = (f_1(a), f_2(a)) \in Im(f)$, we have also $(f_1(a), f_2(a)) \in Im(g)$. Since g is idempotent, by Theorem 2.1, we have $f_2(a) \in g^{-1}(f_1(a), f_2(a))$ and therefore $(f + g)(a) = g(f_2(a)) = (f_1(a), f_2(a)) = f(a)$ and this shows $f + g = f$. In a similar way we show $g + f = g$. Hence, $\underline{S}$ is a left-zero semigroup. □

Our next aim is to characterize rectangular bands.

Lemma 3.1. *Let $f, g \in cO_A^n$. Then $f + g + f = f$ iff $g_2(a) \in f_2^{-1}(a)$ for all $a \in Im(f_2)$.*

Proof. Assume that $f+g+f=f$ and $a \in Im(f_2)$. Then there are elements $x \in A$ and $i \in \{1,\ldots,n\}$ such that $f(x)=(i,a)$. Then we also have $(f+g+f)(x)=(i,a)$ and $(f+(g+f))(x)=(i,a)$

$$\begin{array}{ll} \Rightarrow & (g+f)(f_2(x))=(i,a) \\ \Rightarrow & (g+f)(a)=(i,a) \\ \Rightarrow & f(g_2(a))=(i,a) \\ \Rightarrow & f_2(g_2(a))=a, \end{array}$$

and therefore $g_2(a) \in f_2^{-1}(a)$.
If conversely $g_2(a) \in f_2^{-1}(a)$ for all $a \in Im(f_2)$, then there is an element $j \in \{1,\ldots,n\}$ and an element $y \in A$ such that $f(y)=(j,a)$ and then

$$\begin{array}{rcl} (f+g+f)(y) & = & (f+(g+f))(y) \\ & = & (g+f)(f_2(y)) \\ & = & (g+f)(a) \\ & = & f(g_2(a)) \\ & = & (j,a) \\ & = & f(y). \end{array}$$

This shows that $f+g+f=f$. □

The set $Rec := \bigcup_{f,g \in cO_A^n} \{\{f,g\} \mid f+g+f=f \text{ and } g+f+g=g\}$ is in general not a semigroup as the following counter-example shows. We consider the six element set $A=\{a,b,c,d,e,f\}$ and cooperations $\alpha,\beta,\gamma \in cO_A^n$ defined by

$$\alpha(\{b,c,d\})=\{(i,c)\}, \alpha(\{a,e,f\})=\{(j,f)\};$$

$$\beta(\{a,b,f\})=\{(k,a)\}, \beta(\{c,d,e\})=\{(l,d)\};$$

$$\gamma(\{a,b,c\})=\{(m,b)\}, \gamma(\{d,e,f\})=\{(o,e)\}$$

for some $i,j,k,l,m,o \in \{1,\ldots,n\}$.
Since

$$c \in \alpha^{-1}(i,c), f \in \alpha^{-1}(j,f), a \in \beta^{-1}(k,a),$$

$$d \in \beta^{-1}(l,d), b \in \gamma^{-1}(m,b), e \in \gamma^{-1}(o,e),$$

the cooperations α,β,γ are idempotent. Since

$$\beta_2(c)=d \in \alpha_2^{-1}(c), \beta_2(f)=a \in \alpha_2^{-1}(f), \gamma_2(c)=b \in \alpha_2^{-1}(c)$$

and

$$\gamma_2(f) = e \in \alpha_2^{-1}(f)$$

we have $\alpha + \beta + \alpha = \alpha$ and $\alpha + \gamma + \alpha = \alpha$. For $a, d \in Im(\beta_2)$, we have

$$\alpha_2(a) = f \in \beta_2^{-1}(a), \alpha_2(d) = c \in \beta_2^{-1}(d),$$

$$\gamma_2(a) = b \in \beta_2^{-1}(a), \gamma_2(d) = e \in \beta_2^{-1}(d).$$

This shows

$$\beta + \alpha + \beta = \beta \text{ and } \beta + \gamma + \beta = \beta.$$

For $b, e \in Im(\gamma_2)$, we have

$$\alpha_2(b) = c \in \gamma_2^{-1}(b), \alpha_2(e) = f \in \gamma_2^{-1}(e), \beta_2(b) = a \in \gamma_2^{-1}(b)$$

and

$$\beta_2(e) = d \in \gamma_2^{-1}(e).$$

This shows

$$\gamma + \alpha + \gamma = \gamma \text{and} \gamma + \beta + \gamma = \gamma.$$

Since $\alpha + \beta : A \to A^{\sqcup n}$ is an n-ary cooperation with

$$(\alpha + \beta)(\{b, c, d\}) = \{(l, d)\}$$

and

$$(\alpha + \beta)(\{a, e, f\}) = \{(k, a)\},$$

for $a, d \in Im((\alpha + \beta)_2)$, we have

$$\gamma_2(a) = b \notin (\alpha + \beta)_2^{-1}(a) \text{ and } \gamma_2(d) = e \notin (\alpha + \beta)_2^{-1}(d).$$

Therefore

$$(\alpha + \beta) + \gamma + (\alpha + \beta) \neq \alpha + \beta \text{ and } \alpha + \beta \notin Rec.$$

Then we obtain the following characterization of rectangular bands.

Theorem 3.1. *Let $\underline{S} \subseteq E(cO_A^n)$ be a subsemigroup. Then $\underline{S}$ is a rectangular band iff for all $f, g \in S$ we have $g_2(a) \in f_2^{-1}(a)$ for all $a \in Im(f_2)$ and $f_2(b) \in g_2^{-1}(b)$ for all $b \in Im(g_2)$.*

Proof. If $\underline{S}$ is a rectangular band, then for any $f, g \in S$ we have $f+g+f = f$ and $g+f+g = g$ and by Lemma 3.2, we obtain $g_2(a) \in f_2^{-1}(a)$ for all $a \in Im(f_2)$ and $f_2(b) \in g_2^{-1}(b)$ for all $b \in Im(g_2)$.

If conversely these conditions are satisfied, then for any two elements $f, g \in S$ we have $f+g+f = f$ and $g+f+g = g$. □

Now we characterize normal bands.

Lemma 3.2. *Let* $f, h, l, g \in cO_A^n$. *Then*

$$f+h+l+g = f+l+h+g \text{ iff } ((l_2 \circ h_2)(a), (h_2 \circ l_2)(a)) \in Ker(g)$$

for all $a \in Im(f_2)$.

Proof. Assume that $f+h+l+g = f+l+h+g$ and $a \in Im(f_2)$. Then there are elements $b \in A$ and $i \in \{1, \ldots, n\}$ such that $f(b) = (i, a)$. Since

$$(f+h+l+g)(b) = (f+l+h+g)(b)$$

we get

$$((f+h+l)+g)(b) = ((f+l+h)+g)(b)$$

$$\Rightarrow g((f+h+l)_2(b)) = g((f+l+h)_2(b))$$

$$\Rightarrow ((f+h+l)_2(b), (f+l+h)_2(b)) \in Ker(g).$$

Then

$$\begin{aligned} (f+l+h)_2(b) &= ((f+l)+h)_2(b) \\ &= h_2((f+l)_2(b)) \\ &= h_2(l_2(f_2(b))) \\ &= (h_2 \circ l_2)(f_2(b)) \\ &= (h_2 \circ l_2)(a) \end{aligned}$$

and similarly

$$(f+h+l)_2(b) = (l_2 \circ h_2)(a).$$

Therefore we get

$$((h_2 \circ l_2)(a), (l_2 \circ h_2)(a)) \in Ker(g).$$

Conversely, we assume that

$$((h_2 \circ f_2)(a), (l_2 \circ h_2)(a)) \in Ker(g)$$

for all $a \in Im(f_2)$. Let $b \in A$. Then

$$f_2(b) \in Im(f_2)$$

and by our assumption we have $((h_2 \circ l_2)(f_2(b)), (l_2 \circ h_2)(f_2(b))) \in Ker(g)$. Then

$$\begin{aligned} & g((h_2 \circ l_2)(f_2(b))) = g((l_2 \circ h_2)(f_2(b))) \\ \Rightarrow \quad & g(h_2((l_2 \circ f_2)(b))) = g(l_2((h_2 \circ f_2)(b))) \\ \Rightarrow \quad & (h+g)((l_2 \circ f_2)(b)) = (l+g)((h_2 \circ f_2)(b)) \\ \Rightarrow \quad & (l+h+g)(f_2(b)) = (h+l+g)(f_2(b)) \\ \Rightarrow \quad & (f+l+h+g)(b) = (f+h+l+g)(b). \end{aligned}$$

Therefore, $f+l+h+g = f+h+l+g$. □

Altogether we have:

Theorem 3.2. *Let $\underline{S} \subseteq E(cO_A^n)$ be a subsemigroup. Then $\underline{S}$ is a normal band iff for all $f, h, l, g \in S$ we have $((l_2 \circ h_2)(a), (h_2 \circ l_2)(a)) \in Ker(g)$ for all $a \in Im(f_2)$.*

To characterize semilattices we need conditions which guarantee commutativity.

Theorem 3.3. *Let $\underline{S}$ be a subsemigroup of $(cO_A^n; +)$ and assume that $S \subseteq E(cO_A^n)$. Then $\underline{S}$ is a semilattice iff for any two cooperations $f, g \in S$ the following conditions hold:*

(i) $Im(f+g) = Im(g+f)$,

(ii) *For each* $(i,b) \in Im(f+g)$ *we have* $f_2(x) \in g^{-1}(i,b) \Leftrightarrow g_2(x) \in f^{-1}(i,b)$.

Proof. Assume that $\underline{S}$ is a semilattice. Then $f+g = g+f$ implies $Im(f+g) = Im(g+f)$. Let $(i,b) \in Im(f+g)$ and $f_2(x) \in g^{-1}(i,b)$. Then $f(g_2(x)) = (g+f)(x) = (f+g)(x) = g(f_2(x)) = (i,b)$. Therefore $g_2(x) \in f^{-1}(i,b)$. Similarly, we show that $g_2(x) \in f^{-1}(i,b)$ implies that $f_2(x) \in g^{-1}(i,b)$.

Conversely, we assume that for any $f, g \in S$ conditions (i) and (ii) are satisfied. To show that $f+g = g+f$ we assume that $x \in A$. Then there exists a pair $(i,b) \in Im(f+g)$ such that $(f+g)(x) = (i,b)$. Since $(f+g)(x) = g(f_2(x))$ and $(f+g)(x) = (i,b)$, we get $g(f_2(x)) = (i,b)$ and then $f_2(x) \in g^{-1}(i,b)$. Condition (ii) implies $g_2(x) \in f^{-1}(i,b)$. Therefore $(g+f)(x) = f(g_2(x)) = (i,b) = (f+g)(x)$. This implies $f+g = g+f$ and $\underline{S}$ is a semilattice. □

4. Green's relations $\mathcal{L}$ and $\mathcal{R}$

Let $(S;\circ)$ be a semigroup. Then Green's relations $\mathcal{L}$ and $\mathcal{R}$ are defined on $(S;\circ)$ as follows:

$$(a,b) \in \mathcal{L} :\Leftrightarrow a = b \vee \exists c,d \in S(c \circ a = b \wedge d \circ b = a)$$

$$(a,b) \in \mathcal{R} :\Leftrightarrow a = b \vee \exists e,f \in S(a \circ e = b \wedge b \circ f = a)$$

for $a,b \in S$. Clearly $\mathcal{L}$ and $\mathcal{R}$ are equivalence relations on S. We want to characterize Green's relations $\mathcal{L}$ and $\mathcal{R}$ on $(cO_A^n;+)$. It is well-known (see [4]) that for transformation semigroups there holds: $(f,g) \in \mathcal{L}$ iff $Im(f) = Im(g)$ and $(f,g) \in \mathcal{R}$ iff $Ker(f) = Ker(g)$. For n-ary cooperations we have

Theorem 4.1. *Let $f,g \in Reg(cO_A^n)$. Then $(f,g) \in \mathcal{L}$ iff $Im(f) = Im(g)$ and $(f,g) \in \mathcal{R}$ iff $Ker(f) = Ker(g)$.*

Proof. If $(f,g) \in \mathcal{L}$ and $f \neq g$, there are n-ary cooperations $h,l \in cO_A^n$ such that $h + f = g$ and $l + g = f$. Let $a \in A$. Then

$$f(a) = (l+g)(a) = g(l_2(a)) \in Im(g)$$

and thus $Im(f) \subseteq Im(g)$. In the corresponding way from $g = h + l$ we get $Im(g) \subseteq Im(f)$, altogether $Im(g) = Im(f)$. If $f = g$, then $Im(f) = Im(g)$.

Conversely, assume that $Im(f) = Im(g)$, For each $(i,a) \in Im(f) = Im(g)$ we choose $d_a \in f^{-1}((i,a))$ and $d'_a \in g^{-1}((i,a))$ and define cooperations $h,l \in cO_A^n$ in the following way:

$$h(x) = (i,d'_a) \text{ for all } x \in f^{-1}((i,a)) \text{ and}$$

$$l(y) = (i,d_a) \text{ for all } y \in g^{-1}((i,a)).$$

The cooperations h and l are well-defined and we have

$$(h+g)(x) = g(h_2(x)) = g(d'_a) = (i,a) = f(x)$$

and

$$(l+f)(y) = f(l_2(y)) = f(d_a) = (i,a) = g(y)$$

and thus

$$h + g = f \text{ and } l + f = g \text{ and then } (f,g) \in \mathcal{L}.$$

As we have seen, this proof did not need that f and g are regular. Therefore, the first part of the proposition holds also for arbitrary $f, g \in cO_A^n$.

Assume now that $(f,g) \in \mathcal{R}$. If $f = g$, then $Ker(f) = Ker(g)$. Assume that $f \neq g$. Then there are $u, v \in cO_A^n$ such that $f = g + u$ and $g = f + v$. Let $(a,b) \in Ker(f)$. Then

$$f(a) = f(b) \Rightarrow f_2(a) = f_2(b) \Rightarrow v(f_2(a)) = v(f_2(a)) \Rightarrow$$

$$(f+v)(a) = (f+v)(b) \Rightarrow g(a) = g(b) \Rightarrow (a,b) \in Ker(g).$$

This shows $Ker(f) \subseteq Ker(g)$. In a similar way we show $Ker(g) \subseteq Ker(f)$ and altogether we have equality.

Conversely, assume that $Ker(f) = Ker(g)$. We denote the equivalence relation $Ker(f)$ $(= Ker(g))$ by π. We form the quotient set A/π. Since A is finite consisting of k elements, there is an integer p with $1 \leq p \leq k$ such that $|A/\pi| = p$. Since

$$|A/\pi| = |Im(f)| = |Im(g)|$$

we can write

$$Im(f) = \{(j_1, a_1), \ldots, (j_p, a_p)\}, Im(g) = \{(k_1, b_1), \ldots, (k_p, b_p)\}$$

for some $j_1, \ldots, j_p, k_1, \ldots, k_p \in \{1, \ldots, n\}$ and $a_1, \ldots, a_p, b_1, \ldots, b_p \in A$. From each equivalence class with respect to π we choose one element $d_i \in A$ for $i = 1, \ldots, p$. For the elements $d_1, \ldots, d_p$ we have then $f(d_i) = (j_i, a_i)$ and $g(d_i) = (k_i, b_i)$. There is no pair of elements $(j_i, a_i), (j_l, a_l) \in Im(f)$ with $1 \leq i < l \leq p$ such that $a_i = a_l$, but $j_i \neq j_l$ and there is no pair of elements $(k_i, b_i), (k_l, b_l)$ with $1 \leq i < l \leq p$ such that $b_i = b_l$ but $k_i \neq k_l$ since f and g are regular (see Theorem 2.1). Now we define $\alpha, \beta \in cO_A^n$ by $\alpha(a_i) = (k_i, b_i)$ for all $i = 1, \ldots, p$ and $\alpha(x) = (r, x)$ for all $x \in A \backslash \{a_1, \ldots, a_p\}$ and for some $r \in \{1, \ldots, n\}$ and $\beta(b_i) = (j_i, a_i)$ for all $i = 1, \ldots, p$ and $\beta(y) = (s, y)$ for all $y \in A \backslash \{b_1, \ldots, b_p\}$ and for some $s \in \{1, \ldots, n\}$. Because of the regularity of f and g the cooperations α, β are well-defined. We prove that $f + \alpha = g$ and $g + \beta = f$. Let $c \in A$. Then there is an element $i \in \{1, \ldots, p\}$ such that $(c, d_i) \in \pi$ and we have

$$(f+\alpha)(c) = \alpha(f_2(c)) = \alpha(f_2(d_i)) = \alpha(a_i) = (k_i, b_i) = g(d_i) = g(c)$$

and thus $f + \alpha = g$.

In a similar way, we show $g + \beta = f$. Hence $(f,g) \in \mathcal{R}$. □

The regularity of f and g is not needed to prove that $(f,g) \in \mathcal{L}$ iff $Im(f) = Im(g)$. Therefore this is true for any $f, g \in cO_A^n$. But for the

second proposition of Theorem 4.1, we need the assumption that both f and g are regular. Indeed, if $f \in cO_A^n \setminus Reg(cO_A^n)$ and $g \in cO_A^n$ such that $Ker(f) = Ker(g)$ and $f \neq g$, then we will show that $(f,g) \notin \mathcal{R}$. Since $f \in cO_A^n \setminus Reg(cO_A^n)$, there are elements $a, b, x \in A$ and there are integers $i \neq j \in \{1, \ldots, n\}$ such that $f(a) = (i,x)$ and $f(b) = (j,x)$. Let $\alpha \in cO_A^n$. Then $(f+\alpha)(a) = \alpha(f_2(a)) = \alpha(x)$ and $(f+\alpha)(b) = \alpha(f_2(b)) = \alpha(x)$. There follows $(f+\alpha)(a) = (f+\alpha)(b)$. Since $(a,b) \notin Ker(f)$ and $Ker(f) = Ker(g)$, we have also $g(a) \neq g(b)$ and thus $g \neq f + \alpha$ for all $\alpha \in cO_A^n$ and then $(f,g) \notin \mathcal{R}$.

References

1. B. Csákány, *Completeness in coalgebras*, Acta Sci. Math., 48(1985), 75-84.
2. K. Denecke, K. Saengsura, *Cohyperidentities and M-solid classes of coalgebras*, (preprint 2006).
3. K. Denecke, K. Saengsura, *Menger Algebras and Clones of cooperations*, Algebra Colloquium, 15:2 (2008), 223–234.
4. J. M. Howie, *Fundamentals of Semigroup Theory* (Oxford Science Publications, Clarendon Press Oxford, 1995).

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 83–94)

ALGEBRAS DERIVED BY SURJECTIVE HYPERSUBSTITUTIONS

K. DENECKE and R. SRITHUS

Institute of Mathematics, Universität Potsdam,
Am Neuen Palais, 14415 Potsdam, Germany
E-mail: kdenecke@rz.uni-potsdam.de

Hypersubstitutions map operation symbols to terms of the corresponding arities. Using hypersubstitutions from each algebra one can derive new algebras. We compare properties of congruence lattices, of subalgebra lattices and of clones of the derived algebras with the corresponding properties of the starting algebras. If both algebras have the same clone of term operations, then they have the most properties in common. This is for instance the case if the hypersubstitution is surjective. We characterize surjective hypersubstitutions and prove that for finite types surjective hypersubstitutions are bijective. This will be applied to i-closed varieties of n-ary type.

Keywords: Hypersubstitution; Clone; Derived algebra; i-closed variety.

1. Preliminaries

Let $(f_i)_{i\in I}$ be a sequence of operation symbols where f_i is n_i-ary. We assume that $n_i \geq 1$ for all $i \in I$. The sequence $\tau = (n_i)_{i\in I}$ is called a *type*. Let $X_n := \{x_1, \ldots, x_n\}$ be an n-element alphabet. Let $W_\tau(X_n)$ be the set of all n-ary terms of type τ and let $W_\tau(X) := \bigcup_{n\geq 1} W_\tau(X_n)$ be the set of all terms of type τ.
On the sets $W_\tau(X_n)$ and $W_\tau(X_m)$ an operation S_m^n :

$$W_\tau(X_n) \times W_\tau(X_m)^n \longrightarrow W_\tau(X_m)$$

is defined inductively as follows:

(i) $S_m^n(x_j, t_1, \ldots, t_n) := t_j$, $x_j \in X_n, t_1, \ldots, t_n \in W_\tau(X_m)$,
(ii) $S_m^n(f_i(s_1, \ldots, s_{n_i}), t_1, \ldots, t_n) :=$
$f_i(S_m^n(s_1, t_1, \ldots, t_n), \ldots, S_m^n(s_{n_i}, t_1, \ldots, t_n))$
with $f_i(s_1, \ldots, s_{n_i}) \in W_\tau(X_n)$, $t_1, \ldots, t_n \in W_\tau(X_m)$.

With these operations and the variables as nullary operations we form the multi-based (heterogeneous) algebra

$$clone\tau := ((W_\tau(X_n))_{n\in\mathbb{N}^+}; (S_m^n)_{m,n\in\mathbb{N}^+}, (x_i)_{i\leq n, n\in\mathbb{N}^+})$$

(with $\mathbb{N}^+ := \mathbb{N} \setminus \{0\}$). This algebra is an abstract clone, i.e. it satisfies the clone axioms

(C1) $\tilde{S}_m^p(\tilde{Z}, \tilde{S}_m^n(\tilde{Y}_1, \tilde{X}_1, \ldots, \tilde{X}_n), \ldots, \tilde{S}_m^n(\tilde{Y}_p, \tilde{X}_1, \ldots, \tilde{X}_n)) \approx$
$\tilde{S}_m^n(\tilde{S}_n^p(\tilde{Z}, \tilde{Y}_1, \ldots, \tilde{Y}_p), \tilde{X}_1, \ldots, \tilde{X}_n), (m, n, p \in \mathbb{N}^+)$
(C2) $\tilde{S}_m^n(\tilde{e}_i^n, \tilde{X}_1, \ldots, \tilde{X}_n) \approx \tilde{X}_i$, $(m, n \in \mathbb{N}^+, 1 \leq i \leq n)$
(C3) $\tilde{S}_n^n(\tilde{Y}, \tilde{e}_1^n, \ldots, \tilde{e}_n^n) \approx \tilde{Y}$ $(n \in \mathbb{N}^+)$.

Here $\tilde{Z}, \tilde{Y}_1, \ldots, \tilde{Y}_p, \tilde{X}_1, \ldots, \tilde{X}_n$ are variables for terms, $\tilde{S}_m^n$, $\tilde{S}_m^p$, $\tilde{S}_n^p$, $\tilde{S}_n^n$ are operation symbols and $\tilde{e}_i^n$ are symbols for variables.

Sometimes it is necessary to measure the complexity of a term. A very common complexity measure is the depth of a term defined by the following steps:

(i) $depth(x_j) := 0$ if $x_j \in X$,
(ii) $depth(f_i(t_1, \ldots, t_{n_i})) := max\{depth(t_1), \ldots, depth(t_{n_i})\} + 1$.

Every term can be visualized by a tree diagram. The *depth* of t is then the maximum length of all pathes going from the root to one of the leaves of the tree. Considering only pathes going from the root to leaves labeled by the variable x_j we get the $j - depth$, $depth_j(t)$ of the term t.

A hypersubstitution σ is a mapping which sends operation symbols to terms preserving the arities. Based on a hypersubstitution one can inductively define a mapping $\hat{\sigma} : W_\tau(X) \longrightarrow W_\tau(X)$ by the following steps:
$\hat{\sigma}[x] := x$ if $x \in X$ is a variable and $\hat{\sigma}[f_i(t_1, \ldots, t_{n_i})] := S_m^{n_i}(\sigma(f_i), \hat{\sigma}[t_1], \ldots, \hat{\sigma}[t_{n_i}])$ where f_i is n_i-ary and $t_1, \ldots, t_{n_i} \in W_\tau(X_n)$.

The last equation shows that $\hat{\sigma}$ is an endomorphism of the algebra $clone\tau$ if we regard $\hat{\sigma} = (\hat{\sigma}_n)_{n\in\mathbb{N}^+}$ as a sequence of mappings where $\hat{\sigma}_n : W_\tau(X_n) \longrightarrow W_\tau(X_n)$.

Hypersubstitutions can be composed by $\sigma_1 \circ_h \sigma_2 := \hat{\sigma_1} \circ \sigma_2$ where $\circ$ is the usual composition of mappings. Together with the identity hypersubstitution $\sigma_{id} : f_i \mapsto f_i(x_1, \ldots, x_{n_i})$ we get a monoid $\mathcal{H}yp(\tau) = (Hyp(\tau); \circ_h, \sigma_{id})$.

Let $\mathcal{A} = (A; (f_i^A)_{i\in I})$ be an algebra of type τ. Every term of type τ induces a term operation t^A on the algebra $\mathcal{A}$. Let $W_\tau(X_n)^{\mathcal{A}}$ be the set of all n-ary term operations of the algebra $\mathcal{A}$. If we define a superposition of operations on A by

$$S_m^{n,A}(f^A, g_1^A, \ldots, g_n^A)(a_1, \ldots, a_m) := \\ f^A(g_1^A(a_1, \ldots, a_m), \ldots, g_n^A(a_1, \ldots, a_m))$$

where f^A is n-ary and $g_1^A, \ldots, g_n^A$ are m-ary, then we obtain another clone

$$T(\mathcal{A}) := ((W_\tau(X_n)^{\mathcal{A}})_{n\in\mathbb{N}^+}; (S_m^{n,A})_{m,n\in\mathbb{N}^+}, (e_i^{n,A})_{i\leq n, n\in\mathbb{N}^+})$$

where $e_i^{n,A}$ are the n-ary projection mappings on the i-th components. It is easy to see that this clone is generated by the fundamental operations $\{f_i^A \mid i \in I\}$ of the algebra $\mathcal{A}$. Let $T^{(n)}(\mathcal{A})$ be the set of all n-ary term operations of $\mathcal{A}$. The set $T^{(n)}(\mathcal{A})$ is closed under the $(n+1)$-ary operation $S_n^{n,A}$ and contains all n-ary projections. Therefore one can form an algebra $(T^{(n)}(\mathcal{A}); S_n^{n,A}, e_1^{n,A}, \ldots, e_n^{n,A})$ of type $(n+1, 0, \ldots, 0)$.

Let $Sub(\mathcal{A})$ be the subalgebra lattice of $\mathcal{A}$, $Con(\mathcal{A})$ be the lattice of congruence relations of $\mathcal{A}$, $V(\mathcal{A})$ be the variety generated by $\mathcal{A}$ and let $\mathfrak{L}(V(\mathcal{A}))$ be the subvariety lattice of the variety $V(\mathcal{A})$.

Two algebras $\mathcal{A}$ and $\mathcal{B}$ of possibly different types are called *rationally equivalent* (*term equivalent*) if $\mathcal{A}$ and $\mathcal{B}$ have the same universe and the same term operations, i.e. if $T(\mathcal{A}) = T(\mathcal{B})$. As a consequence we have $Sub(\mathcal{A}) = Sub(\mathcal{B})$, $Con(\mathcal{A}) = Con(\mathcal{B})$ and the subvariety lattices $\mathfrak{L}(V(\mathcal{A}))$ and $\mathfrak{L}(V(\mathcal{B}))$ are isomorphic.

Hypersubstitutions can be applied to equations as well as to algebras of type τ. Let $s \approx t$ be an identity satisfied in the algebra $\mathcal{A}$, i.e. $s^{\mathcal{A}} = t^{\mathcal{A}}$. The identity $s \approx t$ is said to be satisfied as a *hyperidentity* in the algebra $\mathcal{A}$ if $\hat{\sigma}[s] \approx \hat{\sigma}[t]$ are identities in $\mathcal{A}$ for all hypersubstitutions σ.

Applying a hypersubstitution σ to an algebra $\mathcal{A}$ of type τ gives the algebra $\sigma(\mathcal{A}) := (A; (\sigma(f_i)^{\mathcal{A}})_{i\in I})$ which is called *algebra derived* from $\mathcal{A}$ by the hypersubstitution σ. There is the following connection between the term operation induced by a term t on the algebra $\mathcal{A}$ and the term operation induced by the same term on the algebra $\sigma(\mathcal{A})$ (see e.g. [6]).

Proposition 1.1. *Let $t \in W_\tau(X)$ be a term of type τ. Let σ be a hypersubstitution of type τ and let $\mathcal{A}$ be an algebra of type τ. Then $t^{\sigma(\mathcal{A})} = \hat{\sigma}[t]^{\mathcal{A}}$.*

Proof. The proof is by induction on the complexity (*depth*) of the term t. If $t = x_j \in X_n$ is a variable, then

$$x_j^{\sigma(\mathcal{A})} = e_j^{n,A} = x_j^{\mathcal{A}} = \hat{\sigma}[x_j]^{\mathcal{A}}.$$

If $t = f_i(t_1, \ldots, t_{n_i})$ and if we inductively assume that $t_k^{\sigma(\mathcal{A})} = \hat{\sigma}[t_k]^{\mathcal{A}}$ for all $1 \leq k \leq n_i$, then

$$\begin{aligned} & t^{\sigma(\mathcal{A})} \\ &= S_m^{n_i,\mathcal{A}}(f_i^{\sigma(\mathcal{A})}, t_1^{\sigma(\mathcal{A})}, \ldots, t_{n_i}^{\sigma(\mathcal{A})}) \\ &= S_m^{n_i,\mathcal{A}}(\sigma(f_i)^{\mathcal{A}}, \hat{\sigma}[t_1]^{\mathcal{A}}, \ldots, \hat{\sigma}[t_{n_i}]^{\mathcal{A}}) \\ &= (S_m^{n_i,\mathcal{A}}(\sigma(f_i), \hat{\sigma}[t_1], \ldots, \hat{\sigma}[t_{n_i}]))^{\mathcal{A}} = \hat{\sigma}[t]^{\mathcal{A}}. \end{aligned}$$

□

We gave the simple proof of this proposition here since the property plays an important rule in what follows.

In [4] the following formula for the depth of $\hat{\sigma}[t]$ with $t = f_i(t_1, \ldots, t_{n_i})$ was proved:

$$\begin{aligned} & depth(\hat{\sigma}[t]) \\ &= max\{depth_k(\sigma(f_i)) + depth(\hat{\sigma}[t_k]) \mid 1 \leq k \leq n_i, x_k \in var(\sigma(f_j))\}. \end{aligned}$$

2. Term Operations of Derived Algebras

Let $\mathcal{B}$ be an algebra of type τ and let σ be a hypersubstitution of type τ. By $\sigma(Sub(\mathcal{B}))$ we understand the set of all algebras derived by σ from subalgebras of $\mathcal{B}$, i.e.

$$\sigma(Sub(\mathcal{B})) := \{\sigma(\mathcal{C}) \mid \mathcal{C} \subseteq \mathcal{B}\}.$$

Directly from the definition of a derived algebra we obtain the following simple propositions:

Proposition 2.1.

(i) The clone of $\sigma(\mathcal{B})$ is a subclone of the clone of $\mathcal{B}$: $\mathcal{T}(\sigma(\mathcal{B})) \subseteq \mathcal{T}(\mathcal{B})$.
(ii) The congruence lattice of $\mathcal{B}$ is a sublattice of the congruence lattice of $\sigma(\mathcal{B})$: $Con(\mathcal{B}) \subseteq Con(\sigma(\mathcal{B}))$.
(iii) $\sigma(Sub(\mathcal{B}))$ forms a sub-meet-semilattice of $Sub(\sigma(\mathcal{B}))$: $\sigma(Sub(\mathcal{B})) \subseteq Sub(\sigma(\mathcal{B}))$.

Proof.

(i) Let t be an arbitrary term of type τ and $t^{\sigma(\mathcal{B})} \in T(\sigma(\mathcal{B}))$. Because of $T(\sigma(\mathcal{B})) = W_\tau(X)^{\sigma(\mathcal{B})}$ any element of $T(\sigma(\mathcal{B}))$ can be written in this form. By Proposition 1.1 we have $\hat{\sigma}[t]^{\mathcal{B}} \in T(\sigma(\mathcal{B}))$. Since $\hat{\sigma}[t] \in W_\tau(X)$, we get $\hat{\sigma}[t]^{\mathcal{B}} \in T(\mathcal{B})$ and thus $\mathcal{T}(\sigma(\mathcal{B})) \subseteq \mathcal{T}(\mathcal{B})$.

(ii) Let θ be a congruence relation on $\mathcal{B}$. Then for every n-ary term operation $t^{\mathcal{B}}$ of $\mathcal{B}$ and for all elements $a_1, \ldots, a_n, b_1, \ldots, b_n \in B$ we have

$$(a_1, b_1) \in \theta, \ldots, (a_n, b_n) \in \theta \Rightarrow (t^{\mathcal{B}}(a_1, \ldots, a_n), t^{\mathcal{B}}(b_1, \ldots, b_n)) \in \theta.$$

Therefore for every $i \in I$ we have

$$(\sigma(f_i)^{\mathcal{B}}(a_1, \ldots, a_{n_i}), \sigma(f_i)^{\mathcal{B}}(b_1, \ldots, b_{n_i})) = (f_i^{\sigma(\mathcal{B})}(a_1, \ldots, a_{n_i}), f_i^{\sigma(\mathcal{B})}(b_1, \ldots, b_{n_i})) \in \theta$$

and thus $\theta \in Con(\sigma(\mathcal{B}))$. This means $Con(\mathcal{B}) \subseteq Con(\sigma(\mathcal{B}))$. The sublattice property is clear as well.

(iii) Let $\mathcal{C} \in \sigma(Sub(\mathcal{B}))$. Then there is a subalgebra $\mathcal{C}' \subseteq \mathcal{B}$ such that $\mathcal{C} = \sigma(\mathcal{C}')$. From $\mathcal{C}' \subseteq \mathcal{B}$ we obtain that $\sigma(\mathcal{C}')$ is also a subalgebra of $\sigma(\mathcal{B})$; i.e. $\mathcal{C} = \sigma(\mathcal{C}') \in Sub(\sigma(\mathcal{B}))$ and then $\sigma(Sub(\mathcal{B})) \subseteq Sub(\sigma(\mathcal{B}))$. Let $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2) \in \sigma(Sub(\mathcal{B}))$. We show that $\sigma(\mathcal{A}_1) \cap \sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_1 \cap \mathcal{A}_2) \in \sigma(Sub(\mathcal{B}))$. In fact, the universe of both algebras, of $\sigma(\mathcal{A}_1) \cap \sigma(\mathcal{A}_2)$ and of $\sigma(\mathcal{A}_1 \cap \mathcal{A}_2)$ is the intersection $A_1 \cap A_2$. Since $\sigma(\mathcal{A}_1 \cap \mathcal{A}_2)$ and $\sigma(\mathcal{A}_1) \cap \sigma(\mathcal{A}_2)$ are both subalgebras of $\sigma(\mathcal{B})$ they must be equal. This shows that $\sigma(Sub(\mathcal{B}))$ is a meet-semilattice and therefore a submeet-semilattice of the meet-semilattice $Sub(\sigma(\mathcal{B}))$. □

It is very natural to ask for which algebra $\mathcal{B}$ and for which hypersubstitution σ do we have equality in one of the propositions (i)-(iii). Clearly, if $\mathcal{B}$ is a two-element algebra, then $Con(\mathcal{B}) = \{\triangle_B, B^2\}$ and then $Con(\sigma(\mathcal{B})) = Con(\mathcal{B})$ since $Con(\sigma(\mathcal{B}))$ is also two-element.

Let $\mathcal{O}(B)$ be the clone of all operations defined on B, then for any algebra $\mathcal{B} = (B; (f_i^B)_{i \in I})$ the clone $\mathcal{T}(B)$ is a subclone of $\mathcal{O}(B)$. The least subclone of $\mathcal{O}(B)$ is the clone $\mathcal{J}_B$ of all projection mappings defined on B.

Assume that $\mathcal{T}(B)$ is a minimal element in the lattice of all subclones of $\mathcal{O}(B)$ and that $T(\sigma(\mathcal{B})) \neq J_B$. Then we must have $T(\sigma(\mathcal{B})) = T(\mathcal{B})$, i.e. the algebras $\mathcal{B}$ and $\sigma(\mathcal{B})$ are rationally equivalent. Then we have also $Con(\mathcal{B}) = Con(\sigma(\mathcal{B}))$ and $Sub(\mathcal{B}) = Sub(\sigma(\mathcal{B}))$. In this case the properties of the algebra $\mathcal{B}$ lead to the equation $T(\sigma(\mathcal{B})) = T(\mathcal{B})$.

Proposition 2.2. *Let $\mathcal{B}$ be an arbitrary non-trivial algebra of type $\tau = (n_i)_{i \in I}, n_i \geq 1$ for all $i \in I$. If σ is surjective, meaning that the extension $\hat{\sigma}$ is surjective, then $T(\sigma(\mathcal{B})) = T(\mathcal{B})$.*

Proof. We noticed already that $W_\tau(X)^{\mathcal{B}} = T(\mathcal{B})$. Let $t^{\mathcal{B}} \in T(\mathcal{B})$. Then there is a term $s \in W_\tau(X)$ such that $t = \hat{\sigma}[s]$. From this equation we get $t^{\mathcal{B}} = \hat{\sigma}[s]^{\mathcal{B}} = s^{\sigma(\mathcal{B})} \in T(\sigma(\mathcal{B}))$ by Proposition 1.1. This shows $T(\mathcal{B}) \subseteq T(\sigma(\mathcal{B}))$ and together with Proposition 2.1 (i) we have equality. □

If conversely $T(\mathcal{B}) = T(\sigma(\mathcal{B}))$, then for any $t^{\mathcal{B}} \in T(\mathcal{B}) = T(\sigma(\mathcal{B})) = W_\tau(X)^{\sigma(\mathcal{B})}$ there is a term $s \in W_\tau(X)$ such that $t^{\mathcal{B}} = s^{\sigma(\mathcal{B})} = \hat{\sigma}[s]^{\mathcal{B}}$. But this means that $t \approx \hat{\sigma}[s]$ is an identity in $\mathcal{B}$ but in general $t \neq \hat{\sigma}[s]$.

Clearly, we have

Proposition 2.3. *Let $\mathcal{B}$ be a non-trivial algebra and let σ be a hypersubstitution of type τ. Then $T(\mathcal{B}) = T(\sigma(\mathcal{B}))$ if and only if for any term $t \in W_\tau(X)$ there exists a term $s \in W_\tau(X)$ such that $t \approx \hat{\sigma}[s]$ is an identity in $\mathcal{B}$.*

Proposition 2.2 motivates to look at all surjective hypersubstitutions. Proposition 2.3 shows once more the influence of the properties of the algebra $\mathcal{B}$. In the next section we want to characterize all surjective hypersubstitutions of type τ.

3. Surjective Hypersubstitutions of Type τ

We noticed already that the heterogeneous algebra

$$clone\tau := ((W_\tau(X_n))_{n\in\mathbb{N}^+}; (S_m^n)_{m,n\in\mathbb{N}^+}, (x_i)_{i\leq n, n\in\mathbb{N}^+})$$

is an abstract clone, i.e. satisfies the clone axioms (C1), (C2), (C3).

Let I_n be the set of all indices from I such that f_i is n-ary and let $F^n := \{f_i(x_1, \ldots, x_n) \mid i \in I_n\}$ and $F := (F^n)_{n\in\mathbb{N}^+}$. Clearly, $F' := \{f_i(x_1, \ldots, x_n) \mid i \in I\} = \bigcup_{n\in\mathbb{N}^+} F^n$.

By the definition of terms the sequence F generates $clone\tau$. Another important property of $clone\tau$ is that it is free in the heterogeneous variety of clones (defined by (C1), (C2), (C3)), freely generated by the sequence F.

We noticed also that extensions $\hat{\sigma}$ of hypersubstitutions σ are endomorphisms of $clone\tau$. Conversely, as a consequence of the freeness of $clone\tau$ and the fact that F is a free generating system, any endomorphism of $clone\tau$ is the extension of a hypersubstitution (see [1]). Let $ker\sigma$ be the kernel of $\hat{\sigma}$, i.e. $ker\sigma := \{(s,t) \mid s,t \in W_\tau(X) \wedge \hat{\sigma}[s] = \hat{\sigma}[t]\}$ and let $Im\sigma$ be the image of $\hat{\sigma}$, i.e. $Im\sigma := \{t \mid t \in W_\tau(X) \wedge \exists s \in W_\tau(X)(\hat{\sigma}[s] = t)\}$. Then σ is called *injective* or *surjective* or *bijective* if $\hat{\sigma}$ has these properties. If $Bij(\tau)$ and $Surj(\tau)$ denote the sets of all bijective and surjective hypersubstitutions, then $Bij(\tau)$ and $Surj(\tau)$ form submonoids of $\mathcal{H}yp(\tau)$ since the composition of two bijective or of two surjective mappings is bijective or surjective, respectively.

Since σ preserves arities, it can be regarded as a sequence $(\sigma_n)_{n\in\mathbb{N}^+}$ of mappings $\sigma_n : F^n \to W_\tau(X_n)$. The extension $\hat{\sigma}$ can also be regarded as a

sequence $(\hat{\sigma}_n)_{n\in\mathbb{N}^+}$, $\hat{\sigma}_n : W_\tau(X_n) \to W_\tau(X_n)$. Therefore the image $Im\sigma$ can also be regarded as a sequence $(Im\sigma)' := (Im\hat{\sigma}_n)_{n\in\mathbb{N}^+}$, $Im\hat{\sigma}_n \subseteq W_\tau(X_n)$.

Clearly σ is injective iff $ker\sigma = \triangle_{W_\tau(X)}$. Since F is a generating system of $clone\tau$ we have:

Proposition 3.1. *A hypersubstitution σ of type τ is surjective iff $F' \subseteq Im\sigma$.*

Proof. If $F' \subseteq Im\sigma$, then for the subalgebras of $clone\tau$ generated by F and $(Im\sigma)'$ we have $\langle F\rangle_{clone\tau} \subseteq \langle (Im\sigma)'\rangle_{clone\tau}$, i.e. $W_\tau(X) \subseteq Im\sigma \subseteq W_\tau(X)$ since $(Im\sigma)'$ is a (heterogeneous) subalgebra of $clone\tau$. Here we used that $\hat{\sigma}$ is an endomorphism of $clone\tau$. Altogether, $Im\sigma = W_\tau(X)$ and σ is surjective. If conversely σ is surjective, then $F' \subseteq Im\sigma$ is clear. □

(Proposition 2.3 is true for arbitrary (homogeneous and heterogeneous) algebras, for generating systems and for endomorphisms).

Bijective hypersubstitutions were characterized in [3] (see also [6]).

Let $\mathfrak{B}$ denote the set of all bijections on the set $\{f_i \mid i \in I\}$ preserving the arities, i.e. $\mathfrak{B}' := (\mathfrak{B}_n)_{n\in\mathbb{N}^+}$ is a sequence with $\bigcup_{n\in\mathbb{N}^+} \mathfrak{B}_n = \mathfrak{B}$ where $h \in \mathfrak{B}_n$ iff $h : \{f_i \mid i \in I_n\} \longrightarrow \{f_i \mid i \in I_n\}$ is bijective. Let S_n be the set of all permutations on the set $\{1,\ldots,n\}$ for $n \in \mathbb{N}^+$. We set $A := \bigcup_{1\leq n\in\mathbb{N}^+} S_n$ and $\mathfrak{P} := \{p \in A^I \mid p(i) \in S_{n_i} \text{ for } i \in I\}$. Then $Bij(\tau)$ can be characterized by the following theorem:

Theorem 3.1. *Let $\tau = (n_i)_{i\in I}$ be a type with $n_i \geq 1$ for all $i \in I$. For each $\sigma \in Hyp(\tau)$ the following statements are equivalent:*

(i) $\sigma \in Bij(\tau)$.

(ii) There are elements $h \in \mathfrak{B}$ and $p \in \mathfrak{P}$ such that $\sigma(f_i) = h(f_i)(x_{p(i)(1)},\ldots,x_{p(i)(n_i)})$ for all $i \in I$.

Not every surjective hypersubstitution is bijective. As a counterexample we consider the type $\tau = (n_i)_{i\in\mathbb{N}^+} = (i)_{i\in\mathbb{N}^+}$.

We consider the hypersubstitution which is defined by $\sigma(f_{2i}) := f_i(x_1,\ldots,x_i)$ for $i \in \mathbb{N}^+$ and $\sigma(f_{2i+1}) := f_{2i+1}(x_1,\ldots,x_{2i+1})$ for $i \in \mathbb{N}^+$. Since $f_i(x_1,\ldots,x_i) \in W_\tau(\{x_1,\ldots,x_i,\ldots,x_{2_{i+1}}\})$, σ is arity-preserving and thus a hypersubstitution. By Proposition 3.1 we get that σ is surjective, but not injective since for example for $i = 3$ we have $\sigma(f_6) = f_3(x_1,x_2,x_3) = \sigma(f_3)$ and then also $\hat{\sigma}[f_6(x_1,x_2,\ldots,x_6)] = f_3(x_1,x_2,x_3) = \hat{\sigma}[f_3(x_1,x_2,x_3)]$, but $f_6(x_1,x_2,\ldots,x_6) \neq f_3(x_1,x_2,x_3)$.

Surjective hypersubstitutions can be characterized as follows:

Theorem 3.2. *Let $\tau = (n_i)_{i\in I}$ be a type with $n_i \geq 1$ for all $i \in I$. Then for each $\sigma \in Hyp(\tau)$ the following statements are equivalent:*

(i) $\sigma \in Surj(\tau)$.

(ii) *For each $i \in I$ there is an element $j \in I$ such that $n_j \geq n_i$ and such that $\hat{\sigma}[f_j(y_1,\ldots,y_{n_j})] = f_i(x_1,\ldots,x_{n_i})$ for some variables $y_1,\ldots,y_{n_j} \in X_{n_j}$ with $|\{y_1,\ldots,y_{n_j}\}| = n_i$ for all $i \in I$.*

Proof. (ii)⇒(i) By Proposition 3.1 and by (ii) we get that σ is surjective.

(i)⇒(ii) Since σ is surjective, for every $i \in I$ there is a term t_i such that $\hat{\sigma}[t_i] = f_i(x_1,\ldots,x_{n_i})$. $Depth(t_i) = 0$ is impossible since otherwise $t_i = x_k$ is a variable and then $\hat{\sigma}[x_k] = x_k$. Therefore $t_i = f_j(s_1,\ldots,s_{n_j})$ and assume that t_i is a term with minimal depth and $\hat{\sigma}[t_i] = f_i(x_1,\ldots,x_{n_i})$. Then $\hat{\sigma}[t_i] = S^{n_j}_{n_i}(\sigma(f_j),\hat{\sigma}[s_1],\ldots,\hat{\sigma}[s_{n_j}]) = f_i(x_1,\ldots,x_{n_i})$ and because of $depth(\hat{\sigma}[t_i]) = max\{depth_k(\sigma(f_j)) + depth(\hat{\sigma}[s_k]) \mid 1 \leq k \leq n_j, x_k \in var(\sigma(f_j))\} = 1$ we must have $\sigma(f_j) \in X_{n_j}$ or $\sigma(f_j) = f_i(x_{j_1},\ldots,x_{j_{n_i}})$ with $x_{j_1},\ldots,x_{j_{n_i}} \in X_{n_j}$ and $|\{x_{j_1},\ldots,x_{j_{n_i}}\}| = n_i$ and thus $n_j \geq n_i$ and $depth(\hat{\sigma}[s_k]) \leq 1$ for all $1 \leq k \leq n_j$. If $\sigma(f_j) \in X_{n_j}$, let say $\sigma(f_j) = x_r$ for some $1 \leq r \leq n_j$, then $\hat{\sigma}[t_i] = \hat{\sigma}[s_r] = f_i(x_1,\ldots,x_{n_i})$ where $depth(s_r) < depth(t_i)$. This contradicts the minimality of the depth of t_i. Therefore there are variables $x_{l_1},\ldots,x_{l_{n_i}} \in X_{n_j}$ with $|\{x_{l_1},\ldots,x_{l_{n_i}}\}| = n_i$ and with $\sigma(f_j) = f_i(x_{l_1},\ldots,x_{l_{n_i}})$. Then we have $f_i(x_1,\ldots,x_{n_i}) = \hat{\sigma}[t_i] = S^{n_j}_{n_i}(f_i(x_{l_1},\ldots,x_{l_{n_i}}),\hat{\sigma}[s_1],\ldots,\hat{\sigma}[s_{n_j}])$. From this equation there follows that $x_1 = \hat{\sigma}[s_{l_1}]$, $x_2 = \hat{\sigma}[s_{l_2}],\ldots,x_{n_i} = \hat{\sigma}[s_{l_{n_i}}]$. For each $i \in \{1,\ldots,n_j\} \setminus \{l_1,\ldots,l_{n_j}\}$ the term s_i can be an arbitrary element of $W_\tau(X_{n_j})$.

Now we consider the following term $\tilde{t}_i = f_j(y_1,\ldots,y_{n_j})$ where $y_{l_1} = x_1,\ldots,y_{l_{n_i}} = x_{n_i}$ and $y_k = x_1$ if $k \in \{1,\ldots,n_j\} \setminus \{l_1,\ldots,l_{n_i}\}$. Then

$$\begin{aligned}
&\hat{\sigma}[\tilde{t}_i] \\
&= \hat{\sigma}[f_j(y_1,\ldots,y_{n_j})] \\
&= S^{n_j}_{n_i}(\sigma(f_j),\hat{\sigma}[y_1],\ldots,\hat{\sigma}[y_{n_j}]) \\
&= S^{n_j}_{n_i}(f_i(x_{l_1},\ldots,x_{l_{n_i}}),\hat{\sigma}[y_1],\ldots,\hat{\sigma}[y_{n_j}]) \\
&= f_i(x_1,\ldots,x_{n_i}).
\end{aligned}$$

□

Now we assume that I is finite. Then $m := max\{n_i \mid i \in I\}$ exists.

Let $W_\tau(X_m)_1 := \{t \in W_\tau(X_m) \mid depth(t) = 1\}$. Then by a similar proof we show that the following propositions are equivalent:

Lemma 3.1. *Let $\tau = (n_i)_{i\in I}$ be a type with $n_i \geq 1$ for all $i \in I$ and*

assume that I is finite. Then for each $\sigma \in Hyp(\tau)$ the following statements are equivalent:

(i) $\sigma \in Surj(\tau)$.

(ii) σ *is bijective on* $W_\tau(X_m)_1$ *where* m *is the maximum of all arities* $n_i, i \in I$ *(i.e. the restriction* $\hat{\sigma}|_{W_\tau(X_m)_1}$ *is bijective).*

Proof. (ii)⇒(i) If σ is bijective on $W_\tau(X_m)_1$, then for every $f_i(x_1, \ldots, x_{n_i}) \in F'$ there is a term $t \in W_\tau(X_m)_1$ such that $\hat{\sigma}[t] = f_i(x_1, \ldots, x_{n_i})$, i.e. $F' \subseteq Im\sigma$ and by Proposition 3.1 we get that σ is surjective.

(i)⇒(ii) Let $t \in W_\tau(X_m)_1$. Then there are elements $i \in I$ and $x_{i_1}, \ldots, x_{i_{n_i}} \in X_m$ such that $t = f_i(x_{i_1}, \ldots, x_{i_{n_i}})$. We consider now the term $f_i(x_1, \ldots, x_{n_i})$. Theorem 3.2 implies that there are elements $j \in I$ and $y_1, \ldots, y_{n_j} \in X_{n_j} \subseteq X_m$ such that $n_j \geq n_i$, $|\{y_1, \ldots, y_{n_j}\}| = n_i$ and $\hat{\sigma}[f_j(y_1, \ldots, y_{n_j}] = f_i(x_1, \ldots, x_{n_i})$. Then $S_m^{n_j}(\sigma(f_j), y_1, \ldots, y_{n_j}) = f_i(x_1, \ldots, x_{n_i})$ and there are $x_{l_1}, \ldots, x_{l_{n_j}} \in X_{n_j} \subseteq X_m$ such that $\sigma(f_j) = f_i(x_{l_1}, \ldots, x_{l_{n_i}})$. So $S_m^{n_j}(\sigma(f_j), y_1, \ldots, y_{n_j}) = S_m^{n_j}(f_i(x_{l_1}, \ldots, x_{l_{n_i}}), y_1, \ldots, y_{n_j})$ and then we obtain $y_{l_1} = x_1, \ldots, y_{l_{n_i}} = x_{n_i}$. Let $\tilde{t} = f_j(\tilde{y}_1, \ldots, \tilde{y}_{n_j})$ such that $\tilde{y}_{l_1} = x_{i_1}, \ldots, \tilde{y}_{l_{n_i}} = x_{i_{n_i}}$ and $\tilde{y}_k = y_k$ if $k \in \{1, \ldots, n_j\} \setminus \{l_1, \ldots, l_{n_i}\}$. Then $\tilde{y} \in W_\tau(X_m)_1$ and

$$\begin{aligned}
&\hat{\sigma}[\tilde{t}] \\
&= \hat{\sigma}[f_j(\tilde{y}_1, \ldots, \tilde{y}_{n_j})] \\
&= S_m^{n_j}(\sigma(f_j), \tilde{y}_1, \ldots, \tilde{y}_{n_j}) \\
&= S_m^{n_j}(f_i(x_{l_1}, \ldots, x_{l_{n_i}}), \tilde{y}_1, \ldots, \tilde{y}_{n_j}) \\
&= f_i(\tilde{y}_{l_1}, \ldots, \tilde{y}_{l_{n_i}}) \\
&= f_i(x_{i_1}, \ldots, x_{i_{n_i}}).
\end{aligned}$$

This shows that $\hat{\sigma}$ is surjective on $W_\tau(X_m)_1$. Since $W_\tau(X_m)_1$ is finite, so σ is bijective. □

Then we obtain:

Corollary 3.1. *Let* $\tau = (n_i)_{i \in I}$ *be a type with* $n_i \geq 1$ *for all* $i \in I$ *and assume that* I *is finite. Then for each* $\sigma \in Hyp(\tau)$ *the following statements are equivalent:*

(i) $\sigma \in Surj(\tau)$.

(ii) For each $i \in I$ *there is a uniquely determined element* $j \in I$ *such that* $n_j \geq n_i$ *and such that* $\hat{\sigma}[f_j(y_1, \ldots, y_{n_j})] = f_i(x_1, \ldots, x_{n_i})$ *for some variables* $y_1, \ldots, y_{n_j} \in X_{n_j}$ *with* $|\{y_1, \ldots, y_{n_j}\}| = n_i$.

Proof. (ii)⇒(i) follows from Theorem 3.1.
(i)⇒(ii) If $\sigma \in Surj(\tau)$, then σ is bijective on $W_\tau(X_m)_1$ and this means that for every $f_i(x_1, \ldots, x_{n_i})$ there is exactly one j and there are variables $y_1, \ldots, y_{n_j} \in X_{n_j}$ such that $\hat{\sigma}[f_j(y_1, \ldots, y_{n_j})] = f_i(x_1, \ldots, x_{n_i})$. Further we have $n_j \geq n_i$ since $\hat{\sigma}$ is arity preserving. □

As a consequence we get that for finite types we have $Bij(\tau) = Surj(\tau)$.

Theorem 3.3. *Let $\tau = (n_i)_{i \in I}$ be a finite type and let σ be a hypersubstitution. Then σ is surjective iff σ is bijective.*

Proof. If σ is bijective, then σ is also surjective.

Assume that σ is surjective. Then by Lemma 3.1, we obtain that σ is bijective on the set $W_\tau(X_m)_1$. Now we consider $\hat{\sigma}$ as a sequence $(\hat{\sigma}_n)_{n \in \mathbb{N}^+}$ of mappings $\hat{\sigma}_n : W_\tau(X_n) \to W_\tau(X_n)$. Let $W_\tau(X_m)_1^n$ be the set of all n-ary terms contained in $W_\tau(X_m)_1$. Clearly $W_\tau(X_m)_1^m = W_\tau(X_m)_1$ since all elements in $W_\tau(X_m)_1$ are m-ary. By the definition of terms we know that every m-ary term can be regarded as n-ary term if $n \geq m$ and then $W_\tau(X_m)_1^n = W_\tau(X_m)_1^m$ for all $n \geq m$. Since $\hat{\sigma}$ is bijective on $W_\tau(X_m)_1$ and preserves arities, so $\hat{\sigma_n}|W_\tau(X_m)_1^n$ is also bijective on $W_\tau(X_m)_1^n$. We consider the sequences $(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$ and $\hat{\sigma}|W_\tau(X_m)_1 = \hat{\sigma_n}|(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$. Then the first sequence generates the free heterogeneous algebra $clone\tau$. But we know that $\hat{\sigma}|(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$ is bijective on $(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$ since $\hat{\sigma_n}|W_\tau(X_m)_1^n$ is bijective on $W_\tau(X_m)_1^n$ for all $n \in \mathbb{N}^+$. So $\hat{\sigma}|(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$ can be extended to a bijection $\hat{\sigma}$ on $(W_\tau(X_m)_1^n)_{n \in \mathbb{N}^+}$ and then $\hat{\sigma}$ is bijective on $W_\tau(X)$. □

4. i-closed Varieties

Now we consider a finite type τ_n where every operation symbol has the same arity $n \geq 1$. Let V be a variety of type τ_n and let Id^nV be the set of all identities satisfied in V consisting of n-ary terms. Let V_n be the variety defined by the equations Id^nV, i.e. $V_n = ModId^nV$. Clearly, $V \subseteq V_n$.

In [5], the concept of an i-closed variety was defined in the following way:

Definition 4.1. Let V be a variety and $\mathcal{B}$ be an algebra of type τ_n. Then V is called *i-closed* if whenever $\mathcal{A} \in V$ and $T(\mathcal{A}) \cong T(\mathcal{B})$, then $\mathcal{B} \in V$.

Let $\mathcal{M}$ be a submonoid of the monoid $\mathcal{H}yp(\tau)$ of all hypersubstitutions of type τ. A variety V is said to be *M-solid* if V contains all derived algebras $\sigma(\mathcal{A})$ for every $\sigma \in \mathcal{M}$. In [5] it was proved:

Theorem 4.1. *Let V be a variety of type τ_n which is the model class of its n-ary identities, i.e. $V = ModId^n V$. Then V is i-closed iff it is $Surj(\tau_n)$-solid.*

Now Theorem 3.3 gives:

Corollary 4.1. *Let V be a variety of a finite type τ_n which is the model class of its n-ary identities, i.e. $V = ModId^n V$. Then V is i-closed iff it is $Bij(\tau_n)$-solid.*

As an example we consider the variety $CI = Mod\{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1\}$ of commutative and idempotent groupoids.

Proposition 4.1. *The variety CI of all commutative and idempotent groupoids is i-closed.*

Proof. The type $\tau = (2)$ has the required form and we can apply Theorem 4.1. From $\{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1\} \subseteq Id^2CI$ there follows $CI \supseteq CI_2$ and therefore $CI = V_2$. By Theorem 3.1, we obtain $Bij(2) = \{\sigma_{id}, \sigma_{x_2x_1}\}$ where σ_t is the hypersubstitution mapping the binary operation symbol f to the term t. Since $\hat{\sigma}_{x_2x_1}[\{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1\}] = \{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1\}$, the variety CI is $Bij(2)$-solid. Therefore CI is i-closed. □

Let M be the monoid of all regular hypersubstitutions, i.e. all hypersubstitutions sending f to a binary term t containing x_1 and x_2. We remark that for this M all M-solid varieties of commutative idempotent groupoids were studied in [2]. Such M-solid varieties are called *regular-solid*. The variety $V_{1,2} = Mod\{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1, (x_1x_2)x_2 \approx x_1x_2\}$ is the least non-trivial regular-solid variety of commutative idempotent groupoids and $V_{2,2} = Mod\{x_1x_2 \approx x_2x_1, x_1^2 \approx x_1, x_1x_2^2 \approx x_2x_1^2\}$ is the greatest one. The lattice of all non-trivial regular-solid varieties of commutative idempotent groupoids is the interval $[V_{1,2}, V_{2,2}]$. Here $x_1x_2^2$ means $(x_1x_2)x_2$.

References

1. K. Denecke, *Menger Algebras and Clones of Terms*, East-West Journal of Mathematics, Vol.5, No.2 (2003), 179-193.
2. K. Denecke and P. Jampachon, *Regular-solid Varieties of Commutative and Idempotent Groupoids* in: Algebras and Combinatorics, An International Congress, ICAC'97, Hongkong, Springer (1999), 177-188.
3. K. Denecke, J. Koppitz and St. Niwczyck, *Equational Theories generated by Hypersubstitutions of type* (n), Int. Journal of Algebra and Computation, Vol. 12, No.6 (2002), 867-876.

4. K. Denecke and S. L. Wismath, *Hyperidentities and Clones*, (Gordon and Breach Science Publishers 2000).
5. J. Jampachon, *M-solid Varieties and Menger algebras of terms*, (Dissertation, Potsdam, 2007).
6. J. Koppitz and K. Denecke, *M-solid Varieties of Algebras*, Springer 2006.
7. J. Płonka, *Proper and inner hypersubstitutions of varieties*, Proceedings of the International Conference Sommer School on General Algebra and Ordered Sets, Olomouc (1994), 106-116.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 95–118)

CONTINUOUS COALGEBRA ENDOMORPHISMS OF SOME COMPLETE ULTRAMETRIC HOPF ALGEBRAS *

BERTIN DIARRA

Laboratoire de Mathématiques, Université Blaise Pascal, Complexe Scientifique des Cézeaux, 63 177 AUBIÈRE Cedex, France
E-mail: Bertin.Diarra@math.univ-bpclermont.fr

Due to properties of the topological tensor product of ultrametric Banach spaces, the algebraic notion of coalgebra has a natural ultrametric counterpart. An important class of ultrametric Banach coalgebras is provided by the spaces $\mathcal{C}(G, K)$ of continuous functions from a totally discontinuous compact group G with values in a complete ultrametric valued field K. The coproduct on $\mathcal{C}(G, K)$ is induced by the law of multiplication of the group G.

Related to ombral calculus is the characterization of the endomorphisms of a coalgebra. For a wide class of ultrametric Banach colagebras the monoid of the continuous coalgebra endomorphisms is anti-isomorphic to the monoid of continuous and weakly continuous algebra endomorphims of dual algebra. Hence, we recover earlier results obtained on the coalgebra $\mathcal{C}(\mathbb{Z}_p, K)$, where $\mathbb{Z}_p$ is the additive group of the ring of p-adic integers and K is a complete valued field of the field extension of p-adic numbers $\mathbb{Q}_p$. In addition for such ground fields K and more generally for K of residue characteristic p, we consider the coalgebra $\mathcal{C}(V_q, K)$, where V_q is the infinite compact monothetic subgroup of the group of units of K generated by q which is not a root of unit such that $|q^h - 1| < 1$, for h an integer ≥ 1. One obtains, as for formal power series, a procedure of substitution on the algebra of p-adic bounded measures $M(V_q, K)$ which gives the continuous and weak*-continuous algebra endomorphisms of $M(V_q, K)$ and which in turn gives the continuous coalgebra endomorphisms of $\mathcal{C}(V_q, K)$.

Keywords: Coalgebra endomorphism; Hopf algebra; Banach spaces.

1. Introduction

Let K be a complete ultrametric valued field.

Let us recall ([16]) that for two ultrametric Banach spaces over K, the topological tensor product of E and F is the completion $E\widehat{\otimes}F$ of the tensor

*2000*Mathematics subject classification*: Primary 16W30 Secondary, 05A40, 05A30, 26E30, 46S10.

product $E \otimes F$ with respect to the tensor norm that is defined for $z \in E \otimes F$ by setting $\|z\| = \|z\|_{\otimes} = \inf_{\sum x_i \otimes y_i = z} \max_i \|x_i\| \|y_i\|$.

A remarkable fact in ultrametric functional analysis is that if X a totally discontinuous (= zero-dimensional) compact topological space and E is an ultrametric Banach space, then for the Banach spaces of continuous functions $\mathcal{C}(X, E)$ and $\mathcal{C}(X, K)$ the canonical linear map Π_E from $\mathcal{C}(X, K) \widehat{\otimes} E$ into $\mathcal{C}(X, E)$ such that $\Pi_E(f \otimes x)(s) = f(s)x$, for $s \in X$ and $x \in E$ is an isometrical isomorphism of Banach spaces. Furthermore, if Y is an other zero-dimensional compact topological space then the linear map Π form $\mathcal{C}(X, K) \widehat{\otimes} \mathcal{C}(Y, K)$ into $\mathcal{C}(X \times Y, K)$ such that $\Pi(f \otimes g)(s, t) = f(s)g(t)$, for $s \in X$ and $t \in Y$ is again an isometrical isomorphism of Banach spaces.

A K- vector space is said to be a *Banach coalgebra*, if H is an ultrametric Banach space and if there exists a continuous linear operator $c : H \longrightarrow H \widehat{\otimes} H$ = topological tensor product (the coproduct) and a continuous linear form $\sigma : H \longrightarrow K$ (the counit) , such that

(i) $(c \otimes id) \circ c = (id \otimes c) \circ c$ (coassociativity) and

(ii) $(id \otimes \sigma) \circ c = id = (\sigma \otimes id) \circ c$. By transposition, the coproduct induces on the strong dual H' of H a structure of a *unitary normed algebra*, the product being the *convolution* given by $\mu \star \nu = (\mu \otimes \nu) \circ c$ and the unit is σ, with $\|\mu \star \nu\| \leq \|c\| \, \|\mu\| \|\nu\|$.
An important class of ultrametric Banach coalgebras is provided by the spaces $\mathcal{C}(G, K)$ of continuous functions from a totally discontinuous compact group G with values in K. More precisely, let ρ be the linear map from $\mathcal{C}(G, K)$ into $\mathcal{C}(G \times G, K)$ such that $\rho(f)(s, t) = f(st)$, then the coproduct on $\mathcal{C}(G, K)$ is $c = \Pi^{-1} \circ \rho$ with Π the above defined map for $X = Y = G$. The counit is the Dirac measure at the neutral element $\sigma(f) = f(e)$. In fact, if one assumes only that G is a totally discontinuous compact monoid then one again have a coalgebra structure on $\mathcal{C}(G, K)$ and this coalgebra is a bialgebra, i.e. the coproduct and the counit are algebra homomorphisms of $\mathcal{C}(G, K)$ into $\mathcal{C}(G, K) \widehat{\otimes} \mathcal{C}(G, K)$ (resp. into K). When G is a group, considering the linear endomorphism η of $\mathcal{C}(G, K)$ such that $\eta(f)(s) = f(s^{-1})$, one verifies that

(iii) $m \circ (1_H \otimes \eta) \circ c = k \circ \sigma = m \circ (\eta \otimes 1_H) \circ c$, where k is the canonical injection of K into $\mathcal{C}(G, K)$.
The map η is called the inversion (or antipode) of $\mathcal{C}(G, K)$. The coalgebra $\mathcal{C}(G, K)$ is then a complete ultrametric Hopf algebra. (see for instance [5])

The dual Banach algebra $M(G,K)$ of the coalgebra $\mathcal{C}(G,K)$ is the space of bounded measures of G with values in K, and the product being the usual convolution of measures.

Definition 1.1. A continuous linear endomorphism φ of the ultrametric Banach coalgebra H is said to be a *coalgebra endomorphism* if one has $c \circ \varphi = (\varphi \otimes \varphi) \circ c$ and $\sigma \circ \varphi = \sigma$.

It is readily seen that the continuous linear operator ${}^t\varphi$ on H' obtained by transposition of φ, is a continuous algebra endomorphism of the algebra H'.

Let us notice that if the duality $\langle H', H \rangle$ is separating and if H' is an integral domain, then any continuous linear endomorphism φ of H such that $c \circ \varphi = (\varphi \otimes \varphi) \circ c$ is the nul operator or is such that $\sigma \circ \varphi = \sigma$.

The duality theory of ultrametric Banach spaces is not so easy as in classical analysis. It may happens that, for non-spherically complete field K that the dual of certain Banach spaces is reduced to the null space ([16]). Nevertheless within a wide class of Banach spaces, considering on the strong dual H' of H the weak* topology $\sigma^*(H', H)$, one can show that the operation of transposition induces a bijection between continuous linear endomorphisms of H and continuous linear endomorphisms H' that are at the same time $\sigma^*(H', H)$-continuous. As a result, one also obtain a bijection between continuous coalgebra endomorphism of H and continuous algebra endomorphisms of H' that are at the same time $\sigma^*(H', H)$-continuous. Examples are provided by coalgebras which are pseudoreflexive Banach spaces, as for instance the coalgebras $\mathcal{C}(G,K)$.

In particular, we recover earlier results obtained on the coalgebra $\mathcal{C}(\mathbb{Z}_p, K)$, where $\mathbb{Z}_p$ is the additive group of the ring of p-adic integers and K is a complete valued field, extension of the field of p-adic numbers $\mathbb{Q}_p$. In addition for such ground fields K and more generally for K of residue characteristic p, we consider the coalgebra $\mathcal{C}(V_q, K)$, where V_q is the infinite compact monothetic subgroup of the group of units of K generated by q which is not a root of unit such that $|q^h - 1| < 1$, for h an integer ≥ 1. One obtains, as for formal power series, a procedure of substitution on the algebra of p-adic bounded measures $M(V_q, K)$ which gives the continuous and weak*-continuous algebra endomorphisms of $M(V_q, K)$ and which in turn gives the continuous coalgebra endomorphisms of $\mathcal{C}(V_q, K)$.

2. Weak*-continuous linear operators

The few reminders, on duality on ultrametric Banach spaces, which follow will be useful in the sequel.

Let E be an ultrametric Banach space over the complete valued field K.

The weak* topology $\sigma^*(E', E)$ on the Banach space E' dual of E is the locally convex toplogy on E' having as a fondamental system of neighbourhoods of zero the sets :

$W(0;\ x_1, \cdots, x_n, \varepsilon) = \{x' \in E' \ :\ |\langle x', x_i\rangle| < \varepsilon, 1 \leq i \leq n\}$, where $x_i \in H$ and $0 < \varepsilon \in \mathbb{R}$. Put $V = \langle x_1, \cdots, x_n\rangle$ be the vector subspace of E spanned by $(x_i)_{1\leq i\leq n}$, one has that the subspace $V^\perp = \{x' \in E' \ / \ \langle x', x\rangle = 0,\ \forall x \in V\}$ of E', i.e. the orthogonal of V in H', is contained in $W(0;\ x_1, \cdots, x_n, \varepsilon), \forall \varepsilon > 0$.

Proposition 2.1. *Let E be an ultrametric Banach space such that any finite dimensional subspace V of E has a topological complement in E. For the weak* topology on the dual space E' of E, one has that the dual space of the locally convex vector space $(E', \sigma^*(E', E))$ is equal to E.*

Proof. Let $f : E' \longrightarrow K$ be a $\sigma^*(E', E)$-continuous linear form. There exists a real number $\eta > 0$ and a finite family of elements $x_1, \cdots, x_{n_\epsilon} \in E$ such that for $x' \in W(0; x_1, \cdots, x_{n_\epsilon}, \eta_\varepsilon)$ one has $|f(x')| < 1$.

The subsapce $V_1 = \langle x_1, \cdots, x_{n_1}\rangle$ of E is such that $V_1^\perp \subset W(0; x_1, \cdots, x_{n_1}, \eta)$. Hence for $x' \in V_1^\perp$ and $\lambda \in K$, one has $|f(\lambda x')| < 1 \Longrightarrow |f(x')| < |\lambda^{-1}|,\ \forall \lambda \in K \setminus \{0\}$. Therefore $f(x') = 0, \forall x' \in V_1^\perp$, that is $V_1^\perp \subset ker f$.

By hypothesis, there exists a continuous linear projection p_1 from E onto V_1. For any (continuous) linear form v_1' from V_1 onto K, setting $x' = v_1' \circ p_1$, one obtains a continuous linear form from E into K whose restriction to V_1 is v_1'. It follows that the canonical map $\omega : E' \longrightarrow V_1'$ which associates to $x' \in E'$ its restriction $x'_{|V_1}$ is surjective. Obviously $ker\omega = V_1^\perp \subset ker f$ and $E'/V_1^\perp = V_1'$. Reducing modulo $V_1^\perp$, one obtains the linear from $\widetilde{f} : E'/V_1^\perp = V_1' \longrightarrow K$ such that $\widetilde{f}(\overline{x'}) = f(x')$, where for $x' \in E'$, $\overline{x'}$ = the class of x' modulo $V_1^\perp$.

Since $V_1 = \langle x_1, \cdots, x_{n_1}\rangle$ is finite dimensional, one has $(E'/V_1^\perp)' = V_1'' = V_1$, and $\widetilde{f} = \sum_{i=1}^{n} \alpha_i x_i$. One then has in the duality $\langle V_1, V_1'\rangle : f(\overline{x'}) = \sum_{i=1}^{n} \alpha_i \langle x_i, \overline{x'}\rangle$. However, for $x', y' \in E'$ such that for $x' - y' \in V_1^\perp$, one has

$\langle x'-y', x_i\rangle = 0, \forall i, 1 \leq i \leq n_1$. It follows that $\langle \sum_{i=1}^{n_1} \alpha_i x_i, x'\rangle = \langle \sum_{i=1}^{n_1} \alpha_i x_i, y'\rangle$ and $f(x') = \widetilde{f}(\overline{x'}) = \langle \sum_{i=1}^{n_1} \alpha_i x_i, x'\rangle$. Setting $x = \sum_{i=1}^{n_1} \alpha_i x_i$, one has $f(x') = \langle x', x\rangle, \forall x' \in E'$.

Let us remind ([16]) that an ultrametric Banach space is pseudoreflexive if the canonical map j_E of E into its bidual E'' is isometric, i.e. for any $x \in E$, one has $\|x\| = \sup_{x' \neq 0} \frac{|\langle x', x\rangle|}{\|x'\|}$. This condition is equivalent to say that E is norm polar according to [13]. □

Theorem 2.1. *Let E be an ultrametric Banach space. If E is pseudoreflexive, then any finite dimensional linear subspace of E has a topological complement in E.*

Proof. We proceed the proof by induction as in [13].

Let $x \in E$, $x \neq 0$. Then by hypothesis, one has $\|x\| = \sup_{x' \neq 0} \frac{|\langle x', x\rangle|}{\|x'\|}$. Hence, one sees that for a real number $r > 1$, there exits $x'_0 \in E'$ such that $\|x\| < r\frac{|\langle x'_0, x\rangle|}{\|x'_0\|}$. Setting for $y \in E$, $q_x(y) = \frac{\langle x'_0, y\rangle}{\langle x'_0, x\rangle} x$, one defines a continuous linear projection of E onto $K.x$ with norm $\|q_x\| \leq r$. Then one has the topological direct sum $E = K \cdot x \oplus \ker q_x$, with $\ker q_x = \ker x'_0$.

Let V be a finite dimensional subspace of E . Assume that any subspace of E with finite dimension $< \dim V$ has a topological complement in E. Let V_1 be a subspace of V such that $\dim V_1 = \dim V - 1$.

Let $p_1 : E \to E$ be a continuous linear projection from E onto V_1.

Fix $x_0 \in V \setminus V_1$, then one has $x_0 = p_1(x_0) + (id - p_1)(x_0) = p_1(x_0) + a$, with $a = x_0 - p_1(x_0) \in V \setminus V_1$ and $p_1(a) = 0$

Consider as above, the projection q_a of E onto $K \cdot a$.

Setting $\pi = q_a \circ (id - p_1)$, one again has a projection of E onto $K \cdot a$.

Furthermore $\|\pi\| \leq \|q_a\|\|id - p\| \leq \|q_a\|\|p\|$ and $\pi \circ p = 0$.

It is then readily seen that the continuous linear endomorphism $P = p_1 + \pi$ of E is a projection of E onto V and V has a topological complement in E equal the space $(id - P)(E)$. □

Corollary 2.1. *Let E be a pseudoreflexive ultrametric Banach space. Then the dual space of E' endowed with the weak* topology, that is $(E', \sigma^*(E', E))'$, is equal to E.*

Remark 2.1.

(i) For more informations on the duality theory, see for instance [16] and for complementation of subspaces, see the more recent paper [13] by C. Perez-Garcia and W.H. Schikhof.

(ii) Let us add here that if E and F are two pseudoreflexive Banach spaces, then their topological tensor product $E \otimes F$ is pseudoreflexive. The converse is also true ([13]).
If the ground field K is spherically complete, then by the Ingleton theorem, any ultrametric Banach space over K is pseudoreflexive.

(iii) By the proof of Theorem 2.1, one can prove that the condition in Proposition 1 can be restricted to a one-dimensional subspace and this is equivalent to say that any linear form on any one-dimensional space can be extended as a continuous linear form on all the space.

We are now ready to give an application of the above facts on duality theory, to the set of coalgebra endomorphisms of the class of Banach ultrametric coalgebras which are pseudoreflexive spaces

Theorem 2.2. *Let H be an ultrametric Banach coalgebra which is a pseudoreflexive normed space. Then the transposition of continuous linear operators induces a bijective correspondence between the set $End.cog(H)$ of the continuous coalgebra endomorphisms of H and the set $Alg_{\sigma^*}(H)$ of the continuous algebra endomorphisms of the normed algebra H' dual of H which are at the same time $\sigma^*(H', H)$-continuous.*

Proof. We first assume that E and F are Banach spaces which satisfy the condition of Proposition 2.1.

Let E and F be two ultrametric Banach spaces. If $u : E \longrightarrow F$ is a continuous linear operator, then by transposition, one obtains a norm-continuous operator ${}^t u : F' \longrightarrow E'$ that is weak*-continuous (E' and F' each endowed with its weak* topology. If furthermore E and F satisfy the condition of Proposition 1, if $v : F' \longrightarrow E'$ is a norm-continuous linear operator which is weak*-continuous, then one sees that ${}^t v : E'' \longrightarrow F''$ gives by restriction a continuous linear operator ${}^t v : E = (E', \sigma^*(E', E))' \longrightarrow F = (F', \sigma(F', F))'$.

Suppose that the Banach coalgebra H is a pseudoreflexive normed space.Then since $H\widehat{\otimes}H$ is also pseudoreflexive, one sees by transposition that ${}^{tt}c : H = (H', \sigma^*(H', H))' \longrightarrow ((H\widehat{\otimes}H)', \sigma^*((H\widehat{\otimes}H)', H\widehat{\otimes}H))' = H\widehat{\otimes}H$ is equal to the restriction of the map ${}^{tt}c : H'' \longrightarrow (H\widehat{\otimes}H)''$ and therefore coincides with c.

Let ψ be a norm-continuous algebra endomorphism of H' that is also weak*-continuous. Then by transposition, one has the continuous linear endomorphism ${}^t\psi$ of $H = (H', \sigma^*(H', H))'$.

For $x \in H$ and $x', y' \in H'$, one has

$$\begin{aligned}
\langle x' \otimes y', {}^t(\psi \otimes \psi) \circ c(x)\rangle &= \langle (\psi \otimes \psi)(x' \otimes y'), \sum_{i \geq i} x_i \otimes y_i\rangle \\
&= \sum_{i \geq 1} \langle \psi(x') \otimes \psi(y'), x_i \otimes y_i\rangle \\
&= \sum_{i \geq 1} \langle \psi(x'), x_i\rangle \langle \psi(y'), y_i\rangle \\
&= \sum_{i \geq 1} \langle x', {}^t\psi(x_i)\rangle \langle y', {}^t\psi(y_i)\rangle \\
&= \langle x' \otimes y', ({}^t\psi \otimes {}^t\psi)(\sum_{i \geq 1} x_i \otimes y_i)\rangle \\
&= \langle x' \otimes y', ({}^t\psi \otimes {}^t\psi) \circ c(x)\rangle.
\end{aligned}$$

It follows that ${}^t(\psi \otimes \psi) \circ c = ({}^t\psi \otimes {}^t\psi) \circ c$. But, by definition if ψ is an algebra endorphism, one has $\psi \circ {}^t c = {}^t c \circ (\psi \otimes \psi)$. Since ${}^{tt}c_{|H} = c$, one obtains by transposition that $c \circ {}^t\psi = {}^t(\psi \otimes \psi) \circ c = ({}^t\psi \otimes {}^t\psi) \circ c$.

On the other hand, since any continuous coalgebra endomorphism φ of H is such that its transpose ${}^t\varphi$ is a norm-continuous endomorphism of the complete algebra H' that is equally weak* -continuous, one obtains the bijective correspondence by noticing that ${}^{tt}\varphi = \varphi$ and ${}^{tt}\psi = \psi$. □

Corollary 2.2. *Let H be an ultrametric Banach coalgebras which is a pseudoreflexive normed space. The sets $End.cog(H)$ and $Alg_{\sigma^*}(H)$ with the law of composition of operators are monoids and the transposition of operators induces an anti-isomorphism of monoids.*

Remark 2.2. Here are some examples of pseudoreflexive normed spaces ([16])

(1) If the ground field K is spherically complete, then any ultrametric K-Banach space is pseudoreflexive.

(2) If X is a totally discontinuous compact topological space, then the space of continuous functions is pseudoreflexive as any Banach space of bounded continuous functions. If G is a totally discontinuous compact monoid, then Theorem 2.2 applies to the Banach coalgebra $\mathcal{C}(G, K)$.

(3) Any free Banach space (resp. Banach of countable type, i.e. containing a dense subspace of denumerable dimension) is pseudoreflexive. Theorem 2.2 can also be applied to the coalgebras of divided power ([6]).

We can now state the following corollary.

Corollary 2.3. *Let G be a totally discontinuous compact group (or more generally a totally discontinuous compact monoid). Then the monoid $End.cog(\mathcal{C}(G,K))$ of the continuous coalgebra endomorphisms of the Banach coalgebra $\mathcal{C}(G,K)$ is anti-isomorphic to the monoid $Alg_{\sigma^*}(M(G,K))$ of the norm-continuous algebra endomorphisms of the Banach algebra of bounded measures $M(G,K)$ which are at the same time weak*-continuous.*

3. Two examples of p-adic Hopf algebras

Let p be a prime number. We shall consider in the sequel that the ground field K is a complete valued field extension of the field of p-adic numbers $\mathbb{Q}_p$ (or more generally of residue characteristic p). We will consider the case where the group $G = \mathbb{Z}_p$ and the case where $G = V_q$ an infinite compact monothetic subgroup of the group of units of K.

3.1. *The divided powers Banach coalgebra of the p-adic continuous functions on $\mathbb{Z}_p$*

Assume that K is a valued field extension of $\mathbb{Q}_p$. We notice that the following statement is also true if the field is generally assumed to be of residue characteristic p.

The Banach algebra $\mathcal{C}(\mathbb{Z}_p,K)$ is a Hopf algebra. The coproduct c is such that $\Pi \circ c(f)(s+t) = f(s,t)$ and the counit is $\sigma(f) = f(0)$.

It is well known that the sequence of the binomial polynomial $B_0(s) = 1$, $B_n(x) = \binom{x}{n}$, $n \geq 1$ is an orthonormal basis of $\mathcal{C}(\mathbb{Z}_p,K)$ (the Mahler basis). Any continuous function $f \in \mathcal{C}(\mathbb{Z}_p,K)$ can be expanded as an uniformly convergent series $f = \sum_{n\geq 0} a_n B_n$, with $\|f\| = \sup_{n\geq 0} |a_n|$. Moreover $a_n = \Delta^n(f)(0)$, where $\Delta = \tau_1 - id$ and τ_1 is the operator of translation by $1 : \tau_1(f)(s) = f(s+1)$.

One can easily deduce from the Chu-Vandermonde identities $B_n(s+t) = \sum_{i+j=n} B_i(s)B_j(t)$ that $c(B_n) = \sum_{i+j=n} B_i \otimes B_j$. One says in this case that the sequence $(B_n)_{n\geq 0}$ is a powers divided sequence of polynomials of the coalgebra $\mathcal{C}(\mathbb{Z}_p,K)$.

It is readily seen that the dual Banach algebra $M(\mathbb{Z}_p,K)$ of $\mathcal{C}(\mathbb{Z}_p,K)$ with the convolution $\mu \star \nu = (\mu \otimes \nu) \circ c$ is isometrically isomorphic to the Banach algebra $K\langle X\rangle$ of the formal power series of bounded cœfficients

$S = \sum_{n\geq 0} b_n X^n$ with norm $\|S\| = \sup_{n\geq 0} |a_n|$. It is well known that the norm on $K\langle X\rangle$ is multiplicative (see for instance [14]). Furthermore, the closed ball of $K\langle X\rangle$ of center 0 and radius 1 is equal to the ring $\Lambda[[X]]$ of formal power series with cœfficients in the ring of valuation Λ of K. If $\mathcal{M}$ is the maximal ideal of Λ, then the set $\mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ is the unique maximal of the ring $\Lambda[[X]]$.

One identifies the algebra $M(\mathbb{Z}_p, K)$ with $K\langle X\rangle$ and puts $\sigma^* = \sigma^*(K\langle X\rangle, \mathcal{C}(\mathbb{Z}_p, K))$ the weak* topology on $K\langle X\rangle$.

Lemma 3.1. *Let φ be a continuous coalgebra endomorphism of $\mathcal{C}(\mathbb{Z}_p, K))$. Then the continuous algebra endomorphism ${}^t\varphi$ of $K\langle X\rangle$ is such that the formal power series ${}^t\varphi(X) = \sum_{n\geq 0} b_n X^n$ belongs to the maximal ideal $\mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ of $\Lambda[[X]]$.*

Proof. The transpose ${}^t\varphi$ of the continuous coalgebra endomorphism φ of $\mathcal{C}(\mathbb{Z}_p, K)$ is a continuous algebra endomorphism of $K\langle X\rangle$ with ${}^t\varphi(1) = 1$. Moreover, for any integer $n \geq 0$, one has $\|\varphi(X)\|^n = \|{}^t\varphi(X^n)\| \leq \|{}^t\varphi\|\|X^n\| = \|{}^t\varphi\|$. It follows that $\|{}^t\varphi(X)\| \leq 1$.

Put ${}^t\varphi(X) = \sum_{i\geq 0} b_n X^n$; then $\|{}^t\varphi(X)\| = \sup_{i\geq 0} |b_i| \leq 1$. Furthermore ${}^t\varphi(X^n) = {}^t\varphi(X)^n = \sum_{i\geq 0} b_i(n) X^i$, with $b_i(n) = \sum_{i_1+\cdots+i_n=i} b_{i_1}\cdots b_{i_n}$ and $\|{}^t\varphi(X^n)\| = \sup_{i\geq 0} |b_i(n)| \leq 1$. Since $1 = {}^t\varphi(1)$, one has $b_i(0) = \delta_{0,i}$(the Kroneker symbol).

The endomorphism ${}^t\varphi$ of $K\langle X\rangle$ obtained by transposition is σ^*-continuous. Hence, for any continuous function $f : \mathbb{Z}_p \longrightarrow K$, one has $\lim_{n\to+\infty} \langle {}^t\varphi(X^n), f\rangle = \lim_{n\to+\infty} \langle X^n, \varphi(f)\rangle = 0$. It follows that for any integer $\ell \geq 0$, one has $\lim_{n\to+\infty} \langle {}^t\varphi(X^n), B_\ell\rangle = 0$, $\lim_{n\to+\infty} b_\ell(n) = 0$, $\forall \ell \geq 0$.

In particular $\lim_{n\to+\infty} \langle {}^t\varphi(X^n), B_0\rangle = \lim_{n\to+\infty} b_0(n) = \lim_{n\to+\infty} b_0^n = 0$. Hence $|b_0| < 1$ and ${}^t\varphi(X)$ belongs to $\mathcal{N}$. □

For any integer $m \geq 0$, the expansion $\varphi(B_m) = \sum_{\ell\geq 0} \alpha_{m,\ell} B_\ell$ is such that $\alpha_{m,n} = \langle X^n, \varphi(B_m)\rangle = \langle {}^t\varphi(X^n), B_\ell\rangle = b_m(n), \forall n \geq 0$, i.e., $\varphi(B_m) = \sum_{\ell\geq 0} b_m(\ell) B_\ell$. As a consequence, one sees that $\|\varphi\| = 1$.

Notice that $c \circ \varphi(B_0) = (\varphi \otimes \varphi) \circ c(B_0) = \varphi(B_0) \otimes \varphi(B_0)$. This means

that $c(B_0)$ is a grouplike element of the coalgebra $\mathcal{C}(\mathbb{Z}_p, K)$. The grouplike elements of $\mathcal{C}(\mathbb{Z}_p, K)$ are readily seen to be the continuous characters of $\mathbb{Z}_p$ into K and are of the form $\chi_\alpha = B_0 + \sum_{n\ell \geq 1} \alpha^\ell B_\ell$, with $\alpha \in K$ such that $|\alpha| < 1$ ([7]).

In particular setting $\varphi(B_0) = \chi_\alpha$, one has $b_0 = \alpha$.

Obviously, any $S = \sum_{n\geq 0} a_n X^n \in K\langle X\rangle$ is σ^*-continuous. Clearly, one has in $K\langle X\rangle$, the σ^*-convergent series ${}^t\varphi(S) = \sum_{n\geq 0} a_n {}^t\varphi(X)^n$. Putting $u =^t \varphi(X)$. Then we can substitute u in the series S and we set ${}^t\varphi(S) = S \diamond u$.

What we have established for ${}^t\varphi$ is true for any norm-continuous algebra endomorphismψ of $K\langle X\rangle$ which is at the same time σ^*-continuous. In other words, one has $\psi(X) = u = \alpha + \sum_{n\geq 1} b_n X^n \in \mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ and for any $S \in K\langle X\rangle$, one has $\psi(S) = S \diamond u$.

Conversely ([7]) , if u belongs to the maximal ideal $\mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ of $\Lambda[[X]]$, then one can perform the substitution by u inside any formal power series S with bounded coefficients. Hence, by setting $\psi_u(S) = S \diamond u$, one has a norm-continuous algebra endomorphism of the Banach algebra $K\langle X\rangle$ that is also σ^*-continuous.

Let us notice that the law of substitution $\diamond$ induced on $\mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ is a structure of monoid with unit X.

Summing-up, we have the following theorem:

Theorem 3.1. *Let K be a complete valued field, extension of the field of p-adic numbers. Then*

(i) With the law of composition of linear operators, the monoid $End.cog(\mathcal{C}(\mathbb{Z}_p, K)$ of the continuous coalgebra endomorphisms of $\mathcal{C}(\mathbb{Z}_p, K)$ is anti-isomorphism to the monoid $Alg\sigma^*(K\langle X\rangle)$ of the continuous algebra endomorphism of the Banach algebra $K\langle X\rangle$ that are also weak-continuous.*

(ii) Moreover, the monoids $End.cog(\mathcal{C}(\mathbb{Z}_p, K)$ and $\mathcal{N} = \mathcal{M} + X\Lambda[[X]]$ are isomorphic.

The reader is referred to [7] for more details and complements on this example.

3.2. *The Hopf algebra $C(V_q, K)$ for K of residue characteristic p*

A topological group is said to be monothetic if it contains a dense subgroup that is generated by a single element.

In the sequel, we assume that K is a complete valued field, extension of the field of p-adic numbers $\mathbb{Q}_p$, or more generally a field of *residue characteristic* p, that is, K may be of characteristic p.

Let V be a compact monothetic subgroup of the group of units $U = \{a \in K, |a| = 1\}$ of the field K with residue characteristic p. Then in the residue field $\overline{K}$ of K, the residue set $\overline{V}$ is a finite subgroup of the multiplicative group of the non-zero elements of $\overline{K}$. Hence $\overline{V}$ is contained in a finite field $\mathbb{F}_{p^b} \subseteq \overline{K}$, and $b \geq 1$ is such that p^b is the least power of p such that $\overline{V} \subset \mathbb{F}_{p^b}$.

Let q be a topological generator of V which has the order h of $\overline{q} \in \overline{K}$ divides $\ell - 1$. Then we put $\zeta = \lim_{\ell \to +\infty} q^{p^{b\ell}}$ to be the Teichmüller representative of $\overline{q}$ in U. Now one has $|q - \zeta| < 1$ and $|q^h - 1| < 1$.

Let us notice that if $p = 2$, and $\overline{V} \subset \mathbb{F}_2$, then $\overline{q} = 1$, $h = 1$ and $\zeta = 1$.

We also set V_q equal the closure of the set $\{q^m, m \in \mathbb{N}\}$.

Put $q_1 = \zeta^{-1} q$. Then one sees $q_1 \in V_q$ and $|q_1 - 1| < 1$. For any $x \in \mathbb{Z}_p$ the series $q_1^x = \sum_{n \geq 0} \binom{x}{n} (q_1 - 1)^n$ converges, even uniformly with respect to x. Here, $\binom{x}{n}$ is the nth binomial polynomial function and if K is of characteristic p, $\binom{x}{n}$ will designate again the class of $\binom{x}{n}$ in the finite field with p elements $\mathbb{F}_p \subset K$.

Proposition 3.1. *Let K be a complete ultrametric valued field of residue characteristic $p \neq 0$, and V a monothetic compact subgroup of the group of units U of K. Let $q \in U$ be a topological generator of the monothetic compact subgroup V of U. Then $V = V_q =$ the adherence of $\{q^n \ / \ n \in \mathbb{N}\}$ in U.*

Moreover, one has $V_q = \bigcup_{0 \leq j \leq h-1} \zeta^j V_{q_1}$, where ζ is the Teichmüller representative of the residue class of q and $q_1 = \zeta^{-1} q$ is such that $|q_1 - 1| < 1$.

Proof. See[8]. □

We assume in what follows that the unit q is such that there exists an integer $h \geq 1$ such that $|q^h - 1|$ and that it is *not a root of unit*. Then the monothetic group V_q generated by q is an infinite compact group.

The Banach algebra $C(V_q, K)$ is then a complete Banach Hopf algebra.

Let us define the sequence of polynomials $(z-1)^{(n)} \in K[z]$ such that $(z-1)^{(0)} = 1$ and $(z-1)^{(n)} = (z-1)\cdots(z-q^{n-1})$, for $n \geq 1$.

It is associated with this sequence of polynomials the sequence of polynomials $Q_n(z) = \dfrac{(z-1)^{(n)}}{(q^n-1)^{(n)}}$.

Let us notice here that the q-integers are defined by setting $[n]_q = \dfrac{q^n-1}{q-1}$, the q-factorials being $[0]_q! = 1$ and $[n]_q! = [1]_q \cdots [n]_q$. Furthermore the q-binomial cœfficients are $\begin{bmatrix} n \\ j \end{bmatrix}_q = \dfrac{[n]!}{[n-j]_q![j]_q!}$, $0 \leq j \leq n$. One sees that $\begin{bmatrix} n \\ j \end{bmatrix}_q = Q_n(q^j)$

Let τ_q be the translation operator of $\mathcal{C}(V_q, K)$ defined by setting for $f \in \mathcal{C}(V_q, K)$, $\tau_q(f)(s) = f(qs)$, $s \in V_q$.

One also has the symbolic powers of the operator $\tau_q - id$ defined by setting $D_q^{(n)} = (\tau_q - id)^{(n)} = (\tau_q - id)\cdots(\tau_q - q^{n-1}id)$.

One verifies that $D_q^{(j)}(Q_n)(s) = q^{-j(n-j)}s^j Q_{n-j}(s), \forall 0 \leq j \leq n$. Hence $D_q^{(n)}(Q_n)(s) = s^n$ and $D_q^{(j)}(Q_n)(s) = 0, \forall j \geq n+1$.

Let X be the linear operator defined on $\mathcal{C}(V_q, K)$ such that for $f \in \mathcal{C}(V_q, K)$, $X(f)(s) = sf(s)$, $s \in V_q$. For any integer $n \geq 0$, one has $X^n(f)(s) = s^n f(s)$. In particular $X^n(1)(s) = s^n$ and one can write $D_q^{(j)}(Q_n) = q^{-j(n-j)}X^j(Q_{n-j}) = q^{-j(n-j)}X^j(1)Q_{n-j}$.

Theorem 3.2. *The sequence of polynomials $(Q_n)_{n\geq 0}$ is an orthonormal basis of the ultrametric Banach space $\mathcal{C}(V_q, K)$ of the continuous functions of V_q with valued in K.*

In other words, for any element $f \in \mathcal{C}(V_q, K)$, one has $f = \sum_{n\geq 0} a_n Q_n, a_n \in K, \lim_{n\to+\infty} a_n = 0$ *and* $\|f\| = \sup_{n\geq 0} |a_n|$.

Moreover, $a_n = D_q^{(n)}(f)(1)$.

Proof. See [8]. □

(i) As an application, one sees that for any continuous function $f : V_q \longrightarrow K$, and $s \in V_q$, one has $\tau_s(f) = \sum_{n\geq 0} D_q^{(n)}(\tau_s(f))(1)Q_n$, where $\tau_s(f)(t) = f(st)$. Since the operators τ_s and $D_q^{(n)}$ commute, one obtains that $\tau_s(f) = \sum_{n\geq 0} D_q^{(n)}(f)(s)Q_n$, that is $\rho(f)(s,t) = \sum_{n\geq 0} D_q^{(n)}(f)(s)Q_n(t)$. From this, one deduces that the coproduct of the Banach coalgebra

$\mathcal{C}(V_q,K)$ is given by $c(f)=\sum_{n\geq 0} D_q^{(n)}(f)\otimes Q_n$. In particular, on the above basis $(Q_n)_{n\geq 0}$, called the Van Hamme basis, one has $c(Q_m)=\sum_{n=0}^{m} D_q^{(n)}(Q_m)\otimes Q_n=\sum_{n+k=m} q^{-nk}X^n(Q_k)\otimes Q_n$.

(ii) Let us remind as a consequence of a general result on ultrametric Banach coalgebras ([6]), that the algebra $M(V_q,K)$ of bounded measures, that is the Banach dual $\mathcal{C}(V_q,K)'$ of the coalgebra $\mathcal{C}(V_q,K)$ is isometrically isomorphic to the algebra $W(V_q,K)$ of the continuous comodule endomorphisms of $\mathcal{C}(V_q,K)$; this space of linear operators coincides with the continuous linear endomorphisms u of $\mathcal{C}(V_q,K)$ that commute with the translations : $u\circ\tau_s=\tau_s\circ u, \forall s\in V_q$. More precisely, one associates to $\mu\in M(V_q,K)$ a unique element $(id\otimes\mu)\circ c\in W(V_q,K)$ and reciprocally, there corresponds to $u\in W(V_q,K)$ a unique measure $\mu=\sigma\circ u\in M(V_q,K)$, where σ is the counit $\sigma(f)=f(1)$.

We only verify here that setting for $u\in W(V_q,K)$, $\theta_1(u)=\sigma\circ u$, noticing by $\star$ the convolution product on $M(V_q,K)$, one has $\theta(u)\star\theta(v)=\theta_1(u)\circ\theta_1(v)$. Indeed, $v\in W(V_q,K)\Longleftrightarrow c\circ v=(id\otimes v)\circ v$. But for $f\in\mathcal{C}(V_q,K)$, one has $c(f)=\sum_{j\geq 1} f_j\otimes g_j \Longrightarrow v(f)=\sum_{j\geq 1} f_j\cdot(\sigma\circ v)(g_j)$. Then, one sees that $(\theta(u)\star\theta(v))(f)=\big((\sigma\circ u)\otimes(\sigma\circ v)\big)\circ c(f)=\sum_{j\geq 1}\sigma\circ u(f_j)\cdot\sigma\circ v)(g_j)=\sigma\circ u\big(\sum_{j\geq 1} f_j\cdot\sigma\circ v(g_j)\big)=(\sigma\circ u)\circ v(f)=\sigma\circ(u\circ v)(f)=\theta_1(u\circ v)(f)$.

Considering the dual family $(Q_n')_{n\geq 0}$ of the orthonormal basis $(Q_n)_{n\geq 0}$ of $\mathcal{C}(V_q,K)$, for $m\geq 0$ one has $(id\otimes Q_n')\circ c(Q_m)=\sum_{\ell+k=m} q^{-\ell k}X^\ell(Q_k)\otimes <Q_n',Q_\ell>=q^{-n(m-n)k}X^n(Q_{m-n})$, for $0\leq n\leq m$ and if $n\geq m+1$, one has $(id\otimes Q_n')\circ c(Q_m)=0$. It follows that $(id\otimes Q_n')\circ c=D_q^{(n)}$ and therefore $Q_n'=\sigma\circ D_q^{(n)}=(\varepsilon_q-\sigma)\star\cdots\star(\varepsilon-q^{n-1}\sigma)=(\varepsilon-\sigma)^{(n)}$, where ε_q is the Dirac measure $\varepsilon_q(f)=f(q)$.

(iii) We also know that any difference operator, i.e. any element $u\in W(V_q,K)$ can be written as a strong uniformly convergent series $u=\sum_{n\geq 1}\alpha_n D_q^{(n)}$, moreover $\|u\|=\sup_{n\geq 1}|\alpha_n|$. And in the same way, any bounded measure $\mu\in M(Vq,K)$ can be expanded as a weak* convergent series $\mu=\sum_{n\geq 1} b_n(\varepsilon-\sigma)^{(n)}$, with $\|\mu\|=\sup_{n\geq 0}|b_n|$.

More on the convolution product of $M(V_q, K)$.

One has in the ring of polynomials $K[z]$ the followings identities:

$$\begin{aligned}(z-1)^{(n)}(z-1)^{(k)} &= \sum_{i+j=n+k} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (z-1)^{(i)} \\ &= \sum_{i+j=n} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (z-1)^{(k+i)} \\ &= \sum_{i+j=k} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (z-1)^{(n+i)}.\end{aligned}$$

It follows that if

$\mu_1 = \sum_{n\geq 0} b_n^1(\varepsilon_q - \sigma)^{(n)}$ and $\mu_2 = \sum_{n\geq 0} b_n^2(\varepsilon_q - \sigma)^{(n)}$ are two elements of $M(V_q, K)$, one has

$$\begin{aligned}\mu_1 \star \mu_2 &= \sum_{n\geq 0}\sum_{k\geq 0} b_n^1 b_k^2 (\varepsilon_q - \sigma)^{(n)} \star (\varepsilon_q - \sigma)^{(k)} \\ &= \sum_{n\geq 0}\sum_{k\geq 0} b_n^1 b_k^2 \sum_{i+j=n} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (\varepsilon_q - \sigma)^{(k+i)} \\ &= \sum_{i,j}\sum_{k\geq 0} b_{i+j}^1 b_k^2 [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} i+j \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (\varepsilon_q - \sigma)^{(k+i)} \\ &= \sum_{\ell\geq 0} \gamma_\ell (\varepsilon_q - \sigma)^{(\ell)}\end{aligned}$$

with

$$\begin{aligned}\gamma_\ell &= \sum_{\substack{j\geq 0 \\ \ell}} \sum_{\substack{k+i=\ell \\ k}} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} i+j \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q b_{i+j}^1 b_k^2 \\ &= \sum_{k=0}^{\ell}\sum_{j=0} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} \ell-k+j \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q b_{\ell-k+j}^1 b_k^2.\end{aligned}$$

3.2.1. *Substitution in $M(V_q, K)$.*

Let φ be a continuous coalgebra endomorphism of the Banach coalgebra $C(V_q, K)$. Then the transpose ${}^t\varphi$ of φ is a continuous algebra endomorphism of the Banach algebra $M(V_q, K)$, that is, at the same time weak* continuous. However, the weak* basis $((\varepsilon_q - \sigma)^{(n)})_{n\geq 0}$ converge weakly towards 0. Hence we can see that in the weak* topology, $\lim_{\sigma^*, n\to+\infty} {}^t\varphi((\varepsilon_q - \sigma)^{(n)}) = 0$. Put ${}^t\varphi(\varepsilon_q) = \nu$. Then one has ${}^t\varphi((\varepsilon_q - \sigma)^{(n)}) = {}^t\varphi(\varepsilon_q - \sigma) \star \cdots \star {}^t\varphi(\varepsilon_q - q^{n-1}\sigma) = (\nu - \sigma) \star \cdots \star (\nu - q^{n-1}\sigma) = (\nu - \sigma)^{(n)}$ and the sequence $((\nu - \sigma)^{(n)})_{n\geq 0}$ converges weakly towards 0. Therefore, for any bounded measure μ expanded as a weak* convergent series $\mu = \sum_{n\geq 0} b_n(\varepsilon_q - \sigma)^{(n)}$, one has ${}^t\varphi(\mu) = \sum_{n\geq 0} b_n(\nu - \sigma)^{(n)}$.

In other words, one can substitute ν in the weak* expansion of μ and one sets ${}^t\varphi(\mu) = \mu \diamond \nu$. In fact for any norm continuous algebra endomorphism ψ of the Banach algebra $M(V_q, K)$ which is at the same time weakly continuous, setting $\nu = \psi(\varepsilon_q)$, one defines as above a sequence of bounded measures $\psi((\varepsilon_q - \sigma)^{(n)}) = (\nu - \sigma)^{(n)}$ which converges weakly towards 0. Hence one can perform the substitution of ν in the weak* expansion of any measure μ.

Lemma 3.2. *Let ψ be a norm-continuous algebra endomorphism of the Banach algebra which is also weakly continuous. Put $\psi(\varepsilon_q) = \nu$. Then, the sequence of measures $(\nu - \sigma)^{(n)} = \psi((\varepsilon_q - \sigma)^{(n)})$ converges weakly towards 0 and for any measure $\mu = \sum_{n\geq 0} b_n(\varepsilon_q - \sigma)^{(n)} \in MV_q, K)$ one has by substitution $\psi(\mu) = \sum_{n\geq 0} b_n(\nu - \sigma)^{(n)} = \mu \diamond \nu$.*

Conversely, let $\nu \in M(V_q, K)$ be such that the sequence $((\nu - \sigma)^{(n)})_{n\geq 0}$ converges weakly towards 0, then for any $\mu = \sum_{n\geq 0} b_n(\varepsilon_q - \sigma)^{(n)} \in M(V_q, K)$, the series $\mu \diamond \nu = \sum_{n\geq 0} b_n(\nu - \sigma)^{(n)}$ is weak - convergent.*

Furthermore the map $\psi : \mu \longrightarrow \mu \diamond \nu$ is a continuous algebra endomorphism of the Banach algebra $M(V_q, K)$ which is also weakly continuous.

Proof. It remains only to prove that if ν satisfies the required condition, then the mapping $\psi : \mu \longrightarrow \mu \diamond \nu$ which, is easily seen to be linear continuous both for the norm topology, and the weak* topology is an algebra homomorphism.

However the rule of the multiplication of the polynomials $(z-1)^{(n)}$ gives here

$$\begin{aligned}(\nu - \sigma)^{(n)} \star (\nu - \sigma)^{(k)} &= \sum_{i+j=n+k} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q (\nu - \sigma)^{(i)} \\ &= \sum_{i+j=n} [j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q.\end{aligned}$$

Then suppose that

$$mu_1 = \sum_{n\geq 0} b_n^1(\varepsilon_q - \sigma)^{(n)} \text{ and } \mu_2 = \sum_{n\geq 0} b_n^2(\varepsilon_q - \sigma)^{(n)},$$

are two elements of $M(V_q, K)$. Then one has

$$(\mu_1 \diamond \nu) \star (\mu_2 \diamond \nu) = \sum_{n\geq 0} \sum_{k\geq 0} b_n^1 b_k^2 (\nu - \sigma)^{(n)} (\nu - \sigma)^{(k)} = \sum_{\ell \geq 0} c_\ell (\nu - \sigma)^{(\ell)},$$

with $c_\ell = \sum_{k=0}^{\ell}\sum_{j=0}^{k}[j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} l-k+j \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q b^1_{\ell-k+j} b^2_k$.

But, we have seen above that the coefficients γ_ℓ of the measure $\mu_1 \star \mu_2$ are such that $\gamma_\ell = \sum_{k=0}^{\ell}\sum_{j=0}^{k}[j]_q!(q-1)^j q^{\frac{j(j-1)}{2}} \begin{bmatrix} l-k+j \\ j \end{bmatrix}_q \begin{bmatrix} k \\ j \end{bmatrix}_q b^1_{\ell-k+j} b^2_k$. Hence $\gamma_\ell = c_\ell$. It follows that $(\mu_1 \diamond \nu) \star (\mu_2 \diamond \nu) = (\mu_1 \star \mu_2) \circ \nu$. □

What we have just proved is that there is a *bijective correspondence* between the set of the continuous algebra endomorphisms of the Banach algebra $W(V_q, K)$ that are also weakly continuous and the bounded measures ν such that the sequence sequence $(\nu - \sigma)^{(n)})_{n\geq 0}$ converges weakly towards 0.

We recall that p^b is the least power of p such that in the residue class field $\overline{K}$ of K. Hence one has $\overline{q} \in \mathbb{F}_{p^b} \subseteq \overline{K}$, where $\mathbb{F}_{p^b}$ is the finite field with p^b elements.

Proposition 3.2. *Let $\nu \in M(V_q, K)$. Then the sequence $(\nu - \sigma)^{(n)})_{n\geq 0}$ converges weakly towards 0 if and only if the sequence $(\nu^{hp^{bk}})_{k\geq 0}$ converges weakly towards σ.*

Proof. The monomials z^n and the symbolic monomials $(z-1)^{(n)}$ are linked by the formulas $z^n = \sum_{j=0}^{n} \begin{bmatrix} n \\ j \end{bmatrix}_q (z-1)^{(j)}$ and $(z-1)^{(n)} = \sum_{j=0}^{n}(-1)^{n-j} \begin{bmatrix} n \\ j \end{bmatrix}_q q^{\frac{(n-j)(n-j-1)}{2}} z^j$.

Let $\nu \in M(V_q, K)$ be such that $\lim_{\sigma^*, n\to+\infty}(\nu - \sigma)^{(n)} = 0$. Then by considering the associated algebra endomorphism ψ of $M(V_q, K)$ defined by $\psi(\mu) = \mu \diamond \nu$, one has $\psi(\varepsilon_q) = \nu$. For any integer $k \geq 0$, one sees that $\psi(\varepsilon_{q^{hp^{bk}}}) = \nu^{hp^{bk}}$. Since $\lim_{k\to+\infty} q^{hp^{bk}} = 1$ and $\varepsilon_1 = \sigma$, one sees that $\lim_{\sigma^*, k\to+\infty} \nu^{hp^{bk}} = \sigma$.

Conversely, assume that $\lim_{\sigma^*, k\to+\infty} \nu^{hp^{bk}} = \sigma$.

Since $\sum_{j=0}^{n}(-1)^{n-j} \begin{bmatrix} n \\ j \end{bmatrix}_q q^{\frac{(n-j)(n-j-1)}{2}} = 0$,

one sees that $(\nu - \sigma)^{(n)} = \sum_{j=0}^{n}(-1)^{n-j} \begin{bmatrix} n \\ j \end{bmatrix}_q q^{\frac{(n-j)(n-j-1)}{2}} (\nu^j - \sigma)$

and for any continuous function $f : V_q \longrightarrow K$, one has

$$\langle(\nu-\sigma)^{(n)},f\rangle=\sum_{j=0}^{n}(-1)^{n-j}\begin{bmatrix}n\\j\end{bmatrix}_q q^{\frac{(n-j)(n-j-1)}{2}}\langle(\nu^j-\sigma),f\rangle.$$

Then, by an appropriate splitting of the index summation $0\leq j\leq n$. One can easily prove as in the proof Theorem 12 of [6] (see also [8]) that $\lim_{n\to+\infty}\langle(\nu-\sigma)^{(n)},f\rangle=0$.

Note

(1) Let ν be an element of $M(V_q,K)$. Then one deduces from the identities $\nu^n=\sum_{j=0}^{n}\begin{bmatrix}n\\j\end{bmatrix}_q(\nu-\sigma)^{(j)}$ and $(\nu-\sigma)^{(n)}=\sum_{j=0}^{n}(-1)^{n-j}\begin{bmatrix}n\\j\end{bmatrix}_q q^{\frac{(n-j)(n-j-1)}{2}}\nu^j$ that $\sup_{n\geq0}\|\nu^n\|=\sup_{n\geq0}\|(\nu-\sigma)^{(n)}\|$.

(2) Let ψ be a continuous algebra endomorphism of $M(V_q,K)$. Since ε_q is an invertible element of $M(V_q,K)$, the measure $\nu=\psi(\varepsilon_q)$ is also invertible.

Moreover, for any integer $n\geq1$, one has

$$\nu^n=\psi(\varepsilon_{q^n})\Longrightarrow\|\nu^n\|\leq\|\psi\|\Longrightarrow\rho(\nu)=\lim_{n\to+\infty}\sup\|\nu^n\|^{\frac{1}{n}}\leq1.$$

By the same way, $\rho(\nu^{-1})\leq1$. Hence, one obtains $\rho(\nu)=1$.

Assume furthermore that ψ is also weakly continuous. Then $\psi(\mu)=\mu\diamond\nu=\sum_{n\geq0}b_n(\nu-\sigma)^{(n)}\Longrightarrow\|\psi(\mu)\|\leq\sup_{n\geq0}|b_n|\|(\nu_\sigma)^{(n)}\|\leq\|\mu\|\sup_{n\geq0}\|(\nu-\sigma)^{(n)}\|$. It follows that $\|\psi\|=\sup_{n\geq0}\|(\nu-\sigma)^{(n)}\|=\sup_{n\geq0}\|\nu^n\|$. □

Remark 3.1. Let $q\in K$ be such that $|q-1|<1$ and not a root of unit. Then the map $z\longrightarrow q^z=\sum_{n\geq0}(q-1)^n\binom{z}{n}$ is an isomorphism of the compact additive group $\mathbb{Z}_p$ onto the compact multiplicative group V_q : $\binom{z}{n}$ is taken to be its residue class if K is of characteristic p. With such q, the Hopf algebras $\mathcal{C}(\mathbb{Z}_p,K)$ and $\mathcal{C}(V_q,K)$ are isometrically isomorphic. It follows that the Banach algebras $M(\mathbb{Z}_p,K)$ and $M(V_q,K)$ are also isometrically isomorphic. Hence, as for the norm of $M(\mathbb{Z}_p,K)$, that of $M(V_q,K)$ is multiplicative. Then, as a consequence, one sees that if ψ is a continuous endomorphism of the Banach algebra $M(V_q,K)$ then $\|\psi\|=1$. If in addition ψ is weakly continuous, then the measure $\nu=\psi(\varepsilon_q)$ is such that $\|\nu\|=1$.

Proposition 3.3. *Let ν_1 and ν_2 be two elements of $M(V_q,K)$ such that one has the weak limits :* $\lim_{\sigma^*,k\to+\infty}\nu_1^{hp^{bk}}=\sigma=\lim_{\sigma^*,k\to+\infty}\nu_2^{hp^{bk}}$.

Then ν_1 and ν_2 can be substituted to ε_q in any bounded measure μ expanded with respect to the weak basis $(\varepsilon_q - \sigma)^{(n)})_{n\geq 0}$ and one has $\mu \diamond (\nu_1 \diamond \nu_2) = (\mu \diamond \nu_1) \diamond \nu_2$.*

For the law of substitution $\diamond$, the set $\mathcal{S}$ of the bounded measures ν such that $\lim_{\sigma^*, k\to+\infty} \nu^{hp^{bk}} = \sigma$ is a monoid with unit ε_q.

Proof. Applying the previous proposition, one can use the substitution by ν_1 and ν_2. Let ψ_1 and ψ_2 be the continuous algebra endomorphisms of the unitary Banach algebra $M(V_q, K)$ associated with ν_1 and ν_2 respectively. then for $\mu \in M(V_q, K)$, one has $\psi_2 \circ \psi_1(\mu) = \psi_1(\mu) \diamond \nu_2 = (\mu \diamond \nu_1) \diamond \nu_2$. Obviously, the substitution in the measure ε_q by $\nu \in \mathcal{S}$ leaves ν unchanged. Hence $\psi_1(\varepsilon_q) = \varepsilon_q \diamond \nu_1 = \nu_1$. Also $\psi_2 \circ \psi_1(\varepsilon_q) = \psi_1(\varepsilon) \diamond \nu_2 = \nu_1 \diamond \nu_2$. It follows that $\nu_1 \diamond \nu_2$ is the measure associated with the algebra endomorphism $\psi_2 \circ \psi_1$. Therefore $\psi_2 \circ \psi_1(\mu) = \mu \diamond (\nu_1 \diamond \nu_2)$ and one as $(\mu \diamond \nu_1) \diamond \nu_2 = \mu \diamond (\nu_1 \diamond \nu_2)$. It is then clear that the law of substitution $\diamond$ is associative on $\mathcal{S}$.

It is readily seen that $\lim_{\sigma^*, k\to+\infty} \varepsilon_q^{hp^{bk}} = \varepsilon_1 = \sigma$, i.e. σ belongs to $\mathcal{S}$.

Notice that one has $\mu \diamond \varepsilon_q = \mu$, i.e. the identity map I is the continuous algebra endomorphism of $M(V_q, K)$ associated with ε_q. In particular, for $\nu \in \mathcal{S}$, one has $\nu \diamond \varepsilon_q = \nu$ and ε_q is the unity of $\mathcal{S}$, because $\varepsilon_q \diamond \nu = \nu$. □

Note

The counit σ belongs to $\mathcal{S}$. For any $\nu \in \mathcal{S}$, one has $\sigma \diamond \nu = \sigma$ and $\nu \diamond \sigma = c_0 \cdot \sigma$, where c_0 is the cœfficient of order 0 in the weak* expansion of ν.

Corollary 3.1. *Let $Alg_{\sigma^*}(M(V_q, K))$ be the set of the continuous algebra endomorphisms of the Banach algebra $M(V_q, K)$ that are also weak* continuous. For the law of composition of linear operators, $Alg_{\sigma^*}(M(V_q, K))$ is a monoid that is anti-isomorphic to the monoid $(\mathcal{S}, \diamond)$.*

Proof. This corollary follows immediately from Proposition 3.2, 3.3. □

Note

(i) Let $\psi \in Alg_{\sigma^*}(M(V_q, K))$, Then ψ is an automorphism of algebra if and only if $\nu = \psi(\varepsilon_q)$ is an invertible element of the monoid $(\mathcal{S}, \diamond)$.

(ii) $C(V_q, K)$ is a Banach space of countable type. If K is a *non-spherically complete* valued field, then the Banach spaces $C(V_q, K)$ and $M(V_q, K)$ are reflexive. This is a consequence of a well known theorem of M. van der Put ([16]). In this case one has $Alg_{\sigma^*}(M(V_q, K)) = Alg(M(V_q, K))$ which is the set of the continuous endomorphisms of $M(V_q, K)$.

Cœfficients of the symbolic powers $(\nu-\sigma)^{(n)}$ of a measure ν.

Let $\nu=\sum_{\ell\geq 0} c_\ell(\varepsilon_q-\sigma)^{(\ell)}$ be the expansion of the measure ν with respect to the weak* basis $((\varepsilon_q-\sigma)^{(\ell)})_{\ell\geq 0}$. For any integer $j\geq 0$, one also has $\nu-q^j\sigma=(c_0-q^j)\sigma)+\nu_0$, with $\nu_0=\sum_{\ell\geq 1} c_\ell(\varepsilon_q-\sigma)^{(\ell)}$. One has $(\nu-\sigma)^{(n)}=\prod_{j=0}^{n-1}((c_0-q^j)\sigma+\nu_0)=\sum_{\ell\geq 0} c_\ell(n)(\varepsilon_q-\sigma)^{(\ell)}$, with $c_0(n)=\prod_{j=0}^{n-1}(c_0-q^j)=(c_0-1)^{(n)}$.

The coefficients of the measures $(\nu-\sigma)^{(n)}$ can be obtained by using induction as follows. Set $\nu-q^n\sigma=(c_0-q^n)\sigma+\sum_{k\geq 1} c_k(\varepsilon_q-\sigma)^{(k)}$.

Then, by applying the formula of the product of two measures, one has $(\nu-\sigma)^{(n+1)}=(\nu-\sigma)^{(n)}\star(\nu-q^n\sigma)=\sum_{\ell\geq 0} c_\ell(n+1)(\varepsilon_q-\sigma)^{(\ell)}$ and $c_\ell(n+1)=$

$$c_\ell(n)(c_0-q^n)+\sum_{k=1}^{\ell}\sum_{j=0}^{k}[j]_q!(q-1)^j q^{\frac{j(j-1)}{2}}\begin{bmatrix}l-k+j\\ n\end{bmatrix}_q\begin{bmatrix}k\\ j\end{bmatrix}_q c_{\ell-k+j}(n)c_k.$$

For instance $c_1(n+1)=(c_0-q^n+(q-1)c_1)c_1(n)+c_1c_0(n)$.
If $c_0=1$, then $c_1(n)=0,\forall n\geq 1$. Hence, for $n\geq 2$, one has

$$c_1(n)=(1-q^n+(q-1)c_1)c_1(n-1),$$

and

$$c_1(n)=c_1\prod_{j=1}^{n-1}(1-q^j+(q-1)c_1).$$

Furthermore, if $c_1\neq 0$, then one can obtain $c_1(n)=\frac{1}{q-1}\cdot((1+(q-1)c_1)-1)^{(n)}$.

Remark 3.2. Let $\nu^n=\sum_{\ell\geq 0}\gamma_\ell(n)(\varepsilon_q-\sigma)^{(\ell)}$ be the expansion of the nth power of ν. Then one has :

$$c_\ell(n)=\sum_{j=0}^{n}(-1)^{n-j}q^{\frac{(n-j)(n-j-1)}{2}}\begin{bmatrix}n\\ j\end{bmatrix}_q\gamma_\ell(j) \text{ and } \gamma_\ell(n)=\sum_{j=0}^{n}\begin{bmatrix}n\\ j\end{bmatrix}_q c_\ell(j).$$

Note

On can prove that the cœfficients $\gamma_\ell(n)$ of the expansion of ν^n, $n\geq 2$ are given by the formulas:

$$\gamma_\ell(n)=\sum_{j_1\geq 0}\cdots\sum_{j_n\geq 0}\sum_{i_1+\cdots+i_n=\ell}\alpha_q(i_1,\cdots,i_n;j_1,\cdots,j_n)=$$

$\sum_{j_1\geq 0}\cdots\sum_{j_n\geq 0}\sum_{i_1+\cdots+i_n=\ell}\prod_{s=0}^{n}[j_s]_q!(q-1)^{j_s}q^{\frac{j_s(j_s-1)}{2}}\begin{bmatrix}i_1+\cdots i_{s-1}\\ j_s\end{bmatrix}_q\begin{bmatrix}i_s+j_s\\ j_s\end{bmatrix}_q$. The sums are finite, since for $2\leq s\leq n$, we have $\begin{bmatrix}i_1+\cdots i_{s-1}\\ j_s\end{bmatrix}_q=0, \forall j_s\geq \ell+1$, and in this case, $i_1+\cdots+i_{s-1}\leq \ell<\ell+1\leq j_s$.

Notice that $\gamma_0(n)=c_0^n$. As a corollary of Proposition 3.2, we have the following proposition

Proposition 3.4. *Let* $\nu^n=\sum_{\ell\geq 0}\gamma_\ell(n)(\varepsilon_q-\sigma)^{(\ell)}$ *be the expansion of the nth power of the measure* ν. *Then* ν *belong to* $\mathcal{S}$ *if and only if* $\lim_{k\to+\infty}\gamma_0(hp^{bk})=1$ *and* $\lim_{k\to+\infty}\gamma_\ell(hp^{bk})=0,\ \forall\ell\geq 1$.

Note

We observe that $\gamma_0(n)=c_0^n$. If $\nu\in\mathcal{S}$, then one has $\lim_{k\to+\infty}c_0^{hp^{bk}}=1$. This is equivalent to say that c_0 belongs to the set $\mathcal{W}=\bigcup_{0\leq j\leq h-1}\zeta^j(1+\mathcal{M})$, where $\mathcal{M}$ is the maximal ideal of the ring of valuation Λ of K and ζ is a primitive root of unit of order $h\in K$.

3.2.2. *The coalgebra endomorphisms of* $\mathcal{C}(V_q,K)$.

Let $End.cog(\mathcal{C}(V_q,K))$ be the continuous coalgebra endomorphims of the ultrametric complete Hopf algebra $\mathcal{C}(V_q,K)$. With the law of composition of linear operators, $End.cog(\mathcal{C}(V_q,K))$ becomes a monoid.

Theorem 3.3. *Any element* φ *of* $End.cog(\mathcal{C}(V_q,K))$ *is uniquely determined by the measure* $\nu={}^t\varphi(\varepsilon_q)$.

Moreover, the mapping $\varphi\longrightarrow{}^t\varphi(\varepsilon_q)$ *is an isomorphism of the monoid* $(End.cog(\mathcal{C}(V_q,K)),\circ)$ *onto the monoid* $(\mathcal{S},\diamond)$.

Proof. Being a free Banach space and also a space of bounded functions, $\mathcal{C}(V_q,K)$ is pseudreflexive Banach space. Hence, the operation of transposition of operators induces an anti-isomorphism of the monoid $End.cog(\mathcal{C}(V_q,K))$ onto the monoid $Alg_{\sigma^*}(M(V_q,K))$. Then, by applying the Corollary 3.1, we see that the monoids $(End.cog(\mathcal{C}(V_q,K)),\circ)$ and $(\mathcal{S},\diamond)$ are isomorphic. □

Corollary 3.2. *Let* φ *be an element of the monoid* $(End.cog(\mathcal{C}(V_q,K))$ *and* $\nu={}^t\varphi(\varepsilon_q)$ *be its corresponding element in* $\mathcal{S}$.

Let $(Q_n)_{n\geq 0}$ *be the Van Hamme basis of* $C(V_q, K)$ *and* $\varphi(Q_m) = \sum_{k\geq 0} \alpha_k(m) Q_k$ *be the expansion of* $\varphi(Q_m)$ *with respect to this basis.*

If $(\nu - \sigma)^{(n)} = {}^t\varphi((\varepsilon_q - \sigma)^{(n)}) = \sum_{\ell\geq 0} c_\ell(n)(\varepsilon_q - \sigma)^{(\ell)}$ *is the expansion of the n-th symbolic power of* ν *with respect to the weak* basis* $(\varepsilon_q - \sigma)^{(n)}$ *of* $M(V_q, K)$, *then we have* $\alpha_n(m) = c_m(n)$, $\forall n, m \geq 0$.

Proof. Indeed for the integers m and $n \geq 0$, we have

$$\begin{aligned}
\alpha_n(m) &= \langle (\varepsilon_q - \sigma)^{(n)}, \varphi(Q_m)\rangle \\
&= \langle {}^t\varphi((\varepsilon_q - \sigma)^{(n)}), Q_m\rangle \\
&= \langle (\nu - \sigma)^{(n)}, Q_m\rangle \\
&= \langle (\sum_{\ell\geq 0} c_\ell(n)(\varepsilon_q - \sigma)^{(\ell)}, Q_m\rangle \\
&= c_m(n).
\end{aligned}$$

□

Examples of elements of $\mathcal{S}$

(1) *The Dirac measures* ε_s.

Let $s \in V_q$. The one has $s = \zeta^j \cdot q_1^x$, where ζ is a primitive root of unit of order h, $q_1 = \zeta^{-1}q$ and x is a p-adic integer. One also has $\lim_{k\to+\infty} s^{hp^{bk}} = 1$. Hence, for any continuous function $f : V_q \longrightarrow K$, we have

$$\lim_{k\to+\infty} \langle \varepsilon_{s^{hp^{bk}}}, f\rangle = \lim_{k\to+\infty} f(s^{hp^{bk}}) = f(1) = \langle \sigma, f\rangle.$$

Therefore ε_s belongs to $\mathcal{S}$.

For $s, t \in V_q$, we have

$$\tau_s(f)(t) = f(st) = \sum_{n\geq 0} D_q^{(n)}(f)(t) Q_n(s).$$

Hence

$$\tau_s = \sum_{n\geq 0} Q_n(s) D_q^{(n)} \text{ and } \varepsilon_s = \sum_{n\geq 0} Q_n(s)(\varepsilon_q - \sigma)^{(n)}.$$

(a) For m an integer ≥ 0, we have $\varepsilon_{q^m} = \sum_{j=0}^{m} \begin{bmatrix} m \\ j \end{bmatrix}_q (\varepsilon_q - \sigma)^{(j)}$.

(b) If m is an integer ≥ 1, then we have :

$$\varepsilon_{q^{-m}} = \sum_{n\geq 0} (-1)^n q^{\frac{-n(2m+n-1)}{2}} \begin{bmatrix} m+n-1 \\ n \end{bmatrix}_q (\varepsilon_q - \sigma)^{(n)}.$$

In particular $\varepsilon_{q^{-1}} = \sum_{n\geq 0} (-1)^n (\varepsilon_q - \sigma)^{(n)}$.

(c) Since for any integer $n \geq 0$, we have $\varepsilon_{s^n} = \sum_{\ell \geq 0} Q_\ell(s^n)(\varepsilon_q - \sigma)^{(\ell)}$, i.e. $\gamma_{s,\ell}(n) = Q_\ell(s^n)$. We also have $(\varepsilon_s - \sigma)^{(n)} = \sum_{\ell \geq 0} c_{s,\ell}(n)\varepsilon_q - \sigma)^{(\ell)}$, with $c_{s,\ell}(n) = \sum_{j=0}^{n} (-1)^{n-j} q^{\frac{(n-j)(n-j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q Q_\ell(s^j)$.
For $s = q^m$, $m \geq 0$, we deduce that :
$c_{q^m,\ell}(n) = \sum_{j=0}^{n} (-1)^{n-j} q^{\frac{(n-j)(n-j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \begin{bmatrix} mj \\ l \end{bmatrix}_q$.
Hence $c_{q^m,\ell}(n) = 0$, $\forall \ell > mn$ and $(\varepsilon_{q^m} - \sigma)^{(n)} = \sum_{\ell=0}^{mn} c_{q^m,\ell}(n)(\varepsilon_q - \sigma)^{(\ell)}$ is a polynomial function in the "variable" ε_q.

(d) In any case, the coalgebra endomorphism φ_s corresponding to ε_s is given by $\varphi_s(Q_n) = \sum_{\ell \geq 0} c_{s,n}(\ell) Q_\ell$.

(2) *Let $\nu \in \mathcal{S}$ and $\alpha \in \mathcal{W} = \bigcup_{0 \leq j \leq h-1} \zeta^j(1 + \mathcal{M})$. Then $\alpha \cdot \nu$ belongs to $\mathcal{S}$.*

Indeed for any continuous function $f : V_q \longrightarrow K$, we have :
$\lim_{k \to +\infty} < (\alpha \cdot \nu)^{hp^{bk}}, f >= \lim_{k \to +\infty} \alpha^{hp^{bk}} \cdot < \nu^{hp^{bk}}, f >= \sigma(f)$.

- For $\alpha \in \mathcal{W}$, one has $\alpha \cdot \sigma \in \mathcal{S}$ and for any integer $n \geq 0$, one has $(\alpha \cdot \sigma - \sigma)^{(n)} = (\alpha - 1)^{(n)} \cdot \sigma$. Furthermore for any measure $\mu = \sum_{n \geq 0} b_n(\varepsilon_q - \sigma)^{(n)}$, one has $\mu \diamond (\alpha \cdot \sigma) = \left(\sum_{n \geq 0} b_n(\alpha - 1)^{(n)} \right) \cdot \sigma$.
- If $\alpha \in \mathcal{W}$ and $\nu \in \mathcal{S}$, with previous notations, one obviously has $(\alpha \cdot \nu)^n = \sum_{\ell \geq 0} \alpha^n \gamma_\ell(n)(\varepsilon_q - \sigma)^{(\ell)}$ and $(\alpha \cdot \nu - \sigma)^{(n)} = \sum_{\ell \geq 0} c_\ell^\alpha(n)(\varepsilon_q - \sigma)^{(\ell)}$ with $c_\ell^\alpha(n) = \sum_{j=0}^{n} (-1)^{n-j} q^{\frac{(n-j)(n-j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \alpha^j \gamma_\ell(j)$. In particular, for $\nu = \varepsilon_s$, one has $c_{s,\ell}^\alpha(n) = \sum_{j=0}^{n} (-1)^{n-j} q^{\frac{(n-j)(n-j-1)}{2}} \begin{bmatrix} n \\ j \end{bmatrix}_q \alpha^j Q_\ell(s^j)$.
- For α and $\beta \in \mathcal{W}$, one has $(\alpha \cdot \varepsilon_q) \diamond (\beta \cdot \varepsilon_q) = (\alpha\beta) \cdot \varepsilon_q$. Hence $(\alpha \cdot \varepsilon_q)$ is reversible, i.e. invertible in $\mathcal{S}, \diamond)$ with reverse $\alpha^{-1} \cdot \varepsilon_q$.
The corresponding algebra endomorphism $\psi : \mu \longrightarrow \psi(\mu) = \mu \diamond (\alpha \cdot \varepsilon_q)$ of $M(V_q, K)$ is an automorphism.

Remark 3.3. Let $\nu = \sum_{l\geq 0} c_l(\varepsilon_q - \sigma)^{(l)} \in \mathcal{S}$. Then ν is reversible if and only if $\nu_1 = c_0^{-1} \cdot \nu$ is reversible. In this case,we use $\nu^{<-1>}$ to denote the reverse of ν. Then, we have $\nu^{<-1>} = \nu_1^{<-1>} \diamond (c_0^{-1} \cdot \varepsilon_q)$.

If $\nu = \sigma + \sum_{\ell\geq 1} c_\ell(\varepsilon_q - \sigma)^{(\ell)} \in \mathcal{S}$ is reversible, then we have $c_1(1) = c_1 \neq 0$ and $c_1(n) = \frac{1}{q-1} \prod_{j=0}^{n-1}(1 - q^j + (q-1)c_1)$. The reverse $\nu^{<-1>} = \sum_{\ell\geq 0} d_\ell(\varepsilon_q - \sigma)^{(\ell)} \in \mathcal{S}$. is such that $d_0 = 1$ and d_n can be obtained by the infinite system of linear equations :

$$(\mathrm{I}) \begin{cases} \sum_{n\geq 1} d_n c_1(n) = 1 \\ \sum_{n\geq 1} d_n c_\ell(n) = 0, \ \ell \geq 2 \end{cases}$$

References

1. G. Boole (edited by J. F. Moulton), *Calculus of finite differences*, 4th edition (Chelsa, New York).
2. N. Bourbaki, *Fonctions d'une variable réelle* - Chap. VI - Diffusion (C.C.L.S., Paris, 1976).
3. L. Comtet, *Analyse combinatoire - Tome Premier*, Collection SUP - Presses Universitaires de France (Paris 1970).
4. B. Diarra, *Bases de Mahler et autres*, Séminaires d'Analyse - Université Blaise Pascal (1994-95) Exposé 16- MR, 98e : 46093.
5. B. Diarra, *Algèbres de Hopf et fonctions presque périodiques ultramétriques*, Rivista di Matematica pura ed applicata, 17 (1996), 113-132.
6. B. Diarra, *Complete ultrametric Hopf algebras which are free Banach spaces*, in p-adic functional analysis, edited by W. H. Schikhof, C. Perez-Garcia, J. Kąkol, Lecure notes in pure and applied mathematics , vol. 192, Marcel Dekker Inc., New York (1997), 61-80.
7. B. Diarra, *The continuous coalgebra endomorphisms of* $C(\mathbb{Z}_p, K)$, Bull. Belg. Math. Soc. - supplement - 7 (December 2002), 63-79.
8. B. Diarra, *Ultrametric q-calculus*, in Ultrametric functional analysis, edited by B. Diarra, A. Escassut, 8 A. K. Katsaras, L. Narici, Contemporary Mathematics, vol. 384, AMS (2005), 63-78.
9. A. Escassut, *Analytic elements in p-adic analysis* (World Scientific Publishing, Singapore, 1995).
10. L. Van Hamme, *Jackson's interpolation formula in p-adic analysis*, Proceedings of the Conference on p-adic analysis, Report 7806, Nijmigen (June 1978), 119-125.
11. L. Van Hamme, *Continuous operators which commute with translations, on the the space of continuous functions on* $\mathbb{Z}_p$, In "p-adic functional analysis",

edited by J. M. Bayod, N. De Grande-De Kimpe and J. Martínez-Maurica, Marcel Dekker, New-York (1991), 75-88.
12. R. Morris, Editor, *Umbral calculus and Hopf algebras*, Contemporary Mathematics, Vol. 6, AMS, Providence (1978).
13. C. Perez-Garcia and W. H. Schikhof , *Finite-dimensional orthocomplemented subspaces in p-adic normed spaces* - In Ultrametric functional analysis - *Cont. Math.* 319- AMS - (2003), 281-298.
14. M. van der Put, *Difference operators over p-adic fields*, Math. Ann. 198, (1972), 189-203.
15. A. M. Robert, *A Course in p-adic analysis*, GTM 198 (Springer 2000) .
16. A. C. M. van Rooij, *Non-archimedean analysis*, Marcel Dekker, Inc (New-York, 1978).
17. G. C. Rota, *Finite operator calculus* (Academic Press, New York, 1975).
18. W. H. Schikhof, *Ultrametric Calculus. An introduction to p-adic analysis*, (Cambridge University Press, Cambridge, 1984).
19. A. Verdoodt, *Normal bases for non-archimedean spaces of continuous functions*, Publicacions Matemàtiques, 37 (1993), 403-427.
20. A. Verdoodt, *The use of operators for the construction of normal bases for the space of continuous functions on V_q*, Bull. Belg. Math. Soc. 1 (1994), 685-699.
21. A. Verdoodt, *Bases and operators for the space of continuous functions defined on a subset of $\mathbb{Z}_p$*, Thesis, Vrije Universiteit Brussel (1995).

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 119–128)

θ^*-PAIRS AND THE STRUCTURE OF FINITE GROUPS*

HAIHUI FENG and XIUYUN GUO

Department of Mathematics, Shanghai University, Shanghai 200444, P. R. China

Let H, C and D be subgroups of a finite group G. We call the pair (C, D) a θ^*-pair for H if it satisfies the following conditions: (i) $D = (C \cap H)_G$, $\langle C^g, H\rangle = G$ for every $g \in G$, (ii) $KH < G$ for every $K/D \lhd G/D$ with $K/D < C/D$. In this paper, we obtain several results on the maximal θ^*-pair which implies G to be solvable, supersolvable and nilpotent.

Keywords: θ^*-pair; Solvable group; Supersolvable group; Nilpotent group.

1. Introduction

The relationship between the properties of subgroups of a finite group G and the structure of G has been studied extensively. In 1990, Mukherjee and Bhattacharya first introduced the concept of θ-pair associated with a maximal subgroup of a finite group in [1] and characterized the structure of the groups. Since then, many results on the structure of a finite group G were obtained by using this concept(see [1, 2, 3, 4]). Because the definition of θ-pair is only for maximal subgroups of a finite group, Xianhua Li and Shiheng Li gave the concept of θ-pair for any subgroup of a finite group in [7] and obtained some new results. However, further research requires additional conditions. We now introduce the concept of θ^*-pair for any subgroup of a finite group which is special case of θ-pair in [1]. By using the concept of θ^*-pair, we can generalize some known theorems and obtain some new results.

Throughout the paper, all groups are finite and all unexplained notation and terminologies are standard [6].

Firstly, we recall the concept of θ-pair associated with a proper subgroup

*The research was partially supported by the National Natural Science Foundation of China(10771132), SGRC(GZ310), the Research Grant of Shanghai University and Shanghai Leading Academic Discipline Project(J50101).

of a finite group given in [7].

Definition 1.1. Given a proper subgroup H of a group G, we call $(A,\ B)$ a θ-pair for H in G if

(i) $A \leq G$, $\langle H,\ A\rangle = G$ and $B = (A \cap H)_G$,

(ii) if A_1/B is a proper subgroup of A/B and $A_1/B \lhd G/B$, then $G \neq \langle H,\ A_1\rangle$.

As an improvement, we introduce the following concept

Definition 1.2. Let H, C and D be subgroups of a group G. We call the pair $(C,\ D)$ a θ^*-pair for H in G if

(1) $D = (C \cap H)_G$ and $\langle C^g,\ H\rangle = G$ for every $g \in G$,

(2) $KH < G$ for every $K/D \lhd G/D$ with $K/D < C/D$.

Remark 1.1. We can easily observe that a θ^*-pair for a subgroup must be a θ-pair. However, the converse is not true. For example, let $G = S_3$ and $H = \langle(12)\rangle$. Then, it is easy to check that $(\langle(23)\rangle,\ 1)$ is a θ-pair of H. However, $(\langle(23)\rangle,\ 1)$ is not a θ^*-pair for H.

A θ^*-pair $(C,\ D)$ is said to be maximal if there is no θ^*-pair $(C_1,\ D_1)$ such that C is a proper subgroup of C_1. A θ^*-pair $(C,\ D)$ is said to be normal if C is normal in G.

It is clear that if $H = G$, then $(1,\ 1)$ is the unique θ^*-pair for H. If $H = 1$, then $(G,\ 1)$ is the unique θ^*-pair for H. Therefore we always assume that H is a proper subgroup of G in the following discussion.

2. Preliminaries

We need the following elementary results.

Lemma 2.1. *Let H be a subgroup of a group G.*

(1) If $(C,\ D)$ is a maximal normal θ^*-pair for H, then $D = H_G$.

(2) Let $(C,\ D)$ and $(A,\ B)$ be both θ^*-pairs for H with $C \leq A$. If $C \lhd G$, then $A \lhd G$ and $A/B \cong C/D$.

(3) If $(C,\ H_G)$ is a normal θ^*-pair for H, then it is a maximal θ^*-pair for H.

Proof. (1) We consider the pair $(CH_G,\ H_G)$. If $(CH_G,\ H_G)$ is not a θ^*-pair for H, then there exists a normal subgroup K of G such that $H_G < K < CH_G$ and $G = KH$. Thus $K = (C \cap K)H_G$ and $G = (C \cap K)H$. Since $D \leq C \cap K$ and $C \cap K < C$, it follows that $(C,\ D)$ is not a θ^*-pair for H

in G, which is a contradiction. Hence $(CH_G,\ H_G)$ is a θ^*-pair for H. The maximality of $(C,\ D)$ implies that $CH_G = C$ and therefore $D = H_G$.

(2) If $D = B$, then $B \leq C$ and therefore $A = C$. If $D < B$, then $C < CB$ and $CB \lhd G$. Noting that $H(BC) = G$ and $B \leq BC \leq A$, we see that $A = BC$ and therefore $(A,\ B)$ is a normal θ^*-pair for H. Observe that $C \cap B \leq (C \cap H)_G = D$, we have $D = C \cap B$. Hence $A/B \cong C/D$.

(3) It is clear by (1) and (2). □

Remark 2.1. Lemma 2.1 (2) implies that a maximal normal θ^*-pair for a subgroup H of a group G is a maximal θ^*-pair for H. However, a maximal θ^*-pair is not necessarily a normal θ^*-pair. For example, let $H = S_4$, $K = Q_8$ and $G = H \times K$. We denote H_2 a Sylow 2-subgroup of H and A a subgroup of K with order 4. It is easy to check that $(H_2 \times K,\ 1 \times A)$ is a maximal θ^*-pair for subgroup $\langle(123)\rangle \times A$. However, $H_2 \times K$ is not normal in G.

Lemma 2.2. *Let H be a subgroup of a group G and N a normal subgroup of G with $N \leq H$.*

(1) If $(C,\ D)$ is a θ^*-pair for subgroup H and $N \leq D$, then $(C/N,\ D/N)$ is a θ^*-pair for H/N and the converse is true. In particular, $(C,\ D)$ is a maximal θ^*-pair for H if and only if $(C/N,\ D/N)$ is a maximal θ^*-pair for H/N.

(2) If $(C,\ D)$ is a maximal θ^*-pair for subgroup H with $N \not\leq D$, then there exists a normal θ^*-pair $(A,\ B)$ such that $N \leq B$ and A/B is isomorphic to a subgroup of a homomorphic image of CN/DN.

Proof. (1) This part obvious by the definition of θ^*-pair.

(2) Since $N \not\leq D$, it follows that $N \not\leq C$ and $C < CN$. Let $B = (CN \cap H)_G$. By the maximality of $(C,\ D)$, we see that $(CN,\ B)$ is not a θ^*-pair for H. Therefore, there exists a normal subgroup A/B of G/B with $A/B < CN/B$ such that $(A,\ B)$ is a normal θ^*-pair for H. Noting that $DN \leq B$, we have A/B is isomorphic to a subgroup of a homomorphic image of CN/DN. □

Lemma 2.3. *Let H be a subgroup of a group G. Suppose $(C,\ 1)$ is a maximal θ^*-pair for H. If N is the unique minimal normal subgroup of G and $N \not\leq C$, then C is a maximal subgroup of CN.*

Proof. If C is not maximal in CN, then there exists a maximal subgroup K of CN such that C is a proper subgroup of K. By the maximality of

$(C,\ 1)$, $(K,\ 1)$ is not a θ^*-pair for H. Noting that $G = \langle K^g,\ H\rangle$ for every $g \in G$, we see that there exists a normal subgroup L of G with $L < K$ such that $(L,\ 1)$ is a θ^*-pair for H. The uniqueness of N implies that $N \leq L$ and therefore $CN \leq K$, a contradiction. Hence C is a maximal subgroup of CN. □

Lemma 2.4. *Let H be a subgroup of G. Then all the conjugated subgroups of H in G have the same θ^*-pairs as H.*

Proof. It is clear by the definition of θ^*-pair. □

We need the following terminology(see [12] P.130).

A class $\mathscr{K}$ of groups is closed if for every $X \in \mathscr{K}$: (i) The homomorphic images of groups from $\mathscr{K}$ are in $\mathscr{K}$, (ii) The subgroups of groups from $\mathscr{K}$ are in $\mathscr{K}$, (iii) Direct products of groups from $\mathscr{K}$ are in $\mathscr{K}$.

Lemma 2.5. *Let $\mathscr{K}$ be a closed class and H a subgroup of a group G, then the following statements are equivalent .*

(1) There exists a maximal θ^*-pair $(T,\ H_G)$ for H such that $T/H_G \in \mathscr{K}$.
(2) There exists a maximal θ^*-pair $(C,\ D)$ for H such that $C/D \in \mathscr{K}$.

Proof. It is clear that (1) implies (2).

Now suppose that there exists a maximal θ^*-pair $(C,\ D)$ for H such that $C/D \in \mathscr{K}$. Let us consider the pair $(CH_G,\ H_G)$. If it is a θ^*-pair for H, then $C = CH_G$ and $D = H_G$ by the maximality of $(C,\ D)$. Hence $(C,\ D)$ is the required pair. If $(CH_G,\ H_G)$ is not a θ^*-pair for H, then there exists a proper normal subgroup K/H_G of G/H_G with $K/H_G < CH_G/H_G$ such that $(K,\ H_G)$ is a θ^*-pair for H. Since $CH_G/H_G \cong C/C \cap H_G \cong (C/D)/(C \cap H_G/D)$ and K/H_G is a subgroup of CH_G/H_G, it follows that $K/H_G \in \mathscr{K}$. By Lemma 2.1(3), $(K,\ H_G)$ is a maximal θ^*-pair for H. Thus $(K,\ H_G)$ is the required pair. □

Lemma 2.6. *Let H be a subgroup of a group G, $(C,\ D) = (C,\ 1)$ a maximal θ^*-pair for H and C a nilpotent subgroup. If N is the unique minimal normal subgroup of G and N is non-solvable, then $C \cap N$ is a Sylow 2-subgroup of N.*

Proof. By the hypothesis, it is clear that $C \not\lhd G$ and $N \not\leq C$. By Lemma 2.3, C is a maximal subgroup of $E = CN$.

If $C \cap N = 1$, let P be a Sylow p-subgroup of C for a prime divisor p of $|C|$. If $P \lhd E$, then $P \leq C_G(N) \lhd G$, a contradiction. Thus we may suppose that $P \not\lhd E$. The maximality of C implies that $N_E(P) = C$ and N is a p'-group. Now the p-group P acts on the p'-group N with $C_N(P) = 1$. By Theorem 6.2.2 [6] , N has the unique P-invariant Sylow q-subgroup Q for each prime q dividing the order of N. Then for each $c \in C$, we have $(Q^c)^P = Q^{Pc} = Q^c$ and the Sylow q-subgroup Q^c is also P-invariant. Hence the uniqueness of Q implies that Q is C-invariant. By the maximality of C, we have $E = CQ$ and $Q \lhd E$. It follows that E is solvable and so is N, a contradiction.

Now let $C \cap N > 1$ and P a Sylow p-subgroup of $C \cap N$ for a prime divisor p of $|C \cap N|$. Since C is nilpotent and $C \cap N \lhd C$, $P \lhd C$. Moreover, since C is maximal in $E = CN$, P must be a Sylow p- subgroup of N. Assume that $p > 2$. Denote by $J(P)$ the Thompson-subgroup of P. Then we have $1 < Z(J(P)) \lhd C$. Noting that $Z(J(P))$ is not normal in N, we have $N_E(Z(J(P))) = C$ by the maximality of C in E. Hence $N_N(Z(J(P))) = N_E(Z(J(P))) \cap N = C \cap N$ is nilpotent. By Theorem 8.3.1 of [6], N is p-nilpotent, a contradiction. Therefore, P is a Sylow 2-subgroup of N. □

3. Main results

Theorem 3.1. *Let G be a group. If there exists a maximal θ^*-pair (C, D) such that C/D is nilpotent for each 2-maximal subgroup H of G, then G is solvable.*

Proof. Assume that the result is not true and let G be a counterexample with minimal order. If G is a non-abelian simple group, then $(G, 1)$ is the unique maximal θ^*-pair for each 2-maximal subgroup H of G. By the hypothesis, G is nilpotent, a contradiction. Therefore G is not a simple group. Let N be a minimal normal subgroup of G. For each 2-maximal subgroup H/N of G/N, by our hypothesis, there exists a maximal θ^*-pair (C, D) of H such that C/D is nilpotent. If $N \leq D$, then we have $(C/N, D/N)$ is a maximal θ^*-pair for H/N by the Lemma 2.2(1) and $(C/N)/(D/N) \cong C/D$ is nilpotent. If $N \not\leq D$, then there exists a normal θ^*-pair (A, B) for H such that A/B is isomorphic to a subgroup of a homomorphic image of CN/DN by the Lemma 2.2(2). By Lemma 2.1, we can suppose that $(A/N, B/N)$ is a maximal θ^*-pair for H/N and $(A/N)/(B/N)$ is nilpotent. The minimality of G implies that G/N is solvable. Since the class of all solvable groups is a saturated formation, we can assume that N is the unique minimal normal subgroup of G and $\Phi(G) = 1$.

Let M be a maximal subgroup of G with $N \not\leq M$. Then $G = MN$, $M_G = 1$ and $C_G(N) = 1$. For each maximal subgroup H of M, we have $H_G = 1$. By our hypothesis, there exists a maximal θ^*-pair $(C,\ 1)$ for H such that C is nilpotent. By Lemma 2.3, we have C is a maximal subgroup of $E = CN$. We claim that

(1) C is a Sylow 2-subgroup of G and

(2) C is a maximal subgroup of G.

Let $C = C_2 \times C_{2'}$, where C_2 is the Sylow 2-subgroup of C and $C_{2'}$ is the Hall $2'$-subgroup of C. Since C is maximal in E and nilpotent, by Theorem 1 in [8] , $C_{2'}$ is normal in E. Since N is a direct product of some non-abelian simple groups, we see that $C_{2'} \cap N = 1$ and therefore $C_{2'} \leq C_G(N) = 1$. Hence C is a 2-subgroup of G. The maximality of $(C,\ 1)$ implies that C is a Sylow 2-subgroup of G.

Let $S = C \cap N$. Then S is a Sylow 2-subgroup of N and $S > 1$. It follows that $C \leq N_G(S) < G$. Suppose that K is a maximal subgroup of G such that $C \leq N_G(S) \leq K$. If $C < K$, then we have $K_G > 1$ by the maximality of $(C,\ 1)$ and it follows that $N \leq K$. By applying Frattini argument, we have $G = NN_G(S) \leq K$, a contradiction. Hence C is a maximal subgroup of G. Thus, our claim is established. Let C_1 be a maximal subgroup of C. Then, by our hypothesis, there exists a maximal θ^*-pair $(T,\ 1)$ for C_1 such that T is nilpotent. By the above discussion, T is a Sylow 2-subgroup of G. By the Sylow's Theorem, there exists an element $g \in G$ such that $C_1 \leq T^g$. It follows that $G = \langle T^g,\ C_1 \rangle = T^g$, a contradiction. This completes the proof. □

Theorem 3.2. *A group G is nilpotent if and only if there exists a maximal θ^*-pair $(C,\ D)$ such that C/D is nilpotent for each 2-maximal subgroup H of G.*

Proof. The necessity holds trivially. Conversely, assume that the result is not true and G is a counterexample with minimal order. By Theorem 3.1, G is solvable. Let N be a minimal normal subgroup of G. Then N is an elementary abelian p-group for some prime divisor p of $|G|$. By Lemma 2.2, G/N satisfies the hypothesis of the Theorem and thus G/N is nilpotent, by the counterexample of G. Since the class of all nilpotent groups is a saturated formation, we can suppose that N is the unique minimal normal subgroup of G and $\Phi(G) = 1$. Hence, there exists a maximal subgroup M of G such that $G = MN$, $M \cap N = 1$ and $M_G = 1$. If $p||M|$, let $P \in Syl_p(M)$, then $N_G(P) > M$ since $M \cong G/N$ is nilpotent. Thus $P \lhd G$. This leads to a contradiction. It hence follows that M is a Hall p'-subgroup of G and

N is Sylow p-subgroup of G. For each maximal subgroup H of M, we have $H_G = 1$. Thus by our hypothesis, there exists a maximal θ^*-pair $(C,\ 1)$ for H such that C is nilpotent. If C is a p-group, then $C \leq N$. This implies $G = \langle C,\ H\rangle \leq NH < NM = G$, a contradiction. Thus there exists a prime $q \neq p$ such that $q||C|$. If $p \nmid |C|$, then there exists an element $g \in G$ such that $C^g \leq M$ and $G = \langle C^g,\ H\rangle \leq M$, a contradiction. Therefore, we have $C = C_p \times C_{p'}$, where C_p and $C_{p'}$ are both non-trivial. If $N \leq C$, then $C_{p'} \leq C_G(N) = N$, a contradiction. Hence $C \not\trianglelefteq G$ and $N \not\leq C$. By the Lemma 2.3, C is a maximal subgroup of $E = CN$ and $C_{p'}$ is a p'-subgroup of G. Therefore, there exists an element $g \in G$ such that $C_{p'}^g \leq M$. Without loss of generality we can suppose $C_{p'} \leq M$. Then we have $N_G(C_{p'}) > C$. By the maximality of $(C,\ 1)$, $N_G(C_{p'})$ contains a normal subgroup of G and this implies $N \leq N_G(C_{p'})$. Hence, we have $C_{p'} \leq C_G(N) = N$, which is the final contradiction. □

Theorem 3.3. *Let G be a group, q the largest prime of $|G|$. If there exists a maximal θ^*-pair $(C,\ D)$ such that C/D is q-closed and solvable for each maximal subgroup of each Sylow subgroup of G, then G is solvable.*

Proof. Suppose that the result is not true and G is a counterexample with minimal order. It is obvious by our hypothesis that G cannot be simple. Let N be a minimal normal subgroup of G. If N is solvable, then N is an elementary abelian p-group for some prime divisor p of $|G|$. By Lemma 2.2, it is easy to check that G/N satisfies the hypothesis of the Theorem. Thus G/N is solvable by the counterexample of G and it follows that G is solvable, a contradiction. Hence we can suppose that N is nonsolvable. For each prime $p||G|$ and each maximal subgroup P_1 of Sylow p-subgroup P of G, there exists a maximal θ^*-pair $(C,\ 1)$ for P_1 such that C is q-closed and solvable. For any minimal normal subgroup N, $(CN,\ 1)$ is not a θ^*-pair for P_1. Then there exists a normal subgroup K of G with $K < CN$ such that $(K,\ 1)$ is a θ^*-pair for P_1. It is obvious that $G/K \cong P_1/K \cap P_1$ is solvable and so is G/KN. Since $KN/N \leq CN/N \cong C/C \cap N$ is solvable, G/N is solvable. By the counterexample of G, we can suppose that N is the unique minimal normal subgroup of G. For any prime $p||N|$, let $P_0 \in Syl_p(N)$. If P_0 is a Sylow p-subgroup of G, then we have $(P_1)_G = 1$ for each maximal subgroup P_1 of P_0. If $(G,\ 1)$ is a θ^*-pair for P_1, then it is the unique maximal θ^*-pair and G is solvable, a contradiction. Thus there exists a proper normal subgroup K of G such that $(K,\ 1)$ is a maximal θ^*-pair for P_1. Noting that $N \leq K$, we have $G = KP_1 \leq KN = K$, a

contradiction. Hence N does not contain any Sylow p-subgroup of G. Let $P \in Syl_p(G)$, $p \neq q$. For each maximal subgroup P_1 of P, there exists a maximal θ^*-pair $(C,\ 1)$ for P_1 such that C is q-closes and solvable. Thus $C \not\lhd G$ and C is not a p-group. Moreover, we see that $(CN,\ 1)$ is not a θ^*-pair for P_1 and there exists a normal subgroup K of G such that $(K,\ 1)$ is a maximal θ^*-pair for P_1 and $G = CNP_1$. Obviously q is a prime divisor of $|C|$ and C is maximal in $E = CN$ by Lemma 2.3. Let $Q \in Syl_q(C)$. Then $Q \lhd C$. If $q||N|$, then $N_E(Q) > C$ and $Q \lhd E$. Since $N \cap Q = 1$, $Q \leq C_G(N) = 1$, a contradiction. Hence, $q \nmid |N|$ and Q is a Sylow q-subgroup of G. Let $Q_1 \lessdot Q$. Then there exists a maximal θ^*-pair $(C_1,\ 1)$ for P_1 such that C_1 is q-closes and solvable. Thus, we have $G = C_1NQ_1$ and $q||C_1|$ since $q \nmid |N|$. Let $Q_2 \in Syl_q(C_1)$. Then $Q_2 \lhd C_1$. Thus, we have $N_G(Q_2) > C_1$ and $N_G(Q_2)$ contains a normal subgroup of G. It follows that $N \leq N_G(Q_2)$ and $Q_2 \leq C_G(N) = 1$, a contradiction. □

Theorem 3.4. *A group G is supersolvable if and only if there exists a maximal θ^*-pair $(C,\ D)$ such that C/D is supersolvable for each maximal subgroup of each Sylow subgroup of G.*

Proof. The necessity holds trivially. Conversely, by Theorem 3.3 G is solvable. Suppose that the result is not true and G is a counterexample with minimal order. By our hypothesis, G cannot be simple. Let N be a minimal normal subgroup of G. Then N is an elementary abelian p-group for some prime divisor p of $|G|$. It is easy to check that G/N satisfies the hypothesis of the Theorem, hence G/N is supersolvable by the counterexample of G. Since the class of all supersolvable groups is a saturated formation, we can suppose that N is the unique minimal normal subgroup of G and $\Phi(G) = 1$. Thus there exists a maximal subgroup M of G such that $G = MN$, $M_G = 1$, $N \cap M = 1$ and $N = F(G) = C_G(N)$. Suppose that N is a Sylow p-subgroup of G. Let P be a maximal subgroup of N. Then $P_G = 1$. If $(G,\ 1)$ is a θ^*-pair for P, then $(G,\ 1)$ is the unique maximal θ^*-pair and G is supersolvable, a contradiction. Thus there exists a proper normal subgroup K of G such that $(K,\ 1)$ is a θ^*-pair for P. However, $N \leq K$ implies $G = K$, a contradiction. Hence N is not a Sylow p-subgroup of G. Suppose that q is the largest prime divisor of $|G|$ and $Q \in Syl_q(G)$. Then $QN/N \lhd G/N$. If $q = p$, then we have $Q \lhd G$ and $Q \leq F(G) = N$. This contradiction implies $q \neq p$. Since G/N is supersolvable, we can suppose that $Q_1 \lessdot Q$ such that $Q_1N/N \lhd G/N$. By our hypothesis, there exists a maximal θ^*-pair $(C,\ 1)$ for Q_1 such that C is supersolvable. Suppose that $N \leq C$. If $q||C|$, then $Q_0 \lhd C$ for each $Q_0 \in Syl_q(C)$. This implies $Q_0 \leq C_G(N) = N$,

a contradiction. If $q \nmid |C|$, then we have $G/N = \langle C/N,\ Q_1N/N\rangle = CQ_1/N$ since $G = \langle C,\ Q_1\rangle$. This implies $|G|_q = |Q_1| < |Q| = |G|_q$, a contradiction. Thus we have $N \not\leq C$ and $C < CN$. By the maximality of $(C,\ 1)$, $(CN,\ 1)$ is not a θ^*-pair for Q_1. Hence there exists a normal subgroup K of G such that $K < CN$ and $(K,\ 1)$ is a θ^*-pair for Q_1. Moreover, $G = CNQ_1$ implies $q||C|$. Let Q_0 be a Sylow q-subgroup of C. Then $Q_0 \lhd C$. By the definition of $(C,\ 1)$, we can suppose that $Q_0 \leq M$. It is clear that $N_G(Q_0) > C$. Thus $N_G(Q_0)$ contains a normal subgroup of G. It follows that $N \leq N_G(Q_0)$ and $Q_0 \leq C_G(N) = N$, a contradiction. □

Theorem 3.5. *A group G is nilpotent if and only if there exists a maximal θ^*-pair $(C,\ D)$ such that C/D is nilpotent for each maximal subgroup of each Sylow subgroup of G.*

Proof. The necessity holds trivially. Conversely, suppose that the conclusion is false and G is a counterexample with minimal order. It is obvious by the hypothesis that G cannot be simple. Let N be a minimal normal subgroup of G.

By Theorem 3.3, G is solvable and N is an elementary abelian p-group for some prime divisor p of $|G|$. It is easy to check that G/N satisfies the hypothesis of the Theorem. Thus G/N is nilpotent by the minimality of G. Since the class of all nilpotent groups is a saturated formation, we can suppose that N is the unique minimal normal subgroup of G and $\Phi(G) = 1$. Then there exists a maximal subgroup M of G such that $G = MN$, $M_G = 1$, $N \cap M = 1$ and $N = F(G) = C_G(N)$. If N is not a Sylow p-subgroup of G, then we let $P/N \in Syl_p(G/N)$. Hence we have $P/N \lhd G/N$. It follows that $P \lhd G$ and $P \leq F(G) = N$, a contradiction. If N is the Sylow p-subgroup of G, then let $P \lessdot N$, and consequently $P_G = 1$. If $(G,\ 1)$ is a θ^*-pair for P, then $(G,\ 1)$ is the unique maximal θ^*-pair and G is nilpotent, a contradiction. Hence there exists a proper normal subgroup K of G such that $(K,\ 1)$ is a θ^*-pair for P. However, $N \leq K$ implies $G = K$, a contradiction. This completes the proof. □

References

1. N. P. Mukherjee, P. Bhattacharya, On theta pairs for a maximal subgroup, *Proc. Amer. Math. Soc.* 109(1990), 589-596.
2. Xiuyun Guo. On theta pairs for maximal subgroup, *Comm. Algebra* 22(1994) No.4, 4653-4659.
3. A. Ballester-Bolinches, Yaoqing Zhao, On maximal subgroups of finite groups and theta pairs, *Comm. Algebra* 24(1996) No.13, 4199-4209.

4. Shirong Li, A Note on Theta Pairs for Maximal Subgroups, *Comm. Algebra* 26(1998)No.12, 4277-4284.
5. Shirong Li, Yaoqing Zhao, On θ-pairs for maximal subgroups, *J. Pure Appl. Algebra* 147(2000), 133-142.
6. D. Gorenstein, Finite Groups, *Chelsea, New York*,1980
7. Xianhua Li, Shiheng Li, Theta pairs and the structure of finite groups, *Siberian Math. J.* 45(2004)No.3, 557-561.
8. J. S. Rose, On finite insolvable groups with nilpotent maximal subgroups, *J. Algebra* 48(1977), 182-196.
9. B. Huppert, Endliche Gruppen I, *Berlin*, 1967.
10. K. Doerk, T. Hawkes, Finite Soluble Groups, Walter de Gruyter, *Berlin*, 1992.
11. M. Weinstein(Ed.), Between Nilpotent and Solvable, *Polygonal Publishers, Washington, NJ*, 1982.
12. H. Kurzweil, B. Stellmacher, The Theory of Finite Groups, *New York*, 2003.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 129–155)

STABILITY OF THE THEORY OF EXISTENTIALLY CLOSED S-ACTS OVER A RIGHT COHERENT MONOID S*

JOHN FOUNTAIN

Department of Mathematics, University of York,
Heslington, York YO10 5DD, UK
E-mail: jbf1@york.ac.uk

VICTORIA GOULD

Department of Mathematics, University of York,
Heslington, York YO10 5DD, UK
E-mail: varg1@york.ac.uk

Let L_S denote the language of (right) S-acts over a monoid S and let Σ_S be a set of sentences in L_S which axiomatises S-acts. A general result of model theory says that Σ_S has a model companion, denoted by T_S, precisely when the class $\mathcal{E}$ of existentially closed S-acts is axiomatisable and in this case, T_S axiomatises $\mathcal{E}$. It is known that T_S exists if and only if S is right coherent. Moreover, by a result of Ivanov, T_S has the model-theoretic property of being stable.

In the study of stable first order theories, superstable and totally transcendental theories are of particular interest. These concepts depend upon the notion of type: we describe types over T_S algebraically, thus reducing our examination of T_S to consideration of the lattice of right congruences of S. We indicate how to use our result to confirm that T_S is stable and to prove another result of Ivanov, namely that T_S is superstable if and only if S satisfies the maximal condition for right ideals. The situation for total transcendence is more complicated but again we can use our description of types to ascertain for which right coherent monoids S we have that T_S is totally transcendental and is such that the U-rank of any type coincides with its Morley rank.

Keywords: Monoid, S-act, Stable, Superstable, Type, Rank.

*For Professor L.A. Bokut on the occasion of his 70th birthday

The authors are grateful to a careful referee for many insightful comments. The second author completed this paper during a visit to C.A.U.L. partially funded by FCT and FEDER, Project ISFL-1-143. She would like to thank C.A.U.L. and Gracinda Gomes for their kind hospitality.

1. Introduction

In this paper we are concerned with the investigation of stability properties of certain complete theories of S-acts. We emphasise that we are taking the algebraist's approach in the sense that our first aim is to associate stability properties of T_S with algebraic properties of the monoid S, our further investigations then focussing on the latter.

Stability properties (see Sections 2 and 5 for the relevant definitions) arose from the question of how many models a theory (a set of sentences of a first order language) has of any given cardinality. The seminal work of Shelah shows that an unstable theory, indeed a non-superstable theory, has 2^λ models of cardinality λ for any $\lambda > |T|$ [30]. The philosophy then is that, in these cases, there are too many models to attempt to classify by means of a sensible structure theorem. It is reasonable therefore for the algebraist to consider for a given class of algebras 'how stable' is the theory associated with it, before embarking on the search for structure or classification theorems.

For a monoid S, a *(right) S-act* is simply a set A upon which S acts on the right with the identity of S acting as the identity map on A. Associated with S is the first order language L_S for S-acts. We denote by Σ_S the set of sentences axiomatising S-acts, and refer to Σ_S as the *theory* of S-acts. Further details are given in Section 2. We can think of an S-act as being analogous to a module over a ring; this observation inspires our characterisation of types and our approach to stability and superstability, after which more significant differences arise between the situation for modules and that for acts.

The model theory of modules has been and continues to be extensively investigated (see [28]), yielding both structure results for modules and giving concrete realizations of model theoretic concepts. In contrast, only a few studies have been made of the model theory of S-acts. Some results in the latter theory are close parallels of corresponding results for modules. As indicated above, there are, however, several major differences between the two theories. Essentially, these differences arise since right congruences on monoids cannot be determined by right ideals (as is the case for rings). For the model theorist, this means that atomic formulae without parameters cannot be replaced by formulae involving parameters.

Given any R-module M over a ring R, or any S-act A over a monoid S, we can consider the set of all sentences (in the appropriate language) that are true in M or in A. These theories are exactly the complete theories of R-modules or S-acts, where a theory T is *complete* if for any sentence

ϕ of the language, $\phi \in T$ or $\neg\phi \in T$. A notable difference between the model theory of modules and that of S-acts is that, as demonstrated by Mustafin [23], for some monoids S, there are S-acts which have unstable theories whereas all complete theories of modules are stable. Mustafin goes on to describe all monoids S for which every S-act has a stable theory or superstable theory. The thrust of his later papers in this area is to move toward a description of those monoids S over which all S-acts are ω-stable [3,24]. On the other hand Stepanova [31] has characterised monoids such that all *regular* S-acts have stable, superstable or ω-stable theories.

Rather than imposing conditions on the theories $\mathrm{Th}(A)$ for *all* S-acts A over a given S, we are concerned here with theories of existentially closed S-acts: we now explain our motivation. An important notion of model theory is that of *model companion*. For a theory T one defines in a natural way the notion of an existentially closed model of T and we denote the class of existentially closed models of T by $\mathcal{E}(T)$. If T is an inductive theory, such as Σ_S, then T has a model companion if and only if $\mathcal{E}(T)$ is axiomatisable. In this case, $\mathrm{The}(\mathcal{E}(T))$ is a model companion of T. Following Wheeler [32] the notion of right coherence for monoids was introduced in [11] where it is shown that the theory of all S-acts (for fixed S) has a model companion T_S if and only if S is right coherent. It follows that the models of T_S are precisely the existentially closed S-acts, and further, that T_S is a complete theory so that $T_S = \mathrm{Th}(M)$ for any existentially closed S-act M. Ivanov [17] argues that T_S is a normal theory (see [25]) and hence stable.

Given that T_S is stable it is natural to investigate conditions under which it satisfies the stronger stability properties of being superstable, ω-stable or totally transcendental. In [4] the corresponding questions in module theory are posed and answered. This work both inspired and heavily influenced the present paper. For a right coherent ring R, the model companion of the theory of all R-modules is denoted by T_R. Properties such as stability are dependent upon the number of types (the details of which are given in Section 3). In [4] types are characterised by pairs consisting of a right ideal of R and an R-homomorphism. This is the key to a thorough analysis of complete types and so to finding for which rings R the theory T_R is superstable or totally transcendental.

In a ring R, a right congruence is determined by a right ideal, but as remarked above, this is not true for monoids in general. For this reason, in the case of right S-acts, complete types are characterised by triples consisting of a right ideal of S, a right congruence on S and an S-morphism. It is this result which allows us to translate model theoretic properties of T_S

into algebraic properties of S and hence to apply the theory of semigroups. An immediate consequence is that we can easily find upper bounds for the number of types. This enables us to deduce Ivanov's result [17, Proposition 1.4] that the theory T_S is stable. Further, if every right ideal of S is finitely generated, then T_S is superstable, and if in addition S is countable and has at most $\aleph_0$ right congruences, then T_S is ω-stable.

To obtain the converse of these results we use the U-rank of types and the fact that a complete theory is superstable if and only if the U-rank of each type is defined (see [27]). Our approach is similar to but slightly more complicated than that of Bouscaren. The end results are that T_S is superstable if and only if every right ideal of S is finitely generated and that for a countable S, T_S is ω-stable if and only if S has at most $\aleph_0$ right congruences and every right ideal of S is finitely generated. The superstability result is also a straightforward consequence of [17, Theorem 2.4]. In these results there is of course the underlying assumption that S is right coherent, for this is needed for the theory T_S to exist. Right coherence does not follow from the property that every right ideal is finitely generated as shown by Example 3.1 in [12]. The equivalent results for modules are that superstability and total transcendence of T_R are both equivalent to R being right noetherian.

Another important rank of types is the Morley rank. This is used to define the concept of total transcendence, a complete theory T being totally transcendental if and only if every type has Morley rank. Morley rank is always greater than U-rank, so that a totally transcendental theory is certainly superstable. In fact a countable theory T is totally transcendental if and only if it is ω-stable [27].

For a complete theory T of modules, the Morley rank of a type (when it exists) coincides with the U-rank of the type [28]. This is not the case for S-acts and we find necessary and sufficient conditions on S for the theory T_S to be totally transcendental with the Morley rank of any type being equal to its U-rank. The final section of the paper is devoted to a study of monoids which satisfy these conditions. If S is such a monoid and is weakly periodic, then S is finite. On the other hand, the infinite cyclic monoid satisfies the conditions.

The structure of this paper is as follows. In Section 2 we outline the basics of model theory we require; since the new work of this article is almost entirely semigroup theoretic, we keep these details to a minimum. We also give some details concerning S-acts over a monoid S. In Section 3 we discuss types and, crucially, show how a type of T_S over an S-act A is

associated with what we call an A-triple. It is this result which allows us to translate arguments from model theory into algebra. We omit most proofs, since the ideas are rather straightforward and may be thought of as being inherent in the work of Ivanov [17]; full details appear in the notes [10]. The next section outlines how we may use our description of types to capture U-rank and the superstability result of [17]. Sections 3 and 4 may be regarded as a survey. The new material begins in Section 5 where we discuss Morley rank and find a criterion for a right coherent monoid S such that T_S is totally transcendental and the U-rank of any type of T_S coincides with its Morley rank. In our final section we investigate the monoids satisfying this criterion.

2. Preliminaries

This paper is intended to be accessible to algebraists with some familiarity with the basic ideas of first-order logic and, with the exception of the final section, only a very little semigroup theory. We recommend [6] and [9] for the former and [15] for the latter. Full accounts of the stability theory we use can be found in the books [1,4,19,26–28]; we extract the key ideas and main results which we need. Any unreferenced results may be found in these texts.

We begin with some brief details concerning S-acts. Further details may be found in the comprehensive [18].

Let S be a monoid. A *(right) S-act* is a set A on which S acts on the right, that is, there is a map $\cdot$ from $A \times S$ to A satisfying :

$$(a \cdot s) \cdot t = a \cdot (st) \text{ and } a \cdot 1 = a$$

for all $s, t \in S, a \in A$, where $\cdot$ maps (a, s) to $a \cdot s$. We usually write as for $a \cdot s$. Clearly we can think of the elements of S as unary operation symbols and A as a unary algebra in the sense of universal algebra. We thus have all the standard concepts and results of universal algebra at our disposal (see, for exàmple [21]). In particular, we have *S-subacts*, *S-morphisms*, *congruences on S-acts* and *quotient S-acts A/ρ* where A is an S-act and ρ is a congruence on A. For an S-subact B of an S-act A, the relation ρ_B is defined by $a_1 \rho_B a_2$ if and only if $a_1 = a_2$ or a_1, a_2 are both in B. It is easy to see that ρ_B is a congruence on A; the quotient S-act A/ρ_B is usually denoted by A/B and is called the *Rees quotient* of A by B. We differ from standard semigroup terminology in that we make the convention that the empty set $\emptyset$ is an S-subact of every S-act.

For any congruence ρ on an S-act A we denote the ρ-class of an element a of A by $a\rho$. For an S-morphism $f : A \to B$ we denote by $\mathrm{Ker}f$ the congruence on A determined by

$$(a, b) \in \mathrm{Ker}f \text{ if and only if } f(a) = f(b).$$

The multiplication in a monoid S makes S itself into a right S-act. The S-subacts of S are called *right ideals* of S and S-act congruences on S are called *right congruences* on S, to distinguish them from semigroup congruences on S.

The category of S-acts and S-morphisms has arbitrary products and coproducts. Another property enjoyed by this category which is useful for our purposes is the *strong amalgamation property*. This asserts that if A, B are S-acts with common S-subact U, then there is an S-act C and injective S-morphisms $f : A \to C, g : B \to C$ such that $f|U = g|U$ and $f(A) \cap g(B) = f(U)$.

Let I be a right ideal of a monoid S and ρ be a right congruence on S. The *ρ-closure* of I, denoted by $I\rho$, is defined by

$$I\rho = \{s \in S : s \,\rho\, t \text{ for some } t \in I\}.$$

It is easy to see that $I\rho$ is a right ideal of S containing I and that $(I\rho)\rho = I\rho$. We say that a right ideal J of S is *ρ-saturated* if $J\rho = J$; thus $I\rho$ is ρ-saturated for any right ideal I. If ν, ρ are right congruences on S and $\nu \subseteq \rho$, then any ρ-saturated right ideal is also ν-saturated.

When I is a ρ-saturated right ideal of S we say that the pair (I, ρ) is a *congruence pair*. We denote by $\mathcal{C}(S)$ or $\mathcal{C}$ the set of all congruence pairs of S.

This paper is concerned with one aspect of the model theory of S-acts. Let L be a first order language. A class $\mathcal{U}$ of L-structures is *axiomatisable* if there is a set of sentences Π of L such that an L-structure U lies in $\mathcal{U}$ if and only if every sentence of Π is true in U, that is, $U \models \pi$. We use the standard notation that if $\phi(x_1, ..., x_n)$ is a formula of L, then the free variables of $\phi(x_1, ..., x_n)$ lie in $\{x_1, ..., x_n\}$. If T is a theory in L (that is, a set of sentences of L, which without loss of generality we may assume to be closed under deduction), then models of T will be denoted by letters M, N, P; we use the same notation for their universes. If M is an L-structure then $\mathrm{Th}(M)$ is the set of sentences true in M; if M is a model of a theory T then certainly $T \subseteq \mathrm{Th}(M)$. The letters $A, B, \ldots$ are used for subsets of models. For a set A, the language $L(A)$ is obtained from L by adding a new constant symbol to L for each element a of A. Again, we follow the usual practice and do not distinguish elements of A from the constants of $L(A)$

which they label. We may denote an $L(A)$-structure by $(B, a)_{a\in A}$, where $A \subseteq B$, so that if T is a theory in L and $A \subseteq M \models T$, $\mathrm{Th}(M, a)_{a\in A}$ is the set of sentences in $L(A)$ true in M.

The language L_S of the theory of S-acts consists of a unary function symbol f_s for each element s of S. We follow the usual convention and write as for $f_s(a)$. Clearly the class of S-acts is axiomatised by the set of sentences

$$\Sigma_S = \{(\forall x)(x1 = x)\} \cup \{(\forall x)((xs)t = x(st)) : s, t \in S\}.$$

An *equation* over an S-act A is an atomic formula of $L_S(A)$ and has one of the forms:

$$xs = xt,\ xs = yt,\ xs = a$$

where $s, t \in S$ and $a \in A$. An *inequation* over A is simply the negation of an equation over A.

A set Σ of equations and inequations over A is *consistent* if Σ has a solution in some S-act containing A. An S-act A is *existentially closed* if every consistent finite set of equations and inequations over A has a solution in A. Since the class of S-acts is inductive, that is, is closed under unions of chains, every S-act is contained in an existentially closed S-act.

To say that the theory of all S-acts has a *model companion* is equivalent to saying that the class of all existentially closed S-acts is axiomatisable by a theory T_S; then T_S is the required model companion. In [11] it is proved that T_S exists if and only if S is right coherent, where a monoid S is *right coherent* if for any finitely generated right congruence ρ on S, every finitely generated S-subact of S/ρ is finitely presented. This result was generalised to varieties of S-acts in [17]. A careful study of right coherence for S-acts is made in [12].

Given two existentially closed S-acts A, B it is certainly the case that A, B can be embedded in an S-act C (the coproduct of A and B for example) and C can be embedded in an existentially closed S-act. It follows from this and the model completeness of T_S that T_S (when it exists) is complete (Proposition 3.1.9 of [6]). That is, for any sentence ϕ of L_S either $\phi \in T_S$ or $\neg\phi \in T_S$; equivalently, $T_S = \mathrm{Th}(M)$ for any of its models M. Since the theory of all S-acts is universal and as T_S is actually the model completion of this theory [11], we have by Theorem 13.2 in [29] that T_S admits elimination of quantifiers. These properties, not all used explicitly here, ensure that T_S is precisely the kind of theory most amenable to the application of stability theory.

3. Types

The notion of a *type* is crucial to our investigations of stability properties of T_S. To define types, it is useful to employ the so-called monster model of a theory; justification of its existence (which uses the notion of saturation) and use can be found in [5]. Let T be a complete theory in L. The *monster model* of T is a model $\mathbf{M}$ of T such that all models of T are elementary substructures of $\mathbf{M}$ and all sets of parameters are subsets of $\mathbf{M}$.

Let A be a subset of $\mathbf{M}$ and let $c \in \mathbf{M}$. Then

$$\mathrm{tp}(c/A) = \{\phi(x) \in L(A) : \mathbf{M} \models \phi(c)\}$$

is a *(complete 1-)type over* A. Clearly $\mathrm{tp}(c/A)$ is a complete set $p(x)$ of sentences of $L(A, x)$ such that a model exists for $p(x) \cup \mathrm{Th}(\mathbf{M}, a)_{a \in A}$. Conversely, if $p(x)$ is a set of formulae satisfying these conditions, then properties of $\mathbf{M}$ (concerned with saturation) give that $p(x) = \mathrm{tp}(b/A)$ for some $b \in \mathbf{M}$. The *Stone space* $S(A)$ of A is the collection of all types over A; $S(A)$ is equipped with a natural topology, which comes into play in the definition of Morley rank (see Section 5).

For a cardinal κ, T is *κ-stable* if for every subset A of a model of T with $|A| \leq \kappa$ we have $|S(A)| \leq \kappa$. If T is κ-stable for some infinite κ, then T is *stable* and T is *superstable* if T is κ-stable for all $\kappa \geq 2^{|T|}$. If T is not stable, then it is said to be *unstable*. Morley argued that a theory T in a countable language is ω-stable if and only if T is κ-stable for every infinite κ [22].

From now on we shall concentrate on the theory T_S for a fixed right coherent monoid S. The purpose of this section is to give a straightforward characterisation of types over S-acts. We do not present the proofs, as they involve quite standard concepts. Some of these ideas appear implicitly in [17]; explicit proofs may be found in the unpublished notes [10] of the authors. Ivanov [17] shows that T_S is a stable theory, and also characterises those monoids S such that T_S is superstable. By making the characterisation of types explicit, we have both an alternative approach to these results of Ivanov, and a solid tool with which to characterise ranks of types, needed for our later discussions.

If A is an S-act, then an *A-triple* is a triple (I, ρ, f) such that $(I, \rho) \in \mathcal{C}$ and $f : I \to A$ is an S-morphism with $\mathrm{Ker} f = \rho \cap (I \times I)$. We denote the set of all A-triples by $\mathcal{T}(A)$.

Let $\mathcal{T} = (I, \rho, f)$ be an A-triple and let $\Sigma_{\mathcal{T}}$ be the union of the following sets of formulae of $L_S(A)$:

$$\{xs = a : a = f(s), s \in I\}, \{xs \neq a : s \notin I, a \in A\},$$

$$\{xs = xt : (s,t) \in \rho\}, \{xs \neq xt : (s,t) \notin \rho\}.$$

An easy argument using quantifier elimination and the fact that the class of S-acts has the strong amalgamation property yields the following.

Lemma 3.1. *[10] Let A be an S-act and let $\mathcal{T}$ be an A-triple. Then there is an embedding of A into an existentially closed S-act E, and an element $c \in E$ such that $tp(c/A) = p_{\mathcal{T}}$ is the unique type over A containing $\Sigma_{\mathcal{T}}$.*

Conversely, given $p \in S(A)$ we obtain an A-triple $\mathcal{T}_p$.

Lemma 3.2. *[10] Let p be a type over an S-act A. Let*

$$I_p = \{s \in S : xs = a \in p \text{ for some } a \in A\},$$

$$\rho_p = \{(s,t) \in S \times S : xs = xt \in p\},$$

and

$$f_p : I_p \to A \text{ be defined by } f_p(s) = a \text{ where } xs = a \in p.$$

Then $\mathcal{T}_p = (I_p, \rho_p, f_p)$ is an A-triple.

The next result is crucial. Essentially, it allows us to translate arguments involving types, and ranks thereof, into arguments internal to our monoid S.

Proposition 3.1. *[10] The maps*

$$p \mapsto \mathcal{T}_p, \quad \mathcal{T} \mapsto p_{\mathcal{T}}$$

are mutually inverse bijections between $S(A)$ and $\mathcal{T}(A)$.

The corollary below is an immediate consequence of the proposition.

Corollary 3.1. *[10]*

(1) Let A be an S-act and let $p, q \in S(A)$. Then $p = q$ if and only if $I_p = I_q, \rho_p = \rho_q$ and $f_p = f_q$.
(2) There is a bijection between the set of right congruences on S and $S(\emptyset)$.
(3) For any congruence pair (I, ρ) on S there is an S-act A and a type p over A with $I_p = I$ and $\rho_p = \rho$.
(4) Let p be a type over an S-subact A of B. Then there is a type q over B such that $I_p = I_q, \rho_p = \rho_q$ and $f_q = jf_p$ where $j : A \to B$ is the inclusion map.

Let A be an S-act and I be a right ideal of S. The number of S-morphisms from I to A is at most $|A|^{|S|}$, the number of right ideals of S is at most $2^{|S|}$ and the number of right congruences on S is at most $2^{|S|^2}$. Hence the number of A-triples is at most $2^{|S|}2^{|S|^2}|A|^{|S|}$. Thus, if we take $\kappa = \max\{\aleph_0, 2^{|S|}\}$ and $|A| \leq \kappa$, then $|\mathcal{T}(A)| \leq \kappa$ and, in view of Proposition 3.1, $|S(A)| \leq \kappa$.

Now consider an arbitrary subset B of the S-act $\mathbf{M}$. It is easy to see that $|S(B)| = |S(A)|$, where A is the S-subact of $\mathbf{M}$ generated by B (indeed, the Stone spaces are homeomorphic, see [1,19]). We can therefore deduce that the theory T_S is stable.

We can do better than this when every right ideal of S is finitely generated, that is, when S is *weakly right noetherian*. Then, for any right ideal I, the number of S-morphisms from I to A is at most $\max\{\aleph_0, |A|\}$ so that there are no more than $2^{|S|}\max\{\aleph_0, |A|\}$ A-triples. Hence for any infinite cardinal κ with $2^{|S|} \leq \kappa$ we have that if $|A| \leq \kappa$, then $|S(A)| \leq \kappa$. Now $|T_S| = \max\{\aleph_0, |S|\}$ so that T_S is superstable [17].

If we assume that S has at most $\max\{\aleph_0, |S|\}$ right congruences in addition to being weakly right noetherian, then we see that the number of A-triples is at most $\max\{\aleph_0, |S|\}^2\max\{\aleph_0, |A|\}$. Thus for any infinite cardinal κ with $|S| \leq \kappa$ we have that if $|A| \leq \kappa$, then $|S(A)| \leq \kappa$. Hence, for a countable S which is weakly right noetherian and has only countably many right congruences we have that T_S is ω-stable. In particular, T_S is ω-stable for any finite monoid S.

A monoid S is *right noetherian* if every right congruence on S is finitely generated; since every right ideal of S is determined by a right congruence, it follows that such a monoid is weakly right noetherian. Moreover, every right noetherian monoid is right coherent [12]. Thus if S is a countable, right noetherian monoid, then T_S is ω-stable.

If S is countably infinite and T_S is ω-stable, then $|S(\emptyset)| \leq \aleph_0$ so that by Corollary 3.1, S has only countably many right congruences.

The following result summarises the above discussion; (1) and (2) are also consequences of results in [17].

Proposition 3.2. *[10,17] Let S be a right coherent monoid. Then*

(1) the theory T_S is stable;
(2) if S is weakly right noetherian, then T_S is superstable;
(3) if S is weakly right noetherian and has at most $\max\{\aleph_0, |S|\}$ right congruences, then T_S is κ-stable for all κ with $\max\{\aleph_0, |S|\} \leq \kappa$;
(4) if S is countable, then if S is weakly right noetherian and has at most

$\aleph_0$ *right congruences,* T_S *is* ω*-stable;*

(5) if S *is finite, then* T_S *is* ω*-stable;*

(6) if S *is countable and right noetherian, then* T_S *is* ω*-stable;*

(7) if S *is countable and* T_S *is* ω*-stable, then* S *has at most* $\aleph_0$ *right congruences.*

The converses of (2) and (4) of the above proposition will be obtained in Section 4.

By an *extension* of a type p in $S(A)$ we mean a type q in $S(B)$ where A is an S-subact of B and $p \subseteq q$. The proof of the following result follows easily from Lemma 3.1.

Proposition 3.3. *[10] Let* A *be an* S*-subact of* B, $p \in S(A)$ *and* $q \in S(B)$. *Then* q *is an extension of* p *if and only if*

(i) $I_p \subseteq I_q$,
(ii) $f_q|I_p = f_p$,
(iii) $f_q^{-1}(A) = I_p$
(iv) $\rho_p = \rho_q$.

A consequence of Proposition 3.3 is that if p and q are as in Corollary 3.1 (4), then q is an extension of p.

For the final result of this section we again make use of the fact that the class of S-acts has the strong amalgamation property.

Proposition 3.4. *[10] Let* A *be an* S*-act and* $p \in S(A)$. *Let* J *be a* ρ_p*-saturated right ideal containing* I_p. *Then there is an* S*-act* B *containing* A *and an extension* q *of* p *in* $S(B)$ *such that* $I_q = J$. *Moreover,* B *can be chosen to be existentially closed.*

4. U-rank and superstability of T_S

Rank notions are an important tool in determining stability properties of theories. In this section we relate the U-rank of a type p, introduced by Lascar in [20], to what we call the ρ_p-rank of the right ideal I_p. It is then straightforward to prove the converse of (2) of Proposition 3.2. As in Section 3, we omit most arguments, which may be found in detail in [10].

First we recall the *foundation* rank on a set $\mathcal{S}$ partially ordered by $\leq$. We define subclasses $\mathcal{S}_\alpha$ of $\mathcal{S}$ for each ordinal α by transfinite induction:

(i) $\mathcal{S}_0 = \mathcal{S}$;
(ii) $\mathcal{S}_\alpha = \bigcap\{\mathcal{S}_\beta : \beta < \alpha\}$, if α is a limit ordinal;

(iii) $x \in \mathcal{S}_{\alpha+1}$ if and only if $x < y$ for some $y \in \mathcal{S}_\alpha$.

We thus obtain a nested sequence of subclasses of $\mathcal{S}$ indexed by the ordinals. The *foundation rank* of $x \in \mathcal{S}$, denoted by $\mathrm{R}(x)$, can now be defined as follows:

If $x \in \mathcal{S}_\alpha$ for all ordinals α, then we write $\mathrm{R}(x) = \infty$. Otherwise, $\mathrm{R}(x) = \alpha$ where α is the (unique) ordinal such that $x \in \mathcal{S}_\alpha \setminus \mathcal{S}_{\alpha+1}$; in this case we say that x *has R-rank.*

The convention that $\alpha < \infty$ for all ordinals α simplifies the statements of the following standard proposition (see for example [27], p. 35).

Proposition 4.1.

(i) For any $x \in \mathcal{S}$ and any ordinal α,

$$R(x) \geq \alpha \text{ if and only if } x \in \mathcal{S}_\alpha.$$

(ii) Let $x, y \in \mathcal{S}$ where $x < y$. If $R(y)$ is an ordinal then $R(x) > R(y)$. Moreover, if $R(x)$ is an ordinal then so is $R(y)$.

(iii) For any $x \in \mathcal{S}$, $R(x)$ is an ordinal if and only if there are no infinite chains of the form

$$x = x_0 < x_1 < \ldots .$$

For the first application of foundation rank, consider a right congruence ρ on S and put

$$\mathcal{S} = \mathcal{S}_\rho = \{J : (J, \rho) \in \mathcal{C}\}.$$

The relation $\leq$ is taken as the usual inclusion order of right ideals. If $J \in \mathcal{S}_\rho$ then $\mathrm{R}(J)$ is said to be the *ρ-rank* of J and is written as $\rho\text{-R}(J)$.

Corollary 4.1. *Let $(I, \rho) \in \mathcal{C}$. Then ρ-$R(I)$ is an ordinal if and only if S has the ascending chain condition on ρ-saturated right ideals containing I.*

Our second application of foundation rank is to obtain the *U-rank* $\mathrm{U}(p)$ of a type $p \in S(A)$, where $A \subseteq \mathbf{M} \models T$ and T is a complete, stable theory in a first order language L. First we review some definitions associated with types of T; for more details the reader can consult one of the standard texts.

If $p \in S(A)$, where $A \subseteq \mathbf{M}$, then the *class* of p, written $\mathrm{cl}(p)$, is the set

$$\mathrm{cl}(p) = \{\phi(x, y_1, \ldots, y_n) \in L : \text{ for some } a_1, \ldots, a_n \in A, \phi(x, a_1, \ldots, a_n) \in p\}$$

and C_p is the set

$$C_p = \{\mathrm{cl}(q) : p \subseteq q, q \in S(M), A \subseteq M \models T\}.$$

It is a fact that C_p has a unique minimum element (under inclusion) denoted by $\beta(p)$. Clearly, if $p \in S(M)$ where $M \models T$, then $\mathrm{cl}(p) = \beta(p)$. For $A \subseteq B$ and an extension $q \in S(B)$ of p, it is obvious that $\beta(p) \subseteq \beta(q)$. Then q is a *non-forking extension* of p if $\beta(p) = \beta(q)$; otherwise, q is a *forking extension* of p.

Put

$$\mathcal{S} = \{\beta(p) : p \in S(A) \text{ for some } A \subseteq \mathbf{M}\}.$$

Clearly $\mathcal{S}$ is partially ordered by set inclusion. The *U-rank* of $p \in S(A)$, denoted $\mathrm{U}(p)$, is the foundation rank of $\beta(p)$. If $\mathrm{U}(p)$ is an ordinal, then we say that p *has U-rank*. Clearly, in our discussion of U-rank, we can assume that all types are over L-substructures of models of T.

Our objective in this section is to characterise those monoids S for which T_S is superstable or ω-stable. In other words, our goal is to prove the converses of (2) and (4) of Proposition 3.2. In fact, the converse of (4) follows easily from that of (2) so our effort is directed towards showing that if T_S is superstable, then S is weakly right noetherian. To do this, we use the characterisation of superstable theories in terms of U-rank of types. Then, by associating the U-rank of a type $p \in S(A)$ with ρ_p-$\mathrm{R}(I_p)$, we are able to achieve our goal.

Theorem 4.1. *[20] Let T be a complete, stable theory in a first order language. Then T is superstable if and only if all types have U-rank.*

Turning our attention to the theory T_S, we have the following characterisation of forking. Recall from Proposition 3.3 that if q is an extension of a type p, then $I_p \subseteq I_q$ and $\rho_p = \rho_q$.

Lemma 4.1. *[10] Let $A \subseteq B$ be S-acts, and let $q \in S(B)$ be an extension of $p \in S(A)$. Then q is a forking extension of p (equivalently, $U(p) > U(q)$) if and only if $I_p \subset I_q$.*

From Proposition 3.4 we know that if for an S-act A we have $p \in S(A)$ and $I_p \subset J$ for some ρ_p-saturated right ideal J, then there is an S-act $B \supseteq A$ and $q \in S(B)$ with $p \subseteq q$. From Lemma 4.1, $\mathrm{U}(p) > \mathrm{U}(q)$.

We can now associate the U-rank of types over T_S with ranks assigned to members of $\mathcal{C}$.

Proposition 4.2. *[10] For any S-act A and $p \in S(A)$,*

$$UR(p) = \rho_p\text{-}R(I_p).$$

Corollary 4.2. *For any S-act A and $p \in S(A)$, p has U-rank if and only if the set of ρ_p-saturated right ideals containing I_p satisfies the ascending chain condition.*

Part (1) of the following theorem is also a consequence of [17] (Theorem 2.4).

Theorem 4.2. *Let S be a right coherent monoid.*

(1) The theory T_S is superstable if and only if S is weakly right noetherian.
(2) If S is countable, then the theory T_S is ω-stable if and only if S is weakly right noetherian and has only countably many right congruences.

Proof.

(1) If S is weakly right noetherian, then T_S is superstable by (2) of Proposition 3.2. Alternatively, this follows from Theorem 4.1 and Corollary 4.2.
Conversely, if T_S is superstable, then applying Corollary 4.2 to the type in $S(\emptyset)$ corresponding to the identity congruence gives that S is weakly right noetherian.
(2) Suppose that S is countable. If T_S is ω-stable, then it is superstable by [22] and so by (1), S is weakly right noetherian. Also we must have $|S(\emptyset)| \leq \aleph_0$ and hence by Corollary 3.1, the number of right congruences on S is countable. The converse is (4) of Proposition 3.2. □

This theorem allows us to give examples of monoids to illustrate the various possibilities. Thus if $S = \{1\} \cup I$ where 1 acts as an identity and I is an infinite set with $ab = a$ for all $a, b \in I$, then I is a right ideal of S which is not finitely generated; moreover, it is easy to see that S is right coherent. Hence T_S exists, but is not superstable.

On the other hand, T_G is superstable for any group G. But, for example, the group of rationals Q has $2^{|Q|}$ subgroups (and hence $2^{|Q|}$ (right) congruences) so that T_Q is not ω-stable.

Both the infinite cyclic group and the quasi-cyclic group $Z(p^\infty)$ (p a prime number) have $\aleph_0$ subgroups so they provide specific examples of infinite groups G such that T_G is ω-stable.

Of course, for any *finite* monoid S we have that T_S is ω-stable.

5. Total transcendence of T_S

Having considered U-rank of types in the previous section we now turn our attention to another rank, the Morley rank $\mathrm{MR}(p)$, of a type p. This rank is used to define totally transcendental theories; to be precise a complete theory T is *totally transcendental* if and only if for all subsets A of models of T, all types over A have Morley rank.

For a countable theory T, it is a fact that T is totally trancendental if and only if T is ω-stable [22]. There are, however, uncountable theories T which are not totally transcendental but are κ-stable for all κ with $|T| \leq \kappa$.

When T is a theory of modules, if p is a type over a subset of a model of T such that $\mathrm{MR}(p)$ is defined, then $\mathrm{MR}(p) = \mathrm{U}(p)$ [28]. For S-acts, however, the picture is different and rather subtle. In this section we investigate those monoids S for which $\mathrm{MR}(p) = \mathrm{U}(p) < \infty$ for all types p over subsets of models of T_S, introducing a condition (MU). We also refer the reader to [16], where Ivanov presents a condition bearing some resemblance to (MU) that will imply $\mathrm{MR}(p) = \mathrm{U}(p)$. In a subsequent article [13] the second author builds on the techniques developed here to consider the more general question of for which monoids S do we have $\mathrm{U}(p) \leq \mathrm{MR}(p) < \infty$. Our algebraic characterisation of such monoids allows us to give examples of S such that T_S is totally transcendental but is such that $\mathrm{U}(p) < \mathrm{MR}(p)$ for some type p. We remark that for a complete, stable theory T, if $p \in S(A)$ and $q \in S(B)$ with $A \subseteq B$, $p \subseteq q$ and $\mathrm{MR}(p)$ an ordinal, then $\mathrm{U}(p) = \mathrm{U}(q)$ if and only if $\mathrm{MR}(p) = \mathrm{MR}(q)$ [27].

The two conditions on monoids used in the characterisation theorem are the right noetherian property (that is, all right congruences are finitely generated) and the condition (MU) which we now explain. Let S be a monoid and let (I, ρ) be a congruence pair, that is, $(I, \rho) \in \mathcal{C}$. We say that (I, ρ) is *critical* if there is a finite subset K of $(S \times S) \setminus \rho$ such that for all right congruences θ which saturate I, contain ρ and agree with ρ on I (i.e. $\theta \cap (I \times I) = \rho \cap (I \times I)$) we have

$$K \subseteq (S \times S) \setminus \theta \text{ implies } \rho = \theta \text{ or } \theta\text{-R}(I) < \rho\text{-R}(I).$$

We then say that S *satisfies* (MU) if every congruence pair of S is critical.

Note that for any right congruence ρ, the congruence pair (S, ρ) is critical. In the very special case where S is a group, to show that S satisfies (MU) we need only show that $(\emptyset, \rho)$ is critical for every right congruence ρ. In this case, for *any* right congruence θ, we have that $\theta\text{-R}(\emptyset) = 1$. Thus to show that $(\emptyset, \rho)$ is critical, we need to find a finite set $K \subseteq (S \times S) \setminus \rho$ such that if $\rho \subseteq \theta$ and $K \subseteq (S \times S) \setminus \theta$, then $\rho = \theta$.

For any right coherent monoid S, if $(I, \rho) \in C$ and $\{s\rho : s \notin I\}$ is finite it is then easy to see that the pair (I, ρ) is critical.

Lemma 5.1. *For any right ideal I of a monoid S with S/I finite, every congruence pair (I, ρ) is critical. In particular, every finite monoid satisfies (MU).*

We now consider a useful sufficient condition for a monoid to satisfy (MU).

Proposition 5.1. *Let $C_r(S)$ be the lattice of right congruences of a monoid S. If $C_r(S)$ satisfies the minimal condition and each $\rho \in C_r(S)$ has only a finite number of covers, then S satisfies (MU).*

Proof. Let (I, ρ) be a congruence pair. If $S = I$, then we have already noted that the pair is critical. Otherwise, ρ cannot be universal since I is ρ-saturated and so the set of right congruences strictly containing ρ contains minimal members which are covers of ρ. Let $\rho_1, \ldots, \rho_t$ be these covers. For each $i \in \{1, \ldots, t\}$ choose a pair (a_i, b_i) in $\rho_i \setminus \rho$. Now put

$$K = \{(a_1, b_1), \ldots, (a_t, b_t)\}.$$

Suppose that $\theta \in C_r(S)$ and $\rho \subseteq \theta$. If $\rho \neq \theta$, then it follows from the minimal condition that $\rho_i \subseteq \theta$ for some i. Thus $(a_i, b_i) \in \theta$ and consequently K is not contained in $(S \times S) \setminus \theta$. Hence the pair (I, ρ) is critical and consequently S satisfies (MU). □

For groups, the converse of Proposition 5.1 is true as we now demonstrate.

Proposition 5.2. *A group G satisfies (MU) if and only if the lattice $L(G)$ of subgroups of G satisfies the minimal condition and every subgroup has only finitely many covers in $L(G)$.*

Proof. Suppose that G satisfies (MU) and let

$$\rho_1 \supseteq \rho_2 \supseteq \ldots$$

be a decreasing sequence of right congruences. Put $\rho = \bigcap\{\rho_i : i \in \omega\}$. By assumption, $(\emptyset, \rho)$ is critical and so there is a finite set K such that $K \subseteq (G \times G) \setminus \rho$ and if $K \subseteq (G \times G) \setminus \rho_m$, then $\rho = \rho_m$. If $(a, b) \in K$, then $(a, b) \notin \rho_t$ for some t and since K is finite, it follows that for some m we do have $K \subseteq (G \times G) \setminus \rho_m$. Hence $\rho_m = \rho_{m+1} = \ldots$ and $C_r(G)$ satisfies the minimal condition.

In view of the minimal condition, every $\rho \in C_r(G)$ except $G \times G$ actually does have covers. If $\{\rho_\lambda : \lambda \in \Lambda\}$ is the set of covers of ρ, then $\rho_\lambda \cap \rho_\mu = \rho$ for each $\lambda, \mu \in \Lambda$ with $\lambda \neq \mu$. Hence, if $(a, b) \notin \rho$, then (a, b) can belong to at most one of the covers of ρ. Since $(\emptyset, \rho)$ is critical, there is a finite set K such that $K \subseteq (G \times G) \setminus \rho$ and K is not contained in $(G \times G) \setminus \rho_\lambda$ for any cover ρ_λ of ρ. But any given pair in K is in at most one cover of ρ and so there are only finitely many covers of ρ.

Now use the fact that the lattice of right congruences on a group is isomorphic to the lattice of subgroups. □

We note that the quasi-cyclic group $Z(p^\infty)$ where p is a prime number satisfies the conditions of Proposition 5.2 and thus satisfies (MU). On the other hand the infinite cyclic group does not satisfy the minimal condition for subgroups and hence does not satisfy (MU). It is, in fact, easy to show that the congruence pair $(\emptyset, \iota)$ is not critical in this case.

We have introduced the condition (MU) to help in our discussions of Morley rank. To define the latter we use make use of the natural topology on Stone spaces of types.

Let T be a complete theory and let $A \subseteq \mathbf{M}$. Then $S(A)$ may be made into a topological space by specifying the sets

$$\langle \phi(x) \rangle = \{p \in S(A) : \phi(x) \in p\}$$

as a basis of open sets, where $\phi(x)$ is a formula of $L(A)$. The space $S(A)$ has a basis of clopen sets $\langle \phi(x) \rangle$, and is compact and Hausdorff.

If T is a theory which has elimination of quantifiers (for example, T_S), then a routine argument gives that the sets $\langle \theta(x) \rangle$ where $\theta(x)$ is a conjunction of atomic and negated atomic formulae form a basis for the topology of $S(A)$.

Let T be a complete theory in a first order language L and let A be a subset of a model of T. Subsets $MR^\alpha(A)$ of $S(A)$ are defined by induction on the ordinal α as follows:

(i) $MR^0(A) = S(A)$.

(ii) If α is a limit ordinal, then

$$MR^\alpha(A) = \bigcap \{MR^\beta(A) : \beta < \alpha\}.$$

(iii) For any α, $MR^{\alpha+1}(A) = MR^\alpha(A) \setminus X^\alpha$, where

$$X^\alpha = \{p \in MR^\alpha(A) : \text{ for all } B \supseteq A \text{ and all extensions } q \text{ of } p \text{ on } B,$$

$$q \notin MR^\alpha(B) \text{ or } q \text{ is isolated in } MR^\alpha(B)\}.$$

We may take B to be an L-substructure of a model of T.

For $p \in S(A)$, the *Morley rank* of p is $\mathrm{MR}(p)$ where, if $p \in MR^{\alpha}(A)$ for all α, then $\mathrm{MR}(p) = \infty$ and otherwise $\mathrm{MR}(p)$ is α where $p \in MR^{\alpha}(A) \setminus MR^{\alpha+1}(A)$. If $\mathrm{MR}(p) < \infty$, then we say that p *has Morley rank.*

It is a standard result that for all types p, $\mathrm{U}(p) \leq \mathrm{MR}(p)$ [27]; we need this in the proof of the main result of this section. We first note that for any type p over an S-act A, $\mathrm{MR}(p) = 0$ if and only if $I_p = S$, that is $\mathrm{U}(p) = 0$. For if $I_p = S$ and $p \subseteq q$ where $q \in S(B)$, then since $1 \in I_q$, $x = b \in q$ for some $b \in B$ and $\{x = b\}$ isolates q in $S(B)$. Thus $p \notin MR^1(A)$ so that $\mathrm{MR}(p) = 0$. The converse is clear.

We can now state the main result of this section; the interested reader may wish to compare our result with that of [17], where some conditions are given which imply that for a relevant type p, $\mathrm{MR}(p) = \mathrm{U}(p)$.

Theorem 5.1. *For every type p over an S-act A, $MR(p) = U(p) < \infty$ if and only if S is right noetherian and satisfies (MU).*

Proof. Suppose first that the condition on ranks of types holds. Let (I, ρ) be a congruence pair. By Corollary 3.1, there is an S-act A and a type p over A with $I_p = I, \rho_p = \rho$. Let the associated A-triple be (I, ρ, f) and let p have Morley rank α. Then there is an open set U in $S(A)$ such that $p \in U$ and $\mathrm{MR}(q) < \alpha$ for all q in $U \setminus \{p\}$. Let $U = \langle \phi(x) \rangle$ where $\phi(x)$ is a conjunction of sets of formulae:-

$$\{xr_i = a_i : i \in \Lambda_1\}, \{xs_j = xt_j : j \in \Lambda_2\},$$

$$\{xu_k \neq xv_k : k \in \Lambda_3\}, \{xw_\ell \neq b_\ell : \ell \in \Lambda_4\}$$

where the index sets $\Lambda_1, ..., \Lambda_4$ are all finite. Since $p \in \langle \phi(x) \rangle$, each r_i is a member of I and each pair (s_j, t_j) is in ρ.

Let θ be any right congruence on S which saturates I, properly contains ρ and agrees with ρ on I. Then (I, θ, f) is an A-triple; let $\overline{p}$ be the associated type over A. Certainly each pair (s_j, t_j) is in θ since $\rho \subseteq \theta$. Thus we see that the sets $\{xr_i = a_i : i \in \Lambda_1\}$ and $\{xs_j = xt_j : j \in \Lambda_2\}$ are contained in $\overline{p}$. If the formula $xw_\ell = b_\ell$ is in $\overline{p}$ for some $\ell \in \Lambda_4$, then $w_\ell \in I$ and $f(w_\ell) = b_\ell$ and consequently, $xw_\ell = b_\ell$ is in p, a contradiction. Thus each inequation $xw_\ell \neq b_\ell$ is in $\overline{p}$ and we see that $\phi(x) \in \overline{p}$ if and only if $xu_k \neq xv_k$ is in $\overline{p}$ for each $k \in \Lambda_3$.

Let $K = \{(u_1, v_1), ..., (u_m, v_m)\}$; since $xu_k \neq xv_k$ is in p we certainly have that $K \subseteq (S \times S) \setminus \rho$. If $K \subseteq (S \times S) \setminus \theta$, then we have $\phi(x) \in \overline{p}$ so that $\overline{p} \in \langle \phi(x) \rangle$ and hence $\mathrm{MR}(\overline{p}) < \mathrm{MR}(p)$. But $\mathrm{U}(\overline{p}) = \mathrm{MR}(\overline{p})$ and $\mathrm{U}(p) =$

$\mathrm{MR}(p)$ so that $\theta\text{-R}(I) < \rho\text{-R}(I)$. Thus (I, ρ) is critical and hence S satisfies (MU).

To see that S is right noetherian we consider the case $I = \emptyset$. Let σ be the right congruence on S generated by $\{(s_j, t_j) : j \in \Lambda_2\}$. Certainly $\sigma \subseteq \rho$ and if p_1 is the type over $\emptyset$ associated with σ, then clearly $p_1 \in \langle \phi(x) \rangle$. Hence, using our assumption on ranks,

$$\mathrm{MR}(p) = \mathrm{U}(p) = \rho\text{-R}(\emptyset) \leq \sigma\text{-R}(\emptyset) = \mathrm{U}(p_1) = \mathrm{MR}(p_1) \leq \mathrm{MR}(p),$$

that is, $\mathrm{MR}(p_1) = \mathrm{MR}(p)$. By the choice of $\langle \phi(x) \rangle$, we have that $p = p_1$ so that $\rho = \sigma$ and ρ is finitely generated.

Conversely, suppose that S is right noetherian and satisfies (MU). By Theorem 4.2, T_S is certainly superstable so that for every S-act A, every type p in $S(A)$ has U-rank. We show by induction that for every p, $\mathrm{MR}(p) = \mathrm{U}(p)$.

If $\mathrm{U}(p) = 0$, then $I_p = S$ and so, as already noted, $\mathrm{MR}(p) = 0$.

Now let $p \in S(A)$ and $\mathrm{U}(p) = \alpha$ and suppose that for all S-acts B and all types $q \in S(B)$ with $\mathrm{U}(q) < \alpha$ we have $\mathrm{MR}(q) = \mathrm{U}(q)$. Let $I = I_p$, $\rho = \rho_p$. Certainly $\mathrm{U}(p) \leq \mathrm{MR}(p)$ so we have $p \in MR^{\alpha}(A)$ and we wish to show that $p \notin MR^{\alpha+1}(A)$, that is, for every S-act B containing A and every extension q of p over B we want either $q \notin MR^{\alpha}(B)$ or q is isolated in $MR^{\alpha}(B)$.

So let $q \in S(B)$ where B is an extension of A and $q|A = p$. Suppose that $q \in MR^{\alpha}(B)$. We have to find an open set U such that $MR^{\alpha}(B) \cap U = \{q\}$. By Proposition 3.3, we have $I \subseteq I_q$ and $\rho = \rho_q$. Now $\alpha \leq \mathrm{MR}(q)$ and so by the inductive assumption we cannot have $\mathrm{U}(q) < \alpha$. But $\mathrm{U}(q) \leq \mathrm{U}(p) = \alpha$ so that $\mathrm{U}(q) = \alpha$. Now by the definition of U-rank, we must have that q is a non-forking extension of p and so by Lemma 4.1, $I = I_q$.

As S is right noetherian, $I = \bigcup\{w_i S : i \in \Lambda\}$ for some finite set Λ and ρ is generated by a finite subset H of $S \times S$. For each $i \in \Lambda$, let $a_i = f_q(w_i)$. By assumption, the pair (I, ρ) is critical. Let K be the finite subset of $(S \times S) \setminus \rho$ required in the definition of criticality and let $\xi(x)$ be the formula obtained by taking the conjunction of the following sets of formulae:

$$\{xw_i = a_i : i \in \Lambda\}, \{xs = xt : (s,t) \in H\}, \{xu \neq xv : (u,v) \in K\}.$$

Then $q \in \langle \xi(x) \rangle$. Let $r \in \langle \xi(x) \rangle$ and suppose that $\mathrm{MR}(r) \geq \alpha$. Our aim is to show that $r = q$ and this will complete the proof that $\mathrm{U}(p) = \mathrm{MR}(p)$ and hence prove the result by induction.

Note that $I \subseteq I_r$ and $\rho \subseteq \rho_r$ so that I_r is ρ-saturated. If $I \neq I_r$, then, by Proposition 3.4, q has an extension $\overline{q}$ with $I_{\overline{q}} = I_r$ and by Lemma 4.1,

$\overline{q}$ is a forking extension of q. Hence, $\mathrm{U}(\overline{q}) < \mathrm{U}(q) = \alpha$. By Proposition 3.3, $\rho_{\overline{q}} = \rho$ and thus

$$\mathrm{U}(r) = \rho_r\text{-R}(I_r) \leq \rho_{\overline{q}}\text{-R}(I_{\overline{q}}) = \mathrm{U}(\overline{q}) < \alpha.$$

The inductive assumption gives $\mathrm{MR}(r) < \alpha$, a contradiction, so that we may suppose that $I = I_r$.

Since f_q and f_r agree on the set of generators $\{w_i : i \in \Lambda\}$ of I, it follows that $f_q = f_r$ and $\rho_r \cap (I \times I) = \ker f_r = \ker f_q = \rho \cap (I \times I)$.

If $\rho \neq \rho_r$, then as $K \subseteq (S \times S) \setminus \rho_r$ we have $\rho_r\text{-R}(I_r) < \rho\text{-R}(I)$ so that $\mathrm{U}(r) < \mathrm{U}(q) = \alpha$ and the inductive assumption gives $\mathrm{MR}(r) < \alpha$, a contradiction. Thus $\rho_r = \rho$ and, as $f_r = f_q$, Corollary 3.1 now gives $r = q$ as desired. □

We have noted already that the infinite cyclic group does not satisfy (MU) although, of course, it is (right) noetherian. On the other hand the group $Z(p^\infty)$ is not (right) noetherian but does satisfy (MU). Thus the two conditions in the theorem are independent. Furthermore, these observations also show that there are monoids S such that T_S is totally transcendental (ω-stable) but such that for some S-act A there is a type p in $S(A)$ with $\mathrm{U}(p) < \mathrm{MR}(p)$.

We can be more precise with our two examples. For any group G and any type p over a G-set A we have $I_p = G$ or $I_p = \emptyset$. In the former case $\mathrm{U}(p) = \mathrm{MR}(p) = 0$ and in the latter case $\mathrm{U}(p) = 1$. It is not difficult to see that if $p \in S(\emptyset)$ (so that necessarily $I_p = \emptyset$), then for any G-set A there is exactly one extension p_A of p in $S(A)$ with $I_{p_A} = \emptyset$. A simple argument using transfinite induction shows that for all ordinals $\alpha \geq 1$, $\mathrm{MR}(p) \geq \alpha$ if and only if $\mathrm{MR}(p_A) \geq \alpha$ for all G-sets A. It follows that $\mathrm{MR}(p) = \alpha$ if and only if $p \in MR^\alpha(\emptyset)$ and p is isolated in $MR^\alpha(\emptyset)$. Moreover, $\mathrm{MR}(p) = \mathrm{MR}(p_A)$ for all G-sets A.

It is now not difficult to show that for the infinite cyclic group G with generator g, if p_n is the type in $S(\emptyset)$ corresponding to the subgroup generated by g^n, then $\mathrm{MR}(p_n) = 1$ for $n \geq 1$ and $\mathrm{MR}(p_0) = 2$. Thus $\mathrm{U}(p_0) < \mathrm{MR}(p_0)$.

Similarly, if $G = Z(p^\infty)$ is regarded as the group of all p^n-th roots of unity for all $n \geq 1$ and if for each n, p_n is the type in $S(\emptyset)$ corresponding to the subgroup generated by a primitive p^n-th root of one, then $\mathrm{MR}(p_n) = 1$. For the type p_∞ in $S(\emptyset)$ corresponding to G itself we find that $\mathrm{MR}(p_\infty) = 2$ so that $\mathrm{U}(p_\infty) < \mathrm{MR}(p_\infty)$.

6. Right noetherian monoids which satisfy (MU)

The main result of the preceding section makes it natural to consider the monoids of the title. As the condition (MU) is rather complicated it is far from clear precisely which monoids satisfy (MU). Of course, any finite monoid is right noetherian and also, by Lemma 5.1, satisfies (MU). One of the main results of this section shows that the converse is true for an extensive class of monoids, namely the weakly periodic monoids. However, not every right noetherian monoid which satisfies (MU) is finite. We will show that an infinite example is the free commutative monoid on one generator.

Our first objective is to show that (right) noetherian groups which satisfy (MU) are finite. To this end we need the lemma below which can be deduced from König's Lemma, but which is very easy to prove directly in much the same way that König's Lemma is proved.

Lemma 6.1. *Let Y be a lattice satisfying the finite chain condition. If every member of Y has only finitely many covers, then Y is finite.*

Proof. Since Y satisfies the descending chain condition, it has a least element x_0. If Y is infinite, then since x_0 has only finitely many covers, x_0 has a cover x_1 such the filter above x_1 is infinite. But x_1 has only finitely many covers, so there must be one of these, say x_2, such that the filter above x_2 is infinite. Continuing in this way we find an infinite chain

$$x_0 < x_1 < x_2 < \dots$$

of elements of Y, contradicting the ascending chain condition. □

Corollary 6.1. *Let G be a right noetherian group which satisfies (MU). Then G is finite.*

Proof. By Proposition 5.2, the lattice $\mathcal{L}(G)$ of subgroups of G satisfies the minimal condition and every subgroup has only finitely many covers in $\mathcal{L}(G)$. Since $\mathcal{L}(G)$ also satisfies the maximal condition, it has the finite chain condition and by Lemma 6.1, $\mathcal{L}(G)$ is finite. As pointed out on pp.170-171 of [2], it follows easily that G is finite. □

The next stage in our argument is to show that any subgroup of a monoid which is right noetherian and satisfies (MU) inherits these properties. To do this we utilise some classical semigroup theory, in particular, basic results about Green's relations $\mathcal{L}, \mathcal{R}$ and $\mathcal{H}$. The relation $\mathcal{L}$ is defined on a monoid S by the rule that for any $a, b \in S$, $a \mathcal{L} b$ if and only if $Sa = Sb$.

The relation $\mathcal{R}$ is defined dually; $\mathcal{H} = \mathcal{L} \cap \mathcal{R}$. Note that $\mathcal{L}$ ($\mathcal{R}$) is a right (left) congruence. Details may be found in any of the standard texts. We recommend [15].

Lemma 6.2. *If the monoid S is right noetherian, then so is every subgroup.*

Proof. Let G be a subgroup of S. For any right congruence ρ on G, let $\overline{\rho}$ denote the right congruence *on* S generated by ρ. If $a, b \in S$ and $a\,\overline{\rho}\,b$, then $a = b$ or there exists a sequence

$$a = c_1 t_1, d_1 t_1 = c_2 t_2, \ldots, d_\ell t_\ell = b,$$

where $(c_i, d_i) \in \rho, 1 \leq i \leq \ell$. Notice in particular that $a\,\mathcal{L}\,b$. Suppose now that $a, b \in G$. We claim that $\overline{\rho} \cap (G \times G) = \rho$. Let e be the identity of G. Then we certainly have

$$a = c_1(et_1), d_1(et_1) = c_2(et_2), \ldots, d_\ell(et_\ell) = b.$$

Taking inverses in G we have

$$et_1 = c_1^{-1} a \in G.$$

This gives that $a\,\rho\,d_1(et_1)$. Now

$$et_2 = c_2^{-1} d_1(et_1) \in G,$$

so that $a\,\rho\,d_2(et_2)$. Continuing in this manner we obtain $a\,\rho\,b$. Thus G is $\overline{\rho}$-saturated and $\overline{\rho} \cap (G \times G) = \rho$ as required. It is now easy to see that if S is right noetherian, so also is G. □

Lemma 6.3. *If the monoid S is right noetherian and satisfies (MU), then so does every maximal subgroup.*

Proof. Let G be a maximal subgroup of S, so that G is a group $\mathcal{H}$-class. We already know from Lemma 6.2 that G is (right) noetherian. Suppose now that S satisfies (MU). To show that G satisfies (MU) we need only prove that the pair $(\emptyset, \rho)$ is critical for any right congruence ρ on G.

Let e be the identity of G, let

$$I = \cup\{SaS : SaS \subset SeS\} \text{ and } J = SeS.$$

Then I and J are ideals of S. From Theorem 1.3 of [14] we know that the principal factor J/I is completely 0-simple or completely simple. Let $\overline{\rho}$ be defined as in Lemma 6.2; since $\rho \subseteq \mathcal{L}$ and $\mathcal{L}$ is a right congruence, we have that $\overline{\rho} \subseteq \mathcal{L}$. Thus any ideal of S is $\overline{\rho}$-saturated. Let ν_I be the Rees

congruence associated with I, so that for any $a, b \in S$, $a \nu_I b$ if and only if $a = b$ or $a, b \in I$. Since I is $\overline{\rho}$-saturated and ν_I-saturated, it is clear that $\tilde{\rho} = \overline{\rho} \cup \nu_I$ is a right congruence saturating I. Moreover, for any $a, b \in S$, if $a \neq b$ and $a \tilde{\rho} b$, then either $a, b \in I$ or $a, b \in J \setminus I$. In the latter case, we have $a \overline{\rho} b$ and so, since J/I is completely (0)-simple, it follows that $a \mathcal{H} b \mathcal{R} e$. Consequently, any right ideal containing I is $\tilde{\rho}$-saturated. Thus if θ is any right congruence on G, then $\tilde{\rho}\text{-R}(I) = \tilde{\theta}\text{-R}(I)$.

The congruence pair $(I, \tilde{\rho})$ is critical; let $K \subseteq (S \times S) \setminus \tilde{\rho}$ be a finite set of pairs guaranteed by the fact that $(I, \tilde{\rho})$ is critical. We need to pick a set of pairs of elements of G that will enable us to show that $(\emptyset, \rho)$ is critical.

For any pair

$$(a, b) \in K \cap \mathcal{H} \cap (R_e \times R_e)$$

choose and fix $c = c_{(a,b)} \in J \setminus I$ with $ac, bc \in G$. It follows from the fact that J/I is completely (0)-simple that $(ac, bc) \notin \rho$. We now put

$$H = \{(ac, bc) : (a, b) \in K \cap \mathcal{H} \cap (R_e \times R_e)\},$$

so that $H \subseteq (G \times G) \setminus \rho$.

Let θ be a right congruence on G containing ρ and such that $H \subseteq (G \times G) \setminus \theta$. Certainly $\tilde{\rho} \subseteq \tilde{\theta}$, I is $\tilde{\theta}$-saturated and $\tilde{\rho} \cap (I \times I) = \tilde{\theta} \cap (I \times I)$. If $K \not\subseteq (S \times S) \setminus \tilde{\theta}$, then there exists $(a, b) \in K \cap \tilde{\theta}$. But $(a, b) \notin \tilde{\rho}$, so we are forced to deduce that $a, b \in R_e$ and $a \mathcal{H} b$. Consequently,

$$(ac, bc) \in \tilde{\theta} \cap (G \times G) = \overline{\theta} \cap (G \times G) = \theta.$$

But $(ac, bc) \in H$, a contradiction. Thus $K \subseteq (S \times S) \setminus \tilde{\theta}$. Now by the definition of critical pair, $\tilde{\rho} = \tilde{\theta}$ or $\tilde{\rho}\text{-R}(I) < \tilde{\theta}\text{-R}(I)$. But the latter is impossible by previous comments on saturation of right ideals. We conclude that $\tilde{\rho} = \tilde{\theta}$ and consequently, $\rho = \theta$ as required. □

From Lemmas 6.1, 6.3 we deduce the following.

Theorem 6.1. *If S is a right noetherian monoid which satisfies (MU), then all subgroups of S are finite.*

A semigroup S is *weakly periodic* if for every element s of S there is a positive integer $n = n(s)$ such that $I^2 = I$ where $I = S^1 s^n S^1$. If S is a semigroup which satisfies the minimal condition for principal ideals or for principal right (or left) ideals or if S is periodic, then S is weakly periodic. Regular and eventually regular (some power of any element is regular) semigroups are weakly periodic as are semisimple semigroups, that is, semigroups with no null principal factors.

Corollary 6.2. *If S is a weakly periodic right noetherian monoid which satisfies (MU), then S is finite.*

Proof. By Theorem 6.1, all subgroups of S are finite. Hence by Theorem 2.3 of [14], S is finite. □

Corollary 6.3. *Let S be a right noetherian monoid which satisfies (MU). If the relation $\mathcal{R}$ is a congruence on S and there are only finitely many trivial $\mathcal{R}$-classes, then S is finite.*

Proof. We show that S is weakly periodic so that the result follows from Corollary 6.2. Let $a \in S$ and consider the sequence $S \supseteq aS \supseteq a^2S \supseteq \ldots$. Let $I = \bigcap\{a^iS : i \in \omega\}$, ρ be the Rees right congruence associated with I and ρ_i that associated with a^iS. If $I = \emptyset$, then we take ρ to be ι. The pair (I, ρ) is critical and so there is a finite subset K of $(S \times S) \setminus \rho$ such that for any right congruence θ with $K \subseteq (S \times S) \setminus \theta$ where θ saturates I, agrees with ρ on I and contains ρ, we have either $\rho = \theta$ or $\theta\text{-R}(I) < \rho\text{-R}(I)$. Since K is finite, $K \subseteq (S \times S) \setminus \rho_n$ for some n. By hypothesis, $a^pS = I$ for some p, or there is an element a^m with $n \leq m$ whose $\mathcal{R}$-class is non-trivial.

In the latter case, suppose that $a^mS \neq I$. Let x, y be distinct elements in the $\mathcal{R}$-class of a^m and let ν be the right congruence generated by the set $\rho \cup \{(x, y)\}$. It is easy to see that if $(u, v) \in \nu$ and $u \neq v$, then $u, v \in a^mS$ and either $u \mathcal{R} v$ or $u, v \in I$. Thus $\rho \subset \nu \subseteq \rho_m$ and hence $K \subseteq (S \times S) \setminus \nu$. Furthermore, ν saturates I and agrees with ρ on I and consequently, $\nu\text{-R}(I) < \rho\text{-R}(I)$. But all right ideals which contain I are both ρ-saturated and ν-saturated since as noted above, if $(u, v) \in \nu$ and $u, v \notin I$, then $u \mathcal{R} v$. Hence $\nu\text{-R}(I) = \rho\text{-R}(I)$, a contradiction. It follows that if $a \in S$ then the descending chain of principal right ideals $S \supseteq aS \supseteq a^2S \supseteq \ldots$ is finite. Thus $a^qS = I$ for some q so that $a^qS = a^{q+1}S = \ldots$. Hence $a^q = a^{2q}s$ for some $s \in S$ and so $a^qS = (a^qS)^2$. It follows that $Sa^qS = (Sa^qS)^2$, and S is weakly periodic. □

On a commutative monoid the relations $\mathcal{H}, \mathcal{R}$ and $\mathcal{L}$ coincide and $\mathcal{R}$ is automatically a congruence. The following result is thus an immediate consequence of Corollary 6.3.

Corollary 6.4. *Let S be a noetherian commutative monoid which satisfies (MU). If S has only finitely many trivial $\mathcal{R}$-classes, then S is finite.*

We now give an example of an infinite noetherian commutative monoid which satisfies (MU). Of course, in view of Corollary 6.4, our example must

have infinitely many trivial $\mathcal{R}$-classes.

Proposition 6.1. *The additive monoid* $\mathbb{N}$ *of non-negative integers is noetherian and satisfies (MU).*

Proof. It is well known and easy to show directly that $\mathbb{N}$ is noetherian. If I is a non-empty ideal of $\mathbb{N}$, then $\mathbb{N}/I$ is finite so that it follows from Lemma 5.1 that any congruence pair (I, ρ) is critical. It remains to consider pairs $(\emptyset, \rho)$. If $\rho = \iota$, then $\iota\text{-R}(\emptyset) = \omega$. When $\rho \neq \iota$, let r, m be the smallest integers such that $(r, r+m) \in \rho$ and $m \geq 1$. In fact, from page 137, Exercise 5 of [7] we know that ρ is generated by $(r, r+m)$. It is then easy to see that $\rho\text{-R}(\emptyset)$ is finite so that $(\emptyset, \iota)$ is critical by choosing $K = \emptyset$. Further, putting

$$K = \{(s, s+n) : 0 \leq s \leq r, 0 \leq n \leq m\} \setminus \{(r, r+m)\},$$

it is clear that $K \subseteq (S \times S) \setminus \rho$. But if $\rho \subset \theta$, then $K \cap \theta \neq \emptyset$ and consequently the pair $(\emptyset, \rho)$ is critical. Thus $\mathbb{N}$ satisfies (MU). □

In our final result we show that $\mathbb{N}$ is the only infinite commutative cancellative principal ideal monoid which is both noetherian and satisfies (MU).

Proposition 6.2. *Let* S *be a commutative, cancellative principal ideal monoid. Then* S *is noetherian and satisfies (MU) if and only if* S *is a finite group or is isomorphic to* $\mathbb{N}$.

Proof. Suppose that S is noetherian and satisfies (MU). If S is finite, then since it is cancellative, it must be a group.

If S is infinite, then by Corollary 6.4, S must have infinitely many trivial $\mathcal{H}$-classes. Let a be a unit of S so that $a \mathcal{H} 1$. For any element $c \in S$, we have $ac \mathcal{H} c$ since $\mathcal{H}$ is a congruence on S. If $a \neq 1$ then $ac \neq c$ since S is cancellative and so H_c is non-trivial unless $a = 1$. Thus the group of units of S is trivial. It follows from Theorem 12 of [8] that S is isomorphic to $\mathbb{N}$. □

References

1. J. T. Baldwin, *Fundamentals of stability theory*, (Springer-Verlag, 1988).
2. G. Birkhoff, *Lattice Theory*, 3rd. Edition (American Math. Soc., Providence, R. I., 1967).
3. V. S. Bogomolov and T. G. Mustafin, *Description of commutative monoids over which all polygons are ω-stable*, Algebra and Logic **28** (1989), 239–247.
4. E. Bouscaren, *Modules existentiellement clos : types et modèles premiers*, (Thèse 3ème cycle, Université Paris VIII, Paris, 1979).

5. E. Bouscaren (Ed.), *Model Theory and Algebraic Geometry*, (Springer, Lecture Notes in Mathematics, 1999).
6. C. C. Chang and H. K. Keisler, *Model Theory*, (North-Holland, Amsterdam, 1973).
7. A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups* Vol.II (American Math. Soc., Providence, R. I., 1967).
8. M. P. Dorofeeva, V. L. Mannepalli and M. Satyanarayana, *Prüfer and Dedekind monoids*, Semigroup Forum **9** (1975), 294–369.
9. H. B. Enderton, *A mathematical introduction to logic*, (Academic Press, New York, 1972).
10. J. B. Fountain and V. A. R. Gould, *Stability of the theory of existentially closed S-sets over a right coherent monoid S*, `www-users.york.ac.uk/$\ sim$vargl/stabilitynotes.ps`.
11. V. A. R. Gould,*Model companions of S-systems*, Quart. J. Math., Oxford **38** (1987), 189-211.
12. V. A. R. Gould,*Coherent monoids*, J. Australian Math. Soc., **53** (1992), 166–182.
13. V. A. R. Gould, *A notion of rank for right congruences on semigroups*, Comm. in Algebra **33** (2005), 4631–4656.
14. E. Hotzel, *On semigroups with maximal conditions*, Semigroup Forum **11** (1975/6), 337-362.
15. J. M. Howie, *An introduction to semigroup theory*, (Academic Press, London, 1976).
16. A. Ivanov, *Simple existentially closed extensions of unoids*, Math. Notes **44** (1988), 724–728.
17. A. Ivanov, *Structural problems for model companions of varieties of polygons*, Siberian Math. J. **33** (1992), 259–265.
18. M. Kilp, U. Knauer and A. V. Mikhalev, *Monoids, Acts and Categories*, (Water De Gruyter, Berlin New York, 2000).
19. D. Lascar, *Stability in model theory* (Longman, London, 1987).
20. D. Lascar and B.Poizat, *An introduction to forking*, J. Symb. Logic **44** (1979), 330-350.
21. R. N. Mckenzie, G. F. Mcnulty and W. F. Taylor, *Algebras, lattices, varieties* Vol.I (Wadsworth, Belmont 1987).
22. M. D. Morley, *Categoricity in power*, Trans. American Math. Soc. **114** (1965), 514-538.
23. T. G. Mustafin, *Stability of the theory of polygons*, Tr. Inst. Mat. Sib. Otd. (SO) Akad. Nauk SSSR **8** (1988), 92-108 (in Russian); translated in Model Theory and Applications, American Math. Soc. Transl. 2 **295** 205–223, (Providence R.I. 1999).
24. T. G. Mustafin and B. Poizat, *Polygones, Math. Logic Quart.* **41** (1995), 93–110.
25. A. Pillay, *Countable models of stable theories*, Proc. American Math. Soc. **89** (1983), 666–672.
26. A. Pillay, *An introduction to stability theory* (Oxford University Press, 1983).
27. A. Pillay, *Geometric stability theory*, (Oxford Logic Guides 32, Clarendon

Press, 1996).
28. M. Prest, *Model theory and modules* (LMS Lecture Notes 130, Cambridge University Press, 1988).
29. G. E. Sacks, *Saturated model theory* (W.A. Benjamin, Reading, Mass., 1972).
30. S. Shelah, *Classification theory and the number of non-isomorphic models* (North-Holland, Amsterdam, 1978).
31. A. A. Stepanova, *Monoids with stable theories for regular polygons*, Algebra and Logic **40** (2001), 239–254.
32. W. H. Wheeler, *Model companions and definability in existentially complete structures*, Israel J. Math. **25** (1976), 305–330.

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 157–178)

PAPER-FOLDING, POLYGONS, COMPLETE SYMBOLS, AND THE EULER TOTIENT FUNCTION: AN ONGOING SAGA CONNECTING GEOMETRY, ALGEBRA, AND NUMBER THEORY*†

PETER HILTON

State University of New York, Binghamton,
Binghamton, NY 13902-6000

JEAN PEDERSEN* and BYRON WALDEN**

Santa Clara University,
Santa Clara, California 95053-0290
**E-mail: jpedersen@scu.edu*
***E-mail: bwalden@scu.edu*

The Greeks understood, around 350 B.C., how to construct, with Euclidean tools, regular N-gons for $N = 2^c N_0$, where $N_0 = 1, 3, 5$ or 15. Two thousand years later, Gauss proved that a Euclidean construction of a regular N-gon is possible if and only if $N = 2^c \times$ (product of distinct Fermat primes).

Here we are content to constuct arbitrarily close approximations to regular polygons. Our constructions lead to some interesting number theory involving the Euler totient function. For a given odd number b, and a given odd number $a < \frac{b}{2}$, we construct a numerical array, called a symbol and describe an explicit procedure based on the symbol for constructing a regular star $\left\{\frac{b}{a}\right\}$-gon.

We can combine the symbols for a given b to produce a complete symbol where each constituent symbol is called a coach. We present two theorems, the Quasi-Order Theorem and the Coach Theorem, which show how the numbers appearing in the symbol (and hence the steps in our constructions) are governed by the values of $\Phi(b)$ and the quasi-order of b *mod* 2. We then generalize the results to any arbitrary base.

Keywords: Paper-folding; Polygons; Euclidean construction; Quasi-order theorem; Complete symbol; Coach theorem; Euler totient function.

*Dedicated to Martin Gardner, who provided the original inspiration for our paper-folding activities.

†2000 Mathematics Subject Classification. 11A99, 51M99

1. Introduction

As is well-known, roundabout 350 B.C. the Greeks were fascinated with the idea of constructing regular N-gons with Euclidean tools (straight edge and compass). They were successful in constructing regular N-gons for $N = 2^c N_0$, with $N_0 = 1, 3, 5, \text{or } 15$, and $N \geq 3$ (so that the polygon will exist). Naturally the Greeks would have liked to answer the question for every N, but, in fact, it seems that no further progress was made until about 2000 years later, when Gauss (1777 – 1855) completely settled the question by proving that a Euclidean construction of a regular N-gon is possible if and only if N is of the form

$$N = 2^c \text{ (a product of distinct Fermat primes)} .$$

A Fermat number has the form $F_n = 2^{2^n} + 1$. If it is prime we call it a Fermat prime.

Gauss' remarkable theorem tells us when a Euclidean construction is possible, provided we know which Fermat numbers are prime — which we don't! However, we do know that the following Fermat numbers are prime:

$F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$, and $F_4 = 65537$.

The great Swiss mathematician Euler (1707 – 1783) showed that $F_5 (= 2^{32} + 1)$ is not prime and, although many composite Fermat numbers have been identified, to this day no other Fermat numbers have proven to be prime, beyond those listed above. Thus, even with Gauss's contribution, a Euclidean construction of a regular N-gon is known to exist for very few values of N; and, even for these N, we do not know, in all cases, an explicit construction.

In the middle of the 20th century one of the authors (JP) discovered a systematic folding procedure, using straight strips of paper (like adding machine tape), that produced some regular polygons to any desired degree of accuracy.[a] Shortly thereafter she began working with PH and they described a systematic method of producing a regular b-gon, to any desired degree of accuracy, for any number $b \geq 3$. We note here that our objective is distinctly different from what the Greeks and Gauss were attempting to do, in that we systematically fold straight strips of paper, as will be described briefly below, and, once sufficient convergence has taken place,

[a] The paper-folding result came about by trying to construct hexa-flexagons, written about by Martin Gardner in Scientific American (December 1956.) Martin Gardner later wrote about JP's braided polyhedra in his book Wheels, Life, and Other Mathematical Amusements (both of these may be found in [1]).

we use the folded strip to construct an approximation to the desired regular polygon, or star polygon.

Thus JP and PH redefined the Greeks' original problem. They agreed that they would be content to produce approximations to regular b-gons as long as they could depend on the error constantly becoming smaller. This seems reasonable since Euclidean constructions are only perfect in the mind — after all, what is actually produced is a function of how sharp your pencil is, how steady you hold the compass, and how carefully you place the ruler. Thus, even with Euclidean constructions, there are inevitable inaccuracies, due to human error. In the systematic folding procedures which we use, each correct fold that is made cuts any previous error in half — and, as you would expect, this produces very respectable (even if not completely perfect!) regular polygons.

The work constructing regular polygons leads to some fascinating number theory involving the Euler totient function[b]. To deal with some of the mathematical questions arising, JP and PH brought in the third author BW.

In order to set the scene for the number-theory aspect of our story we now give a very abbreviated version of how polygons may be obtained from the folded paper and refer the interested reader to the various references for more details about the actual constructions.

First we need to explain how a suitably creased strip of tape may be folded by what we call the FAT-algorithm to produce a regular convex p-gon. Figure 1(a) shows a strip of paper on which the dotted lines indicate certain special crease lines. Assume that the crease lines at the vertices labeled V_1, V_2, $\cdots$ which are on the top edge, form identical angles of $\frac{\pi}{p}$ (π radians$= 180^\circ$) with the top edge with an identical angle of $\frac{\pi}{p}$ between the two downward crease lines. Now, if you fold first on the longer crease line coming from V_1 and then (twisting the paper in the same spiral direction) on the shorter line coming from V_1 you will see that the top edge of the strip has rotated through an angle of $\frac{2\pi}{p}$. Repeating the process at the points V_2, V_3, V_4, will produce the portion of a regular polygon shown in

[b] The Euler totient function, Φ, counts the number of positive numbers less than a positive number b that are relatively prime to b. It is a well-known result of number theory that if p is prime then $\Phi(p^n) = (p-1)p^{n-1}$. Furthermore, for all mutually prime positive numbers k and ℓ, $\Phi(k \cdot \ell) = \Phi(k)\Phi(\ell)$. Thus we can calculate $\Phi(m)$ for any positive integer m. Another, equivalent, definition of Φ is $\Phi(n) = n\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right)$, where $p_1, p_2, \cdots, p_k$ are the distinct prime factors of n.

Figure 1(b). When this process of folding and twisting, which we call the FAT algorithm, is repeated p times the top edge will have rotated through an angle of 2π and hence a regular p-gon will emerge.

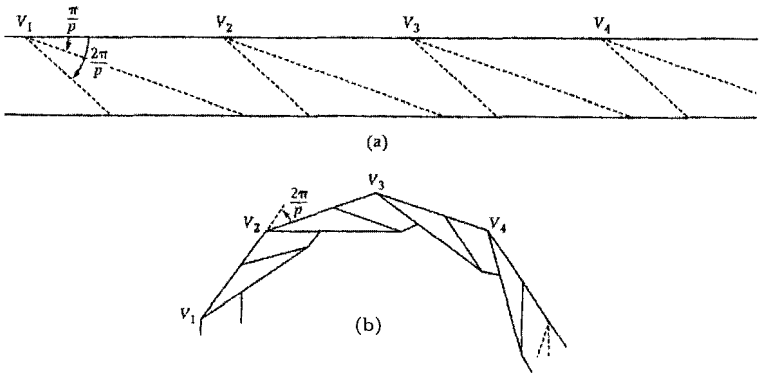


FIGURE 1

Since the 7-gon is the polygon with the smallest number of sides for which no Euclidean construction exists, we show a suitably folded strip of paper in Figure 2(a) that can be used to construct it. This strip of paper may be obtained by starting with an arbitrary fold line going downwards at the left-hand side of the strip. We assume that this line makes an angle of $\frac{2\pi}{7}$ with the top edge of the tape and fold according to the following rules:

(1) Each new crease line goes in the forward (left to right) direction along the strip of paper.
(2) Each new crease line bisects the angle between the last crease line and the edge of the tape from which it emanates.
(3) The bisection of angles at any vertex continues until a crease line produces an angle of the form $\frac{a'\pi}{b}$ where a' is an odd number; then the folding stops at that vertex and recommences at the intersection point of the last crease line with the other edge of the tape.

To prove that the smallest angle on the tape in Figure 2(a) does approach $\frac{\pi}{7}$, simply assume that the original angle indicated as measuring $\frac{\pi}{7}$ on the top left-hand side is $\frac{\pi}{7} + \epsilon$ (where ϵ is any number). Then calculate the measure of each new angle that appears on the tape between a fold line and the edge of the tape. For example, if we were to call the successive new

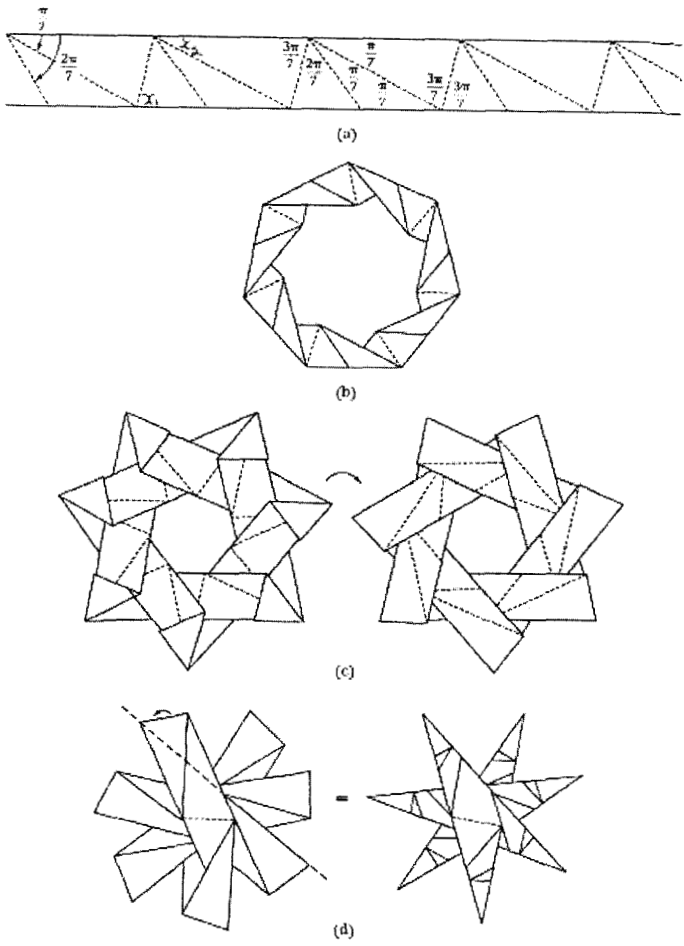


FIGURE 2

angles appearing next to the edge of the tape x_1, x_2, x_3, $\cdots$, then we see immediately, from elementary geometry, that x_1 (the angle formed by the first fold line sloping upwards and the bottom edge of the tape) satisfies

the equation

$$\frac{\pi}{7} + \epsilon + 2x_1 = \pi,$$

from which it follows that $x_1 = \frac{3\pi}{7} - \frac{\epsilon}{2}$. Then the next angle formed between a fold line and an edge of the tape (namely the second small angle along the top of the tape) satisfies the equation

$$\frac{3\pi}{7} - \frac{\epsilon}{2} + 4x_2 = \pi,$$

from which it follows that $x_2 = \frac{\pi}{7} + \frac{\epsilon}{2^3}$. From these two elementary calculations it is clear that, every time the tape is folded UP once and DOWN twice (we call the folded strip (2,1)-tape), the original angle will get closer still to $\frac{\pi}{7}$. This "error correction" method is an example of what we call our optimistic strategy in action. In fact, in this setting, it is always the case that if we assume we HAVE what we want (provided it is a rational multiple of π), and fold according to the 3 rules given above, we always GET what we want!). The proof of this, in general, is given in [2,3,6].

Figure 2(b) shows the FAT 7-gon produced by performing the FAT-algorithm on consecutive vertices along the top of the folded tape shown in Figure 2(a). But, as so often happens in mathematics, we get more than we expected. Figures 2(c), (d) show the regular $\{\frac{7}{2}\}$- and $\{\frac{7}{3}\}$-gons that are produced from this (2,1)-tape by executing the FAT-algorithm on the crease lines that make angles of $\frac{2\pi}{7}$ and $\frac{3\pi}{7}$, respectively, with an edge of the tape (we always orient the folded tape so that the angle we are using, for the FAT algorithm, is along the top edge of the tape). In Figures 2(c), 2(d) the FAT-algorithm was executed on every other suitable vertex along the edge of the tape so that, in (c), the resulting figure, or its flipped version, could be woven together in a more symmetric way and, in (d), the excess could be folded neatly around the points to obtain the figure on the right.

What we have described for $b = 7$ will work for any odd number. When b is an even number it is necessary to make secondary fold lines. The reader will probably see just where these secondary fold lines should be made in the tape of Figure 2(a) in order to produce tape from which it is possible to construct a regular FAT 14-gon. Similarly, more secondary fold lines may be made on this tape to produce the FAT 28-gon, 56-gon, and, in general, the $2^n 7$-gon, $n \geq 1$. Detailed instructions for doing this, and for constructing all star polygons, may be found, for example, in Section 4.5 of [2]. Thus we need only concern ourselves with finding folding procedures for odd numbers b.

We notice, from our example, that the folded tape produced some regular star polygons as well. We will call these regular star polygons $\left\{\frac{b}{a}\right\}$-gons[c], where $a \geq 2$; and we observe that it makes sense to require that $a < \frac{b}{2}$, since to take larger values of a would just result in star polygons where the top edge visits vertices in the opposite order around the bounding b-gon — and each of those star polygons is identical to one already produced. For example, a $\left\{\frac{7}{5}\right\}$-gon is the same as a $\left\{\frac{7}{2}\right\}$-gon.

Encouraged by the observation that having an angle of $\frac{3\pi}{7}$ on the tape at equally spaced intervals allows us to construct a regular star $\left\{\frac{7}{3}\right\}$-gon, we ambitiously try to construct, by similar means, a regular star $\left\{\frac{11}{3}\right\}$-gon. Once again we can obtain the instructions by using our optimistic strategy, which means that we assume that we can fold the desired angle of $\frac{3\pi}{11}$ at A_0 in Figure 3(a), and we adhere to the same three folding principles that we used in constructing the regular 7-gon.

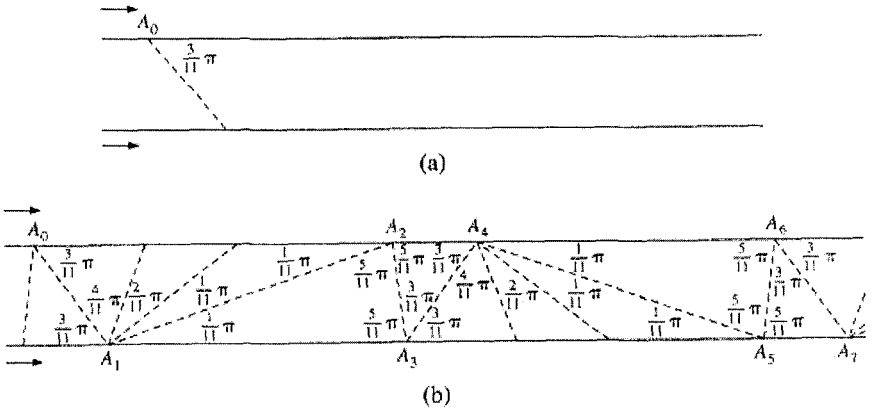


FIGURE 3

Once again the optimistic strategy works; and, using the three rules

[c] One may think of the $\left\{\frac{b}{a}\right\}$-gon as a polygon where the top edge of the tape visits every ath vertex of the bounding b-gon. For example we see in Figure 3(c) that for the $\left\{\frac{7}{2}\right\}$-gon the top edge visits every 2nd vertex of the bounding 7-gon.

above, we get the tape whose angles converge to those shown in Figure 3(b). We could denote this folding procedure by $D^1U^3D^1U^1D^3U^1$, interpreted in the obvious way on the tape — that is, the first exponent "1" refers to the one bisection (producing a crease in a downward direction) at the vertices A_{6n} (for $n = 1, 2, \cdots$) on the top of the tape; similarly the "3" refers to the 3 bisections at the bottom of the tape (producing creases in an upward direction) made at the vertices A_{6n+1}; etc. However, since the folding is duplicated halfway through, we can abbreviate the notation for the folding procedure and write simply $(1, 3, 1)$, with the understanding that we alternately fold from the top and the bottom of the tape as described, with the number of bisections at each vertex running, in order, through the values $1, 3, 1, \cdots$.

The convergence can be shown using an error-correction type of proof similar to that described earlier for the tape that produced the 7-gon. The reader should have no difficulty in supplying the details.

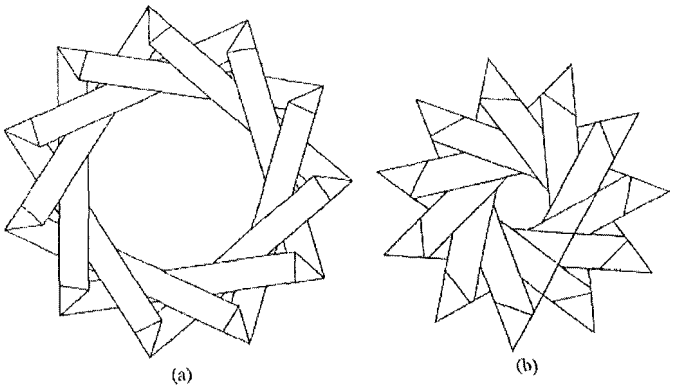


FIGURE 4

Now consider the tape in Figure 3(b). If the FAT-algorithm is performed on consecutive parallel creases of the same length, some $\{\frac{b}{a}\}$-gon will result. If, for example, we use the crease lines $A_{6n}A_{6n+1}$, $0 \leq n \leq 10$, (all of which make an angle of $\frac{3}{11}\pi$ with the top edge of the tape) we can get the $\{\frac{11}{3}\}$-gon shown in Figure 4(a). Likewise, if we perform the FAT-algorithm on

the shortest crease lines emanating downward from A_{6n+4} ($0 \leq n \leq 10$), we can get the $\{\frac{11}{4}\}$-gon shown in Figure 4(b).

In fact, it is possible to fold, from this tape, ALL of the possible regular star 11-gons;[d] namely, the $\{\frac{11}{1}\}$-, $\{\frac{11}{2}\}$-, $\{\frac{11}{3}\}$-, $\{\frac{11}{4}\}$-, and $\{\frac{11}{5}\}$-gon. Notice that there are fold lines of five different lengths on this tape.

Now, in preparation for the number theory, and to enable us to systematically determine the folding procedure for any given a and b, let us look at the patterns in the arithmetic of the computations for this last example where $a = 3$ and $b = 11$. Referring to Figure 3(b) we observe that

the smallest angle to the right of A_n where	is of the form $\frac{a}{11}\pi$	and the number of bisections at the next vertex
$n = 0$	$a = 3$	$= 3$
$n = 1$	$a = 1$	$= 1$
$n = 2$	$a = 5$	$= 1$
$n = 3$	$a = 3$	$= 3$
$n = 4$	$a = 1$	$= 1$
$n = 5$	$a = 5$	$= 1$

We could write this as a shorthand symbol in the following way:

$$(b =)11\left|\begin{matrix}(a_1 =)3 & (a_2 =)1 & (a_3 =)5\\ 3 & 1 & 1\end{matrix}\right| \qquad (1)$$

Notice that (1) only has three entries in the top and bottom row (instead of the six you might have expected from the layout above it); this is because we stop when the next a_{i+1} will be the same as the initial number a_1 in the top row. We then say that the symbol (1) is contracted.

Observe that if we had started with the assumption that our original angle was $\frac{\pi}{11}$ (say, at vertex A_4), we would have gotten identically the same folded tape, but then the symbol would have taken the form

$$(b =)11\left|\begin{matrix}(a_1 =)1 & (a_2 =)5 & (a_3 =)3\\ 1 & 1 & 3\end{matrix}\right|$$

In fact, it should be clear that we can start anywhere (with $a_1 = 1$, 3, or 5), and the resulting symbol (1′), analogous to (1), will be obtained by cyclic permutation of the matrix component of the symbol, placing our choice of a_1 in the first position along the top row.

---

[d] We will consider the regular 11-gon to be a special regular star polygon denoted, more elaborately, as a $\{\frac{11}{1}\}$-gon.

If we followed this procedure in an effort to construct a $\{\frac{33}{9}\}$-gon we would obtain the symbol

$$(b=)33\left|\begin{matrix}(a_1=)9 & (a_2=)3 & (a_3=)15\\ 3 & 1 & 1\end{matrix}\right|$$

Notice that this gives precisely the same folding procedure as that for an $\{\frac{11}{3}\}$-gon in the symbol shown in (1). This means that if you fold the $(3,1,1)$-tape and perform the FAT-algorithm on the tape on the crease lines making an angle of $\frac{9\pi}{33}$ with the top of the tape you will obtain an $\{\frac{11}{3}\}$-gon; because, of course $\frac{33}{9}=\frac{11}{3}$. This is, as we like to say, the most difficult method known to man for reducing fractions! We will avoid this difficulty by requiring that a_1 and b are mutually prime. We may specify this by restricting a_1 so that $\gcd(a_1,b)=1$ and we will then call our symbol reduced.

In general, if we wish to fold a $\{\frac{b}{a}\}$-gon, with b,a odd and mutually prime, $a<\frac{b}{2}$, then we may construct a symbol as follows. We write

$$b\left|\begin{matrix}a_1 & a_2 & \cdots & a_r\\ k_1 & k_2 & \cdots & k_r\end{matrix}\right| \tag{2}$$

where b, a_i are odd, $a_i<\frac{b}{2}$, $\gcd(a_1,b)=1$ (reduced) and

$$b-a_i=2^{k_i}a_{i+1},\ i=1,2,\cdots,r,\ a_{r+1}=a_1(\text{contracted}) \tag{3}$$

In fact, (3) implies that if $\gcd(a_1,b)=1$, then $\gcd(a_i,b)=1$ for $i=1,2,\cdots,r$.

Example 1.1. Suppose we wish to construct all possible star 31-gons. We start by finding instructions for folding a convex 31-gon; thus $b=31$, $a_1=1$ and we construct the symbol

$$(b=)31\left|\begin{matrix}(a_1=)1 & 15\\ 1 & 4\end{matrix}\right|$$

which tells us that by using the $(1,4)$-folding procedure we can produce tape that can be used to construct a FAT 31-gon. In fact, we get more; the $(1,4)$ tape can be used to construct FAT

$$\left\{\frac{31}{1}\right\}-,\ \left\{\frac{31}{2}\right\}-,\ \left\{\frac{31}{4}\right\}-,\ \left\{\frac{31}{8}\right\}-\text{ and }\left\{\frac{31}{15}\right\}-\text{gons.}$$

But this folding doesn't produce a $\{\frac{31}{3}\}$-gon. To obtain the $\{\frac{31}{3}\}$-gon we start with $b = 31$, $a = 3$ obtaining the symbol

$$(b =)31\left|\begin{matrix}(a_1 =)3 & 7\\ 2 & 3\end{matrix}\right|,$$

which tells us that by using the $(2,3)$-folding procedure, we produce $(2,3)$-tape from which we can fold the FAT $\{\frac{31}{3}\}$-gon. Again, we get more than we initially sought, the $(2,3)$-tape can be used to construct FAT

$$\left\{\frac{31}{3}\right\}-,\ \left\{\frac{31}{6}\right\}-,\ \left\{\frac{31}{12}\right\}-,\ \left\{\frac{31}{7}\right\}-,\ \text{and}\ \left\{\frac{31}{14}\right\}-\text{gons.}$$

However, we still don't have a procedure for folding a $\{\frac{31}{5}\}$-gon. So, in order to get the $\{\frac{31}{5}\}$-gon, we construct a 2-symbol with $b = 31$, $a_1 = 5$ obtaining

$$(b =)31\left|\begin{matrix}(a =)5 & 13 & 9 & 11\\ 1 & 1 & 1 & 2\end{matrix}\right|,$$

which tells us that, by using the $(1,1,1,2)$-folding procedure, we produce tape from which we can fold the FAT $\{\frac{31}{5}\}$-gon. Once again, we get bonuses. In fact, the $(1,1,1,2)$-tape can be used to construct FAT

$$\left\{\frac{31}{5}\right\}-,\ \left\{\frac{31}{10}\right\}-,\ \left\{\frac{31}{13}\right\}-,\ \left\{\frac{31}{9}\right\}-,\ \text{and}\ \left\{\frac{31}{11}\right\}-\text{gons.}$$

We can combine all the possible symbols for $b = 31$ into one complete symbol, which we will write $\Sigma(31)$, calling each part a coach, with the number of coaches being denoted by c. The complete symbol $\Sigma(31)$ then takes the form

$$31\left|\begin{matrix}1 & 15\\ 1 & 4\end{matrix}\right|\left|\begin{matrix}3 & 7\\ 2 & 3\end{matrix}\right|\left|\begin{matrix}5 & 13 & 9 & 11\\ 1 & 1 & 1 & 2\end{matrix}\right| \tag{4}$$

What do you notice about the value of the bottom row sum in each coach of (4)? It is the same, namely, 5. Notice, too, that the parity in the number of entries is the same in each coach (namely, in this case, even). Is this an accident? The interested reader should try writing out a few complete $\Sigma(b)$ symbols for odd numbers b of your choice and looking for patterns among the numbers involved. Our first theorem applies to any coach of a complete $\Sigma(b)$ symbol.

Theorem 1.1 (The Quasi-Order Theorem). *Suppose that* a, b *are odd, with* $a = a_1 < \frac{b}{2}$, *and the symbol*

$$b\left|\begin{matrix} a_1 & a_2 & \cdots & a_r \\ k_1 & k_2 & \cdots & k_r \end{matrix}\right| \tag{5}$$

is obtained using the calculation

$$b - a_i = 2^{k_i} a_{i+1} (\textit{with } k_i \textit{ maximal}). \tag{6}$$

Let $\sum_{i=1}^{r} k_i = k$, *and assume that* (5) *is not only contracted* ($a_{r+1} = a_1$), *but also reduced* ($\gcd(a_i, b) = 1$). *Then the quasi-order of* 2 *mod* b *is* k. *That is,* k *is the smallest positive integer such that*

$$2^k \equiv \pm 1 \mod b.$$

In fact, $2^k \equiv (-1)^r \mod b$. *(This means that* b *exactly divides* $2^k - (-1)^r$, *which may be written as* $b|2^k - (-1)^r$.)

The proof of Theorem 1.1 may be found in [3].

Notice that the symbol (1) may now easily be obtained by using (6), where $b = 11$, $a_1 = 3$. As a result we see that $r = 3$, and $k = 5$ and the quasi-order theorem then tells us that $11|2^5 - (-1)^3$ or $11|32 + 1 = 33$, and that 5 is the smallest positive exponent m such that $11|2^m \pm 1$. Again, the quasi-order theorem would tell us, for any coach of the symbol (4), where $b = 31$, $k = 5$, and $r = 2$ or 4, that 31 exactly divides $2^5 - 1$ (which can hardly be considered a surprise!)

The example of $\Sigma(31)$ also gives us a glimpse of what is to come. Notice that $\Phi(31) = 30$, where Φ is the Euler totient function, $c = 3$, and $k = 5$, so that, in this case $\Phi(b) = 2ck$. Here are a couple of other complete symbols for you to examine to see whether the relationships you have observed are just happy accidents, or whether you believe that they must always happen. (Of course, you may construct some for yourself as well, just to make sure we haven't chosen the only ones for which the relation $\Phi(b) = 2ck$ holds.) Thus

$$\Sigma(43): \quad 43\left|\begin{matrix} 1 & 21 & 11 & 3 & 5 & 19 & 7 & 9 & 17 & 13 & 15 \\ 1 & 1 & 5 & 3 & 1 & 3 & 2 & 1 & 1 & 1 & 2 \end{matrix}\right|,$$

$$\Phi(43) = 42,\ k = 7,\ c = 3.$$

$$\Sigma(51): \quad 51\left|\begin{matrix} 1 & 25 & 13 & 19 & 5 & 23 & 7 & 11 \\ 1 & 1 & 1 & 5 & 1 & 2 & 2 & 3 \end{matrix}\right|,$$

$$\Phi(51) = 32,\ k = 8,\ c = 2.$$

$$£(65): \quad 65\left|\begin{array}{c|cc|ccc|cccc} \mathbf{1} & 3 & 31 & 17 & 7 & 29 & 9 & 11 & 27 & 19 & 23 & 21 \\ \mathbf{6} & 1 & 1 & 4 & 1 & 2 & 3 & 1 & 1 & 1 & 1 & 2 \end{array}\right|,$$

$$\$(65) = 48, \; k = 6, \; c = 4.$$

Now we enunciate our theorem about £(6). Let $b > 1$ be an odd number, and let $(6) be the Euler totient function of *b.* Let us form £, the complete symbol of *b;* and let c be the number of coaches in £ *(b).* Finally let *k* be the quasi-order of 2 mod *b.* Then we have

Theorem 1.2 (The Coach Theorem). *(6) = 2c&.

The proof of this theorem appears in [10]. We omit the proof, because it may be obtained, with mild variations, by letting $t = 2$ in the next section.

A particularly interesting example occurs when $b = 641$. If we construct the first coach of £(641) we obtain

$$641\left|\begin{array}{ccccccccc} 1 & 5 & 159 & 241 & 25 & 77 & 141 & 125 & 129 \\ 7 & 2 & 1 & 4 & 3 & 2 & 2 & 2 & 9. \end{array}\right|.$$

We see that $k = 32$, and, since $(641)= 640, we can infer, from the Coach Theorem, that the number of coaches will be 10. However, an even more striking fact in this example is that the Quasi-Order Theorem tells us that since $k = 32$ and $r = 9$, $2^{32} \equiv (—I)^9 \mod 641$, or, equivalently that $641|2^{32} + 1$. Thus we have proved that the Fermat number $F^5 = 2^{2&} 4\text{-}1$ is not prime! We would like to think that Fermat and Euler would have been interested in this argument.

Having come this far it is natural to ask if Theorems 1.1 and 1.2 can be generalized. The answer is "yes" and we cany out the details in Sections 2 and 3.

2. The Generalized Quasi-order Theorem

We should point out that when we generalize the Quasi-order Theorem we are leaving the realm of paper-folding and entering that of pure number theory. But, of course, we are hoping that some of the things that were true in our original theorem will hold in general. We now think of our original case as having base 2, because we were bisecting the angles and it was the power of 2 that we were always factoring out of the difference $b — ai$ to get the next entry in our symbol. We now wish to get analogous results in a general base *t.*

So, how should we generalize the Quasi-Order Theorem? It is interesting, and not altogether surprising, that our main difficulty in generalizing

this theorem to a general base t lies not in proving the generalization but in stating it. For generalization is an art, not an algorithmic procedure, so judicious choices must be made in formulating the generalization. It is particularly striking that the appropriate generalization of the relation $a = a_1 < \frac{b}{2}$ is not, as we originally thought, $a = a_1 < \frac{b}{t}$, but rather the original $a = a_1 < \frac{b}{2}$! We are now ready to formalize the appropriate generalization.

Theorem 2.1 (The General Quasi-Order Theorem). *Suppose b, a are mutually prime, with b prime to the base t, where $t \nmid a$, and $a_1 = a < \frac{b}{2}$. Then construct the contracted ($a_{r+1} = a_1$), reduced ($\gcd(a_i, b) = 1$), t-symbol*

$$b \begin{vmatrix} a_1 & a_2 & \cdots & a_r \\ k_1 & k_2 & \cdots & k_r \\ \epsilon_1 & \epsilon_2 & \cdots & \epsilon_r \\ q_1 & q_2 & \cdots & q_r \end{vmatrix} \tag{7}$$

where

$$q_i b + (-1)^{\epsilon_i} a_i = t^{k_i} a_{i+1}, \quad i = 1, 2, \cdots, r \ \ (a_{r+1} = a_1). \tag{8}$$

Moreover, when $\epsilon_i = 0$, we use $q_i b + a$, $1 \le q_i \le \frac{t}{2} - 1$, and when $\epsilon_i = 1$, we use $q_i b - a$, $1 \le q_i \le \frac{t}{2}$, choosing the smallest q_i such that $q_i b + (-1)^{\epsilon_i} a_i$ has t as a factor. Then the quasi-order of $t \bmod b$ is $k = \sum k_i$. Indeed

$$t^k \equiv (-1)^E \bmod b,$$

where $E = \sum \epsilon_i$.

The proof of Theorem 2.1 appears in [3] where the bottom row of the symbol, involving the q_i, doesn't appear because it isn't essential to the proof. We include the bottom row here because we have discovered that it plays a role when one is "looking for patterns" among the coaches of the complete symbol. This brings up the question: Does The Coach Theorem hold for any base t? Before we try to answer this, let us look at some examples of $\Sigma_t(b)$, where $\Sigma_t(b)$ is the complete symbol of b for the base t.

Example 2.1. Construct $\Sigma_4(17)$ and state what it means in terms of the quasi-order of 4 mod 17.
Solution: For each coach we use the equation (beginning with the smallest available a_i)

$$q_i b + (-1)^{\epsilon_i} a_i = 4^{k_i} \cdot a_{i+1}$$

where $q_i = 1$, or 2, $\epsilon_i = 0$, or 1. A routine approach would be to calculate $b - a_i$, $b + a_i$,and $2b - a_i$ in that order, stopping at the first one that yields at least one factor of 4. Proceeding in this way, and only recording the successive calculations we obtain, for each coach, the calculations below, that are then recorded in the symbol (9).

$$\begin{aligned}
\text{First coach}: &\quad 17 - 1 = 4^2 \cdot 1 \\
\text{Second coach}: &\quad 2 \cdot 17 - 2 = 4^2 \cdot 2 \\
\text{Third coach}: &\quad 17 + 3 = 4 \cdot 5 \\
&\quad 17 - 5 = 4 \cdot 3 \\
\text{Fourth coach}: &\quad 2 \cdot 17 - 6 = 4 \cdot 7 \\
&\quad 17 + 7 = 4 \cdot 6
\end{aligned}$$

so that

$$\Sigma_4(17) = 17 \left| \begin{matrix} 1 & 2 \\ 2 & 2 \\ 1 & 1 \\ 1 & 2 \end{matrix} \right| \left| \begin{matrix} 3 & 5 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{matrix} \right| \left| \begin{matrix} 6 & 7 \\ 1 & 1 \\ 1 & 0 \\ 2 & 1 \end{matrix} \right|, k = 2, c = 4 \tag{9}$$

From any coach in (9) we see that the quasi-order of 4 mod 17 is 2. In fact,

$$4^2 \equiv (-1)^1 \mod 17.$$

Furthermore, $\Phi(17) = 16 = 2 \cdot 2 \cdot 4$, where Φ is the Euler totient function, so the Coach Theorem holds.

Example 2.2. Construct $\Sigma_5(67)$ and state what it means in terms of the quasi-order of 5 mod 67.
Solution: For each coach we use the equation (beginning with the smallest available a_i) $q_i b + (-1)^{\epsilon_i} a_i = 5^{k_i} \cdot a_{i+1}$, where $q_i = 1$ or 2, $\epsilon_i = 0$ or 1.
This time we calculate $b - a_i$, $b + a_i$, $2b - a_i$, and $2b + a_i$ in that order, stopping at the first one that yields at least one factor of 5. The calculations

for the first coach are

$$\begin{aligned} 2\cdot 67+1 &= 5^1\cdot 27\\ 67-27 &= 5^1\cdot 8\\ 67+8 &= 5^2\cdot 3\\ 67+3 &= 5^1\cdot 14\\ 2\cdot 67-14 &= 5^1\cdot 24\\ 2\cdot 67-24 &= 5^1\cdot 22\\ 67-22 &= 5^1\cdot 9\\ 2\cdot 67-9 &= 5^3\cdot 1 \end{aligned}$$

And we can then write down the first coach:

$$67\left|\begin{matrix} 1 & 27 & 8 & 3 & 14 & 24 & 22 & 9\\ 1 & 1 & 2 & 1 & 1 & 1 & 1 & 3\\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1\\ 2 & 1 & 1 & 1 & 2 & 2 & 1 & 2 \end{matrix}\right|_5$$

In a similar way we obtain, for the second and third coaches (always beginning the coach with the smallest a_i available):

$$67\left|\begin{matrix} 2 & 13 & 16 & 6 & 28 & 19 & 23 & 18 & 17 & 4 & 26 & 32 & 7 & 12 & 11 & 29 & 21 & 31 & 33\\ 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2\\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0\\ 1 & 1 & 2 & 2 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 \end{matrix}\right|_5$$

$k = 11$, $c = 3$, and E is always an odd number so that $(-1)^E = -1$. These three coaches comprise $\Sigma_5(67)$ (but there isn't space to write the complete symbol across the page here!).

From any coach of $\Sigma_5(67)$ we see that the quasi-order of 5 mod 67 is 11 and, in fact,

$$5^{11} \equiv -1 \mod 67.$$

Once again The Coach Theorem seems to hold; that is, $\Phi(67) = 66 = 2\cdot 11\cdot 3$.

3. The Generalized Coach Theorem

In this section we enunciate and prove the main theorem of this paper.

Theorem 3.1 (The Generalized Coach Theorem). $\Phi(b) = 2ck$, *where* Φ *is the Euler totient function ,* k *and* c *are as given in Section 2.*

Proof. Let $\mathbb{Z}_b^*$ be the multiplicative group of residues mod b prime to b, so that order $(\mathbb{Z}_b^*) = \Phi(b)$; and let T be the subgroup of $\mathbb{Z}_b^*$ generated by -1 and 2. We first observe that, with $t = 2$, where t is the base,

$$T = \left\{(-1)^i 2^j \mod b;\ 0 \le i \le 1,\ 0 \le j \le k-1\right\}, \tag{10}$$

with k being the quasi-order of 2 mod b. We now modify (10) to accommodate the replacement of base 2 by a general base t; we will then write T_t in place of T, so that $T = T_2$. Thus we start by modifying (10) to handle the case of base t. Then (10) is replaced by

$$T_t = \left\{(-1)^i t^j \mod b;\ 0 \le i \le 1,\ 0 \le j \le k-1\right\},$$

where b is prime to t, k is the quasi-order of t mod b, $t^k \equiv \pm 1 \bmod b$. Note that, independently of the choice of t,

$$|T_t| = 2k,$$

Thus a coach is a t-symbol analogous to the 2-symbol defined on p. 126 of [2]. We will construct the modified φ-symbol analogous to that introduced on p. 131 of [2], and show how the generalized ψ-symbol is related to the generalized φ-symbol. In fact, we will obtain a modified symbol from a given element of $\mathbb{Z}_b^*/T_t$; but it is then an automatic step — which we will describe — to obtain the symbol itself, that is, the coach.

The argument now proceeds in a way similar to the case $t = 2$, as described in [10]. First we show that each element of $\mathbb{Z}_b^*/T_t$ is represented by a number a which is (i) $\not\equiv 0$ mod t, (ii) prime to b, and (iii) less than $\frac{b}{2}$. We claim that it is obvious from the structure of T_t that we may represent an element of $\mathbb{Z}_b^*/T_t$ by a number a' which is (i) $\not\equiv 0$ mod t, (ii) prime to b, and (iii)$'$ less than b. The argument at this point very closely resembles that given in the case $t = 2$ in [10].

Thus let a' be such a number. If $a' < \frac{b}{2}$, there is nothing further to do. But if $a' > \frac{b}{2}$, then, for some q_i, ϵ_i, and t, $\frac{q_i b - t^{k_i} a'}{-(-1)^{\epsilon_i}} < \frac{b}{2}$ and represents the same element ξ of $\mathbb{Z}_b^*/T_t$. However, $b - a'$ is a multiple of t, so we may set $b - a' = t^\ell a$, with $a \not\equiv 0$ mod t, and $\ell \ge 1$. Again a represents the same element of $\mathbb{Z}_b^*/T_t$ as a', and a satisfies conditions (i), (ii), (iii).

We now apply the generalized reverse algorithm φ (that is, the reverse of ψ), given by

$$\frac{q_i b - t^{k_i} a_{i+1}}{-(-1)^{\epsilon_i}} = a_i \tag{11}$$

Thus, writing a_1 for a, we obtain a sequence of numbers $\not\equiv 0 \mod t$

$$a_1, a_2, \cdots, a_r, a_{r+1}, \tag{12}$$

all satisfying conditions (i), (ii), (iii), where

$$a_{r+1} = a_1. \tag{13}$$

Explicitly, the passage from a_1 to a_2 is achieved by repeatedly multiplying a_1 by t until we achieve $t^{\ell_1}a_1 > \frac{b}{2}$, with $\ell_1 \geq 1$ and then set $a_2 = \frac{q_1 b - t^{\ell_1}a_1}{-(-1)^{\epsilon_1}}$; and we continue to generate the entire sequence $a_1, a_2, \cdots, a_r$. Plainly $a_1, a_2, \cdots, a_r$ all represent the same element of $\mathbb{Z}_b^*/T_t$. Now let us insert the multiples of t, $ta_1, t^2a_1, \cdots, t^{\ell_1}a_1$ between a_1 and a_2, and proceed similarly between a_2 and $a_3, \cdots, a_{r-1}$ and a_r, a_r and a_1. The result is now precisely the effect of the generalized φ-algorithm on the modified symbol. The generalized φ-algorithm is, of course, inverse to the generalized ψ-algorithm used in the construction of the symbol. Let us now pause to sum up the steps so far. Starting with a_1 satisfying (i), (ii), and (iii), and representing a given element ξ of $\mathbb{Z}_b^*/T_t$, we then apply the generalized (reverse) φ-algorithm (11) to $a = a_1$, and iterate the applications, obtaining (12), namely,

$$a_1, a_2, \cdots, a_r, a_{r+1},$$

all satisfying conditions (i), (ii), (iii) and with $a_{r+1} = a_1$. Thus, starting with a_1 satisfying conditions (i), (ii), (iii) and representing a given element ξ of $\mathbb{Z}_b^*/T$, we apply to a_1 the reverse φ-algorithm, that is, we construct the sequence

$$a_1, ta_1, \cdots, t^{\ell_1}a_1,$$

such that ℓ_1 is minimal for the property $t^{\ell_1}a_1 > \frac{b}{2}$, (so that $\ell_1 \geq 1$) and set

$$a_2 = \frac{q_1 b - t^{\ell_1}a_1}{-(-1)^{\epsilon_1}}.$$

We begin again with a_2 and again apply the generalized φ-algorithm to obtain a_3. We continue in this way until we reach $\cdots, a_r, a_{r+1}$, with $a_{r+1} = a_1$. This must occur eventually since φ is a permutation, indeed, the permutation inverse to the ψ-algorithm (see p. 116 of [3]), which was used to construct the symbol in the first place.

Indeed, it is easy to see how the symbol may be derived from the element of $\mathbb{Z}_b^*/T_t$ represented by the number a_1 satisfying conditions (i), (ii), (iii). We write down the sequence arising from the process of multiplying by t,

together with (13), to pass from a_1 to a_2, and then proceed as described above, obtaining

$$a_1, ta_1, \cdots, t^{\ell_1}a_1, a_2, ta_2, \cdots, t^{\ell_2}a_2, a_3, \cdots, a_r, ta_r, \cdots, t^{\ell_r}a_r, a_1 \quad (14)$$

It is then not difficult to see that we obtain from (14) a coach as follows. We write down the terms $\neq 0 \mod t$ in (14) as the modified top line of the modified coach, thus

$$a_1, a_2, \cdots, a_r, a_1; \quad (15)$$

and the modified second line of the coach simply lists the number of t-multiples between successive entries (15) in the sequence (14). Thus the second line is

$$\ell_1, \ell_2, \cdots, \ell_r. \quad (16)$$

The result is a modified symbol, from which the true symbol (or coach) is obtained by omitting the repeated a_1 from the start of (15) and then writing each of (15), (16) backwards.

We demonstrate the procedure with the particular, but not special, case $t = 5$, using the first coach of $\Sigma_5(67)$. The reverse symbol, incorporating the facts of (8), using (15), may be written as

$$67\begin{pmatrix} 1 & 9 & 22 & 24 & 14 & 3 & 8 & 27 & 1 \\ & 3 & 1 & 1 & 1 & 1 & 2 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ & 2 & 1 & 2 & 2 & 1 & 1 & 1 & 2 \end{pmatrix}_5. \quad (17)$$

Thus if we start with the element of $\mathbb{Z}_{67}^*/T_5$ represented by the number 1, the sequence (14) is given by the following two lines (to read as a single sequence)

$$\begin{array}{ccccccccccccccccc} & & & & \frac{2\cdot 67-5^3\cdot 1}{-(-1)^1} & & & & \frac{1\cdot 67-5^1\cdot 9}{-(-1)^1} & & & & \frac{2\cdot 67-5^1\cdot 22}{-(-1)^1} & & \frac{2\cdot 67-5^1\cdot 24}{-(-1)^1} \\ & & & & \downarrow & & & & \downarrow & & & & \downarrow & & \downarrow \\ \underline{1}, & 5, & 25, & 125, & \underline{9}, & & 45, & & \underline{22}, & & 110, & & \underline{24}, & 120, & \underline{14}, \\ & & & \frac{1\cdot 67-5^1\cdot 14}{-(-1)^0} & & & & & \frac{1\cdot 67-5^2\cdot 3}{-(-1)^0} & & & \frac{1\cdot 67-5^1\cdot 8}{-(-1)^1} & & \frac{2\cdot 67-5^1\cdot 27}{-(-1)^0} & \\ & & & \downarrow & & & & & \downarrow & & & \downarrow & & \downarrow & \\ & & 70, & \underline{3}, & & 15, & 75, & & \underline{8} & & 40, & \underline{27}, & & \underline{1} & \end{array}$$

From this sequence we can construct the modified symbol (17). Then, by writing (17) in reverse, and omitting the repeated "1" we can re-

cover the first coach of $\Sigma_5(67)$. Of course, each of the numbers 1, 27, 8, 3, 14, 24, 22, 9 represents the same element ξ of $\mathbb{Z}_{67}^*/T_5$.

It is now plain that, in general, the process thus far described sets up a one-one correspondence between the set of coaches and the elements of $\mathbb{Z}_b^*/T_t$.

We are now in a position to complete the proof of Theorem 3.1. We simply have to count the elements in $\mathbb{Z}_b^*/T_t$. But, of course,

$$|\mathbb{Z}_b^*| = \Phi(b)$$

and

$$|T_t| = 2k,$$

as follows easily from (10). Thus

$$c = |\mathbb{Z}_b^*| \Big/ |T_t| = \Phi(b) \Big/ 2k,$$

so that

$$\Phi(b) = 2ck. \tag{18}$$

□

4. Some corollaries

Of course, Theorem 3.1 implies that $\Phi(b)$ is even. However, the standard algorithm for calculating $\Phi(b)$ shows that $\Phi(b)$ is even if b has an odd prime factor, but in general all we have assumed is that b is relatively prime to the base t and $b > 1$. However, it is striking that

$$k \Big| \frac{1}{2}\Phi(b). \tag{19}$$

This follows easily if b is prime, but is not so obvious if b is composite. However, (19) is a trivial consequence of (18).

Let us say that b has a cyclic coach in base t if $c = 1$; that is, if we can obtain only one coach from b. We then have

Theorem 4.1. *b has a cyclic coach in base t if, and only if, $\Phi(b) = 2k$, where k is the quasi-order of t mod b.*

An example is given by $\Sigma_2(21)$. In this case $b = 21$, and $\Phi(21) = 12$, $k = 6$. Other examples are

- $\Sigma_3(70)$, where $\Phi(70) = 24$ and $k = 12$;
- $\Sigma_4(27)$, where $\Phi(27) = 18$ and $k = 9$;
- $\Sigma_{10}(43)$, where $\Phi(43) = 42$ and $k = 21$.

5. Some open questions

We leave the reader with some open questions.

(1) How can you tell if a complete symbol $\Sigma_t(b)$ will have only one coach? How can you predict how many coaches a symbol will have?
(2) How can you tell what the value of r will be for a given coach? Is it the case that the smallest r always occurs in the first coach (where $a_1 = 1$)?
(3) Given b, how are the values of r for different coaches in the complete symbol related? Can you tell in which coach the largest r will occur?
(4) Given b, can you tell which collection of numbers will come together as the top row of a coach?
(5) For a fixed k, can the sequence of numbers $k_1, k_2, \cdots k_r$ appear in the bottom row of two distinct coaches? (The answer is no: see pages 135-6 of [2]) Are there simple a priori necessary and/or sufficient conditions, beyond $\sum k_i = k$, that can tell you whether such a sequence is a bottom row of a coach, short of running the entire computation?
(6) Given the last 3 rows of a generalized Σ_t symbol, can you determine the values of b and a_1?

6. Acknowledgment

The authors would like to thank Peter Ross for his careful scrutiny of this and earlier versions of our manuscript. We also want to thank Nicholas Tran for his help with the technical details during the preparation of this paper for publication.

7. Note on references

References [1] through [9] have illustrations showing how we were able to produce respectable polygons, both convex and star by paper-folding. References [3,8,9] show how to construct Platonic solids that are braided from the straight strips of folded paper, and how to construct other polyhedra including non-convex ones that collapse in interesting ways. Reference [10] concerns complete symbols, and the related theorems, in the case $t = 2$. Reference [11] uses our techniques in a variety of other geometrical constructions involving circles and angle dissections.

References

1. Gardner and Martin, *Martin Gardner's Mathematical Games: The Entire Collection of his Scientific American Columns*, CD-ROM, Mathematical As-

sociation of America, 2005.
2. Hilton Peter, Derek Holton, and Jean Pedersen, *Mathematical Reflections — In a Room with Many Mirrors*, 2nd printing, Springer-Verlag NY, 1998.
3. Hilton Peter, Derek Holton, and Jean Pedersen, *Mathematical Vistas — From a Room with Many Windows*, Springer-Verlag NY, 2002.
4. Hilton Peter and Jean Pedersen, Approximating any regular polygon by folding paper, *Mathematics Magazine* 56 (1983), no. 3, 141-155.
5. Hilton Peter and Jean Pedersen, Folding regular star polygons and number theory, *Mathematical Intelligencer* 7 (1985), no. 1, 15 – 26.
6. Hilton Peter and Jean Pedersen, Geometry in Practice and Numbers in Theory, *Monographs in Undergraduate Mathematics* 16 (1987), 37pp. (Available from Department of Mathematics, Guilford College, Greensboro, North Carolina 27410 U.S.A.)
7. Hilton Peter and Jean Pedersen, Geometry: A gateway to understanding, *The College Mathematics Journal* 24 (1993), no. 4, 298-317.
8. Hilton, Peter and Jean Pedersen, Symmetry in theory—mathematics and aesthetics. Symmetry Cult. Sci. 8 (1997), no. 3-4, 239-263. Electronic version (1999) available at http://members.tripod.com/vismath/hil/ped1.htm (This article discusses 2-period folding procedures and how to make the resulting polygons from the folded tape. You will, however, have to navigate around advertisements.)
9. Hilton Peter and Jean Pedersen, *Build Your Own Polyhedra*, Addison Wesley, 1998.
10. Hilton Peter, Jean Pedersen and Byron Walden, A property of complete symbols: An ongoing saga connecting geometry and number theory, *Homage to a Pied Puzzler*, A K Peters (to appear).
11. Polster, Burkard, Variations on a theme in paper-folding. *Amer. Monthly* 111 (2004), no. 2, 39-47.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 179–187)

KOSZUL ALGEBRAS AND HYPERPLANE ARRANGEMENTS

MICHEL JAMBU

Laboratoire J.A. Dieudonné,
Université de Nice Sophia-Antipolis, UMR 6621, France
E-mail: jambu@unice.fr

This is a survey to apply theory from noncommutative graded algebras to questions about the holonomy algebra and the Orlik-Solomon algebra of a hyperplane arrangement. We first recall the main properties of Koszul algebras and hyperplane arrangements. Then, we focus our interest on the class of hypersolvable arrangements which includes both the fiber-type and the generic arrangements. For these hypersolvable arrangements, the holonomy algebra is Koszul and koszulness of the Orlik-Solomon algebra characterizes the subclass of fiber-type's.

1. Koszul Duality and Koszul Algebras (overall)

Let k be an arbitrary field and let V be an n-dimensional k-vector space ($V \cong k^n$) and let $T(V) = \bigoplus_{n\geq 0} T_n$ be the k-tensor algebra over V where $T_0 \cong k$, $T_1 \cong V$. Then $T(V) \cong k < x_1, \ldots, x_n >$, the free associative k-algebra. Let A be a k-graded algebra, $A = \sum_{p\geq 0} A_p$. Assume that A is connected, i.e. $A_0 = k$ and is generated by A_1. A is "naturally" represented as the factor of the tensor algebra $T(A_1)$ by a homogeneous ideal $I = \sum_{p\geq 2} I_p$.

$$A \cong T(A_1)/I$$

Definition 1.1. A is said to be quadratic if I is generated by $I_2 \subset A_1 \bigotimes A_1$.

Therefore, a quadratic algebra A is determined by a vector space of generators $V = A_1$ and a subspace of quadratic relations $I_2 \subset V \bigotimes V$. Such a *quadratic* algebra is denoted as $A = \{V; I\}$.

Definition 1.2. Let $A = \{V; I\}$ be a quadratic algebra. The quadratic dual or Koszul dual algebra of A is defined by $A^! = \{V^*; I^\perp\}$, where V^* is the dual of V, $I^\perp \subset V^* \bigotimes V^*$ is the orthogonal complement to I with respect to the natural pairing: $\langle v \otimes v', v^* \otimes v'^* \rangle = \langle v, v^* \rangle \langle v', v'^* \rangle$ between $V \bigotimes V$ and $V^* \bigotimes V^*$.

Remark 1.1. $(A^!)^! = A$.

Example 1.1.

$$\begin{aligned} A = &\ k[x_1, \ldots, x_n] \quad \text{(commutative polynomial algebra)} \\ = &\ k\langle x_1, \ldots, x_n \rangle / \langle x_i x_j - x_j x_i \rangle \text{ for } i < j. \end{aligned}$$

Then $A \cong \mathcal{S}_n$ which is the symmetric algebra. A is a quadratic algebra.

$$\begin{aligned} A^! = &\ k\langle y_1, \ldots, y_n \rangle / \langle y_k^2;\ y_i y_j - y_j y_i \rangle \text{ (for i < j)} \\ = &\ \bigwedge(y_1, \ldots, y_n) \quad \text{(exterior algebra)}. \end{aligned}$$

Definition 1.3. Let A be a quadratric algebra,and ${}_Ak$ be the trivial graded left A-module A/A^+ where A^+ is the augmentation ideal $\bigoplus_{p>0} A_p$. A is said to be Koszul if ${}_Ak$ admits a free graded resolution:

$$\cdots \to P^i \to P^{i-1} \to \cdots \to P^1 \to P^0 \to {}_Ak \to 0,$$

where P^i is generated by its components of degree i.

Let denote the following objects:
$E(A) := \mathrm{Ext}_A^*({}_Ak, {}_Ak)$ the graded cohomology algebra of the trivial graded A-module ${}_Ak$.

The *Hilbert series* $H(A,t) := \sum_{n \geq 0} \dim(A_n) t^n$.

The *Koszul complex* of A:

$$\cdots \to K_i \xrightarrow{d_i} K_{i-1} \to \cdots \to K_1 \to K_0 \to {}_Ak \to 0,$$

where K_i free A-modules, $K_i = \mathrm{Hom}_k(A_i^!, A)$ and d_i is defined as $d_i f(a) = \sum_{k=1}^{n} f(x_i a) e_i$, $a \in A_{i-1}^!$, $(x_1, \ldots, x_n)$ is the basis of $A_1^!$, dual basis of the basis $(e_1, \ldots, e_n)$ of A_1.

Theorem 1.1. *Let A be a quadratic algebra. Then the following assertions are equivalent:*

(1) A is Koszul;

(2) $A^!$ *is Koszul;*
(3) $E(A) = A^!$*;*
(4) The Koszul complex of A *is acyclic;*
(5) $H(A,t).H(E(A),-t) = 1$.

Corollary 1.1. *A is Koszul iff* $H(A,t).H(A^!,-t) = 1$.

Example 1.2. $A = k[x_1,\ldots,x_n]$ is Koszul and $H(A,t) = \frac{1}{(1-t)^n}$, $A^! = \bigwedge(y_1,\ldots,y_n)$ is Koszul and $H(A^!,t) = (1+t)^n$.

2. Hyperplane Arrangements

We refer the reader to [6] as a general reference on arrangements.
Let $\mathcal{A}$ be an arrangement of hyperplanes over $\mathbf{C}$ i.e. $\mathcal{A} = \{H_1,\ldots,H_n\}$, where H_i are linear hyperplanes of $\mathbf{C}^l$. Define the complement $M(\mathcal{A}) = \mathbf{C}^l - \bigcup_{i=1}^{n} H_i$ and $L(\mathcal{A})$ the geometric lattice intersection of hyperplanes with reverse order

$$X \leq Y \text{ if } Y \subseteq X.$$

Notice that $\mathrm{rk}(X) = \mathrm{codim}(X)$.

Orlik-Solomon algebra (combinatorially defined)

$$A_k^*(\mathcal{A}) := \bigwedge(e_1,\ldots,e_n)/\mathcal{J} \cong H^*(M(\mathcal{A});k),$$

where $\mathcal{J}$ ideal generated by the relations of the form:

$$\sum_{j=1}^{s}(-1)^{i-1}e_{i_1}\ldots\widehat{e}_{i_j}\ldots e_{i_s}$$

for all $1 \leq i_1 < \ldots < i_s \leq n$ such that $\mathrm{rk}(H_{i_1} \cap \cdots \cap H_{i_s}) < s$.
$A_k^*(\mathcal{A})$ is not necessary quadratic.

Poincaré polynomial

$$\begin{aligned} P(\mathcal{A},t) &:= \sum_i \dim A_k^i(\mathcal{A})t^i \\ &= \sum_{X\in L(\mathcal{A})} \mu(X)(-t)^{\mathrm{codim}(X)} \\ &= P(A_k^*(\mathcal{A}),t) \end{aligned}$$

Quadratic Orlik-Solomon algebra

$$\overline{A_k^*}(\mathcal{A}) := \bigwedge(e_1,\ldots,e_n)/\overline{\mathcal{J}}$$

where $\overline{\mathcal{J}}$ ideal generated by:

$$e_i e_j - e_i e_k + e_j e_k,\ 1 \leq i < j < k \leq n,$$

and

$$\mathrm{rk}(H_i \cap H_j \cap H_k) = 2.$$

$\overline{A_k^*}(\mathcal{A})$ only depends on $L_2(\mathcal{A})$, the elements of codimension 2 of $L(\mathcal{A})$.

Quadratic Poincaré polynomial

$$\overline{P}(\mathcal{A}, t) := P(\overline{A_k^*}(\mathcal{A}), t)$$

Example 2.1. Braid arrangements in $\mathbf{C}^l$.

$\mathcal{A}_l = \{H_{ij} \mid 1 \leq i < j \leq l\}$, where $H_{ij} = \ker(z_i - z_j)$.

Notice that the fundamental group of the complement is isomorphic to the Pure braid group P_l.

Moreover, $A_k^*(\mathcal{A}_l) = \overline{A_k^*}(\mathcal{A}_l)$, the Orlik-Solomon algebra of a braid arrangement is quadratic.

Remark 2.1. There is a linear fibration given by forgetting the last coordinates:

$$\mathbf{C} - \{(l-1) \text{ points}\} \hookrightarrow M(\mathcal{A}_l) \longrightarrow M(\mathcal{A}_{l-1})$$

where $M(\mathcal{A}_l)$ is the complement of the braid arrangement in $\mathbf{C}^l$.

$$\begin{array}{ccc} \mathbf{C}^l & \longrightarrow & \mathbf{C}^{l-1} \\ \cup & & \cup \\ M(\mathcal{A}_l) & \longrightarrow & M(\mathcal{A}_{l-1}) \end{array}$$

Remark 2.2. The Coxeter arrangements $\mathcal{D}_l$, $l \geq 4$ in $\mathbf{C}^l$ are defined by $\{(x_i - x_j), (x_i + x_j),\ 1 \leq i < j \leq n\}$. Then $A_k^*(\mathcal{D}_l) \neq \overline{A_k^*}(\mathcal{D}_l)$, and the Orlik-Solomon algebra is not quadratic.

As a "natural" generalization of braid arrangements, we define the **fiber-type** arrangements.

Definition 2.1 (Falk,Randell). *[1] $\mathcal{A} = \{0\}$ is a fiber-type arrangement in $\mathbf{C}$. The arrangement $\mathcal{A}$ in $\mathbf{C}^l$ is fiber-type if it is strictly linearly fibered with base $M(\mathcal{B})$ the complement of the fiber-type arrangement $\mathcal{B}$ in $\mathbf{C}^{l-1}$.*

Then $\mathcal{A}$ is a fiber-type arrangement iff there is a composition series

$$\mathcal{A}_1 \subset \ldots \subset \mathcal{A}_i \subset \mathcal{A}_{i+1} \subset \ldots \subset \mathcal{A}_l = \mathcal{A}$$

where $\mathrm{rk}\mathcal{A}_1 = 1$ and $(\mathcal{A}_{i+1}, \mathcal{A}_i)$ defines a linear fibration

$$\mathbf{C} - \{|\mathcal{A}_{i+1} - \mathcal{A}_i| \text{ points}\} \hookrightarrow M(\mathcal{A}_{i+1}) \longrightarrow M(\mathcal{A}_i).$$

$\mathcal{A}$ is fiber-type iff the lattice $L(\mathcal{A})$ is *supersolvable.*

Proposition 2.1. *Let $\mathcal{A}$ be a fiber-type arrangement. Then*

$$A_k^*(\mathcal{A}) = \overline{A_k^*}(\mathcal{A}).$$

2.1. *Holonomy Lie Algebra*

Let $\mathrm{Lib}_k(\mathcal{A})$ be the k-graded free Lie algebra over $\{x_1, \ldots, x_n\}$.

Definition 2.2 (Kohno). *[4] The **holonomy Lie** algebra of $\mathcal{A}$ is denoted $\mathcal{G}_k(\mathcal{A})$:*

$$\mathcal{G}_k(\mathcal{A}) := \mathrm{Lib}_k(\mathcal{A})/\mathcal{N},$$

where $\mathcal{N}$ ideal generated by $\left[x_{i_k}, \sum_{j=1}^{s} x_{i_j}\right]$ for $k = 1, \cdots, s$, $1 \leq i_1 < \cdots < i_s \leq n$ such that rk $\bigcap_{j=1}^{s} H_{i_j} = 2$ and it is maximal with this property.

Remark 2.3. Let $H_*\big(M(\mathcal{A})\big)$ be the homology coalgebra with coefficients in k and comultiplication dual to the cup product,

$$\mathcal{G}_k(\mathcal{A}) \cong \mathrm{Lib}_k\big(H_1(M(\mathcal{A}))/\ker\Big(H_2(M(\mathcal{A})) \to \wedge^2 H_1\big(M(\mathcal{A})\big)\Big)\Big).$$

Definition 2.3. The **holonomy algebra** of $\mathcal{A}$ denoted $\mathcal{U}_k(\mathcal{A})$ is the universal enveloping algebra of $\mathcal{G}_k(\mathcal{A})$.

A holonomy algebra is a quadratic algebra.

Lemma 2.1. $\mathcal{U}_k(\mathcal{A}) \cong \big(\overline{A_k^*}(\mathcal{A})\big)^!$.

In the following, we will study a "large" class of arrangements for which $\mathcal{U}_k(\mathcal{A})$ is Koszul. However, let us give an example showing that this result is not always true.

Example 2.2. Let $\mathcal{A}$ be the arrangement defined by the linear forms x, y, z, $x+y, x+z, y+z$. Then $P(\mathcal{A}, t) = 1 + 6t + 12t^2 + 7t^3$ and $\overline{P}(\mathcal{A}, t) = 1 + 6t + 12t^2 + 8t^3 + t^4$. The holonomy algebra $\mathcal{U}_k(\mathcal{A})$ is not Koszul because $(1 - 6t + 12t^2 - 8t^3 + t^4)^{-1}$ has some negative coefficients (eg t^{13}).

Remark 2.4. Let $\big(\Gamma_n\pi_1(M(\mathcal{A}))\big)_{n\geq 1}$ be the Lower Central Series of the fundamental group, defined as follows:

(1) $\Gamma_1\pi_1(M(\mathcal{A})) = \pi_1(M(\mathcal{A}))$,
(2) $\Gamma_{n+1}\pi_1(M(\mathcal{A})) = [\pi_1(M(\mathcal{A})), \Gamma_n\pi_1(M(\mathcal{A}))]$,

$$\mathrm{gr}^*_\Gamma\big(\pi_1(M(\mathcal{A}))\big) := \bigoplus_{i\geq 1} \mathrm{gr}^i_\Gamma\big(\pi_1(M(\mathcal{A}))\big),$$

where $\mathrm{gr}^i_\Gamma\big(\pi_1(M(\mathcal{A}))\big) = \Gamma_{i+1}\pi_1(M(\mathcal{A}))/\Gamma_i\pi_1(M(\mathcal{A}))$.

Then as graded Lie algebras

$$\mathcal{G}^*_{\mathbf{Q}}(\mathcal{A}) \cong \mathrm{gr}^*_\Gamma(\pi_1(X)) \bigotimes \mathbf{Q}.$$

2.2. *Hypersolvable Arrangements*

This is a "large" class of arrangements containing both, the fiber-type ones (whose the complement is a K$[\pi, 1]$-space), and the generic ones (whose the complement is never a K$[\pi, 1]$-space).

Let $\mathcal{B}$ be a subarrangement of $\mathcal{A}$, denote $\overline{\mathcal{B}} = \mathcal{A} - \mathcal{B}$. In the following definition, we denote $\mathrm{rk}(\alpha, \beta)$ as $\mathrm{rk}(H_\alpha \cap H_\beta)$, where $H_\alpha = \ker\alpha$ and $H_\beta = \ker\beta$.

Definition 2.4. [2] $(\mathcal{A}, \mathcal{B})$ is said to be a **solvable extension** if

(1) For any $H_\alpha, H_\beta \in \mathcal{B}$ and any $H_a \in \overline{\mathcal{B}}$, then $\mathrm{rk}(\alpha, \beta, a) = 3$.
(2) Given $H_a, H_b \in \mathcal{B}$, $a \neq b$, there exists $H_\gamma \in \mathcal{B}$ such that $\mathrm{rk}(a, b, \gamma) = 2$. Denote $\gamma = f(a, b)$.
(3) Given distinct elements $H_a, H_b, H_c \in \overline{\mathcal{B}}$, then
$$\mathrm{rk}\big(f(a, b), f(b, c), f(c, a)\big) = 2.$$

Then we can distinguish 2 cases:

(1) There is a fibration:

$$\mathbf{C} - \{|\mathcal{A} - \mathcal{B}| \text{ points}\} \hookrightarrow M(\mathcal{A}) \longrightarrow M(\mathcal{B})$$

called the *fibered case.*
(2) $\mathrm{rk}\mathcal{B} = \mathrm{rk}\mathcal{A}$, called the *singular case.*

In case 2, there exists a *deformation* such that we eliminate singular case in order to get a fibration as in case 1.

Definition 2.5. [2,3] $\mathcal{A}$ is said to be **hypersolvable** if there is a composition series

$$\mathcal{A}_1 \subset \ldots \subset \mathcal{A}_i \subset \mathcal{A}_{i+1} \subset \ldots \subset \mathcal{A}_{l(\mathcal{A})} = \mathcal{A}$$

with $\mathrm{rk}(\mathcal{A}_1) = 1$ and $(\mathcal{A}_{i+1}, \mathcal{A}_i)$ is a solvable extension.

Then, after "enough" deformations, we can eliminate all the fibered cases and we get $\widetilde{\mathcal{A}}$ which is a fiber-type arrangement with the same lattice up to rank 2.

Remark that both fiber-type and generic arrangements are hypersolvable. The arrangement defined (in the Example 2.2) by the defining equation:

$xyz(x+y)(x+z)(y+z) = 0$ is not hypersolvable.

The arrangement defined by the following equation:

$(x+y)(x-y)(z+2y)(z+y)(z-y)(z-2y)z = 0$ is hypersolvable but neither fiber-type nor generic.

The Orlik-Solomon algebra of a hypersolvable arrangement is not necessary quadratic.

Theorem 2.1 (Jambu, Papadima). *[2] Let $\mathcal{A}$ be a hypersolvable arrangement; then $\overline{A_k^*}(\mathcal{A})$ is Koszul, for any field k; therefore $\mathcal{U}_k(\mathcal{A})$ is Koszul.*

Corollary 2.1 (Jambu, Papadima). *[2] Let $\mathcal{A}$ be a fiber-type arrangement; then $A_k^*(\mathcal{A})$ is Koszul.*

Sketch of the (algebraic) proof of the theorem:

(1) $H^*(\bigvee_m S^1; k) \cong T(V)$ is Koszul where $V \cong k^m$.

(2) Suppose $(\mathcal{A}, \mathcal{B})$ solvable, then as $A_k^*(\mathcal{B})$-modules

$$\overline{A_k^*}(\mathcal{A}) \cong \overline{A_k^*}(\mathcal{B}) \otimes H^*(\bigvee_{|\overline{\mathcal{B}}|} S^1; k).$$

(3) Recall that a graded subalgebra B^* of A^* is *normal* if $AB^+ = B^+A$. Then there is a canonical graded algebra projection $\pi : A \to F = A/AB^+$ $(B \hookrightarrow A \to F)$.

(4) **Lemma :** *Suppose B normal subalgebra of A such that A is free as a right B-module and all algebras are quadratic. If B and F are Koszul, then A is Koszul.*

(5) Let point out that $\overline{A_k^*}(\mathcal{B})$ is normal in $\overline{A_k^*}(\mathcal{A})$ which is a free right $\overline{A_k^*}(\mathcal{B})$-module, the quotient $F \cong H^*(\bigvee_{|\overline{\mathcal{B}}|} S^1; k)$ is Koszul. Then $\overline{A_k^*}(\mathcal{A})$ is Koszul if $\overline{A_k^*}(\mathcal{B})$ is Koszul.

Another proof is given using Shelton and Yusvinsky's result [8] saying that the Orlik-Solomon algebra of a fiber-type arrangement is Koszul altogether with the deformations from $\mathcal{A}$ to a fiber-type arrangement $\widetilde{\mathcal{A}}$.

Theorem 2.2 (Jambu, Papadima). *[2](Generalized LCS Formula) Let $\mathcal{A}$ be a hypersolvable arrangement; then for any field k:*

$$\overline{P}(\mathcal{A}, -t) = \prod_{i=1}^{\infty} (1 - t^i)^{\dim \mathcal{G}_k^i(\mathcal{A})}.$$

Proof. Recall that for all arrangements $\mathcal{U}_k(\mathcal{A}) \cong \big(\overline{A_k^*}(\mathcal{A})\big)^!$.

Compute the inverse of the Hilbert series of $\mathcal{U}_k(\mathcal{A})$ by the well-known Poincaré-Birkhoff-Witt theorem:

$$H(\mathcal{U}_k(\mathcal{A}), t)^{-1} = \prod_{i=1}^{\infty} (1 - t^i)^{\dim \mathcal{G}_k^i(\mathcal{A})}.$$

$\overline{A_k^*}(\mathcal{A})$ is Koszul, then $H(\overline{A_k^*}(\mathcal{A}), t).H\big(\big(\overline{A_k^*}(\mathcal{A})\big)^!, -t\big) = +1$. □

Corollary 2.2. *(LCS Formula) Let $\mathcal{A}$ be a fiber-type arrangement. Then*

$$P(\mathcal{A}, -t) = \prod_{i=1}^{\infty} (1 - t^i)^{\dim \mathcal{G}_k^i(\mathcal{A})}.$$

Remark 2.5. Kohno obtained this result for braid arrangements.

Theorem 2.3 (Jambu, Papadima). *[2] Let $\mathcal{A}$ be a hypersolvable arrangement. Then*

$$\mathcal{G}_{\mathbf{Z}}^*(\mathcal{A}) \cong \mathrm{gr}_{\Gamma}^*\Big(\pi_1\big(M(\mathcal{A})\big)\Big)$$

as graded Lie algebras.

The main point of the proof is to show that $\mathcal{G}_{\mathbf{Z}}^*(\mathcal{A})$ is torsion-free as a graded abelian group.

Definition 2.6. $\mathcal{A}$ is said to be a **rational** K$[\pi, 1]$-arrangement if the **Q**-completion of $M(\mathcal{A})$, denoted $\mathbf{Q}_\infty\big(M(\mathcal{A})\big)$, is aspheric.

Equivalently, $\mathcal{A}$ is rational K$[\pi, 1]$ iff the 1-minimal model $\mathcal{M}$ of $M(\mathcal{A})$ satisfies $f^* : H^*(\mathcal{M}) \longrightarrow H^*(M(\mathcal{A}), \mathbf{Q})$ is an isomorphism.

Theorem 2.4 (Papadima, Yuzvinsky). *[7] $\mathcal{A}$ is rational $K[\pi, 1]$ iff $H^*\big(M(\mathcal{A}); \mathbf{Q}\big), (\cong A_{\mathbf{Q}}^*(\mathcal{A}))$ is a Koszul algebra and the LCS formula holds.*

Theorem 2.5 (Jambu, Papadima). *[2] Let $\mathcal{A}$ be a hypersolvable arrangement. Then the following assertions are equivalent:*

(1) $\mathcal{A}$ is fiber-type.
(2) $\mathbf{Q}_\infty(M(\mathcal{A}))$ is aspheric. (ie $A^*_{\mathbf{Q}}(\mathcal{A})$ is Koszul)
(3) The LCS formula holds.

Let us remark that if $\mathcal{A}$ is hypersolvable, then $A^*(\mathcal{A})$ is quadratic iff $\mathcal{A}$ is fiber-type. Therefore for the hypersolvable arrangements, quadraticity of $A^*_{\mathbf{Q}}(\mathcal{A})$ is equivalent to being fiber-type so is equivalent to being rational $\mathrm{K}[\pi, 1]$.

Question : Is quadraticity of $A^*_{\mathbf{Q}}(\mathcal{A})$ sufficient for $M(\mathcal{A})$ being rational $\mathrm{K}[\pi, 1]$?

Example 2.3. $\mathcal{A}$ in $\mathbf{C}^3$ given by the following linear forms $(x, y, z, x + y, x - z, y - z, x + y - 2z)$. $A^*(\mathcal{A})$ is quadratic (but $\mathcal{A}$ is not hypersolvable and therefore not fiber-type). $P(\mathcal{A}, t) = (1+t)(1+6t+10t^2)$. If $\mathcal{A}$ is rational $\mathrm{K}[\pi, 1]$, then $A^*(\mathcal{A})$ is Koszul and $P(\mathcal{A}, -t)^{-1} = H(\mathcal{U}(\mathcal{A}), t)$, therefore $(1 - 6t + 10t^2)^{-1}$ as an infinite formal series has its coefficients integer and nonnegative which implies that $1-6t+10t^2$ has a real root $r \in (0, 1]$ (interesting exercise for undergraduate students following a course on Complex Analysis). Hence we obtain a contradiction.

References

1. M. Falk, R. Randell, *The lower central series of a fiber-type arrangement*, Invent. Math. 82 (1985), 77-88.
2. M. Jambu, S. Papadima, *A generalization of fiber-type arrangements and a new deformation method*, Topology. 37 (1998) 1135-1164.
3. M. Jambu, S. Papadima, *Deformations of hypersolvable arrangements*, Topology and its Appl, 118 (2002) 103-111.
4. T. Kohno, *On the holonomy Lie algebra and the nilpotent completion of the fundamental group of the complement of hypersurfaces*, Nagoya Math. J., 92 (1983) 21-37.
5. T. Kohno, *Série de Poincaré-Koszul associée aux groupes de tresses pures*, Invent. Math., 82 (1985) 57-75.
6. P. Orlik, H. Terao, *Arrangements of Hyperplanes*, Grundlehren Math. Wiss., Vol 300 (Springer-Verlag, Berlin, 1992).
7. S. Papadima, S. Yuzvinsky, *On Rational $K[\pi, 1]$-spaces and Koszul Algebras*, Journal of Pure and Applied Algebra, 144 (1999) 157-167.
8. B. Shelton, S. Yuzvinsky, *Koszul Algebras from Graphs and Hyperplane Arrangements*, Journal of the London Mathematical Society (1997) 56(3), 477-490.

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 189–204)

SOME PROBLEMS IN *PI*-THEORY

ALEXANDER KEMER

Ulyanovsk University,
Ulyanovsk, 432700, Russia
E-mail:kemer_alexander@rambler.ru

AVERYANOV ILYA

Ulyanovsk University,
Ulyanovsk, 432700, Russia
E-mail: ilya@averyanov.org, ilya.averyanov@gmail.com

In this paper, we discuss some problems in PI-theory. Some developments of this theory are presented.

Keywords: Polynomial identity; Trace identity; Variety of the algebras.

1. Basic Definitions

In this section, we give the basic definitions and recall some old results.

Consider the free associative algebra $F\langle X\rangle$ over an associative and commutative ring F generated by a countable set X. The elements of the set X are called variables and the elements of the algebra $F\langle X\rangle$ are called polynomials (depending on the variables from X).

Let A be an arbitrary associative F-algebra and $f = f(x_1, \ldots, x_n) \in F\langle X\rangle$. We say the algebra A satisfies the (polynomial) identity $f = 0$ if for every elements $a_i \in A$, the equality

$$f(a_1, \ldots, a_n) = 0$$

holds in A. Denote by $T[A]$ the set of all polynomials $f \in F\langle X\rangle$ such that $f = 0$ is an identity of the algebra A. Evidently, the set $T[A]$ forms an ideal of the free algebra $F\langle X\rangle$. This ideal is called the ideal of identities of the algebra A. An arbitrary ideal Γ of the algebra $F\langle X\rangle$ is said to be a T-ideal if Γ is an ideal of identities of some algebra. The class of alebras satisfying a given set of identities is called a variety (of algebras). Denote by $T[V]$ the set of all polynomials $f \in F\langle X\rangle$ such that $f = 0$ is an identity of every

algebra from the variety V. The set $T[V]$ also forms a T-ideal. This ideal is called the ideal of identities of the variety V. We say that the algebra A generates the variety V if $T[A] = T[V]$.

Now we recall the concepts of algebras with trace and trace identities.

Let A be an associative algebra with unit over F, R be an associative and commutative F-algebra with a unit. Let $C(A)$ be the center of A and $\pi : R \to C(A)$ be a homomorphism of F-algebras. Put $ar = a\pi(r)$ for $a \in A, r \in R$. Then, it turns the algebra A into an R-algebra. Let $Tr : A \to R$ be an arbitrary R-linear mapping satisfying the property: $Tr(ab) = Tr(ba)$, for every $a, b \in A$. Then we call the set (A, R, π, Tr) an algebra with trace. Sometimes we simply say that A is an algebra with trace, having in mind that the algebra R and the mappings π, Tr are given.

Example 1.1. The full matrix algebra $M_n(F)$ with trace $Tr : M_n(F) \to F$ defined in the usual way is an algebra with trace.

Example 1.2. Let G be the Grassmann algebra with unit generated by a countable set. Let G_0 be the subspace generated by the products of even number of generators and G_1 be the subspace generated by the products of odd number of generators. Denote by $M_{n,k}$ the subalgebra of the algebra $M_{n+k}(G)$ consisting of all matrix of the form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A and D are the submatrices of the sizes $n \times n$ and $k \times k$ respectively with the elements from G_0; B and C are the submatrices of the sizes $n \times k$ and $k \times n$ respectively with the elements from G_1. Define the trace $Tr : M_{n,k} \to G_0$ by putting

$$Tr\begin{pmatrix} A & B \\ C & D \end{pmatrix} = Tr(A) - Tr(D),$$

where $Tr(A)$ and $Tr(D)$ are the usual traces of the matrices over commutative algebra G_0.

Let X be a countable set and $F^{\natural}\langle X\rangle$ be the free associative algebra with unit generated by the set X. Define an equivalence on the semigroup with unit $\langle X\rangle$ generated by the set X, putting $u_1 \sim u_2$ if and only if there exist elements $v, w \in \langle X\rangle$ such that $u_1 = vw, u_2 = wv$. If $u \in \langle X\rangle$ then we put $\overline{u} = \{v \in \langle X\rangle | v \sim u\}$. Denote by $T\langle X\rangle$ the free associative and commutative algebra with unit generated by all elements $Tr(\overline{u})$, where $u \in \langle X\rangle$.

The algebra $\widetilde{F}\langle X\rangle = F^{\sharp}\langle X\rangle \otimes T\langle X\rangle$ is called the free algebra with trace generated by the set X. Identifying the algebras $F^{\sharp}\langle X\rangle \otimes 1$ and $F^{\sharp}\langle X\rangle$, we have the inclusions:

$$X \subseteq F\langle X\rangle \subseteq F^{\sharp}\langle X\rangle \subseteq \widetilde{F}\langle X\rangle.$$

We identify also the algebras $T\langle X\rangle$ and $1 \otimes T\langle X\rangle$. Then an arbitrary element $f \in \widetilde{F}\langle X\rangle$ can be written as an F-linear combination of the elements $u_0 Tr(u_1)\cdots Tr(u_n)$, where $u_i \in \langle X\rangle, n \geq 0$. We call the elements of this form trace monomials and the elements of the algebra $\widetilde{F}\langle X\rangle$ the trace polynomials in the variables from X. We also call the polynomials from $F\langle X\rangle$ the ordinary polynomials.

Let A be an algebra with trace $f = f(x_1, \ldots, x_n) \in \widetilde{F}\langle X\rangle$. We say the algebra A satisfies the trace identity $f = 0$ if for arbitrary $a_1, \ldots, a_n \in A$ the equality $f(a_1, \ldots, a_n) = 0$ is satisfied in A. The ideal

$$\widetilde{T}[A] = \{f \in \widetilde{F}\langle X\rangle | f = 0 \text{ is an idenity of } A\}$$

is called the ideal of trace identities of the algebra A. An arbitrary ideal of the algebra $\widetilde{F}\langle X\rangle$ which is an ideal of the trace identities of some algebra, is called a $\widetilde{T}$-ideal.

Obviously, the ideal of trace identities of an arbitrary algebra contains the ideal of the ordinary identities of the same algebra.

A class of all algebras with trace satisfying a given set of trace identities is called a variety of algebras with trace. If $\widetilde{V}$ is a variety of the algebras with trace then we denote by $\widetilde{T}[\widetilde{V}]$ the ideal of trace identities of $\widetilde{V}$.

Example 1.3. Let A be a square matrix of order n, which elements are considered as variables. It is well-known that if $\mathrm{char} F = 0$ then the coefficients of the characteristic polynomial $X_n(t)$ of the matrix A are F - linear combinations of expressions of the form

$$Tr(A^i)\cdots Tr(A^k).$$

It follows from this and the Cayley - Hamilton theorem that the matrix algebra $M_n(F)$ satisfies the trace identity $X_n(x) = 0$. The identity $Tr(1) = n$ is also the trace identity (of degree 0) of the algebra $M_n(F)$.

Denote by P the set of all ordinary multilinear polynomials of $F\langle X\rangle$. Let Γ be a T-ideal of $F\langle X\rangle$. The intersection $P \cap \Gamma$ is called the multilinear component of Γ. The multilinear component of the ideal of identities of a variety is called the multilinear component of the variety.

2. Prime classical and non-classical varieties

Let $\widetilde{P}_n$ be the set of all multilinear polynomials with trace of degree n depending on the variables $x_1, \ldots, x_n$. We denote also by $\widetilde{P}$ the set of all multilinear polynomials with trace. It follows from the definition of the free algebra with trace that every polynomial $f \in \widetilde{P}_n$ can be written in the unique way as an F-linear combination of the monomials

$$u_0(Tr(1))^l Tr(u_1) \cdots Tr(u_n), \quad u_i \in \langle X \rangle, \quad n, l \geq 0,$$

which belong to $\widetilde{P}_n$ and satisfy the properties:

(1) $u_i \neq 1$ for every $i > 0$;
(2) For every $i > 0$ the least number j, such that x_j occurs in u_{i+1}, is greater than the least number k, such that x_k occurs in u_i.

Denote by K the subalgebra with unit of the algebra $\widetilde{F}\langle X \rangle$ generated by the element $Tr(1)$. Let KS_{n+1} be the group algebra (over K) of the symmetric group of permutations S_{n+1} acting on the set $\{0, 1, \ldots, n\}$. We define the K-linear mapping $\lambda_n : \widetilde{P}_n \to KS_{n+1}$ by putting

$$\lambda_n(x_{i_1} \cdots x_{i_s} Tr(x_{j_1} \cdots x_{j_t}) Tr(x_{k_1} \cdots x_{k_l}) \cdots) = \sigma \in S_{n+1},$$

where σ is a permutation which decomposition into cycles is the following

$$\sigma = (0, i_1, \ldots, i_s)(j_1, \ldots, j_t)(k_1, \ldots, k_l) \ldots .$$

We can see the symbol 0 plays a role of a label which indicates the non-trace part of the monomial. It follows from the definition of the free algebra with trace that the mapping λ_n is the isomorphism of K - modules. If $f \in \widetilde{P}_n, a \in KS_{n+1}$ then we put $fa = \lambda_n^{-1}(\lambda_n(f)a), af = \lambda_n^{-1}(a\lambda_n(f))$. It turns the space $\widetilde{P}_n$ into KS_{n+1}-module.

Example 2.1. Put

$$X_n(x_1, \ldots x_n) = \lambda_n^{-1}(\sum_{\sigma \in S_{n+1}} (-1)^\sigma \sigma).$$

Then we can be proved that the polynomial $X_n(x_1, \ldots, x_n)$ is the full linearization of the Cayley - Hamilton polynomial $X_n(x)$.

The description of the base of trace identetities of the algebra $M_n(F)$ in the case of characterstic 0 was obtained by Yu. P. Razmyslov in 1974 [13]. He proved the following remarkable theorem.

Theorem 2.1. *(see [13]). If char $F = 0$ then every trace identity of the algebra $M_n(F)$ follows from the Cayley - Hamilton identity $X_n(x) = 0$ and identity $Tr(1) = n$.*

In 1995 A.R.Kemer has generalized this theorem:

Theorem 2.2. *(see [3]). Let F be a field of an arbitrary characteristic. Every multilinear trace identity of the algebra $M_n(F)$ follows from the identity $X_n(x_1, \ldots, x_n) = 0$ and identity $Tr(1) = n$.*

We note that A. Zubkov [17] also used the other techniques to obtain the same result later.

In 1985, Yu. P. Razmyslov [14] described the trace identities of the superalgebras $M_{n,k}$ over a field of characteristic 0. We recall the description.

Consider a rectangle Yuong diagram containing $n+1$ rows of length $k+1$. Put $m = (n+1)(k+1)$. Fill this diagram by the numbers $0, 1, \ldots, m$ in the standard order: From top to bottom and from left to right. Denote by $D_{n,k}$ the Young tableau we obtained. Let R be the subgroup of the group S_{m+1} consisting of all permutations σ such that for every i the numbers i and $\sigma(i)$ belongs to the same row of $D_{n,k}$, C is the subgroup of all permutations σ such that for every i the numbers i and $\sigma(i)$ belongs to the same column. Put

$$X_{n,k} = X_{n,k}(x_1, \ldots, x_m) = \lambda^{-1}(\sum_{p \in R, q \in C} (-1)^q pq).$$

Theorem 2.3. *(see [14]). If char $F = 0$ then every trace identity of the algebra $M_{n,k}(F)$ follows from the identities $\sigma X_{n,k} = 0, \sigma \in S_{m+1}$ and identity $Tr(1) = n - k$.*

We remark that in the case $k = 0$ we provided Theorem 2.2. Recently L. Samoilov [15] has obtained a shorter proof of Theorem 2.3.

A T-ideal Γ of the algebra $F\langle X\rangle$ is called a verbally prime if and only if for every T- ideals Γ_1, Γ_2 inclusion $\Gamma_1\Gamma_2 \subseteq \Gamma$ implies one of the inclusions $\Gamma_1 \subseteq \Gamma$ or $\Gamma_2 \subseteq \Gamma$. A T-ideal Γ is called a verbally semiprime if and only if there are no non-trival nilpotent modulo Γ T-ideals. A variety of the algebras is called a prime (semiprime) if the ideal of identities of this variety is verbally prime (semiprime). In the same manner, we define the prime and semiprime varieties of the algebras with trace.

In the case of characteristic zero, the prime varieties were described by A.R.Kemer [2]. The following structure Theorem is valid.

Theorem 2.4. *(see [2]). Let char $F = 0$. 1. For every non-zero T-ideal Γ of the algebra $F\langle X\rangle$ there exists a nilpotent modulo Γ verbally semiprime T-ideal containing Γ. 2. A T-ideal Γ is verbally semiprime if and only if Γ is the intersection of finite number of the verbally prime T-ideals. 3. A*

variety is prime if and only if either this variety is generated by the algebra $M_{n,k}$ or $M_n(G)$, where G is the Grassmann algebra of infinite rank.

The problem of classification of prime varieties in the case of characteristic p is open. First of all, we remark that this problem scarcely can be solved in full sense because the Finite Bases Problem in characteristic p has a negative solution. Perhaps it can be solved at multilinear level.

Take an arbitrary $\gamma \in F$. Then we call a $\widetilde{T}$-ideal $\widetilde{\Gamma}$ γ-classical if

(1) $Tr(1) - \gamma \in \widetilde{\Gamma}$;
(2) For every n a set $\lambda_n(\widetilde{\Gamma} \cap \widetilde{P}_n)$ is a two-sided ideal of the group algebra KS_{n+1} .

A variety $\widetilde{V}$ of the algebras with trace is called γ-classical if the ideal of trace identities of $\widetilde{V}$ is γ-classical. A variety of ordinary algebras V is called a γ-classical if

$$T[V] = \widetilde{T}[\widetilde{V}] \cap F\langle X\rangle$$

for some γ-classical variety of the algebras with trace $\widetilde{V}$.

Using Razmyslov's results [14], it is easy to describe the γ-classical varieties of the algebras with trace in the case of characteristic zero.

Theorem 2.5. *Let char $F = 0, \gamma \in F$. A proper variety of the algebras with trace $\widetilde{V}$ is a γ-classical if and only if 1. The number γ is integer; 2. The variety $\widetilde{V}$ is generated by the algebra $(M_{n,k})$, where $n - k = \gamma$.*

It also can be proved in the case of characteristic zero that the varieties $\mathrm{Var}(M_n(G))$ are non-classical, i.e. they are not γ-classical for any $\gamma \in F$.

An ordinary (non-trace) multilinear polynomial $h = h(x_1, \ldots, x_m)$ is called a trace-killer of the matrix algebra $M_k(F)$ if the algebra $M_k(F)$ satisfies the trace identity of the form

$$h(x_1, \ldots, x_m)Tr(x_{m+1}) = g(x_1, \ldots, x_m, x_{m+1})$$

for some ordinary multilinear polynomial g.

Example 2.2. It is easy to verify that the algebra $M_2(F)$ satisfies the following trace identity

$$[x, y, z]Tr(t) = [[x, y] \circ t, z]$$

($a \circ b = ab + ba$). Consequently, the polynomial $[x, y, z]$ is a trace-killer of the algebra $M_2(F)$.

Consider a proper T-ideal Γ of the free algebra $F\langle X\rangle$ over a field of characteristic p. It was proved in [3] that $T[M_k(F)] \cap P \subseteq \Gamma$ for some k (we recall that P is the set of all multilinear ordinary polynomials). A minimal number k with this property is called the matrix type of Γ. The matrix type of the ideal of identites of the variety is called the matrix type of variety.

Example 2.3. It is easy to prove that the matrix type of the Grassmann algebra G (over F) of infinite rank is $\leq p$. Indeed, if we define the trace in G trivially: $Tr(g) = 0$ for every $g \in G$, then the algebra G satisfies the Cayley - Hamilton trace identity $X_p = 0$, because G satisfies the full linearization of the identity $x^p = 0$. By theorem 2.2 G satisfies every multilinear identity of the matrix algebra $M_p(F)$, i.e. the matrix type of the algebra G is not greater than p. Later, we shall see that the matrix type of G is equal to p.

Let k be the matrix type of a T-ideal Γ. Then we call a T-ideal Γ regular if and only if Γ does not contain some trace-killer of the algebra $M_k(F)$. A variety is called regular if and only if its ideal of identities is regular.

Theorem 2.6. *(see [4]). Every regular prime variety of the matrix type k is k-classical.*

If a prime variety V of the matrix type k is non-classical then by Theorem 2.6, this variety is irregular. It means that the algebras from V satisfy every identity $h = 0$, where the polynomial h is a trace killer of the algebra $M_k(F)$. This condition is very strong. For example, if $k = 2$ then the algebras from V must satisfy the identity $[x, y, z] = 0$. It follows from this immediately that in the case $p \neq 2$, there are no such variety V.

Example 2.4. If char $F \neq 2$ then the verbally prime variety generated by the Grassmann algebra of infinite rank G is non-classical. Indeed, assume on the contrary that $T[G] = \widetilde{\Gamma} \cap F\langle X\rangle$, where $\widetilde{\Gamma}$ is γ-classical $\widetilde{T}$-ideal for some $\gamma \in F$. Since $[x_1, x_2, x_3] \in \widetilde{\Gamma}$. Then $\sigma[x_1, x_2, x_3] = 0$ modulo $\widetilde{\Gamma}$ for every $\sigma \in S(4)$. In particular, if the permutation σ is a cycle $(0, 1, 2)$ then we have the identity modulo $\widetilde{\Gamma}$

$$0 = \sigma[x_1, x_2, x_3]$$

$$= \lambda_3^{-1}((0,1,2)((0,1,2,3) - (0,2,1,3) - (0,3,1,2) + (0,3,2,1)))$$

$$= \lambda_3^{-1}((0,2,3,1) - (1,3) - (0,3,2,1) + (0,3))$$

$$= [x_2, x_3]x_1 - Tr(x_1x_3)Tr(x_2) + x_3Tr(x_1)Tr(x_2).$$

Substituting $x_1 = x_2 = 1$ into this identity, we obtain $\gamma(x_3 - Tr(x_3)) = 0$. Hence, $\gamma = 0$ (if $x_3 - Tr(x_3) \in \widetilde{\Gamma}$ then $[x_3, x_1] \in T[G]$). Substituting $x_1 = 1, x_2 = x, x_3 = y$ into our identity, we obtain the following identity modulo $\widetilde{\Gamma}$

$$[x, y] = Tr(y)Tr(x),$$

This implies that identity $2[x, y] = 0$. Since $charF \neq 2$ then $[x, y] \in T[G]$. We obtain a contradiction.

In conclusion we mention two conjectures about the non-classical prime varieties. Yu.P. Razmyslov [14] has formulated the conjecture about the identities of the algebras $M_n(G)$.

Conjecture 2.1. *If char $F = 0$ then*

$$T[M_n(G)] = (\widetilde{T}[M_{n,n}] + \{Tr(x)\}^{\widetilde{T}}) \cap F\langle X\rangle$$

($\{g\}^{\widetilde{T}}$ is a $\widetilde{T}$-ideal generated by g).

The following conjecture about the multilinear components of the non-classical prime varieties in arbitrary characteristic generalizes the conjecture 1 and also looks quite probable.

Conjecture 2.2. *If V is a non-classical prime variety then*

$$P \cap T[V] = P \cap (\widetilde{T}[\widetilde{V}] + \{Tr(x)\}^{\widetilde{T}})$$

for some 0-classical prime variety $\widetilde{V}$.

3. Prime subvarieties of $Var M_2(F)$

In this Section, we give the description of the multilinear components of the prime subvarieties of the variety $\mathrm{Var}M_2(F)$ generated by the matrix algebra of order 2 over a field F of characteristic $p > 0$ [5].

Let $\widetilde{Z}\langle X\rangle$, $\widetilde{Q}\langle X\rangle$ are the free algebras with trace over the ring of integers Z and over the field of the rational numbers Q respectively. To formulate the description we need the following lemma.

Lemma 3.1. *Let $n = p^k - 2$, l be the maximal number with property: p^l divides $(n+1)!$. Then there exists a uniqely defined modulo $\widetilde{T}[M_2(Q)]$ polynomial $g_k \in \widetilde{Z}\langle X\rangle$ such that*

$$X_{0,n} = p^l g_k$$

modulo $\widetilde{T}[M_2(Q)]$.

We recall the formula

$$\lambda_n(X_{0,n}) = \sum_{\sigma \in S_{n+1}} \sigma.$$

Thus by the lemma, the polynomial $\frac{1}{p^l} X_{0,n}$ is equal to some polynomial g_k with interger coefficients modulo trace identities of the algebra $M_2(Q)$.

Let ϕ be the natural homomorphism $\widetilde{Z}\langle X\rangle \to \widetilde{F}\langle X\rangle$. Put $f_k = \phi(g_k)$. Denote by $\widetilde{\Gamma}_k$ the $\widetilde{T}$-ideal of the algebra $\widetilde{F}\langle X\rangle$ generated by the polynomial f_k, the Cayley - Hamilton polynomial X_2 and $Tr(1) - 2$. Let V_k be the variety corresponding to the T-ideal

$$\Gamma_k = F\langle X\rangle \cap \widetilde{\Gamma}_k.$$

Then we call a subvariety V of $VarM_2(F)$ trivial if either V is a subvariety of the variety of commutative algebras or the multilinear components of the varieties V and $\mathrm{Var}M_2(F)$ are equal.

Theorem 3.1. *(see [5]).*

(1) If V is a non-trivial prime subvariety of $VarM_2(F)$ then for some k the multilinear component of the varieties V and V_k are equal;

(2) For every k there exists a prime subvariety V which multilinear component is equal to the multilinear component of V_k;

(3) If $k < s$ then

$$P \cap \widetilde{\Gamma}_k \supset P \cap \widetilde{\Gamma}_s.$$

It is easy to calculate the polynomial f_1 in any characteristic:

$$f_1 = -X_{p-2}.$$

The prime variety V_1 is known and was found by Yu. P. Razmyslov [13]. The relatively free algebra of countable rank of this variety is very interesting. This algebra satisfies the Engel identity of degree $p - 1$ but it is not Lie nilpotent.

It is also easy to calculate the polynomial f_2 for $p = 2$.

$$f_2 = x_1 \circ x_2 + Tr(x_1)Tr(x_2).$$

The relatively free algebra of countable rank satisfying identities $f_2 = 0$ and $Tr(1) = 0$ is also very interesting and can be considered as the Grassmann algebra in characteristic 2.

Recently T. Antipova, A. Antipov and A.Kemer [6] with a help of computer have calculated the polynomial f_3 for $p = 2$ by using computer. The result is:

$$f_3(x,y,z,t,u,v) = xytvzu + xyuvtz + xyuzvt + xyvutz + xyztuv$$

$$+xyztvu + xyzvut + xtyuvz + xtyvuz + xtuvzy + xtvzuy + xuvtzy$$

$$+xvutzy + xzvuyt + ytuvxz + ytvxzu + yuxtvz + yuxzvt + yuvtxz$$

$$+yuzvtx + yvtxuz + yvutxz + yztuvx + yztvux + yzvutx + txuyvz$$

$$+tyuvzx + tuvzyx + tvzxuy + tvzuyx + uvtzyx + uzxvty + uzvtyx$$

$$+vutzyx + zxtuvy + zxtvuy + zxvuty + ztuvyx + ztvuyx + zuxtvy$$

$$+zuvytx + zvyuxt + zvtxuy + zvuytx + xzutyTr(v) + xtyuzTr(v)$$

$$+ytuzxTr(v) + zuytxTr(v) + xytzTr(uv) + xyuzTr(tv) + xyzuTr(tv)$$

$$+xuzyTr(tv) + xzvtTr(yu) + yuxtTr(zv) + yuxzTr(tv) + yvuzTr(xt)$$

$$+yzuxTr(tv) + txuyTr(zv) + tvzxTr(yu) + uzyxTr(tv) + zxuyTr(tv)$$

$$+ztyxTr(uv) + zuyxTr(tv) + zuvyTr(xt) + xyztTr(u)Tr(v)$$

$$+xztyTr(u)Tr(v)+xtzyTr(u)Tr(v)+yztxTr(u)Tr(v)+ytzxTr(u)Tr(v)$$

$$+tzyxTr(u)Tr(v) + xytTr(zuv) + xytTr(zvu) + xyzTr(tuv)$$

$$+xtyTr(zvu) + xuzTr(yvt) + xzuTr(ytv) + ytxTr(zuv) + tyxTr(zuv)$$

$$+tyxTr(zvu)+uzxTr(yvt)+zyxTr(tvu)+zuxTr(ytv)+xytTr(zu)Tr(v)$$

$$+xyzTr(tu)Tr(v)+xyzTr(uv)Tr(t)+xtyTr(zu)Tr(v)+ytxTr(zu)Tr(v)$$

$$+yuzTr(xt)Tr(v)+tyxTr(zu)Tr(v)+zyxTr(tu)Tr(v)+zyxTr(uv)Tr(t)$$

$$+zuyTr(xt)Tr(v) + xyzTr(t)Tr(u)Tr(v) + zyxTr(t)Tr(u)Tr(v)$$

$$+xyTr(ztuv) + xzTr(ytuv) + xzTr(ytvu) + xzTr(yvut) + yxTr(zvut)$$

$$+ytTr(xzuv) + ytTr(xzvu) + tyTr(xuvz) + tyTr(xvuz) + zxTr(ytuv)$$

$$+zxTr(yuvt) + zxTr(yvut) + xyTr(tvu)Tr(z) + xzTr(ytu)Tr(v)$$

$$+yxTr(tuv)Tr(z)+ytTr(xzu)Tr(v)+zxTr(yut)Tr(v)+tyTr(xuz)Tr(v)$$

$$+xyTr(zt)Tr(uv)+xtTr(yu)Tr(zv)+yxTr(zt)Tr(uv)+txTr(yu)Tr(zv)$$

$$+xyTr(uv)Tr(z)Tr(t)+yxTr(uv)Tr(z)Tr(t)+xyTr(z)Tr(t)Tr(u)Tr(v)$$

$$+yxTr(z)Tr(t)Tr(u)Tr(v) + xTr(ytuvz) + xTr(yzvut) + yTr(xtuvz)$$

$$+yTr(xtvzu) + yTr(xuvtz) + yTr(xuzvt) + yTr(xvutz) + yTr(xztuv)$$

$$+yTr(xztvu)+yTr(xzvut)+zTr(xtvyu)+zTr(xuyvt)+xTr(ztuv)Tr(y)$$

$$+xTr(zvut)Tr(y)+xTr(yuz)Tr(tv)+xTr(yzu)Tr(tv)+xTr(tuv)Tr(yz)$$

$$xTr(tvu)Tr(yz) + yTr(xtz)Tr(uv) + yTr(xzt)Tr(uv) + zTr(yuv)Tr(xt)$$

$$zTr(yvu)Tr(xt) + xTr(tuv)Tr(y)Tr(z) + xTr(tvu)Tr(y)Tr(z)$$

$$+Tr(xytvuz) + Tr(xtyuvz) + Tr(xzuvty) + Tr(xzvuyt)$$

$$+Tr(ytvuz)Tr(x)+Tr(yuvtz)Tr(x)+Tr(yvutz)Tr(x)+Tr(yztuv)Tr(x)$$

$$+Tr(yztvu)Tr(x)+Tr(yzuvt)Tr(x)+Tr(xtvz)Tr(yu)+Tr(xzvt)Tr(yu)$$

$$+Tr(ytuv)Tr(xz)+Tr(ytvu)Tr(xz)+Tr(yuvt)Tr(xz)+Tr(yvut)Tr(xz)$$

$$+Tr(ztuv)Tr(xy)+Tr(zvut)Tr(xy)+Tr(xytz)Tr(u)Tr(v)+Tr(xzty)Tr(u)Tr(v)$$

$$+Tr(ytuz)Tr(x)Tr(v) + Tr(yzut)Tr(x)Tr(v) + Tr(ztuv)Tr(x)Tr(y)$$

$$+Tr(zvut)Tr(x)Tr(y)+Tr(xyt)Tr(zuv)+Tr(xty)Tr(zvu)+Tr(xuz)Tr(yvt)$$

$$+Tr(xzu)Tr(ytv) + Tr(xyt)Tr(zu)Tr(v) + Tr(xyz)Tr(tu)Tr(v)$$

$$+Tr(xty)Tr(zu)Tr(v) + Tr(xzy)Tr(tu)Tr(v) + Tr(ytu)Tr(xz)Tr(v)$$

$$+Tr(yut)Tr(xz)Tr(v) + Tr(ztu)Tr(xy)Tr(v) + Tr(zut)Tr(xy)Tr(v)$$

$$+Tr(xyz)Tr(t)Tr(u)Tr(v) + Tr(xzy)Tr(t)Tr(u)Tr(v)$$

$$+Tr(yzt)Tr(x)Tr(u)Tr(v) + Tr(ytz)Tr(x)Tr(u)Tr(v)$$

$$+Tr(tuv)Tr(x)Tr(y)Tr(z) + Tr(tvu)Tr(x)Tr(y)Tr(z)$$

$$+Tr(ztu)Tr(x)Tr(y)Tr(v) + Tr(zut)Tr(x)Tr(y)Tr(v)$$

$$+Tr(x)Tr(y)Tr(z)Tr(t)Tr(u)Tr(v).$$

We have also calculated the polynomial f_2 for $p = 3$, but this polynomial can not be written down because it contains more then 1000 summands. Finally, we are sure that the other polynomials f_k cannot be calculated by any computer. This result shows that in the case of characteristic p the verbally prime varieties cannot be described in traditional way.

The $\widetilde{T}$-ideals $\widetilde{\Gamma}_k$ can be described in other terms. Let $A \subseteq \widetilde{Z}\langle X\rangle$. We denote by $A_Z^{\widetilde{T}}$ the $\widetilde{T}$-ideal of the algebra $\widetilde{Z}\langle X\rangle$ generated by the set A.

Theorem 3.2.

$$\widetilde{\Gamma}_k = F\phi(\widetilde{Z}\langle X\rangle \cap \frac{1}{p^l}\{X_2, X_{n,0}, Tr(1) - 2\}_Z^{\widetilde{T}})$$

(the numbers n and l are the same as in lemma).

Now we give another approach which looks quite promising.

Let n, k, γ, m be the integers satisfying the properties:

(1) n, k, m are non-negative;
(2) $n - k = \gamma$ modulo p.

Put

$$\widetilde{\Gamma}_{n,k}(\gamma, m) = F\phi(\widetilde{Z}\langle X\rangle \cap \frac{1}{p^m}\{\sigma X_{n,k}, Tr(1) - \gamma | \sigma \in S_{(n+1)(k+1)}\}_Z^{\widetilde{T}}).$$

(the polynomial $X_{n,k}$ was defined ealier).

Theorem 3.3. $\widetilde{\Gamma}_k = \widetilde{\Gamma}_{2,p^k}(2 + p^{l+1}, k + l)$.

This theorem shows that the variety V_k is a slightly transformed variety $\mathrm{Var}M_{2,p^k}$. In conclusion, we formulate the first conjecture about the multilinear components of the classical prime varieties as following:

Conjecture 3.1. *If V is a classical prime variety then for some n, k, γ, l, the multilinear component of the variety V is equal to the multilinear component of the $\widetilde{T}$-ideal $\widetilde{\Gamma}_{n,k}(\gamma, m)$.*

4. Trace-killers for $M_3(F)$

The results mentioned above show that the trace-killers are very important for describing prime varietes of algebras. It is obvious that trace-killers for $M_k(F)$ form a T-ideal.We now denote it by $\Gamma^{(k)}$. For any algebra A we also denote by $T_n[A]$ the ideal $T[A] \cap F\langle x_1, \cdots, x_n\rangle$ of $F\langle x_1, \cdots, x_n\rangle$, $x_1, \cdots, x_n \in X$. We put $\Gamma_n^{(3)} = \Gamma^{(3)} \cap F\langle x_1, \cdots, x_n\rangle$. In this Section we give fine description of $\Gamma^{(3)}$, if *char* $F = 0$, $\Gamma_n^{(3)}$, if *char* $F = p$ with $p > 3$. We use the results of A.R.Kemer [1,2], which imply that if *char* $F = 0$, then there exist a finite dimensional classical algebra A for which $\Gamma^{(3)} = T[A]$; f *char* $F = p$, $p > 3$ then for each n there exist a finite dimensional classical algebra A_n for which $\Gamma_n^{(3)} = T_n[A_n]$.

Let $E = e_1F + e_2F$, $e_i^2 = e_i$ $e_ie_j = 0$ if $i \neq j$, $\widetilde{B} = E *_F \langle x_1, \cdots, x_4\rangle$. Let B be the factor-algebra of $\widetilde{B}$, defined by relations

(a) $e_iue_j[e_jve_iwe_j, e_jae_j] = 0$,$[e_iue_jve_i, e_iae_i]e_iwe_j = 0$;
(b) $e_iue_jve_iwe_j[e_jae_j, e_jbe_j] = 0$,$e_iue_j[e_jae_j, e_jbe_j]e_jve_iwe_j = 0$,
(b′) $[e_iae_i, e_ibe_i]e_iue_jve_iwe_j = 0$, $e_iue_jve_i[e_iae_i, e_ibe_i]e_iwe_j = 0$;
(c) $[e_iue_jve_i, e_iae_jbe_i] = 0$;
(d) $e_iue_jS_4(e_jae_j, e_jbe_j, e_jce_j, e_jde_j) = 0$,
(d′) $S_4(e_jae_j, e_jbe_j, e_jce_j, e_jde_j)e_jue_i = 0$;
(e) $e_iue_j[e_jae_j, e_jbe_j, e_jce_j] = 0$,$[e_jae_j, e_jbe_j, e_jce_j]e_jue_i = 0$;
(f) $e_iue_j[e_jae_j, e_jbe_j][e_jce_j, e_jde_j] = 0$,
(f′) $[e_jae_j, e_jbe_j][e_jce_j, e_jde_j]e_jue_i = 0$;
(g) $[e_iae_i, e_ibe_i]e_iue_j[e_jce_j, e_jde_j] = 0$;
(h) $[e_iue_jve_i, [e_iae_i, e_ibe_i]] = 0$;
(i) $[e_iue_jve_i, e_iae_i, e_ibe_i] + [e_iue_jve_i, e_ibe_i, e_iae_i] = 0$;
(j) $[e_iue_jve_i, e_iae_i, e_ibe_i] = 0$;
(k) $[e_iae_i, e_ibe_i][e_iue_jve_i, e_ice_i] = 0$;
(l) $[e_iue_j[e_jae_j, e_jbe_j]e_jve_i, e_ice_i] = 0$;
(m) $e_ia_1e_jb_1e_ia_2e_jb_2e_ia_3e_j = 0$;

$a, b, c, d, u, v, w, a_k, b_k \in B, \{i, j\} = \{1, 2\}$

One can see, that $e_iEe_i = D_i + R_i, D_i = e_iF$, $D_i \cap R_i = (0)$, R_i is ideal of e_iEe_i. We also put $e_ia_1e_ia_2e_ia_3e_ia_4e_i = 0$ in B for each $a_j \in R_i$ Obviously B is a finite dimensional classical algebra.

Theorem 4.1. *(see [10]) If char* $F = 0$, *then there exist a finite-dimensional local (noncommutative) algebra* C *such that* $\Gamma(3) = T[M_2(F)] \cap T[B] \cap T[C]$.

Theorem 4.2. *(see [10]) If char* $F = p$, $p > 3$, *then for each* n

there exist a finite-dimensional local (noncommutative) algebra C such that $\Gamma_n^{(3)} = T_n[M_2(F)] \cap T_n[B] \cap T_n[C]$.

Corollary 4.1. *If char $F = p$, $p > 3$ then for each n there exist N such that if $f \in T_n[M_2(F)]$, $degf > N$ and f is representable as a sum of products of five or more commutators, then f is a trace-killer for M_3. If char $F = 0$, then there exist N such that if $f \in T[M_2(F)]$, $degf > N$ and f is representable as a sum of products of five or more commutators, then f is a trace-killer for $M_3(F)$.*

5. Conjecture of C.Procesi

Let R_{Q_p} and R_{F_p} be the algebras of generic matrices of order n over ring Q_p and field F_p respectively, where Q_p is the ring of rational numbers with denominators not divisible by p, and F_p is an infinite field of characteristic p. The conjecture of C.Procesi is well-known [11] : the kernel of the canonical epimorphism $R_{Q_p} \to R_{F_p}$ is equal to pR_{Q_p}.

In 1985, W.Shelter [16] gave a negative answer to this conjecture for $n = 2$, $p = 2$. Later A.Kemer [8] obtained the following result: for each prime p, there exists $n \leq p$ such that the conjecture of Procesi is not true for the pair p, n. We also prove that the conjecture of Procesi is true for $n = 3, p > 3$.

Theorem 5.1. *(see [9]) Let R_{Q_p} be a 2-generated algebra of generic 3×3 matrices over ring Q_p, R_{F_p} is a 2-generated algebra of generic 3×3 matrices over an infinite field F_p of characteristic $p > 3$. Then the kernel of the canonical epimorphism $R_{Q_p} \to R_{F_p}$ is equal to pR_{Q_p}.*

6. Matrix type of some algebras

In this Section, we discuss results obtained by the A.R.Kemer in [7].

Denote by t_n the matrix type of the algebra $M_n(G)$, where G is the Grassmann algebra of infinite rank over a field F of characteristic $p \neq 2$. We mentioned above that the algebra G satisfies every multilinear identity of the algebra $M_p(F)$. It follows from this that the algebra $M_n(G)$ satisfies every multilinear polynomial identity of the algebra $M_{np}(F)$ i.e. $t_n \leq pn$. It is also easy to see that the matrix type of the Grassmann algebra is greater than $\frac{p-1}{2}$ because G does not satisfy the Standard identity of degree $p-1$. Thus we have the trivial estimations for t_n

$$\frac{p-1}{2} n < t_n \leq pn.$$

The main result of the paper [7] gave a precise formula for t_n.

Theorem 6.1. $t_n = pn$.

This Theorem is equivalent to the following statement:

There exists a multilinear polynomial $f \in T[M_{np-1}(F)]$ *such that* $f \notin T[M_n(G)]$.

The problem of calculation of such polynomials is very difficult even for $n = 1$. This problem is also connected with the divisibility by p.

Denote by S_m the standard polynomial of degree m:

$$S_m = S_m(x_1, \ldots, x_m) = \sum_{\sigma \in S(m)} (-1)^\sigma x_{\sigma(1)} \cdots x_{\sigma(m)}.$$

For $n = 1$, the statement of Theorem 6.1 can be formulated in the following form.

For sufficiently large even m, *there exists a multilinear polynomial* $f \in Z\langle X\rangle$ *and a number* $a \in Z$ *which is not divisible by* p *such that* $aS_m = p^l f$ *modulo* $T[M_{p-1}(Q)] \cap T[G]$, *where* p^l *is the maximal power of* p *dividing* $m!$.

Actually Theorem 6.1 follows from the following theorem which gives a good estimation for the minimal degree of the standard identity of the algebra $M_n(G)$.

Let d_n be the minimal number m such that the algebra $M_k(G)$ satisfies the standard identity of degree m. Then we have the following theorem:

Theorem 6.2. $2pn \geq d_n > (2n-3)p - 1$.

The problem of calculating of the number d_k is very interesting. We formulate the following conjecture.

Conjecture 6.1. $d_k = (2k-1)p + 1$.

The following statement is formulated as a lemma in [6].

Theorem 6.3. *For even number* $m \geq (3p+1)(p-1)$, *there exists a multilinear polynomial* $f \in Z\langle X\rangle$ *and a number* $a \in Z$, *which is not divisible by* p, *such that* $aS_m = p^l f$ *modulo* $T[M_{p-1}] \cap T[G]$, *where* p^l *is the maximal power of* p *dividing* $m!$

References

1. Kemer A.R. Identities of finitely generated algebras over an infinite field, *Izv. AN SSSR. Ser Mat.* ,1990, vol 54, no 4, pp 726-753 (Russian).

2. Kemer A.R. Ideals of identities of associative algebras, *Amer. Math. Soc. Translations of Math. Monographs*, 1991, Vol 87, Providence, R.I.
3. Kemer A. Multilinear identities of the algebras over a field of characteristic p, *Intern. J. of Algebra and Computations*,1995, no 2, pp 1-9.
4. Kemer A. Multilinear components of the regular prime varieties, *Lecture notes in pure and applied mathematics*, 1998, Vol 198, pp 171-183.
5. Kemer A. Multilinear components of the prime subvarieties of the variety $Var(M_2(F)$, *Algebras and Representations* 2001, Vol. 4, no 1, pp.87-104.
6. Kemer A., Antipova T., Antipov A. On the traditional way of description of the prime varieties in characteristic p, *Analele stiintifice Univ. Ovidius. Constanta*, 2000, no 8.
7. Kemer A. Matrix Type of Some Algebras over a Field of Characteristic p *J. of Algebra* 2002, Vol. 251, no 2,pp. 849-863(15)
8. Kemer A. R. On some problems in PI-theory in characteristic p connects with dividing by p, *Proc. of 3 Intern. Algebra conf., Kluwer Acad. Publish.*, 2003, pp. 53-67.
9. Kemer A.R. Averyanov I.V. Conjecture of Procesi for 2-generated algebra of generic 3×3 matrices *J. Algebra* , 2006, Vol. 299, pp. 151-170.
10. A. Kemer, I. Averyanov Description of the algebras generating the variety of trace-killers *Adv. in Appl. Math.*, 2006, Vol. 37, Issue 3, pp.390-403.
11. Procesi C. The invariant theory of $n \times n$ matrices, *Adv. in Math.*, 1976, Vol 19, pp 306-381.
12. Procesi C. Computing with 2×2 matrices, *J. Algebra*, 1984, Vol. 87, no 2, pp 342-359.
13. Razmyslov Yu. P. Trace identities of the matrix algebras over field of characteristic zero, *Izv. AN SSSR (russian)*, 1974, Vol. 38, no 4, pp 723-756.
14. Razmyslov Yu. P. Trace identities and central polynomials of the matrix superalgebras $M_{n,k}$, *Mat. sbornik (russian)*, 1985, Vol. 128, no 2, pp 194-215.
15. Samoilov L.M., A new proof of Razmyslov's theorem about trace identities of matrix superalgebras, *Fundam. Prikl. Mat.*, 6:4, 2000, pp 1221-1227 (russian).
16. Schelter W. F. On question concerning generic matrices over the integers, *J. Algebra*, 1985, Vol. 96, pp 48-53.
17. Zubkov A. N. On the generalization of the theorem of Procesi - Razmyslov, *Algebra i Logika*, 1996, Vol. 35, no 4, pp. 433-457.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 205–217)

ON IRREDUCIBLE SUBALGEBRAS OF MATRIX WEYL ALGEBRAS* †

P. S. KOLESNIKOV

Sobolev Institute of Mathematics,
Novosibirsk 630090, Russia
E-mail: pavelsk@math.nsc.ru

We observe an application of conformal algebra theory to the following natural problem: describe those subalgebras of the algebra of differential operators with polynomial coefficients (Weyl algebra) that act irreducibly on the space of polynomials.

1. Introduction

Since the celebrated paper [9], the Weyl algebras have become a topic of precise algebraic study. In a formal way, the nth Weyl algebra A_n, $n \geq 1$, can be defined as follows. Let $\Bbbk\langle X_n\rangle$ stands for the free associative algebra generated by the set $X_n = \{p_i, q_i \mid i = 1, \ldots, n\}$, and let Σ_n be a subset of $\Bbbk\langle X_n\rangle$ that consists of $p_ip_j - p_jp_i$, $q_iq_j - q_jq_i$, $q_ip_j - p_jq_i - \delta_{ij}$, where $i, j = 1, \ldots, n$. Then

$$A_n = \Bbbk\langle X_n \mid \Sigma_n\rangle = \Bbbk\langle X_n\rangle/(\Sigma_n).$$

Throughout the paper, we assume $\operatorname{char} \Bbbk = 0$. The algebra A_n is simple, and it has a canonical faithful irreducible representation on the space of polynomials $H_n = \Bbbk[x_1, \ldots, x_n]$. Namely,

$$\begin{aligned} p_if(x_1, \ldots, x_n) &= x_if(x_1, \ldots, x_n), \\ q_if(x_1, \ldots, x_n) &= \frac{\partial f(x_1, \ldots, x_n)}{\partial x_i}, \\ i &= 1, \ldots, n. \end{aligned} \tag{1}$$

Weyl algebras are in some sense the simplest (after finite-dimensional ones) simple noncommutative algebras. Since a linear basis of A_n can be

*Partially supported by RFBR
†1991 Mathematics Subject Classification. 16S50; 14R10; 16S32

chosen in the form $p_1^{a_1} \dots p_n^{a_n} q_1^{b_1} \dots q_n^{b_n}$, $a_i, b_i \geq 0$, the Gel'fand–Kirillov dimension of A_n is equal to $2n$.

Less is known on subalgebras of Weyl algebras. For example, this is an open question whether A_1 contains a proper subalgebra isomorphic to A_1 (the Dixmier problem). This question is known to be equivalent in a certain sense to the famous Jacobian Conjecture, see [2,5].

Also, for example, it had been staying unknown for a long time whether a Weyl algebra (e.g., A_1) contains a non-Noetherian subalgebra. Recently, the problem was solved (positively) in [3].

In this paper, we consider the following task which seems very difficult to complete in general.

Problem 1.1. *Describe those subalgebras of A_n that act irreducibly on the module H_n with respect to the representation* (1).

Note that a similar problem can be stated in a more general form for matrix algebras (subalgebras of $\mathbb{M}_N(A_n)$) with canonical (left) action on $H_n \otimes \Bbbk^N$.

This problem remains unsolved even for $n = N = 1$. However, it is possible to define a special class of subalgebras that occur to be related with so called conformal algebras [12]. In the context of conformal algebras, a special case of Problem 1.1 naturally appears as an analogue of the classical Burnside theorem [6].

This paper is devoted to a discussion of the relations between Problem 1.1 (for $n = 1$) and a structure theory topic of associative conformal algebras. We will also state a conjecture that relates to the case of n-conformal algebras, $n > 1$.

2. Conformal algebras

In this section, we consider a generalization of the notion of a conformal algebra from [12] which is complementary in some sense to the one studied in [1].

Suppose G is a linear algebraic group over an algebraically closed field $\Bbbk$. Denote by H the algebra of regular functions $\Bbbk[G]$ on G. This is an affine commutative algebra equipped with the structure of Hopf algebra dual to the group structure on G. Namely, the coproduct Δ, counit ε, and antipode

S are defined on H as follows:

$$\Delta : H \to H \otimes H, \quad \Delta(f) = \sum_{(f)} f_{(1)} \otimes f_{(2)},$$

$$\sum_{(f)} f_{(1)}(g_1) f_{(2)}(g_2) = f(g_1 g_2),$$

$$\varepsilon(f) = f(e), \quad S(f)(g) = f(g^{-1}),$$

where $f \in H$, $g, g_1, g_2 \in G$, e is the identity of G.

Definition 2.1. A *(G)-conformal algebra* C is a left H-module endowed with $\Bbbk$-bilinear operations $(\cdot\ {}_g\ \cdot)$, $g \in G$, such that

- for every $a, b \in C$ the function $(a\ {}_x\ b) : G \to C$ is regular, i.e., $(a\ {}_x\ b) = \sum_i f_i \otimes c_i \in H \otimes C$, where $(a\ {}_g\ b) = \sum_i f_i(g) c_i$, $g \in G$;
- for every $a, b \in C$, and for every $f \in H$, $g \in G$ we have

$$(fa\ {}_g\ b) = f(g^{-1})(a\ {}_g\ b), \quad (a\ {}_g\ fb) = L_g f(a\ {}_g\ b), \tag{2}$$

where $(L_g f)(x) = f(gx)$.

For $G = \{e\}$, a (G)-conformal algebra is just an algebra over the field $\Bbbk$. If $G = \mathbb{A}^1$ (affine line) then Definition 2.1 coincides with the definition of a conformal algebra from [12] in terms of λ-brackets.

Let us call an $(\mathbb{A}^n)$-conformal algebra by *n-conformal* algebra for simplicity. This is exactly a pseudo-algebra over $H_n = \Bbbk[x_1, \ldots, x_n]$ in the sense of [1].

Example 2.1. Suppose A is an algebra over $\Bbbk$. Then the free H-module $C = H \otimes A$ equipped with the operations

$$(f \otimes a)\ {}_g\ (h \otimes b) = f(g^{-1}) L_g h \otimes ab, \quad f, h \in H,\ a, b \in C,\ g \in G,$$

is called the *current* (G)-conformal algebra over A.

Example 2.2. Suppose A is an H-comodule algebra, i.e., a (non-associative, in general) algebra endowed with a coassociative coaction $\Delta_A : A \to H \otimes A$, $\Delta_A(a) = \sum_{(a)} a_{(1)} \otimes a_{(2)}$. Then the free H-module $C = H \otimes A$ with respect to the operations given by

$$(f \otimes a)\ {}_g\ (h \otimes b) = f(g^{-1}) b_{(1)}(g) L_g h \otimes a b_{(2)}, \quad f, h \in H,\ a, b \in C,\ g \in G,$$

is called the *differential* (G)-conformal algebra over A.

Current conformal algebra is a particular case of a differential one, when $\Delta_A(a) = 1 \otimes a$, $a \in A$. A series of examples of H-comodule algebras is provided by affine G-varieties.

Suppose V is a Zariski closed subset of an affine space, and let the group G act on V continuously. Then the coordinate algebra $A = \Bbbk[V]$ is an H-comodule algebra (Δ_A is dual to the action of G on V). Then the differential (G)-conformal algebra over A is denoted by $\mathrm{Cend}_1^{(G,V)}$ (conformal endomorphisms of 1-generated free A-module). The (G)-conformal algebra of $(N \times N)$-matrices over $\mathrm{Cend}_1^{(G,V)}$ is denoted by $\mathrm{Cend}_N^{(G,V)}$.

The conformal algebra $\mathrm{Cend}_N^{(G,V)}$ has the following natural "geometric" interpretation. Denote by M_N the linear space of all $\Bbbk^N$-valued regular functions on V. Then an element of $\mathrm{Cend}_N^{(G,V)}$ can be thought of as a transformation rule of the space M_N by means of the group G. For example, the left shift transformation L defined by

$$L : g \mapsto L_g, \quad (L_g u)(x) = u(gx), \quad g \in G,\ u \in M_N,\ x \in V,$$

belongs to $\mathrm{Cend}_N^{(G,V)}$ (it plays the role of a unit in this conformal algebra [20]).

It is clear how to define subalgebras, left and right ideals of a (G)-conformal algebra (these are H-submodules closed under all g-products in an appropriate way).

Proposition 2.1. *([18]) The (G)-conformal algebra $\mathrm{Cend}_N^{(G,V)}$ is simple if and only if V is an irreducible G-set.*

The following identity is easy to verify for $\mathrm{Cend}_N^{(G,V)}$:

$$a\ {}_x\ (b\ {}_y\ c) = (a\ {}_x\ b)\ {}_{yx}\ c, \tag{3}$$

$a, b, c \in \mathrm{Cend}_N^{(G,V)}$, $x, y \in G$. A (G)-conformal algebra C which satisfies the relations (3) is said to be associative.

This definition of associativity is coherent to the general operadic approach by [10] assuming the base multi-category (pseudo-tensor category in [4]) is the category of left H-modules $\mathcal{M}^*(H)$ [1].

In this case, the opposite family of g-products on a (G)-conformal algebra C can be defined as follows. Given $a, b \in C$, consider $(a\ {}_x\ b) = \sum_i f_i \otimes c_i \in H \otimes C$ and set

$$\{a\ {}_x\ b\} = \sum_i S(f_{i(2)}) \otimes S(f_{i(1)}) c_i. \tag{4}$$

The operations $\{\cdot\ _g\ \cdot\}$, $g \in G$, defined by (4) satisfy "dual" definition of a (G)-conformal algebra. Namely, $(a\ _x\ b)^{\mathrm{op}} := \{b\ _x\ a\}$ define a structure of a (G^{op})-conformal algebra on the same H-module C.

If the group G is abelian, then the multi-category $\mathcal{M}^*(H)$ is symmetric (see, e.g., [19]), and the following identities have sense:

$$\begin{aligned} \text{(commutativity)} \quad & (a\ _x\ b) = (a\ _x\ b)^{\mathrm{op}}; \\ \text{(anti-commutatvity)} \quad & (a\ _x\ b) = -(a\ _x\ b)^{\mathrm{op}}; \\ \text{(Jacobi)} \quad & a\ _x\ (b\ _y\ c) - b\ _y\ (a\ _x\ b) = (a\ _x\ b)\ _{yx}\ c. \end{aligned}$$

It is also clear how to define "super" versions of commutativity and Jacobi identity. For an arbitrary variety of algebras one may define its (G)-conformal analogue following the general scheme of [17]. However, the group G has to be abelian if the set of defining identities of the variety contains an identity with a permutation of variables.

Proposition 2.2. *Let C be an associative (G)-conformal algebra over an abelian linear algebraic group G. Then the same H-module C with respect to new operations $[a\ _g\ b] := (a\ _g\ b) - (a\ _g\ b)^{\mathrm{op}}$ is a Lie (G)-conformal algebra denoted by $C^{(-)}$.*

Proof. Straightforward computation. □

In the paper [8], the structure theory of Lie $(\mathbb{A}^1)$-conformal algebras over $\Bbbk = \mathbb{C}$ *of finite type* was developed. In this case, $H = \mathbb{C}[x]$, and a (G)-conformal algebra is simply called a conformal one. The finiteness condition for a (G)-conformal algebra C means that C is a finitely generated H-module. As a corollary, the structure theorem for associative conformal algebras of finite type was obtained in [13]. In [21], the analogue of the Wedderburn principal theorem was proved for associative conformal algebras of finite type.

The first steps beyond the finite type case were done in the papers [20] and [6]. In the first one, the simple unital finitely generated conformal algebras of linear growth were described: it turns out that such a conformal algebra is isomorphic to $\mathrm{Cend}_N^{(\mathbb{A}^1,\mathbb{A}^1)}$. In the second paper mentioned above, conformal subalgebras of Cend_1 were described, and the general conjecture on the structure of such subalgebras in Cend_N was stated.

The last conjecture was proved in [14], that led to a structure theory of associative conformal algebras infinite type (but with a finite faithful rep-

resentation, see below). Also, this result allowed to complete classification of simple associative conformal algebras of linear growth [15].

Let us return to the case of an arbitrary linear algebraic group G. Suppose V is an affine G-set and consider the (G)-conformal algebra $\mathrm{Cend}_N^{(G,V)}$, $N \geq 1$. Recall that $\mathrm{Cend}_N^{(G,V)} = \mathrm{Cend}_1^{(G,V)} \otimes \mathbb{M}_N(\Bbbk) \simeq H \otimes \mathbb{M}_N(A)$, where $A = \Bbbk[V]$.

Suppose $a = \sum_i f_i \otimes a_i \in H \otimes \mathbb{M}_N(A)$. Then for every $u \in M_N := A \otimes \Bbbk^N$ and for every $g \in G$ we are given $(a\ {}_g\ u) := \sum_i f_i(g^{-1}) a_i(L_g u) \in M_N$. The family of operations obtained satisfy the properties of Definition 2.1 and also the relation 3. This shows a way to the definition of what is a module over an associative (G)-conformal algebra (c.f. [7]).

Definition 2.2. (c.f. [6]) A conformal subalgebra $C \subseteq \mathrm{Cend}_N^{(G,V)}$ is said to be *irreducible* if M_N contains no nonzero proper A-submodules invariant with respect to $(a\ {}_g\ \cdot)$, $a \in C$, $g \in G$.

Recall that a left ideal of a ring is called *essential* if it has a nonzero intersection with every nonzero left ideal of a ring. The same notion can be applied to (G)-conformal algebras.

Let $V = G$, where G acts on itself by left multiplications. Then $A = H$ and $\mathrm{Cend}_N^{(G,G)} \simeq H \otimes \mathbb{M}_N(H)$.

Theorem 2.1. *([18]) Suppose the base field $\Bbbk$ is algebraically closed. Let C be an irreducible subalgebra of $\mathrm{Cend}_N^{(G,G)}$. Then $C_1 = (1 \otimes H)C$ is an essential left ideal of $\mathrm{Cend}_N^{(G,G)}$.*

For the case $G = \mathbb{A}^1$ Theorem 2.1 was proved in [14]. The generalization to the case $G = \mathbb{A}^n$ appeared in [16].

Finally, note that all left ideals of $\mathrm{Cend}_N^{(G,V)}$ have the following form: $\mathfrak{F}(H \otimes I)$, where I is a left ideal of $\mathbb{M}_N(A)$ and $\mathfrak{F}$ is the formal Fourier transform, i.e.,

$$\mathfrak{F}(h \otimes f \otimes a) = \sum_{(f)} hS(f_{(1)}) \otimes f_{(2)} \otimes a,$$

$h \in H$, $f \in A$, $a \in \mathbb{M}_N(\Bbbk)$.

Throughout the rest of the paper we will mainly consider the case $G = \mathbb{A}^n$, $n \geq 1$. The $(\mathbb{A}^n)$-conformal algebra $\mathrm{Cend}_N^n := \mathrm{Cend}_N^{(\mathbb{A}^n, \mathbb{A}^n)}$ is closely related to the matrix nth Weyl algebra $\mathbb{M}_N(A_n)$. Subalgebras of Cend_N^n give rise to a special class of subalgebras in $\mathbb{M}_N(A_n)$. In the following section, we state a formal description of such subalgebras.

3. TC-algebras

Let us fix a natural number $n \geq 1$ and consider the polynomial algebra $H = H_n = \Bbbk[x_1, \ldots, x_n]$ as a topological algebra with respect to x-adic topology.

Suppose A stands for a (Hausdorff) topological algebra (non-associative, in general) equipped with continuous derivations $\partial_1, \ldots, \partial_n$.

An *A-valued field* is a continuous map $a : H \to A$ which is *translation-invariant* (T-invariant, for short), i.e.,

$$a\frac{\partial}{\partial x_i} = \partial_i a, \quad i = 1, \ldots, n. \tag{5}$$

Denote the set of all A-valued fields by $\mathcal{F}(A)$.

The space of fields $\mathcal{F}(A)$ can be considered as a left H-module with respect to the action defined by

$$(x_i a)(f) = -a\left(\frac{\partial f}{\partial x_i}\right) = -\partial_i a(f), \quad a \in \mathcal{F}(A),\ f \in H.$$

If B is a subspace of A which is invariant under all ∂_i, $i = 1, \ldots, n$, then the space of B-valued fields $\mathcal{F}(B)$ is an H-submodule of $\mathcal{F}(A)$. Conversely, if C is an H-submodule of $\mathcal{F}(A)$ then denote by $\mathcal{A}(C)$ the space of values of fields from C, i.e.,

$$\mathcal{A}(C) = \{a(f) \mid a \in C,\ f \in H\}.$$

This construction is a particular case of so-called *annihilation space* [1] of an H-module.

Note that the following relations always hold:

$$B \supseteq \mathcal{A}(\mathcal{F}(B)), \quad \mathcal{F}(\mathcal{A}(C)) \supseteq C,$$

and also

$$\mathcal{A}(\mathcal{F}(\mathcal{A}(\mathcal{F}(B)))) = \mathcal{A}(\mathcal{F}(B)).$$

Definition 3.1. A topological algebra A with continuous derivations ∂_i, $i = 1, \ldots, n$, is said to be a *TC-algebra* (from "translation invariance", "continuity") if $A = \mathcal{A}(\mathcal{F}(A))$.

It is clear that for any TC-algebra the derivations ∂_i necessarily commute with each other, and each of them is locally nilpotent.

A homomorphism of TC-algebras is a continuous ∂_i-invariant homomorphism of algebras.

Example 3.1. (see [16])

(i) The polynomial algebra $H = \Bbbk[x_1, \dots, x_n]$ itself is an associative TC-algebra with respect to the x-adic topology and $\partial_i = \partial/\partial x_i$, $i = 1, \dots, n$.

(ii) If n is even then the same H with respect to the Poisson bracket

$$\{f, g\} = \sum_{i=1}^{k} \frac{\partial f}{\partial T_i} \frac{\partial g}{\partial T_{k+i}} - \frac{\partial f}{\partial T_{k+i}} \frac{\partial g}{\partial T_i}, \quad n = 2k, \ f, g \in H,$$

is a Lie TC-algebra.

(iii) The Weyl algebra A_n is an associative TC-algebra with respect to the q-adic topology and $\partial_i(a) = ap_i - p_i a$, $a \in A_n$, $i = 1, \dots, n$.

(iv) A left/right ideal B of the algebra A_n is a TC-subalgebra if and only if $B = A_n f(p)$ or $B = f(p)A_n$, respectively. Here $p = (p_1, \cdots, p_n)$.

(v) The Lie algebra $\mathcal{W}_n \subset \mathbb{A}_n^{(-)}$ is also a TC-algebra, as well as its classical subalgebras $\mathcal{S}_n = \{D \in \mathcal{W}_n \mid Dv = 0\}$ and $\mathcal{H}_n = \{D \in \mathcal{W}_n \mid Ds = 0\}$, $v = \mathrm{d}T_1 \wedge \cdots \wedge \mathrm{d}T_n$, $s = \sum_{i=1}^{k} \mathrm{d}T_i \wedge \mathrm{d}T_{k+i}$, $n = 2k$.

It is clear how to expand the examples from (i), (iii), (iv) to matrices over H or over A_n.

Recall that $H = \Bbbk[x_1, \dots, x_n]$; this is the coordinate Hopf algebra for the group $G = \mathbb{A}^n$. Suppose M is a finitely generated H-module. Then $E = \mathrm{End}_{\Bbbk} M$ can be considered as a topological algebra with respect to the finite topology in the sense of [11] and continuous derivations $\partial_i = [\cdot, x_i]$, $i = 1, \dots, n$. Indeed, a sequence $\psi_k \in E$, $k \geq 1$, converges to zero in E if and only if for every finite number of $u_j \in M$, $j = 1, \dots, m$, we have $\psi_k u_j = 0$ for $k \gg 1$. Therefore, if $\psi_k \to 0$ then $\partial_i \psi_k \to 0$ for all $i = 1, \dots, n$.

Proposition 3.1. *If M is a free N-generated H-module then*

$$\mathcal{A}(\mathcal{F}(\mathrm{End}_{\Bbbk} M)) \simeq \mathbb{M}_N(A_n)$$

as TC-algebras.

Proof. Note that $\mathbb{M}_N(A_n)$ is a subalgebra of $\mathrm{End}_{\Bbbk} M$, where the embedding is given by the canonical representation (1). This representation is obviously a continuous open map with respect to the q-adic topology on $\mathbb{M}_N(A_n)$ and the finite topology on $\mathrm{End}_{\Bbbk} M$. Moreover, since the matrix algebra over a TC-algebra is also a TC-algebra, we have

$$\mathcal{A}(\mathcal{F}(\mathrm{End}_{\Bbbk} M)) \supseteq \mathcal{A}(\mathcal{F}(\mathbb{M}_N(A_n))) = \mathbb{M}_N(A_n).$$

Suppose $\varphi \in \mathcal{A}(\mathcal{F}(\mathrm{End}_{\Bbbk} M))$, i.e., $\varphi = a(f)$, $a \in \mathcal{F}(\mathrm{End}_{\Bbbk} M)$, $f \in H$. Consider $\varphi_{m_1, \dots, m_n} = a\left(x_1^{m_1} \dots x_n^{m_n}\right)$.

It follows from (5) that the field a is completely defined by an arbitrary family of values $u^{(i)}_{m_1,\dots,m_n} = \varphi_{m_1,\dots,m_n}(e_i)$, where $\{e_i \mid i = 1,\dots,N\}$ is a basis of M over H. Indeed, since the map $a : H \to \operatorname{End}_{\Bbbk} M$ is continuous, for each i there exist only a finite number of $(m_1,\dots,m_n) \in \mathbb{Z}^n_{\geq 0}$ such that $u^{(i)}_{m_1,\dots,m_n} = \sum_{j=1}^{N} f^{(i,j)}_{m_1,\dots,m_n} e_j \neq 0$, $f^{(i,j)}_{m_1,\dots,m_n} \in H$. Now for every $h = h(x_1,\dots,x_n) \in H$

$$a(h) = \sum_{m_1,\dots,m_n \geq 0} \frac{1}{m_1! \dots m_n!} F_{m_1,\dots,m_n} \frac{\partial^{m_1}}{\partial q_1^{m_1}} \dots \frac{\partial^{m_n}}{\partial q_n^{m_n}} h(q_1,\dots,q_n), \quad (6)$$

where $F_{m_1,\dots,m_n} = \left(f^{(i,j)}_{m_1,\dots,m_n}\right) \in \mathbb{M}_N(\Bbbk[p_1,\dots,p_n])$. In particular, $\varphi = a(f) \in \mathbb{M}_N(A_n)$. □

4. Irreducible TC-subalgebras of matrix Weyl algebras

Let us say that S is a *TC-subalgebra* of $\mathbb{M}_N(A_n)$ if $S = \mathcal{A}(\mathcal{F}(S))$. As before, suppose M is a free N-generated left module over $H = \Bbbk[x_1,\dots,x_n]$. In order to study TC-subalgebras of $\mathbb{M}_N(A_n)$ we will use a conformal algebra structure on the space of all fields over $\mathcal{F}(\operatorname{End}_{\Bbbk} M)$.

Proposition 4.1. *The H-module $\mathcal{F}(\operatorname{End}_{\Bbbk} M)$ can be endowed with an n-conformal algebra structure isomorphic to* Cend^n_N.

Proof. As we have already seen in the proof of Proposition 3.1, an arbitrary field from $\mathcal{F}(\operatorname{End}_{\Bbbk} M)$ is uniquely defined by a finite family of matrices $F_{m_1,\dots,m_n}$, $m_i \geq 0$. Suppose a is such a field as in (6). Then consider

$$\hat{a} = \sum_{m_1,\dots,m_n} \frac{(-x_1)^{m_1}}{m_1!} \dots \frac{(-x_n)^{m_n}}{m_n!} \otimes F_{m_1,\dots,m_n} \in \operatorname{Cend}^n_N .$$

The correspondence $a \mapsto \hat{a}$ is obviously a bijective H-linear map between $\mathcal{F}(\operatorname{End}_{\Bbbk} M)$ and Cend^n_N. □

Therefore, we can identify Cend^n_N with $\mathcal{F}(\operatorname{End}_{\Bbbk} M)$.

For every $a \in \operatorname{Cend}^n_N$ we have two associated families of linear maps $a(f) \in \operatorname{End}_{\Bbbk} M$, $f \in H$, and $(a\ {}_\lambda\ \cdot) \in \operatorname{End}_{\Bbbk} M$, $\lambda \in \mathbb{A}^n$. It is easy to obtain the relation between these families:

$$(a\ {}_\lambda\ \cdot) = \sum_{s \in \mathbb{Z}^n_{\geq 0}} \frac{\lambda_1^{s_1}}{s_1!} \dots \frac{\lambda_n^{s_n}}{s_n!} a(u_s), \quad (7)$$

where $u_s = x_1^{s_1} \dots x_n^{s_n} \in H$. Note that for every fixed $v \in M$ the expression for $(a \,{}_\lambda\, v)$ contains only a finite number of summands.

In a straightforward computation one may get the following explicit expression for $(a \,{}_\lambda\, b) \in \mathrm{Cend}_N^n$, $\lambda \in \mathbb{A}^n$:

$$(a \,{}_\lambda\, b)(h) = \sum_{s \in \mathbb{Z}_{\geq 0}^n} \frac{\lambda_1^{s_1}}{s_1!} \dots \frac{\lambda_n^{s_n}}{s_n!} \sum_{(u_s)} a(u_{s(1)}) b(S(u_{s(2)})h), \tag{8}$$

where $u_s = x_1^{s_1} \dots x_n^{s_n} \in H$. Note that for every fixed a and b the sum contains only a finite number of summands.

Corollary 4.1. *A TC-subalgebra S of $\mathbb{M}_N(A_n)$ is equal to $\mathcal{A}(C)$, where C is a conformal subalgebra of Cend_N^n.*

Proof. Given a TC-subalgebra S, consider $C = \mathcal{F}(S)$. By (8), this is a conformal subalgebra of $\mathcal{F}(\mathrm{End}_{\Bbbk} M)$ which is now identified with Cend_N^n. TC-property implies $S = \mathcal{A}(C)$. □

Proposition 4.2. *A conformal subalgebra C of Cend_N^n is irreducible if and only if the TC-subalgebra $\Bbbk[p_1, \dots, p_n]S \subseteq \mathbb{M}_N(A_n)$, $S = \mathcal{A}(C)$, is irreducible.*

Proof. Suppose C is an irreducible conformal subalgebra, $S = \mathcal{A}(C)$, $S_1 = \Bbbk[p_1, \dots, p_n]S$. Assume $U \subseteq M_N$ is an S_1-invariant subspace. Then $U_1 = S_1 U \subseteq U$ is an H-submodule.

Moreover, if $v \in U_1$ then relation (7) implies $(C \,{}_{(\lambda)}\, v) \subseteq Sv \subseteq U_1$, i.e., U_1 is an H-submodule invariant with respect to $(a \,{}_\lambda\, \cdot)$, $a \in C$, $\lambda \in \mathbb{A}^n$. If $U_1 = 0$ then $U_2 = \{v \in M \mid Sv = 0\} \supseteq U \neq 0$, but it is easy to derive from the properties of S that U_2 is a C-invariant H-submodule. If $U_2 = M$ then $S = 0$ which is impossible.

Conversely, suppose $S_1 = \Bbbk[p_1, \dots, p_n]S$ is irreducible, and assume $0 \neq U \leqslant_H M$ is a C-invariant submodule, i.e., $(a \,{}_\lambda\, v) \in U$ for all $a \in C$, $\lambda \in \mathbb{A}^n$, $v \in U$. For a fixed $v \in U$, choosing different λ's, we obtain from (7) a system of linear equations that can be resolved in $a(u_s)v$, i.e., all $a(f)v \in U$ for all $f \in H$. Hence, U is an S-invariant subspace; since U is an H-submodule, we also have $\Bbbk[p_1, \dots, p_n]SU \subseteq U$. □

The following theorem provides an approach to a solution of the problem stated in the Introduction for the class of TC-subalgebras.

Theorem 4.1. *Let S be an irreducible TC-subalgebra of $\mathbb{M}_N(A_n)$. Then $\Bbbk[p_1, \dots, p_n]S$ is an essential left TC-ideal of $\mathbb{M}_N(A_n)$.*

Proof. By Corollary 4.1, $S = \mathcal{A}(C)$, where C is a conformal subalgebra of Cend_N^n. If S is irreducible then so is $\Bbbk[p_1, \ldots, p_n]S$. By Proposition 4.2, C is an irreducible conformal subalgebra. It follows from Theorem 2.1 for $G = \mathbb{A}^n$ that $(1 \otimes H)C$ is an essential left ideal of Cend_N^n.

It follows from (8) that $\mathcal{A}((1 \otimes H)C)$ is a left ideal of $\mathbb{M}_N(A_n)$ which has a nonzero intersection with every nonzero left TC-ideal of $\mathbb{M}_N(A_n)$. Finally, note that $\mathcal{A}((1 \otimes H)C) = \Bbbk[p_1, \ldots, p_n]S$. □

Corollary 4.2. *An irreducible TC-subalgebra of* $\mathbb{M}_N(A_1)$ *is equal to* $\mathbb{M}_N(A_1)Q$, *where* $Q \in \mathbb{M}_N(\Bbbk[p_1])$, $\det Q \neq 0$.

Proof. Denote p_1 and q_1 by p and q, respectively. The description of left ideals of $\mathrm{Cend}_N = \mathrm{Cend}_N^1$ was found in [6]: all these ideals are of the form $\mathrm{Cend}_{N,Q} = \mathfrak{F}(H \otimes HQ(x))$, where $H = \Bbbk[x]$. Such an ideal is essential if and only if $\det Q \neq 0$.

Therefore, an irreducible TC-subalgebra S of $\mathbb{M}_N(A_1)$ has the following property: $\Bbbk[p]S = \mathcal{A}(\mathrm{Cend}_{N,Q}) = \mathbb{M}_N(A_1)Q(p)$, $\det Q \neq 0$. For $C = \mathcal{F}(S)$, we have $(1 \otimes H)C = \mathrm{Cend}_{N,Q}$.

If the sum

$$S + pS + p^2S + \cdots = \mathbb{M}_N(A_1)Q(p) \tag{9}$$

is direct, then

$$C + (1 \otimes x)C + (1 \otimes x^2)C + \ldots \tag{10}$$

is also direct. In this case, $\mathrm{GKdim}\, C = 0$; these conformal subalgebras were described in [6], the subalgebra S is not irreducible itself.

If the sum (9) is not direct, then so is (10), and it is easy to show (see [14]) that $C = \mathrm{Cend}_{N,Q}$, therefore, $S = \mathbb{M}_N(A_1)Q$. □

Finally, let us state a conjecture that is supposed to be a description of the picture for $n > 1$.

Conjecture 4.1. *Let* S *be a TC-subalgebra of* $\mathbb{M}_N(A_n)$. *Then the following are equivalent:*

(1) S *is irreducible;*
(2) $\mathrm{GKdim}\, S = 2n$ *and* $\Bbbk[p_1, \ldots, p_n]S$ *is an essential left ideal of* $\mathbb{M}_N(A_n)$;
(3) S *is an essential left ideal of* $\mathbb{M}_N(A_n)$.

Acknowledgements

The main results of this paper were presented on the 2nd International Congress in Algebras and Combinatorics in Beijing–X'ian, July 2007. The author is very grateful to Yuqun Chen and Kar Ping Shum for their support in attending the Congress. The author gratefully acknowledges the support of the Pierre Deligne fund based on his 2004 Balzan prize in mathematics.

References

1. Bakalov B., D'Andrea A., Kac V. G. Theory of finite pseudoalgebras, Adv. Math. **162** (2001) no. 1, 1–140.
2. Bass H., Connell E. H., Wright D. The Jacobian conjecture: reduction of degree and formal expansion of the inverse, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 2, 287-330.
3. Bavula V. Solutions to two questions about the Weyl algebras, Proc. Amer. Math. Soc. **133** (2005), 1587–1591.
4. Beilinson A. A., Drinfeld V. G. Chiral algebras, Amer. Math. Soc. Colloquium Publications **51**, AMS, Providence, RI, 2004.
5. Belov-Kanel A., Kontsevich M. The Jacobian conjecture is stably equivalent to the Dixmier conjecture, Mosc. Math. J. **7** (2007), no. 2, 209–218.
6. Boyallian C., Kac V. G., Liberati J. I. On the classification of subalgebras of Cend_N and gc_N, J. Algebra **260** (2003) no. 1, 32–63.
7. Cheng S.-J., Kac V. G. Conformal modules, Asian J. Math. **1** (1997), 181–193.
8. D'Andrea A., Kac V. G. Structure theory of finite conformal algebras, Sel. Math., New Ser. **4** (1998), 377–418.
9. Dixmier J. Sur les algèbres de Weyl, Bull. Soc. Math. France **96** (1968), 209-242.
10. Ginzburg V., Kapranov M. Kozul duality for operads, Duke Math. J. **76** (1994) no. 1, 203–272.
11. Jacobson N. Structure of rings, American Mathematical Society Colloquium Publications **37**, AMS, Providence, RI, 1956.
12. Kac V. G. Vertex algebras for beginners. Second edition, Univ. Lecture Series bf 10, AMS, Providence, RI, 1998.
13. Kac V. G. Formal distribution algebras and conformal algebras, XIIth International Congress in Mathematical Physics (ICMP'97), Internat. Press, Cambridge, MA, 1999, 80–97.
14. Kolesnikov P. S. Associative conformal algebras with finite faithful representation, Adv. Math. **202** (2006) no. 2, 602–637.
15. Kolesnikov P. S. Simple associative conformal algebras of linear growth, J. Algebra **295** (2006) no. 1, 247–268.
16. Kolesnikov P. S. Associative algebras related to conformal algebras, Appl. Categ. Structures, **16** (2008) no.1-2, 167-181.
17. Kolesnikov P. S. Identities of conformal algebras and pseudoalgebras, Comm. Algebra **34** (2006) no. 6, 1965–1979.

18. Kolesnikov P. S. On irreducible algebras of conformal endomorphisms over a linear algebraic group, J. Math. Sci., to appear.
19. Leinster T. Higher operads, higher categories. LMS Lecture Note Series, **298**, Cambridge University Press, Cambridge, 2004.
20. Retakh A. Associative conformal algebras of linear growth, J. Algebra **237** (2001) no. 2, 769–788.
21. Zelmanov E. I. Idempotents in conformal algebras, Proc. of the Third International Algebra Conference, in: Y. Fong et al (Eds.), 2003, 257–266.

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 219–224)

ON THE LENGTH OF CONJUGACY CLASSES AND P-NILPOTENCE OF FINITE GROUPS*†

QINGJUN KONG and XIUYUN GUO

Department of Mathematics, Shanghai University, Shanghai 200444, People's Republic of China

We investigate the influences of lengths of conjugacy classes of finite groups on the structure of finite groups. Some sufficient conditions for a finite group to be p-nilpotent and supersolvable are obtained. Some known results are generalized.

Keywords: Conjugacy classes; P-nilpotent groups; Finite groups.

1. Introduction

All groups considered in this paper are finite. If G is a group, then we use x^G to denote the conjugacy class containing x, $|x^G|$ the length of x^G and Con(G) the set of the conjugacy classes of G.

One of the questions that were studied extensively is what can be said about the structure of the group G if some information is known about the arithmetic structure of Con(G). The answers in many cases were given in many papers. For example, Chillag and Herzog in [1] proved that G is supersolvable and both $|G/F(G)|$ and $|F(G)'|$ are square-free numbers if $|C|$ is a square-free number for each $C \in Con(G)$. Cossey and Wang in [2] showed that G is a solvable p-nilpotent group and the Sylow p-subgroups of $G/O_p(G)$ is of order at most p if there is no conjugacy class whose length of G is divisible by p^2 and q does not divide $p-1$, where p, q are prime divisors of $|G|$. Recently in [3], Liu, Wang and Wei replaced "all conjugacy classes" by "some conjugacy classes" and obtained many interesting results. In this paper, we continue to investigate influences of conjugacy classes of finite

*The research was partially supported by the National Natural Science Foundation of China(10771132), SGRC(GZ310), the Research Grant of Shanghai University and Shanghai Leading Academic Discipline Project(J50101).
†MSC: 20D10; 20D20.

groups on the structure of finite groups. We will concentrate to put our emphasis on the p-part structure of conjugacy class length of p'-element of prime-power-order of G and obtain some theorems on p-nilpotent groups and supersolvable groups. Some results in [1–3] are generalized.

2. Basic definitions and preliminary results

In this section,we give some lemmas which are useful for our main results.

Definition 2.1. An element x of a group G is called a $p'-$element or a $p-$regular element if the order of x is a $p'-$number, that is, $(o(x), p) = 1$.

Lemma 2.1 (1,Lemma 1). *Let $N \trianglelefteq G, x \in N$, and $y \in G$. Then*

(i) $|x^N| \mid |x^G|$.
(ii) $(yN)^{G/N} \mid |y^G|$.

Let p be a prime and Z_p a cyclic group of order p. We denote $(Z_p)^n \rtimes Z_q$ for a Frobenius group with Frobenius kernel $\underbrace{(Z_p \times \cdots \times Z_p)}_{n}$ and Frobennius complement Z_q, and $\underbrace{(Z_p \times \cdots \times Z_p)}_{n}$ as Z_q-module is irreducible.

Lemma 2.2. *Let G be a solvable group and p a prime dividing the order of G. Then G is not p-nilpotent but every proper normal subgroup of G and every proper quotient group of G are p-nilpotent if and only if there exist a prime q and a positive integer n such that $G \simeq (Z_p)^n \rtimes Z_q$.*

Lemma 2.3 (1,Proposition 3). *If the fixed prime $p \| C_G(x)|$ for all $x \in G$, then G is not a nonabelian simple group.*

3. Main results

In this section, we will study the structure of finite groups by using the conjugacy class length of p'-element of prime-power-order. Our first results is the following:

Theorem 3.1. *Let G be a group and p a prime divisor of $|G|$. Suppose that no conjugacy class length of p'-element of prime-power-order of G is divisible by p^2. Then $P/O_p(G)$ is an elementary abelian p-group, where $P \in Syl_p(G)$.*

Proof. Since $(O_p(G/O_p(G)) = 1$ and the quotient group $G/O_p(G)$ satisfies the hypothesis of the theorem by Lemma 2.1, we may assume that $O_p(G) = 1$. Now,set: $U = \Phi(P) \bigcap Z(P)$ and we consider $C_G(U)$. It is easy to prove that there exists an element $x \in G$ such that g centralizes some maximal subgroup of P^x for every p'-element g of prime-power-order of G, and whence g centralizes U^x. On the other hand, it is also easy to prove that g centralizes some G-conjugacy of U for every p-element g, so $\bigcup_{x\in G} C_G(U^x) = \bigcup_{x\in G} C_G(U)^x$ contains all elements of prime-power order of G. As a consequence of the finite simple groups(see [7]), we have $G = \bigcup_{x\in G} C_G(U)^x$. It follows that $G = C_G(U)$, and so $U \leq O_p(G) = 1$. Hence $\Phi(P) = 1$, and therefore P is an elementary abelian p-group. The proof is completed. □

Theorem 3.2. *Let G be a solvable group and p be a prime with $(|G|, p - 1) = 1$. Suppose that no conjugacy class length of p'-element of prime-power order of G is divisible by p^3 and G is $(Z_p \times Z_p) \rtimes Z_q$-free for any prime q with $q \neq p$. Then G is a p-nilpotent group.*

Proof. Assume that the result is false and let G be a counterexample of minimal order. If G is a simple group, then G is a cyclic group of prime order, a contradiction. Hence, G contains proper normal subgroup. For any proper normal subgroup N of G, we have N and G/N are p-nilpotent by Lemma 2.1. Hence, by Lemma 2.2, G is a Frobenius group $(Z_p)^n \rtimes Z_q$. Thus there exists a conjugacy class $C \in Con(G)$ such that $|C| = p^n$. The hypothesis implies that $n \leq 2$ and therefore $|P| \leq p^2$ if $P \in Syl_p(G)$.

If $n = 1$, then $G \simeq Z_p \rtimes Z_q$. Since $(|G|, p - 1) = 1$, it follows that $N_G(P) = C_G(P)$. By Burnside theorem [7;10.1.8] we see that G is p-nilpotent, a contradiction. If $n = 2$, then $G \simeq (Z_p \times Z_p) \rtimes Z_q$, again a contradiction. Our proof is hence completed. □

By using a similar arguments, we obtain the following theorem.

Theorem 3.3. *Let G be a solvable group and p be a prime with $(|G|, p^2 - 1) = 1$. If no conjugacy class length of p'-element of prime-power-order of G is divisible by p^3, then G is p-nilpotent.*

Remark 3.1. In Theorem 3.2 and Theorem 3.3, the condition that G is $(Z_p \times Z_p) \rtimes Z_q$-free and the condition $(|G|, p^2 - 1) = 1$ can not be dropped. For example, let $G = A_4$ be the alternating group of degree 4. Then A_4 has four conjugacy classes C_1, C_2, C_3, C_4 and $|C_1| = 1, |C_2| = 3, |C_3| =$

$4, |C_4| = 4$. Let $p = 2$, obviously $2^3 \nmid |C_i|$, where $i = 1, 2, 3, 4$. but A_4 is not 2-nilpotent.

If p is the smallest prime dividing the order of a group, then we deduce the following Theorem:

Theorem 3.4. *Let G be a group and p be the smallest prime divisor of $|G|$. Suppose that no conjugacy class length of element of G is divisible by p^3 and G is A_4–free. Then G is a p-nilpotent group.*

Proof. Suppose that $G_p \in Syl_p(G)$. If G_p is cyclic, then G is p-nilpotent by Burnside theorem. If $|G_p| = p^2$ and $G_p \simeq (Z_p \times Z_p)$, then $N_G(G_p)/C_G(G_p)$ isomorphic to a subgroup of $GL(2.p)$. If $p > 2$, then it is easy to see that $N_G(G_p) = C_G(G_p)$. Hence G is p-nilpotent by Burnside theorem.

If $p = 2$, then it is easy to see that $N_G(G_p) = C_G(G_p)$ or that $N_G(G_p) \simeq A_4$. But $N_G(G_p) \simeq A_4$ is impossible. Hence $N_G(G_p) = C_G(G_p)$ and therefore G is p-nilpotent.

Now we suppose that $|G_p| \geq p^3$. Since $p^3 \nmid |C|$ for any $C \in Con(G)$ and $|C| = |G : C_G(x)|$ for any $x \in G$, $p||C_G(x)|$. By Lemma 2.3 G is not a simple group. By Lemma 2.1 and by induction, for every proper normal subgroup N of G and every proper quotient group G/N of G, we have N and G/N are p-nilpotent. If G is not p-nilpotent, then by Lemma 2.2, there exist a prime q and a positive integer n such that $G \simeq (Z_p)^n \rtimes Z_q$. Hence, there exists a conjugacy class $C \in Con(G)$ such that $|C| = p^n$. By hypothesis, we have $n \leq 2$, and thereby, $|G_p| \leq p^2$, a contradiction. This shows that G is p-nilpotent. □

Theorem 3.5. *Let A and B be normal subgroups of a group G such that $G = AB$. If G is $(Z_p)^2 \rtimes Z_q$-free for any distinct primes p and q, and $|x^G|$ is cube-free for every element x of $A \bigcup B$, then G is a supersolvable group.*

Proof. The hypotheses are clearly inherited by quotient groups.

We first prove that A and B are supersolvable. We only prove that A is supersolvable. Assume that A is not supersolvable and choose A a counterexample of minimal order. Because the class of supersolvable groups forms a saturated formation , we may suppose that A has a unique minimal normal subgroup N and $\Phi(A) = 1$. Then by Lemma 2.1 and Theorem 3.4, we see that A has a Sylow tower of supersolvable type. Hence A is solvable. Thus we may suppose that N is an elementary abelian p-group and $|N| = p^k$ with $k \geq 1$. If $k = 1$, then G/N is supersolvable by Lemma 2.1. This leads

to A is supersolvable, a contradiction. Hence we may suppose that $k > 1$ and A has no normal subgroup of prime order. Since $\Phi(A) = 1$, there exists a maximal subgroup M of A such that $A = MN$ and $M \bigcap N = 1$. Clearly $M \cong A/N$ is supersolvable. Let Q be a minimal normal subgroup of M. Then $|Q| = q$, a prime. Set $Q = \langle x \rangle$. If $C_A(x) \bigcap N = C_N(x) \neq 1$, then $M < N_A(Q) = A$. Thus Q is a minimal normal subgroup of prime order of A, a contradiction. Hence, $C_A(x) \bigcap N = C_N(x) = 1$ and it follows that $N\langle x \rangle$ is a Frobunius group and by hypotheses, $|N| = p^2$. That is, $N\langle x \rangle$ is Frobenius group $(Z_p \times Z_p) \rtimes Z_q$, a contradiction. Therefore there is no counterexample and A is supersolvable. In the same way, B is also supersolvable.

Next we prove that G is solvable. Since A and B are supersolvable and $G/A = AB/A \cong B/A \bigcap B$, G is solvable.

Now we prove that G is supersolvable. Assume that the result is false. Then, we choose G a counterexample of minimal order. Because the supersolvable groups forms a saturated formation , we may suppose that G has a unique minimal normal subgroup N and $\Phi(G) = 1$. Let $|N| = p^k$ with $k > 1$. Then $F(G) = N = C_G(N)$. If $A' = 1$ or $B' = 1$, G is supersolvable by [AB, theorem 3]. a contradiction. So both A' and B' are nonidentity normal subgroups of G. By the uniqueness of N we can get that $N \leq A'$ and $N \leq B'$. On the other hand, $A' \leq A$ since A is solvable. Thus there exists a subgroup H of A such that H/N is a minimal normal subgroup of G/N. Since G/N is supersolvable, $|H/N| = q$, which is a prime. If $q = p$, then K is a normal p-subgroup of G, which implies $H \leq F(G) = N$, a contradiction. If $q \neq p$, then, by Schur-Zassenhaus Theorem, there is an element x such that $H = N\langle x \rangle$ and $|x| = q$. Suppose that there exists an element $y \neq 1$ in N such that $y \in C_G(x)$. Since N is abelian, $H = N\langle x \rangle \leq C_G(y)$. So $Z(H) > 1$. But $Z(H)$ char $H \trianglelefteq G$. Thus $Z(H) \trianglelefteq G$. Again by the uniqueness of N, we have that $N \leq Z(H)$. This implies $H = N \times \langle x \rangle$, which is a nilpotent normal subgroup of G. Thus $H \leq F(G) = N$, a contradiction. Hence, $C_H(x) = (x)$ and $|x^H| = |H : C_H(x)| = |N| = p^k$ with $k > 1$. If $k = 2$, then $H = N\langle x \rangle$ is a Frobenius group $(Z_p \times Z_p) \rtimes Z_q$, which is contrary to the hypotheses. If $k \geq 3$, again contradicts to the hypotheses. Hence there exists no counterexample and our proof is hence completed. □

References

1. D.Chillag, M.Herzog, On the length of the conjugacy classes of finite groups, *J.Algebra* 131(1990)110-125.
2. J.Cossey, Y.Wang, Remarks on the length of conjugacy classes of finite

groups, *Comm.Algebra*, 27(9)(1999)4347-4353.
3. Xiaolei Liu, Yanming Wang and Huaquan Wei, Notes on the length of conjugacy classes of finite groups, *Journal of Pure and Applied Algebra*, 196(2005)111-117.
4. Yakov Berkovich, Lev Kazarin, Indices of elements and normal structure of finite groups, *J.Algebra*, 283(2005)564-583.
5. J.S.Robinson, A course in the theory of groups, *Spring-verlag, New York, Heidelberg, Berlin*,1980.
6. L.M.Isaacs, Character theory of finite groups, *New York:Academic Press*,1976.
7. B.Fein, W.M.Kantor, M.Schacher, Relative Brauer groups II, *J.Reine Angew.Math.* 328(1981)39-57.
8. A.Ballester-Bolinches, J.Cossey, M.C.Pedraza-Aguilera, On products of finite supersolvable groups, *Comm.Algebra*, 29(7)(2001)3145-3152.
9. R.Baer, Group elements of prime power index, *Trans.Amer.Math.Soc.*, 75(1953)20-47.
10. B.Huppert, Endliche Gruppen I. *Spring-verlag, Berlin Heidelberg, New York*,1967.
11. M.Weinstein, Between nilpotent and solvable, *Polygonal Publishing House*,1982.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 225–242)

COMPUTATIONS WITH FINITE INDEX SUBGROUPS OF $\mathrm{PSL}_2(\mathbb{Z})$ USING FAREY SYMBOLS*

CHRIS A. KURTH

Department of Mathematics,
Iowa State University, Ames, IA 50011, USA
E-mail: kurthc@iastate.edu

LING LONG

Department of Mathematics,
Iowa State University, Ames, IA 50011, USA
E-mail: linglong@iastate.edu

Finite index subgroups of the modular group are of great arithmetic importance. Farey symbols, introduced by Ravi Kulkarni in 1991, are a tool for working with these groups. Given such a group Γ, a Farey symbol for Γ is a certain finite sequence of rational numbers (representing vertices of a fundamental domain of Γ) together with pairing information for the edges between the vertices. They are a compact way of encoding the information about the group and they provide a simple way to do calculations with the group. For example: calculating an independent set of generators and decomposing group elements into a word in these generators, finding coset representatives, elliptic points, and genus of the group, testing if the group is congruence, etc. We will discuss Farey Symbols and explicit algorithms for working with them.

1. Introduction

Modular forms are certain functions defined on the upper half plane displaying certain symmetries under the Möbius transformation action of a finite-index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. The theory of modular forms has been in the central stage of number theory for more than one century and continues to be one of its most exciting areas. Working with modular forms requires knowing information about their underlying groups. There is a vast literature about many aspects of finite index subgroups of the modular groups and their relations with other fields such as combinatorics, algebraic curves.

*2000 Mathematics Subject Classification. 11F06

Interested readers are referred to articles like [1, 2], or a recent survey article by the second author on these groups and their modular forms [9].

Some finite index subgroups of the modular group can be described purely by congruence relations, and as such are called congruence subgroups of the modular group. These groups are relatively easy to work with, as they contain a certain normal subgroup $\Gamma(N)$, such that the quotient of the group by $\Gamma(N)$ is just a subgroup of $\mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})$. Most computational methods for working with modular forms work only for congruence subgroups [14]. Noncongruence subgroups are finite-index subgroups of $(\mathrm{PSL}_2(\mathbb{Z}))$ that cannot be described by congruence relations. Even though they make up the majority of finite-index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ there are few tools for working with them. One tool that works equally well with both congruence and noncongruence groups is the method of Farey symbols introduced by Ravi Kulkarni [6]. In this paper we will recast how to use Farey symbols and some related computational topics. We will discuss some explicit algorithms for working with Farey symbols. The first author has implemented a collection of such algorithms into a free SAGE package called "KFarey". It should be made clear to the readers that "KFarey" is an undergoing project and we will continue to improve the current functions, implement other existing algorithms such as [4, 8], and investigate new algorithms to make "KFarey" useful to a large audience.

2. Subgroups of $\mathrm{PSL}_2(\mathbb{Z})$

Let $\mathrm{SL}_2(\mathbb{Z})$ be the group of 2×2 matrices with integer coefficients and determinant 1, and let $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{I, -I\}$. Let $\mathbb{H}$ be the upper half plane $\mathbb{H} := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. Then $\mathrm{PSL}_2(\mathbb{Z})$ acts faithfully on $\mathbb{H}$ under the action

$$\gamma z = \frac{az + b}{cz + d},$$

where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z})$. The objects of our study will be the finite index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$. For example, the standard congruence groups, $\Gamma(N)$, $\Gamma_1(N)$ and $\Gamma_0(N)$.

If Γ is a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ then the action of Γ partitions $\mathbb{Q} \cup \{\infty\}$ into equivalence classes, where $q_1 \sim q_2$ if $q_1 = \gamma q_2$ for some $\gamma \in \Gamma$. These equivalence classes $\{q\}$ are called the **cusps** of Γ and the width of the cusp $\{q\}$ is $[\mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{Z})}(q) : \mathrm{Stab}_\Gamma(q)]$. We say that the level of Γ is the least common multiple of the cusp widths of Γ.

Recall that a congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ is a subgroup Γ that contains a principal level N congruence subgroup $\Gamma(N)$ for some N. If N is the smallest such N such that this is true, then Γ has level N.

Definition 2.1. Let Γ be a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. For our purposes, a fundamental domain of Γ is a hyperbolic polygon P on $\mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$ such that:

(1) If z is in the interior of P and $\gamma \in \Gamma$, then $\gamma z \in P$ implies $\gamma = I$.
(2) For every $z \in \mathbb{H}$ there is $\gamma \in \Gamma$ such that $\gamma z \in P$.

Lemma 2.1. *Let Γ be a finite index subgroup of $PSL_2(\mathbb{Z})$ and P a hyperbolic polygon. Suppose P is such that:*

(1) If z is in the interior of P and $\gamma \in \Gamma$, then $\gamma z \in P$ implies $\gamma = I$.
(2) For each side e of P, there is $\gamma \in \Gamma$ such that γ maps e to another side of P in an orientation reversing manner.

Then P is a fundamental domain of Γ.

Proof. Let the images γP of P under elements γ of Γ be called P-tiles. By Condition (1) they cannot overlap. Also, given Condition (1), we only need to show that $\mathbb{H} \subseteq \bigcup_{\gamma \in \Gamma} \gamma P$. Suppose this is not true. Then there is $\eta \in \Gamma$ such that ηP has an edge e without a P-tile on the other side. Then $\eta^{-1} e$ is an side of P and the γ of Condition (2) maps $\eta^{-1} e$ to another side of P. Specifically, γ^{-1} maps P to a P-tile adjacent to P across the side e. Then $\eta\gamma^{-1} P$ is a P-tile adjacent to ηP across the side e. Contradiction. □

3. Farey Symbols

3.1. *Special Polygons*

Farey Symbols were introduced by Ravi Kulkarni in 1991 [6] as a compact and efficient way to compute with finite index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$. The idea is to describe the group by a fundamental domain with vertices at certain rational numbers and certain hyperbolic arcs joining these rational numbers. Most of the theory here is summarized from [6].

If x and y are two points on $\mathbb{H} \cup \mathbb{Q}$ then there is a unique circle passing through x and y with center on $\mathbb{Q}$. We say the **hyperbolic arc** joining x and y is the arc of this circle contained in $\mathbb{H} \cup \mathbb{Q}$ joining x and y. We also say the hyperbolic arc joining $x \in \mathbb{H}$ to ∞ is the vertical line segment

$\{x+ti : 0 \leq t \in \mathbb{R}\} \cup \{\infty\}$. We write $H_{x,y}$ for the hyperbolic arc joining x and y. A **hyperbolic polygon** is a polygon composed of hyperbolic arcs.

Through the course of this paper, when a vertex of a hyperbolic arc is in $\mathbb{Q}$ it will always be assumed to be in the form $\frac{a}{b}$ with $a, b \in \mathbb{Z}$, $(a,b) = 1$ and $b > 0$. If the vertex is ∞, we will write it either $\frac{-1}{0}$ or $\frac{1}{0}$ (depending on if it is the leftmost or rightmost element of a Farey sequence).

Let $\rho = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ and let T be the hyperbolic triangle with vertices ρ, ρ^2 and ∞. Then T is a fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})$ ([5] Prop. III.1). Let E_e be the edge joining i to ∞, E_o be the edge joining ρ to ∞, and E_f be the edge joining i to ρ. Then we call an arc A in the upper half plane an **even edge** (resp. **odd edge**, resp. **f-edge**) if $A = \gamma E_e$ (resp. $A = \gamma E_o$, resp. $A = \gamma E_f$) for some $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ (See Figure 1). E_e and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} E_e$ together form a hyperbolic arc from 0 to ∞, and in general even edges come in pairs joining rational numbers $\frac{a}{b}$, and $\frac{a'}{b'}$ with $|a'b - ab'| = 1$ because of the following lemma:

Lemma 3.1. *If $\gamma \in PSL_2(\mathbb{Z})$ and a_1/b_1, a_2/b_2, a_1'/b_1', and a_2'/b_2' are rational numbers in simplest form such that*

$$\gamma(a_1/b_1) = a_1'/b_1', \text{ and } \gamma(a_2/b_2) = a_2'/b_2',$$

then

$$a_2b_1 - a_1b_2 = a_2'b_1' - a_1'b_2' \ .$$

Proof. If $\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ then

$$\gamma(\frac{a_1}{b_1}) = \frac{Aa_1 + Bb_1}{Ca_1 + Db_1} = \frac{a_1'}{b_1'} \text{ and } \gamma(\frac{a_2}{b_2}) = \frac{Aa_2 + Bb_2}{Ca_2 + Db_2} = \frac{a_2'}{b_2'}.$$

So:

$$\begin{aligned} a_1'b_2' - a_2'b_1' &= (Aa_1 + Bb_1)(Ca_2 + Db_2) - (Aa_2 + Bb_2)(Ca_1 + Db_1) \\ &= ADa_1b_2 + BCa_2b_1 - ADa_2b_1 - BCa_1b_2 \\ &= (AD - BC)(a_1b_2 - a_2b_1) \\ &= (a_1b_2 - a_2b_1). \end{aligned}$$

□

So the quantity $a_2b_1 - a_1b_2$ is invariant under transformations in $\mathrm{PSL}_2(\mathbb{Z})$. Note that even edges, odd edges and free edges only map to even edges, odd edges and free edges respectively under transformations $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$.

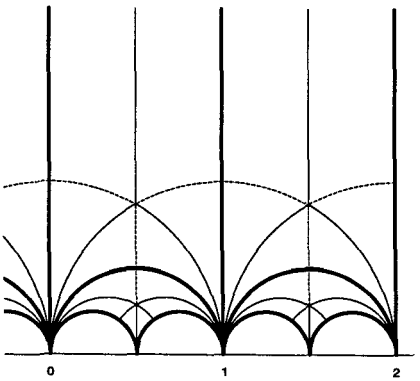


Fig. 1. Even edges are thick, odd edges are thin, and f-edges are dashed

Definition 3.1. A special polygon P is a convex hyperbolic polygon together with a side pairing defined in the following way: The polygon is such that:

(1) The boundary of P consists of even and odd edges.
(2) The even edges of P come in pairs, each pair forming a hyperbolic arc between elements of $\mathbb{Q} \cup \{\infty\}$.
(3) The odd edges of P come in pairs, each pair meeting a vertex with inner angle $\frac{2\pi}{3}$.

The sides of the polygon are denoted as follows:

(1) Each odd edge is called an **odd side**.
(2) As even edges come in pairs, either each edge of the pair is an **even side**, or the union of the two edges (a semicircle) is called a **free side**.

The side pairing on the edges is defined as follows:

(1) Each odd side is paired with the odd side it meets at an angle of $\frac{2\pi}{3}$. This is called an **odd pairing**.
(2) Each even side is paired with the even side with which it forms a semicircular arc. This is called an **even pairing**.
(3) There are an even number of free sides and they are partitioned into sets of two, each called a **free pairing**.

We will always assume that 0 and ∞ are vertices of P.

The sides of a special polygon P have a natural orientation obtained by tracing the perimeter of the polygon in a certain direction. If $\{s, s'\}$

is a side pairing then there is a unique $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ such that γ maps s to s' in an orientation-reversing manner. We call this the **side pairing transformation** associated with the side pairing, and we let Γ_P be the group generated by all the side pairing transformations of P. Note that it doesn't matter which side we pick for s and which for s' because the two possible γ's are inverses of each other. Also note that if s is an even side (resp. odd side) then γ is order 2 (resp. order 3).

Two theorems of Kulkarni are fundamental here:

Theorem 3.1. *([6] Theorem 3.2) If P is a special polygon then P is a fundamental domain for Γ_P. Moreover, the side pairing transformations $\{\gamma_i\}$ are an independent set of generators of Γ_P (i.e. the only relations on the γ_i's are $\gamma_i^2 = 1$ or $\gamma_i^3 = 1$ for any finite-order γ_i's).*

Theorem 3.2. *([6] Theorem 3.3) For every $\Gamma \subset PSL_2(\mathbb{Z})$ of finite index, there is a special polygon P such that $\Gamma = \Gamma_P$.*

Proof. [6] and also follows from the proof of the algorithm in Section 4. □

Note that although it is true that any subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ with fundamental domain F is generated by the transformations that map its edges together, the fact that the set of generators of a special polygon is an independent set of generators is something special to the special polygon. For example, consider the fundamental domain shown in Figure 2. There are six sides, and the three side pairing transformations are $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, and $\begin{pmatrix} 3 & -2 \\ 2 & -1 \end{pmatrix}$ But this is not a independent list of generators because $\begin{pmatrix} 3 & -2 \\ 2 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. A special polygon for $\Gamma(2)$ is shown in Figure 3. The pairing transformations from the special polygon are $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 3 & -2 \\ 2 & -1 \end{pmatrix}$. These are independent generators of $\Gamma(2)$.

3.2. *Farey Symbols*

Recall that the classical Farey sequences F_n are constructed by taking all the rational numbers $0 \leq a/b \leq 1$ with denominator at most n and

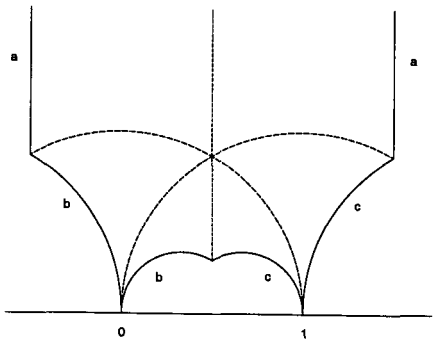


Fig. 2. A fundamental domain for $\Gamma(2)$

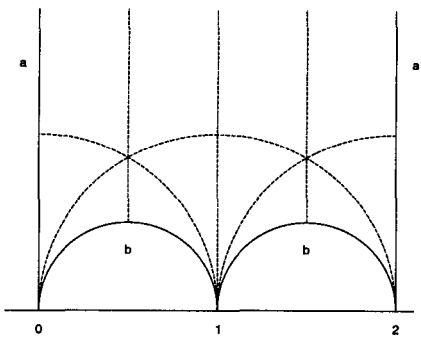


Fig. 3. A special polygon for $\Gamma(2)$

$(a, b) = 1$ and writing them as a finite sequence in ascending order $\{a_0/b_0, \ldots, a_n/b_n\}$. Then for each i we have $a_{i+1}b_i - a_ib_{i+1} = 1$. We are interested in sequences that satisfy this condition.

Definition 3.2. A **generalized Farey sequence** is a finite sequence:

$$\{\frac{-1}{0}, x_0, \ldots, x_n, \frac{1}{0}\}$$

such that:

(1) Each $x_i = a_i/b_i$ is a rational number in reduced form with $b_i > 0$. Additionally, we often consider $x_{-1} = \frac{-1}{0}$ and $x_{n+1} = \frac{1}{0}$.

(2) If we let $a_{-1} = -1$, $b_{-1} = 0$, $a_{n+1} = 1$, and $b_{n+1} = 0$ then

$$a_{i+1}b_i - a_ib_{i+1} = 1 \tag{1}$$

for $-1 \leq i \leq n$.

Note that this definition forces x_0 and x_n to be integers. We will always assume $x_i = 0$ for some i.

Definition 3.3. A **Farey symbol** is a generalized Farey sequence with some additional pairing information. Namely, between each adjacent entries x_{i-1} and x_i we assign a **pairing** p_i which is either a positive integer called a **free pairing** or the symbol "$\circ$" called an **even pairing** or "$\bullet$" called an **odd pairing**. Each integer that appears as a free pairing appears exactly twice in the pairing information.

So if P is a special polygon, let $x_0, \ldots, x_n$ be the vertices of P lying in $\mathbb{Q}$ listed in ascending order. Recall these vertices satisfy $a_{i+1}b_i - a_ib_{i+1} = 1$. Then $\{\frac{-1}{0}, x_0, \ldots, x_n, \frac{1}{0}\}$ is a generalized Farey sequence. We make a Farey symbol out of the generalized Farey sequence by adding the pairing information in the obvious way.

On the other hand, if F is a Farey symbol we can construct a special polygon for F. For adjacent entries of the Farey sequence x_{i-1} and x_i, if p_i is a free pairing or an even pairing we let P have as a side the hyperbolic arc joining x_{i-1} and x_i. Otherwise if it is odd we let γ be the unique element of $\mathrm{PSL}_2(\mathbb{Z})$ such that $\gamma(0) = x_{i-1}$ and $\gamma = x_i$ and join x_{i-1} and x_i by the arcs $\gamma(H_{0,\rho})$ and $\gamma(H_{\rho,\infty})$. Thus we get a hyperbolic polygon which is made into a special polygon by adding pairing information in the obvious way.

Example 3.1. $\Gamma(2)$ has a Farey symbol $-\infty \underset{1}{\smile} \frac{0}{1} \underset{2}{\smile} \frac{1}{1} \underset{2}{\smile} \frac{2}{1} \underset{1}{\smile} \infty$.

3.3. *Generators*

If P is a special polygon for a group Γ then Γ is independently generated by the transformations mapping each side to its paired side. If F is a Farey symbol:

$$-\infty \underset{p_0}{\smile} a_0/b_0 \underset{p_1}{\smile} a_1/b_1 \underset{p_2}{\smile} \cdots \underset{p_{n-1}}{\smile} a_{n-1}/b_{n-1} \underset{p_n}{\smile} a_n/b_n \underset{p_{n+1}}{\smile} \infty ,$$

then we can explicitly give formulas for the γ corresponding to a given side pairing.

Theorem 3.3. *Suppose $(a_i/b_i, a_{i+1}/b_{i+1})$ are two adjacent vertices of F. Then if the pairing between them p_{i+1} is an even pairing, let:*

$$G_{i+1} = \begin{pmatrix} a_{i+1}b_{i+1} + a_ib_i & -a_i^2 - a_{i+1}^2 \\ b_i^2 + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_ib_i \end{pmatrix}.$$

If p_{i+1} is an odd pairing, let:

$$G_{i+1} = \begin{pmatrix} a_{i+1}b_{i+1} + a_ib_{i+1} + a_ib_i & -a_i^2 - a_ia_{i+1} - a_{i+1}^2 \\ b_i^2 + b_ib_{i+1} + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_{i+1}b_i - a_ib_i \end{pmatrix}.$$

And if p_{i+1} is a free pairing that is paired with the side between a_k/b_k and a_{k+1}/b_{k+1}, let:

$$G_{i+1} = \begin{pmatrix} a_{k+1}b_{i+1} + a_kb_i & -a_ka_i - a_{k+1}a_{i+1} \\ b_kb_i + b_{k+1}b_{i+1} & -a_{i+1}b_{k+1} - a_ib_k \end{pmatrix}.$$

Then G_{i+1} is the side transformation corresponding to the pairing p_{i+1}.

Proof. [6] Theorem 6.1. □

3.4. *Group Invariants*

Several invariants of the group Γ can be read off from the Farey symbol F. Firstly, the number of inequivalent order-2 (resp. order-3) elliptic points, e_2 (resp. e_3), is the number of even (resp. odd) pairings in F. Also, the number of free pairings in F (half the number of free edges) is equal to r, the rank of $\pi_1(\Gamma\backslash\mathbb{H})$ (the fundamental group of the uncompactified modular curve).

To discuss the cusps of Γ, note that if (x_i, x_{i+1}) is an edge with an even or odd pairing, then x_i and x_{i+1} are equivalent cusps (since $G_{i+1} \in \Gamma$ maps x_i to x_{i+1}). Likewise, if (x_i, x_{i+1}) and (x_j, x_{j+1}) are paired edges then x_i and x_{j+1} are equivalent cusps and x_j and x_{i+1} are equivalent cusps. This defines an equivalence relation on the vertices of P. The equivalence classes are easy to compute, because the defining equivalences occur in a cyclic patten. So the number of cusps t can be counted as the number of equivalence classes.

For an edge $(\frac{a_i}{b_i}, \frac{a_{i+1}}{b_{i+1}})$ let $\gamma = \begin{pmatrix} a_i & a_{i+1} \\ b_i & b_{i+1} \end{pmatrix}$. So $\gamma^{-1}(x_i) = \infty$ and $\gamma^{-1}(x_{i+1}) = 0$. Then define the **width** of a vertex x_i to be the "width" of γP at ∞. That is:

$$\text{width}(x_i) = \begin{cases} |a_{i-1}b_{i+1} - a_{i+1}b_{i-1}| & \text{if } x_i \text{ is adjacent to no odd edge} \\ |a_{i-1}b_{i+1} - a_{i+1}b_{i-1}| + 1/2 & \text{if } x_i \text{ is adjacent to 1 odd edge} \\ |a_{i-1}b_{i+1} - a_{i+1}b_{i-1}| + 1 & \text{if } x_i \text{ is adjacent to 2 odd edges} \end{cases}.$$

The cusp width of a cusp x of Γ is then the sum of the widths of the vertices of P Γ-equivalent to x.

$\Gamma\backslash\mathbb{H}$ is a genus g orientable surface with t points missing, one for each cusp. The rank of its fundamental group is $r = 2g+t-1$, so we can calculate the genus $g = \frac{r-t+1}{2}$. Moreover, using the Hurwitz formula ([13], Prop. 1.40) we get the index of Γ in $\mathrm{PSL}_2(\mathbb{Z})$, $\mu = 3e_2 + 4e_3 + 12g + 6t - 12$. An even simpler formula for the index comes from noting that $n + 2 = 2r + e_2 + e_3$ where $n+1$ is as in Definition 3.2. This, combined with the previous formula, implies $\mu = 3n + e_3$.

4. Coset Permutation Representation of a Group

Another method of representing groups that will be useful to us in determining if a group is congruence is the coset permutation representation developed by Millington [10], [11]. Let Γ be a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ with $[\mathrm{PSL}_2(\mathbb{Z}) : \Gamma] = \mu$ and $\mathrm{PSL}_2(\mathbb{Z}) = \cup_{i=1}^{\mu} \alpha_i \Gamma$ a coset decomposition with $\alpha_1 = I$. Let F be the standard fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})$. Then $\cup_{i=1}^{\mu} \alpha_i^{-1} F$ is a fundamental domain for Γ. Let

$$E = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

E and V generate $\mathrm{PSL}_2(\mathbb{Z})$, as do L and R. The conversions between them are:

$$E = LR^{-1}L, \quad V = R^{-1}L \tag{2}$$

$$L = EV^{-1}, \quad R = EV^{-2} \tag{3}$$

We have $E^2 = V^3 = 1$. In fact it is well-known that $\mathrm{PSL}_2(\mathbb{Z})$ is isomorphic to the group [12]:

$$\mathrm{PSL}_2(\mathbb{Z}) \cong \langle e, v : e^2 = v^3 = 1 \rangle \tag{4}$$

For each γ in $\mathrm{PSL}_2(\mathbb{Z})$, left multiplication acts on the left cosets of Γ in $\mathrm{PSL}_2(\mathbb{Z})$ by permutation, i.e. there is a homomorphism $\phi : \mathrm{PSL}_2(\mathbb{Z}) \to S_n$, such that if $\phi(\gamma) = \sigma_\gamma$ then $\gamma\alpha_i\Gamma = \alpha_{\sigma_\gamma(i)}\Gamma$. In this way every finite-index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ is associated with a pair of permutations $e = \varphi(E)$ and $v = \varphi(V)$ with $e^2 = v^3 = 1$ which generate a transitive permutation group (transitivity comes from E and V generating $\mathrm{PSL}_2(\mathbb{Z})$). We call (e, v) a **coset permutation representation** of Γ and (l, r) an **LR-representation** of Γ, where $l = \varphi(L)$ and $r = \varphi(R)$. Each form can

be obtained from the other form by the equations (2) and (3). Note that $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ is in Γ if and only if $\gamma\Gamma = \Gamma$, i.e., $\sigma_\gamma(1) = 1$.

On the other hand, suppose e and v are a pair of permutations on μ letters with $e^2 = v^3 = 1$ that generate a transitive permutation group S (such a permutation we call **valid**). Define a homomorphism $\varphi : \mathrm{PSL}_2(\mathbb{Z}) \to S$ such that $\varphi(E) = e$ and $\varphi(V) = v$ (This is well-defined because of (4)). Let $\Gamma = \{\gamma \in \mathrm{PSL}_2(\mathbb{Z}) : \varphi(\gamma)(1) = 1\}$. Then Γ is an index-μ subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. Thus we have a correlation between valid pairs of permutations and finite-index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$. To test if $A \in \mathrm{PSL}_2(\mathbb{Z})$ is in Γ we write A as a word in L and R (Using, essentially, the Euclidean Algorithm) and replace L and R with the permutations l and r. If the resulting permutation fixes 1 then A is in Γ.

If one of the cosets is fixed by e, say $e(i) = i$, it corresponds to an elliptic element in Γ, for $E\alpha_i\Gamma = \alpha_i\Gamma$ means $\alpha_i^{-1}E\alpha_i\Gamma = \Gamma$, meaning $\alpha_i^{-1}E\alpha_i$ (which is order 2) is in Γ. So e_2, the number of inequivalent elliptic elements of order 2 in Γ, is equal to the number of elements fixed by e. Similarly, e_3 is the number of elements fixed by v.

The cusp width of Γ at ∞ is the smallest positive integer n such that $L^n \in \Gamma$. Thus the cusp width at infinity is the order of the cycle in $\varphi(L)$ which contains "1". Likewise, suppose i is in a cycle of length k in $\varphi(L)$, i.e. $L^k\alpha_i\Gamma = \alpha_i\Gamma$, but $L^n\alpha_i\Gamma \neq \alpha_i\Gamma$ for $0 < n < k$. Then $\alpha_i^{-1}L^k\alpha_i \in \Gamma$, but $\alpha_i^{-1}L^n\alpha_i \notin \Gamma$ for $0 < n < k$. If $q = \alpha_i^{-1}\infty$ then $\alpha_i^{-1}L^k\alpha_i q = q$ but $\alpha_i^{-1}L^n\alpha_i q \neq q$ for $0 < n < k$. Thus $\alpha_i^{-1}L^k\alpha_i$ is a generator for the stabilizer of the cusp q, and this cusp has width k.

5. Algorithms

5.1. *Calculating a Farey Symbol*

Recall that T is the standard fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})$, and let T^* be the hyperbolic triangle with vertices ρ, i and ∞ (So $T = T^* \cup (-\overline{T^*})$). $\mathscr{T}T = \{\gamma T : \gamma \in \mathrm{PSL}_2(\mathbb{Z})\}$ is a tessellation of the upper half plane and any finite index subgroup Γ has a fundamental domain which is a simply connected union of $\mathscr{T}$-tiles. Let $\mathscr{T}^* = \{\gamma T^* : \gamma \in \mathrm{PSL}_2(\mathbb{Z})\} \cup \{\gamma(-\overline{T^*}) : \gamma \in \mathrm{PSL}_2(\mathbb{Z})\}$. $\mathscr{T}^*$ is also a tessellation of the upper half plane, and we will construct a fundamental domain for Γ out of $\mathscr{T}^*$-tiles. The starting point for our construction will be the six tiles around an odd vertex. The following lemma shows this is a reasonable starting point:

Lemma 5.1. *Let Γ be a subgroup of $PSL_2(\mathbb{Z})$ with index ≥ 3. Then the stabilizer of $\rho = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ or $\rho - 1 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$ is trivial (i.e. one of these*

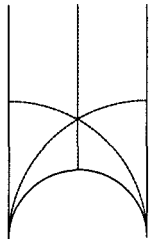


Fig. 4. A hyperbolic triangle

points is not elliptic in Γ*).*

Proof. If the two stabilizers are not trivial then they must be $\Gamma_\rho = \{I, A, A^2\}$ and $\Gamma_{\rho-1} = \{I, B, B^2\}$ where $A = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$. But A and B generate an index-2 subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. So Γ is either the (unique) index-2 subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ or $\mathrm{PSL}_2(\mathbb{Z})$ itself. And if the index of Γ in $\mathrm{PSL}_2(\mathbb{Z})$ is bigger than 2, at least one of A and B cannot be in Γ□

So if Γ is not $\mathrm{PSL}_2(\mathbb{Z})$ or Γ_2, the unique index 2 subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ ([12]), then the hyperbolic triangle with vertices either 0, 1 and ∞, or -1, 0 and ∞ is contained in a fundamental domain of Γ. The triangle is made of 6 $\mathscr{T}^*$-tiles (see Figure 4). We will make a polygon P starting with this triangle, then attach $\mathscr{T}^*$-tiles to P and assign partial pairing information to sides until we get a fundamental domain for Γ (at which point all the pairing information will be filled in). In the algorithm we will say a $\mathscr{T}^*$-tile T is **adjoinable** to P if T is adjacent to a tile of P and if $P \cup T$ is contained in some fundamental domain of Γ. Note that if T is adjacent to P with adjacency edge e and if e cannot be paired with any other edge of P then T is adjoinable.

Algorithm:

(1) If $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ let P be the special polygon with Farey symbol

$$-\infty \underset{\circ}{\smile} 0 \underset{\bullet}{\smile} \infty$$

or if $\Gamma = \Gamma_2$ let P be the special polygon with Farey symbol

$$-\infty \underset{\bullet}{\smile} 0 \underset{\bullet}{\smile} \infty\,.$$

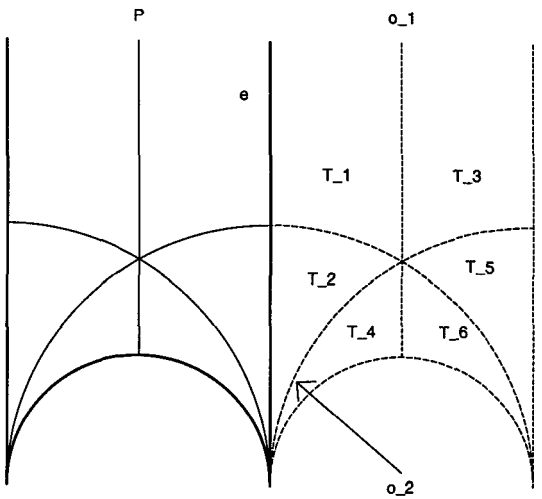


Fig. 5.

In either case return P and terminate.

(2) If $\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$ is not in Γ then let P be the hyperbolic polygon with vertices 0, 1, and ∞. Otherwise let P be the hyperbolic polygon with edges -1, 0 and ∞.

(3) If any of the three sides of P map to each other by a $\gamma \in \Gamma$, assign that pairing to the side. (Note that initially all sides are even sides).

(4) P is now a polygon where every side is either:
   (a) even and already paired.
   (b) odd and already paired.
   (c) even and unpaired.

(5) Pick an unpaired even side e. Figure 5 shows the typical case (The other cases are the same as this case with everything translated by some $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$). Since e is unpaired, T_1 and T_2 must be adjoinable. If o_1 and o_2 are the new odd edges of P after adding T_1 and T_2 to P then either $\gamma o_1 = o_2$ for some $\gamma \in \Gamma$, or there is no such γ. If there is γ pair the two edges and go to Step (3).

(6) If o_1 doesn't pair with o_2 then it doesn't pair with any other side because the only other unpaired sides are odd. So tiles T_3 and likewise T_4 are adjoinable. Each of these tiles has a free edge and the free edges cannot pair with each other (because their common vertex would have

an internal angle of $\frac{4\pi}{3}$, so the pairing transformations would make things overlap), so T_5 and T_6 are adjoinable.

(7) We've now added 6 $\mathscr{T}^*$-tiles to P (One even triangle). If either of the new even edges pair with any of the old unpaired even edges then assign that pairing.

(8) If all the sides of P are paired then we are done. Otherwise go to Step (4).

The output of the algorithm is a special polygon P with $\Gamma_P = \Gamma$. Note that the algorithm must terminate, because a fundamental domain of Γ has hyperbolic area $\frac{\pi}{3}[\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]$ and a single $\mathscr{T}^*$-tile has area $\frac{\pi}{6}$. So for P to be contained in a fundamental domain of Γ it can have at most $2 \cdot [\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]$ $\mathscr{T}^*$-tiles.

To effectively implement the algorithm we use Farey symbols. We need only a way to test for group membership. Note that if p_i/q_i and p_{i+1}/q_{i+1} are two adjacent vertices of the fundamental polygon then the hyperbolic triangle added to the edge $H_{x_i,x_{i+1}}$ in Step (6) is the triangle with vertices p_i/q_i, p_{i+1}/q_{i+1}, and $(p_i + p_{i+1})/(q_i + q_{i+1})$.

So given a finite-index subgroup Γ of $\mathrm{PSL}_2(\mathbb{Z})$, if we have a way to test for group membership we can calculate a Farey symbol by the following algorithm:

Algorithm for calculating a Farey Symbol:

(1) If $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ are in Γ then $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, so return

$$-\infty \underset{\circ}{\smile} 0 \underset{\bullet}{\smile} \infty$$

and terminate. If $\begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$ and $\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$ are in Γ then $\Gamma = \Gamma_2$, so return

$$-\infty \underset{\bullet}{\smile} 0 \underset{\bullet}{\smile} \infty$$

and terminate.

(2) If $\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \notin \Gamma$ then let F be the (partial) Farey symbol:

$$-\infty \smile \tfrac{0}{1} \smile \tfrac{1}{1} \smile \infty\,.$$

Otherwise let F be:

$$-\infty \overset{\frown}{} \frac{-1}{1} \overset{\frown}{} \frac{0}{1} \overset{\frown}{} \infty .$$

(3) For each i with $0 \leq i \leq n+1$, if the pairing between x_{i-1} and x_i is not filled in then check if it can be paired with itself (even or odd pairing), or if it can be paired with another unpaired edge (i.e., check if the appropriate G_i is in Γ). Wherever something can be paired, assign that pairing.
(4) If all edges are now paired, return F and terminate.
(5) If there is still an unpaired edge, say between p_i/q_i and p_{i+1}/q_{i+1}, make a new vertex $(p_i+p_{i+1})/(q_i+q_{i+1})$ with no pairing information on the edges adjacent to it. Go to Step (3).

The output is a Farey symbol for Γ.

5.2. *Group Membership*

The following algorithm described in [7] tests if $A \in \mathrm{PSL}_2(\mathbb{Z})$ is an element of the group corresponding to a Farey symbol F. We will need a lemma about even lines:

Lemma 5.2. *Let l be an even line (a semicircle on the upper half plane with rational endpoints a/b and a'/b' such that $|ab'-a'b|=1$). Let P be a special polygon in $\mathbb{H}$. Then either $l \subset P$ or $l \cap P = \emptyset$.*

Proof. [7] Proposition 2.1. □

Let Γ be a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ and A an element of $\mathrm{PSL}_2(\mathbb{Z})$.

$$A = \begin{pmatrix} c_0 & c_0' \\ d_0 & d_0' \end{pmatrix} \tag{5}$$

A maps the even line $H_{0,\infty}$ to $l = H_{c_0'/d_0', c_0/d_0}$. By the lemma, either $l \subset P$ or it is disjoint from P (except possibly at endpoints). If it is disjoint there is an edge which it is naturally "closest" to (In a sense discussed in [7]. The idea of the algorithm is to translate P across the "closest" edge until P intersects $H_{c/d,c'/d'}$, at which point A will be in Γ if and only if l is the image of $(0,\infty)$ or the an edge paired with $(0,\infty)$. In the actual algorithm we work in the other direction, translating the even line instead of the special polygon.

Algorithm: [7]

Let $k = 0$ and F be a Farey symbol for Γ with 0 as one of its vertices. Without loss of generality, we can assume $\frac{c_k'}{d_k'} < \frac{c_k}{d_k}$.

(1) There are two possibilities: If $\frac{c_k'}{d_k'}$ and $\frac{c_k}{d_k}$ are both vertices of P then terminate. Otherwise we must have $x_i \leq \frac{c_k'}{d_k'} < \frac{c_k}{d_k} \leq x_{i+1}$ with at least one "$\leq$" a strict inequality.

(2) Let g_{i+1} be the generator corresponding to the pairing p_{i+1} (recall this is the transformation mapping $l = H_{c_k'/d_k', c_k/d_k}$ to its paired side). If p_{i+1} is a free or even pairing, let $\alpha_k = g_{i+1}$. If p_{i+1} is an odd pairing, let $m = \frac{a_i + a_{i+1}}{b_i + b_{i+1}}$ where $x_i = \frac{a_i}{b_i}$, $x_{i+1} = \frac{b_{i+1}}{b_{i+1}}$. Then the interval $(\frac{c_k'}{d_k'}, \frac{c_k}{d_k})$ must be between either x_i and m or between m and x_{i+1}. If $\frac{c_k}{d_k} \leq m$, let $\alpha_k = g_{i+1}$. Otherwise let $\alpha_k = g_{i+1}^{-1}$.

(3) Let $\frac{c_{k+1}}{d_{k+1}} = \alpha_k \cdot \frac{c_k}{d_k}$, $\frac{c_{k+1}'}{d_{k+1}'} = \alpha_k \cdot \frac{c_k'}{d_k'}$. Replace k with $k+1$ and go to Step (1).

The algorithm returns $\frac{c_k'}{d_k'}$ and $\frac{c_k}{d_k}$, which are two vertices of P, and a list of α_i's.

Theorem 5.1. *The algorithm terminates, and A is in Γ if and only if one of the following is true:*

(1) $\begin{pmatrix} c_k & c_k' \\ d_k & d_k' \end{pmatrix} = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

(2) $(\frac{c_k'}{d_k'}, \frac{c_k}{d_k})$ *is a free side paired with* $(0, \infty)$.

(3) $\begin{pmatrix} c_k & c_k' \\ d_k & d_k' \end{pmatrix} = \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ *and* 0 *and* ∞ *are adjacent vertices with an even pairing between them.*

Proof. See [7]. □

In addition, if A is in Γ, A can be written as a word in the generators of Γ because $A = \alpha_0^{-1}\alpha_1^{-1}\ldots\alpha_k^{-1}\begin{pmatrix} c_k & c_k' \\ d_k & d_k' \end{pmatrix}$, and each term in that product is one of the generators for F.

5.3. *Coset Representatives*

Let Γ be a group with special polygon P. Let T be the hyperbolic triangle with vertices i, ρ, and ∞. By the construction of P, T is contained in P. The set of $\gamma \in \Gamma$ such that γT is in P is a set of coset representatives of Γ.

Let a_i/b_i and a_{i+1}/b_{i+1} be a vertex of the special polygon, and let $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\varphi = \begin{pmatrix} a_i & a_{i+1} \\ b_i & b_{i+1} \end{pmatrix}$. Then $\varphi^{-1}(a_i/b_i) = \infty$ and $\varphi^{-1}(a_{i+1}/b_{i+1}) = 0$. Let w_i be $|a_{i-1}b_{i+1} - a_{i+1}b_{i-1}|$ if the pairing between a_i/b_i and a_{i+1}/b_{i+1} is not an odd pairing and $|a_{i-1}b_{i+1} - a_{i+1}b_{i-1}| + 1$ if it is. Then w_i is the number of $\mathscr{T}^*$-tiles of the form γT in P. Thus a list of left coset representatives for Γ is $\bigcup_{i=0}^{n} \{T^{-j}\phi_i^{-1} : 0 \leq j < w_i\}$.

5.4. *Congruence Testing*

Let Γ be a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. Lang, Lim and Tan give a test purely in terms of Farey symbols to determine if Γ is a congruence group [7]. Their test relies on Wohlfahrt's Theorem [15] which says that if Γ has level N then Γ is a congruence group if and only if Γ contains $\Gamma(N)$. In Lang, Lim and Tan's test, if Γ has level N one computes a Farey symbol for $\Gamma(N)$, giving a complete set of generators for $\Gamma(N)$. One then checks if each of these generators is contained in Γ using the above algorithm. The difficulty with this algorithm is that the index of $\Gamma(N)$ increases very quickly with N, so if Γ has large level, the calculation of a Farey symbol for $\Gamma(N)$ can be very lengthy, even if Γ has relatively small index.

Another test for congruence was developed by Tim Hsu using Millington's coset permutation representations [3]. If we have an LR-representation of Γ there is a list of relations that are satisfied if and only Γ is congruence.

To calculate an LR-representation from a Farey symbol, use the above algorithm to calculate a list of left coset representatives $\alpha_i \in \mathrm{PSL}_2(\mathbb{Z})$ where $\mathrm{PSL}_2(\mathbb{Z}) = \bigcup_{i=1}^{\mu} \alpha_i \Gamma$. To calculate l, for instance, recall that l is the permutation such that $L\alpha_i\Gamma = \alpha_{l(i)}\Gamma$. So l sends i to the unique j such that $\alpha_j^{-1} L \alpha_i \in \Gamma$. So we run through every $1 \leq i \leq \mu$ and calculate the permutation. r can be calculated similarly. (Actually, although we need l and r it is easier to calculate e and v, because we know beforehand that they are order 2 and 3 respectively. Then $l = ev^{-1}$ and $r = ev^{-2}$).

Knowing l and r we can directly apply Tim Hsu's congruence algorithm [3]. Depending on the order of l, (i.e. the level of Γ) there are different lists of relations of l and r that are satisfied if and only if Γ is congruence. For example, if N is the order of l and N is odd then Γ is a congruence group if and only if $r^2 l^{-\frac{1}{2}}$ is the identity permutation (where $\frac{1}{2}$ is the inverse of 2 modulo N).

6. Implementation

Helena Verrill has a written a MAGMA package for working with Farey symbols for congruence groups. Also, for congruence or noncongruence

groups, the algorithms described above have been implemented by the first author as a collection of functions for SAGE. The package and basic examples may be downloaded at:

`http://www.public.iastate.edu/~kurthc/research/index.html`

References

1. A. O. L. Atkin and H. P. F. Swinnerton-Dyer, *Modular forms on noncongruence subgroups*, Combinatorics (Proc. Sympos. Pure Math., Vol. XIX, Univ. California, Los Angeles, Calif., 1968), Amer. Math. Soc., Providence, R.I. (1971), 1–25.
2. B. Birch, *Noncongruence subgroups, covers and drawings*, The Grothendieck theory of dessins d'enfants (Luminy, 1993), London Math. Soc. Lecture Note Ser., vol. 200, Cambridge Univ. Press, Cambridge (1994), 25–46.
3. T. Hsu, *dentifying congruence subgroups of the modular subgroup*, Proceedings of the American Mathematical Society, 124 (1996), No. 5, 1351–1359.
4. T. Hsu, *Permutation techniques for coset representations of modular subgroups*, Geometric Galois actions, 2, London Math. Soc. Lecture Note Ser., vol. 243, Cambridge Univ. Press, Cambridge (1997), 67–77.
5. N. Koblitz, *Introduction to elliptic curves and modular forms*, second ed., (Springer-Verlag, New York, 1993).
6. R. S. Kulkarni, *An arithmetic geometric method in the study of the subgroups of the modular group*, American Journal of Mathematics, 113 (1991), No. 6, 1053–1133.
7. M. L. Lang, Chong-Hai Lim, and Ser-Peow Tan, *An algorithm for determining if a subgroup of the modular group is congruence*, Journal of the London Mathematical Society, 51 (1995), 491–502.
8. M. L. Lang, *Normalisers of subgroups of the modular group*, J. Algebra, 248 (2002), No. 1, 202–218.
9. L. Long, *Finite index subgroups of the modular group and their modular forms*, arXiv:0707.3315 (2007).
10. M. H. Millington, *On cycloidal subgroups of the modular group*, Proc. London Math. Soc. (3), 19 (1969), 164–176.
11. M. H. Millington, *Subgroups of the classical modular group*, J. London Math. Soc. (2), 1, (1969), 351–357.
12. R. A. Rankin, *Modular forms and functions*, (Cambridge University Press, Cambridge, 1977).
13. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan, No. 11. Iwanami Shoten, Publishers, Tokyo (1971), Kanô Memorial Lectures, No. 1.
14. W. A. Stein, *Modular forms: A computational approach*, (American Mathematical Society 2007).
15. K. Wohlfahrt, *An extension of f. klein's level concept*, Ill. J. Math., 8 (1964), 529–539.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 243–255)

GRÖBNER—SHIRSHOV BASES AND NORMAL FORMS FOR THE COXETER GROUPS E_6 AND E_7*

DENIS LEE

Mechanics and Mathematics Department,
Novosibirsk State University, Novosibirsk, Russia
E-mail: LeeVD@yandex.ru

In this paper, Gröbner—Shirshov bases and normal forms of elements for the Coxeter groups E_6 and E_7 are found. These results support the conjecture in [4] about the general form of Gröbner—Shirshov bases for all Coxeter groups.

Keywords: Gröbner—Shirshov basis; Composition—Diamond Lemma; Normal form; Coxeter group.

1. Introduction

Coxeter groups are abstract groups of reflection symmetries of great importance in geometry, combinatorics and other areas of mathematics. They are usually defined by generators and defining relations. Let S be a totally ordered set of n elements. An $n \times n$ symmetric matrix $M = (m_{ij})$ is called a Coxeter matrix if $m_{ii} = 1$ and $m_{ij} \geq 2$ or ∞ for $i \neq j$. The following (semi)group presentation defines the corresponding Coxeter group: $W = smg\langle S|(s_is_j)^{m_{ij}} = 1, \text{ where } 1 \leq i,j \leq n, m_{ij} \neq \infty\rangle$.

This article deals with the finite Coxeter groups of types E_6 and E_7. The main problem is to obtain the Gröbner—Shirshov bases of the groups and a normal form of their elements.

In [4], Gröbner—Shirshov bases and normal form of the elements were already found for the Coxeter groups of types A_l, B_l, D_l and a general form of the Gröbner—Shirshov bases for all Coxeter groups was conjectured. All results presented here support the conjecture.

This article provides another application of the method of Gröbner—Shirshov bases for semigroups and groups, defined by generators and defining relations. Usually other methods were applied in these cases, partic-

*Mathematics Subject Classification(2000): 16S15, 20F05, 20F55.

ularly: Newman's lemma, the method of elementary reduction and so on. Though we consider the method of Gröbner—Shirshov bases more convenient in many cases. Various specializations of the method are the Euclidean algorithm, the Gaussian elimination algorithm, some results of Gröbner and Newman. Works of A. I. Shirshov [4], H. Hironaka [7] and B. Buchberger [5] considerably generalized this method.

To apply Gröbner—Shirshov method often requires a large amount of calculation. For this reason, computers are very helpful in some cases. A lot of software today is capable of performing such calculations.

In our case, all calculations were reduced to several series of simpler calculations and the main result was derived without the aid a computer. But we used a computer to make additional checks of auxiliary results derived during the work. We expect the main result of this work will be used for solving the same problem for the Coxeter group of type E_8.

2. Gröbner—Shirshov bases

Take a totally ordered set X, the free associative algebra $k\langle X\rangle$ over some field k, and the set X^* of words over X (including the empty word 1). We consider X^* totally ordered too. Also we assume that the order on X^* is monomial, i.e. if $u > v$ then $aub > avb$ for all words u, v, a, b. Denote by $\overline{f}$ the leading word of $f \in k\langle X\rangle$. The polynomial f is *monic* if the coefficient of $\overline{f}$ is equal to 1. We denote by $|u|$ the length of a word u.

For two monic polynomials f, g and a word w, we define their *composition* (see [3], [8]):

$$(f,g)_w = \begin{cases} f - agb, & \text{if } w = \overline{f} = a\overline{g}b, \\ fb - ag, & \text{if } w = \overline{f}fb = a\overline{g},\ |\overline{f}| + |\overline{g}| > |w|. \end{cases}$$

The word w is called the *ambiguity* of f and g. The first type of composition is called *the composition of including g in f*, and the second type of composition — *the composition of intersection of f and g.*

Take some set S of monic polynomials and $f, g \in S$. The composition $(f,g)_w$ is called *trivial relative to S* (more precisely, relative to S and w) if

$$(f,g)_w = \Sigma \alpha_i a_i s_i b_i, \text{ where } \alpha_i \in K, a_i, b_i \in X^*, a_i\overline{s_i}b_i < w,$$

we write this as $(f,g)_w \equiv 0 \pmod{S,w}$) (see [2], [3]).

The reduction $f \mapsto f - agb$, where $\overline{f} = a\overline{g}b$, is called the *elimination of the leading word of the polynomial g in f.* It is easy to verify if the elimination of the leading word of polynomials from S in $(f,g)_w$ reduces $(f,g)_w$ to zero, then $(f,g)_w$ is trivial.

The set S is called *Gröbner—Shirshov* bases if all compositions $(f,g)_w$ of all polynomials $f,g \in S$ are trivial. In the original articles [2] and [3], this set S is called a set closed with respect to compositions.

Lemma 2.1. *(Composition—Diamond Lemma) (see [2], [3], [8], [4]) A set S is a Gröbner—Shirshov basis in $k\langle X\rangle$ if and only if the set of S-reduced words*

$$Red(S) = \{u \in X^*; u \neq a\bar{s}b, s \in S, a, b \in X^*\}$$

is a linear basis of the algebra

$$k\langle X\rangle / Id(S) = \langle X|S\rangle.$$

If $S \subset k\langle X\rangle$ is not a Gröbner—Shirshov basis, then we can add to S all nontrivial compositions of polynomials from S. Having repeated this procedure (possibly infinitely many times) we obtain a Gröbner—Shirshov basis S^{comp}. The process is called *the Buchberger-Shirshov algorithm* (see [8], [5]).

If the set S consists of semigroup relations (i.e. $u - v$, where $u, v \in X^*$), then each nontrivial composition of polynomials from S has the same semigroup form. Hence, S^{comp} consists of semigroup relations too.

Consider that some semigroup $A = smg\langle X|S\rangle$. Then $S \subset k\langle X\rangle$ and we can obtain the Gröbner—Shirshov bases S^{comp}. The set S^{comp} does not depend on the field k and consists of semigroup relations. We will call S^{comp} the Gröbner—Shirshov bases for the semigroup A.

3. Coxeter groups of types E_6 and E_7

Take some Coxeter group W defined as above by a Coxeter matrix:

$W = smg\langle S|(s_is_j)^{m_{ij}} = 1,$ where $1 \leq i,j \leq n, m_{ij} \neq \infty\rangle$.

If $n = 6,7$ or 8, $m_{ii-1} = 3$ for $i \leq n-1$, $m_{nn-3} = 3$ and $m_{ij} = 2$ for $i - j \geq 2$, then the corresponding Coxeter groups are the Coxeter groups of types E_6, E_7 and E_8. In other words, the Coxeter groups of types E_6, E_7 and E_8 have the following defining relations with $l = 6,7,8$ respectively:

$s_i^2 = 1;$

$s_is_{i-1}s_i = s_{i-1}s_is_{i-1}, \quad i \leq l-1;$

$s_ls_{l-3}s_l = s_{l-3}s_ls_{l-3}.$

For $m_{ss'} < \infty$, let us introduce the following notation which we will use to formulate the conjecture mentioned in the introduction:

$m(s,s') = ss'\ldots$

(the word of length $m_{ss'}$, consisting of alternating letters s and s'),

$(m-1)(s,s') = ss'\ldots$
(the word of length $m_{ss'}-1$, consisting of alternating letters s and s').

For example, if $m_{ss'}=2$ then $m(s,s')=ss'$ and $(m-1)(s,s')=s$.

The defining relations of the Coxeter groups have the following form in this notation:

$$s^2=1, \quad m(s,s')=m(s',s), \quad s>s' \tag{1}$$

Let us call two words in S equivalent, if they are equal modulo the commutative relations involved in (1). Two relations $a=b$ and $c=d$ are equivalent if their right and left sides are equivalent correspondingly. Gröbner—Shirshov bases and normal form of elements of A_l, B_l, D_l were obtained in the work [4]. That article proposes the conjecture about a general form of the Gröbner—Shirshov bases of all Coxeter groups. Now we formulate the conjecture here:

Conjecture *(see [4]) Gröbner—Shirshov bases of W consists of (1) and relations equivalent to the following:*

$$\begin{aligned}&(m-1)(s,s')(m-1)(s_1,s_2)\ldots(m-1)(s_{2k-1},s_{2k})(m)(s_{2k+1},s_{2k+2})\\ &=(m)(s,s')(m-1)(s_1,s_2)\ldots(m-1)(s_{2k-1},s_{2k})(m-1)(s_{2k+1},s_{2k+2}),\end{aligned}$$

where $s>s'$, $s_1<s_2$, $\ldots$, $s_{2k-1}<s_{2k}$, $s_{2k+1}<s_{2k+2}$, all adjacent pairs (s,s'), (s_1,s_2), $\ldots(s_{2k-1},s_{2k})$, (s_{2k+1},s_{2k+2}) are distinct, and also we have:

$s_2=s'$ if $m_{ss'}$ is even,
$s_2=s$ if $m_{ss'}$ is odd,
$\ldots$
$s_{2k+2}=s_{2k}$ if $m_{s_{2k-1}s_{2k}}$ is even,
$s_{2k+2}=s_{2k-1}$ if $m_{s_{2k-1}s_{2k}}$ is odd.

The main idea for solving the problem stated above is as follows. First, we determine a set of relations that follow from the defining relations of the Coxeter group. Then we compute the number of all reduced words with respect to this set of relations. This number turns out equal to the number of elements (which is well-known) of the corresponding Coxeter group. By the Composition—Diamond Lemma this equality means that the set of relations obtained is a Gröbner—Shirshov basis. The reduced words describe the normal form of elements in the Coxeter group.

4. Gröbner—Shirshov basis of the Coxeter group E_6

For $l=6,7$ we introduce the following notation:

su = sjS(_3si_4...Sj, where $i <_ I - 3$;

$su\text{-}2 = si;$

$su\text{-}i = 1;$

s»j = SjSi_iSi_2...Sj, where $I - 1 \geq i \geq j;$

$s_{ii+1} = 1$, where $I - 1 _> i.$

Theorem 4.1. *The following relations hold in the Coxeter group* E_6*:*

(1) $s\} = 1.$

(2) $Si+ijSi+i = SiSi+ij, \; l _\leq j _\leq i < l - \backslash .$

(3) $snsi\text{-}2i = si\text{-}_2susi\text{-}2i+i, \; i \leq l - 2.$

(4) $SuSl\text{-}2jSl\text{-}lj = Sl\text{-}iSuSl\text{-}2jSl\text{-}ij+i, \; i < j <_ I - 1.$

(5) $SHSI = si\text{-}_3su, \; i < I - 2.$

(6) $SuSl\text{-}2jSlj = Sl\text{-}.3SliSl\text{-}2jSl_{j+}i, \; i < j \leq I - 2.$

(7) $SuSl\text{-}2jSl\text{-}lSlj = Sl\text{-}3SuSl\text{-}2jSl\text{-}iSlj+l, \; \% < j < I - 1.$

(8) $Su_1Sl\text{-}2jSl\text{-}il\text{-}2Su_2Sl\text{-}2j + l = Sl\text{-}3Su_1Sl\text{-}2j s_{l-1}l-2 s_{li_2} s_{l-2j+2},$
$h < j < I - 2, \; i_2 < j + 1, \; J + 1 < I - 1.$

(9) $Su_1Si\text{-}.2j_1Sl\text{-}ll\text{-}3Sli_2Sl\text{-}2j_2Sl\text{-}.ik_2Sli_3 =$
$s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1}l-3s_{li_2}s_{l-2j_2}s_{l-1k_2}s_{li_3+1},$
$h < ji < l\text{-}3, \; i2 < h < fo, \; h < h\text{-}$

Proof, $sf = 1$ is defining relation of the E_e. Since sy and sj_i are subject of relations *A*5, all relations which hold in *A*& must hold in *EQ* . The relation $s_{i+}ijS_{i+}i = SiSi+ij, \; i + 1 \leq I\text{-}1$ was proved in [4] that dealt with the group ***Ai.***

3.

$SliSl\text{-}2i$

$= s_l s_{l-3} \ldots s_i s_{l-2} s_{l-3} \ldots s_i$

$= s_l s_{l-3} \ldots s_{i+1} s_{l-2} s_{l-3} \ldots s_i s_{i+1} s_i$

$= s_l s_{l-3} \ldots s_{i+1} s_{l-2} s_{l-3} \ldots s_{i+1} s_i s_{i+1}$

$= s_l s_{l-3} \ldots s_{i+2} s_{l-2} s_{l-3} \ldots s_{i+1} s_{i+2} s_{i+1} s_i s_{i+1}$

$= s_l s_{l-3} \ldots s_{i+2} s_{l-2} s_{l-3} \ldots s_{i+2} s_{i+2} s_{i+2} s_i s_{i+1}$

$= \ldots$

$= s_l s_{l-3} s_{l-2} s_{l-3} \ldots s_{i+2} s_i s_{i+1}$

$= s_l s_{l-2} s_{l-3} s_{l-2} \ldots s_{i+2} s_i s_{i+1}$

$= s_{l-2} s_l s_{l-3} s_{l-2} \ldots s_{i+2} s_i s_{i+1}$

$= s_{l-2} s_l s_{l-3} \ldots s_i s_{l-2} s_{l-3} \ldots s_{i+1};$

4. We can prove it similarly to the previous relations;

5. $SuSi = SiSi\text{-}_3 \ldots SiSl = SlS_{l-3} s_l s_{l-4} \ldots s_i = s_{l-3} s_{li};$

6.

$$
\begin{aligned}
&s_{li}s_{l-2j}s_{lj}\\
&= s_{li}s_{l-2}s_{l-3}s_{l}s_{l-4}s_{l-3}s_{l-4}s_{l-5j}s_{l-5j}\\
&= SUS1\text{-}2S1\text{-}3S1S1\text{-}3S1\text{-}4SI\text{-}3SI\text{-}.SJSI\text{-}5J\\
&= \quad SuSl\text{-}2SlSl\text{-}3SlSl\text{-}4Sl\text{-}3Sl\text{-}5jSl\text{-}4j_{+}1\\
&= \quad S1\text{-}3SUS1\text{-}2SI\text{-}3S1SI\text{-}4SI\text{-}3SI\text{-}.5JSI\text{-}4J+I
\end{aligned}
$$

7. Use relation 6:

$$
\begin{aligned}
&s_{li}s_{l-2j}s_{l-1}s_{lj}\\
&= s_{li}s_{l-2j}s_{lj}s_{l-1}\\
&= s_{l-3}s_{li}s_{l-2j}s_{lj+1}s_{l-1}\\
&= s_{l-3}s_{li}s_{l-2j}s_{l-1}s_{lj+1};
\end{aligned}
$$

8. Use relation 7:

$$
\begin{aligned}
&SJ\text{»!}\ S\{\text{-}2jSl\text{-}\ ll\text{-}2Sli_2\ Sl\text{-}2j+1\\
&= s_{li_1}s_{l-2j}s_{l-1}s_{l-2}s_{l}s_{l-3i_2}s_{l-2}s_{l-3j+1}\\
&= s_{li_1}s_{l-2j}s_{l-1}s_{l}s_{l-2}s_{l-3}s_{l-2}s_{l-4i_2}s_{l-3j+1}\\
&= s_{li_1}s_{l-2j}s_{l-1}s_{l}s_{l-3}s_{l-2}s_{l-3}s_{l-4i_2}s_{l-3j+1}\\
&= \ldots\\
&= s_{li_1}s_{l-2j}s_{l-1}s_{lj}s_{l-2j},
\end{aligned}
$$

here we can use relation 7:

$$
\begin{aligned}
&s_{li_1}s_{l-2j}s_{l-1l-2}s_{li_2}s_{l-2j+1}\\
&= s_{l-3}s_{li_1}s_{l-2j}s_{l-1}s_{lj+1}s_{l-2j}\\
&= s_{l-3}s_{li_1}s_{l-2j}s_{l-1}s_{l-2}s_{lj+1}s_{l-2j+2}s_{j}\\
&= s_{l-3}s_{li_1}s_{l-2j}s_{l-1l-2}s_{lj}s_{l-2j+2}.
\end{aligned}
$$

9. This relation can be easily proved similarly to relation 8.

a

Theorem 4.2. *The algorithm of elimination of the leading words of relations in Theorem 4-1 reduces each word in Coxeter group Ee to one of the following form:*

(1) $Si_{mi}S2m_{:l}\text{---}Sl\text{-}lmi\text{-}_1SliSl\text{-}2j, \quad I' < j \leq I - 1,\ 1 \leq m_n \leq U + 1;$

(%) $Si_{mi}S2m2\text{---}Sl\text{-}lm_{l-1}SliSl\text{-}2jSl\text{-}ik, \quad i < j \leq k \leq l - l,\ 1 < m_n \leq n + l;$

(3) $Si_{mi}S2m2\text{-}Sl\text{-}lm,_{}_1SU_1Sl\text{-}2j_1Sl\text{-}lk_1SH_2Sl\text{-}2j_2Sl\text{-}ik_2,$

$\dot{\mathrm{n}} < h < h, *2 < h < h, h < h, \mathrm{fci} > I - 2, 1 \leq \mathrm{m_n} \leq n + 1;$

(4) $Slm_1S2m2\text{-}Sl\text{-}lm_{l-1}SU_1Sl\text{-}2j_1Sl\text{-}ll\text{-}2Su_2Si\text{-}2J2^{S}l\text{-}lk_2,$

$h<ji<l\sim 2,\ i_2 < 32 \leq k_2,\ h + 1 < \mathrm{J2},\ 1 < rn_n \leq n + 1;$

(5) $Si_{mi}S2mi\text{--}Sl\text{-}l_{mi_1}Sli_1Si_2j_1Sl\text{-}ll\text{-}3Su_2Sl\text{-}2j_2Sl\text{-}lk_2SH_{31}$

$h<ji<l\text{-}3, i_2< \boldsymbol{h} <k_2<l,\ 32 \bar{<} 13,\ 1 < \bar{rm}_n < \bar{n} + \mathbf{1};$

Proof. These forms of words can be easily obtained using the algorithm

of elimination of leading words. In addition the result was verified on a computer. □

Theorem 4.3. *The Gröbner—Shirshov basis of the Coxeter group E_6 consists of the relation in Theorem 4.1. A normal form of elements in E_6 is described in Theorem 4.2.*

Proof. Note that each reduced word u can be presented as a product of two words: $u = v \cdot w$, where v is a maximal subword which does not contain the letter s_6, and w is a maximal subword containing the letter s_6. Moreover, the word v runs over all words which are reduced relative to the Gröbner—Shirshov basis of A_5. The number of such words is equal to (5+1)! (see [4]). Thus, in order to calculate the number of all reduced words we have to calculate the number of words w. It is easy to see that in the series of words from Theorem 4.2 the numbers of possibilities for w are: 11, 10, 12, 18, 21. Thus we see that the number of all reduced words is equal to $(11 + 10 + 12 + 18 + 21) \cdot (5 + 1)! = 51840$, which is equal to the order of the group E_6. This means that the series of words in Theorem 4.2 define the normal form of elements in E_6. By the Composition—Diamond Lemma relations in Theorem 4.1 comprise a Gröbner—Shirshov basis of the Coxeter group E_6. □

5. Gröbner—Shirshov basis of the Coxeter group E_7

We can handle the Coxeter group E_7 almost in the same way as E_6. However, this case is much more complicate than the case of E_6. The use of a computer reduces the time of calculations.

Theorem 5.1. *The following relations are valid in the Coxeter group E_7:*

(1) $s_i^2 = 1$.

(2) $s_{li}s_{l-2i} = s_{l-2}s_{li}s_{l-2i+1}$, $i \leq l-2$.

(3) $s_{li}s_{l-2j}s_{l-1j} = s_{l-1}s_{li}s_{l-2j}s_{l-1j+1}$, $i < j \leq l-1$.

(4) $s_{li}s_l = s_{l-3}s_{li}$, $i < l-2$.

(5) $s_{li}s_{l-2j}s_{lj} = s_{l-3}s_{li}s_{l-2j}s_{lj+1}$, $i < j \leq l-2$.

(6) $s_{li}s_{l-2j}s_{l-1}s_{lj} = s_{l-3}s_{li}s_{l-2j}s_{l-1}s_{lj+1}$, $i < j < l-1$.

(7) $s_{li_1}s_{l-2j}s_{l-1l-2}s_{li_2}s_{l-2j+1} = s_{l-3}s_{li_1}s_{l-2j}s_{l-1l-2}s_{li_2}s_{l-2j+2}$,
$i_1 < j < l-2$, $i_2 < j+1$, $j+1 < l-1$.

(8) $s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1k}s_{li_3}$
$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1k}s_{li_3+1}$,
$i_1 < j_1 < l-3$, $i_2 < j_2 < k$, $k \in \{l, l-1\}$, $i_3 = j_1 + 2 < l-1$.

(9) $S_{i_1}Sl\text{-}2j_1Sl\text{-}ll\text{-}ZSu_2Sl\text{-}2J2^{S}l\text{-}1k s_{li_3} s_{l-2j_3}$

$\approx s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1k} s_{li_3} s_{l-2j_3+1}$,

$ii < ji < I - 3$, $i_2 < J2 < k$, $k \pounds \{I - 2,1 - 3\}$, $i_3 < J3$, $h < J3$, $h = h + 3$,

the left side of each relation with $jz = I - 1$ *ends not in* $sjj_3sj_2j_3+i$ *but rather in* SH_3+I-

(10) $Sli_1Sl\text{-}2nSl\text{-}il\text{-}2Su_2Sl\text{-}2j_2Sl\text{-}_1l\text{-}iSu_3Sl\text{-}2j_3$

$= Sj\text{-}sSjii\ SJ\text{-}2J!\ S(_11_3SH_2{}^{s}l\text{-}2j_2\ Sl\text{-}li\text{-}4SH_3\ S(\text{-}2j_3+l,$

$i\backslash < jl < l\text{-}\ 3$, $12 < \dot{3}2 < 1\text{-}\dot{}\ 4$, $l_3 < \dot{J}3$, $J\dot{2} < J3$, $h = 3\dot{l} + 2$,

the left side of each relation with $j_3 = I - 1$ *ends not in* $su_isi\text{-}2j_3+i$, *but rather in* s_{li_3+1}

(11) $Sii_1Si\text{-}2j_1 s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{li_3} s_{l-2}$

$\approx s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{li_3}$,

$h < ji < I - 4$, $*2 < J2 < h < I - 2$.

(12) $Sii_lSl\text{-}2j_lSl\text{-}ll\text{-}4Su_2S_{l-2j_2} s_{l-1} s_{li_3} s_{l-2j_3} s_{l-1}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1} s_{li_3} s_{l-2j_3}$,

$h < ji < l\text{-}\ 4$, $h < h < h < I - 1$, $h < h < I - 1$-

(13) $SH_1S(_2j_1S(_K_4S(i_2Sj_2j_2Sj_i(_2Sji_3S(_2j_3Sj\text{-}l$

$= Sl\text{-}3Sli_1Sl\text{-}2jiSl\text{-}U\text{-}4Sli_2Sl\text{-}2J2^{S}l\text{-}ll\text{-}2SU_3Sl\text{-}2j_3,$

$h < j\ i < l - 4$, $i_2 < J2 < l\sim\ 2$, $i_3 < J3 < I - 1$, $J3 - J2 \geq 2$.

(14) $S(_{il}Sj_2j_1S(_iJ_4S(i_2S(_2j_2S(_i(_3S;i_3S(_2j_3Sj\text{-}l$

$= Si\text{-}3Su_lSl\text{-}2j_1Sl\text{-}U\text{-}4Sli_2Si_2j_2Sl\text{-}U\text{-}3Sli_3Si\text{-}2j_3,$

$k < ji < l\text{-}\ 4$, $i2 < J2 < l\sim\ 3$, $i_3 < j_3 < I - 1$.

(15) $s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-4} s_{li_3} s_{l-2j_3} s_{l-1l-2}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-4} s_{li_3} s_{l-2j_3} s_{l-1}$,

$H < j\backslash < I \sim 4$, $i_2 < J2 < l\sim\ 4$, $i_3 < j_3 < I - 2$.

Proof. Relations 1-8 are analogous to the relations in Theorem 4.1 and they can be proved as in Theorem 4.1. So we will prove here relations 9-15 only.

9. We prove here the case $k = 1 - 2$, the case $k = 1 - 3$— is similar. Use relation 8:

$$s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_{l-2j_3}$$
$$= s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1} s_l s_{l-2} s_{l-3i_3} s_{l-2j_3}$$
$$= s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1} s_l s_{l-2} s_{l-3} s_{l-2} s_{l-4i_3} s_{l-3j_3}$$
$$= s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1} s_{ll-3} s_{l-2i_3} s_{l-3j_3}$$
$$= \ldots$$
$$= s_{li_1} s_{l-2j_1} s_{l-1l-3} s_{li_2} s_{l-2j_2} s_{l-1} s_{lj_3-1} s_{l-2i_3},$$

use relation 8:

$$s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1}s_{ll-3}s_{l-2i_3}s_{l-3j_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1}s_{lj_3}s_{l-2i_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3}s_{l-2j_3+1};$$

10. Use relation 9:

$$s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-4}s_{li_3}s_{l-2j_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{l}s_{l-4}s_{l-3}s_{l-4i_3}s_{l-2j_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{ll-4}s_{l-3}s_{l-5i_3}s_{l-2j_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{ll-4}s_{l-5i_3}s_{l-3}s_{l-2j_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{li_3}s_{l-2j_3}s_{l-2},$$

use relation 9:

$$s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-4}s_{li_3}s_{l-2j_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{li_3}s_{l-2j_3+1}s_{l-2}$$
$$= S/\text{-}3SH!\ SI\text{-}2J!\ Sl\text{-}U\text{-}_3Su_2\ Sl\text{-}_2j_2\ Si\text{-}\ U\text{-}4Sli_3\ S(\text{-}2j_3+1\ 5$$

11. Use relation 8:

$$s_{lh}\ Sl\text{-}2j_1\ Si\text{-}\ lJ\text{-}4\text{«}Jia\ Si\text{-}2j_2\ S(i_3\ \text{«}i\text{-}2$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{l-4}s_{li_2}s_{l-2j_2}s_{li_3}s_{l-2}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-3}s_{l-2j_2}s_{li_3}\ S/\text{-}2$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-2}s_{li_3}s_{l-2}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l}s_{l-2}s_{l-3}s_{l-2}s_{l-4i_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{ll-3}s_{l-2i_3},$$

because of ii $< j\backslash <$ I — 4 and Z = 7, then ji = Z — 5; i.e., we can use relation 8 here:

$$Sih\ Sl\text{-}2j!\ Sl\text{-}U\text{-}4Su_2\ S;_2j_2\ {}^{s}li_3\ {}^{s}l\text{-}2$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l}s_{l-2i_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-2}s_{li_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3};$$

12. Use relation 11:

$$s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{li_3}s_{l-2j_3}s_{l-1}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3}s_{l-1}s_{l-2j_3}s_{l-1}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3}s_{l-1}s_{l-2}s_{l-1}s_{l-3j_3}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3}s_{l-2}s_{l-1}s_{l-2j_3},$$

use relation 11:

$$s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{li_3}s_{l-2j_3}s_{l-1}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3}s_{l-1}s_{l-2j_3}$$
$$= s_{l-3}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{li_3}s_{l-2j_3};$$

13. Use relation 12:

$$s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3}s_{l-2j_3}s_{l-1}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{l}s_{l-2}s_{l-3}s_{l-2}s_{l-4i_3}s_{l-3j_3}s_{l-1}$$
$$= s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{ll-3}s_{l-2}s_{l-3}s_{l-4i_3}s_{l-3j_3}s_{l-1}$$

$= \ldots$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1} s_{lj_3-1} s_{l-2i_3} s_{l-1}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1} s_{lj_3-1} s_{l-2j_3} s_{l-1} s_{j_3-1,i_3}$,

use relation 12:

Sjii *Sl-2ji Sl-ll-iSliz SI-2J2 SI-II-2SH3 Sl-2J3*^{S}*l-l*

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1} s_{lj_3-1} s_{l-2j_3} s_{j_3-1,i_3}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1} s_{lj_3-1} s_{l-2i_3}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_{l-2j_3}$;

14. Use relation 13:

$s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3} s_{l-1}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{l-3} s_{li_3} s_{l-2j_3} s_{l-1}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_l s_{l-2j_3} s_{l-1}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_{l-2} s_{l-1} s_{lj_3}$,

use relation 13:

$s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3} s_{l-1}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_{l-2} s_{lj_3}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-2} s_{li_3} s_l s_{l-2j_3}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3}$;

15. Use relation 14:

$s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-4} s_{li_3} s_{l-2j_3} s_{l-1l-2}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_l s_{l-4} s_{l-3i_3} s_{l-2j_3} s_{l-1l-2}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-3} s_{l-2j_3} s_{l-1l-2}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3} s_{l-2} s_{l-1l-2}$

$= s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3} s_{l-1} s_{l-2} s_{l-1}$,

use relation 14:

$s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-4} s_{li_3} s_{l-2j_3} s_{l-1l-2}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-2j_3} s_{l-2} s_{l-1}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1l-4} s_{li_2} s_{l-2j_2} s_{l-1l-3} s_{li_3} s_{l-3} s_{l-2j_3} s_{l-1}$

$= s_{l-3} s_{li_1} s_{l-2j_1} s_{l-1}$ *l-4Sn*$_2$ *Sl-2j*$_2$ S{_ii_4S(j$_3$ S(_$_2$j$_3$S|_l.

D

Theorem 5.2. T/ie *method of elimination of the leading words for relations in Theorem 5.1 reduces each word in Coxeter group Ej to one of the following form:*

(1) *Sl*$_{mi}$*S2m*$_2$*---Sl-lmi-*$_1$*SliSl-2jSl-lk, i < j < k <l,* _1 < _m$_n$ <_71+ 1;

(2) $s_{1m_1} s_{2m_2} \ldots s_{l-1m_{l-1}} s_{li_1} s_{l-2j_1} s_{l-1k_1} s_{li_2} s_{l-2j_2} s_{l-1k_2}$,

$h < ji < h,i_2 < J2 < h,ji < i2,h$ 6 *{1,1-1},1* $\leq m_n \leq$ n + l;

(3) Si$_{mi}$S$_{2m2}$...Si$_{_lmi_l}$SH$_1$Sj_2j$_1$Si_ii_2Sij$_2$Si_2j$_2$S(_u$_2$,

$h < h < h,i2 < h < ^2, ji + 1 <$ J2,1 $\leq m_n \leq n$ 4-1;

$h < \text{ii} < \text{fci},22 < 32 < k_2,ki\ e\ \{/-3,/-4\}, \underline{l} < m_n \leq \text{n} + \text{l};$

(5) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1k_1}s_{li_2}s_{l-2j_2}s_{l-1k_2}s_{li_3},$
$\text{»i} < \text{ji} < h,i_2 < 32 < k_2,ki,k_2 \in \{1,1-1\},\ j_2 < i_3, \mathbf{jl} < i_2,$
$l \leq \underline{m}_n < n \pm l;$

(6) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1k_1}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3},$
$i_1 < j_1 < k_1, i_2 < j_2 < l-2, k_1 \in \{1,1-1\}, j_{2+1} < j_3, j_1 < i_2,$
$1 \leq \text{Win} \leq n+1;$

(7) $^{s}lmiS2m2\sim\text{-}Sl\text{-}lm_{l-1}Sli_1Sl\text{-}2j_1Sl\text{-}il\text{-}2Sli_2Sl\text{-}_2j_2Sl\text{-}ik_2S_{H3},$
$h < h < I-2,i2 < h < k_2,k_2\ \text{e}\ \{/,/-1\},\ j_2 < i_3,j_x + 1 < j_2,$
$1 \leq \mathbf{m}_n \leq \mathbf{m} + 1;$

(8) $Si_{mi}S2m_2\text{-}Sl\text{-}lm_{l-1}s_{li_1}s_{l-2j_1}s_{l-1l-2}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3},$
$h < h < l\sim 2, i_2 < j_2 < l-2,\ j_2+1 < j_3, j_1+1 < j_2,\ 1 \leq m_n \leq n+1;$

(9) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1k_2}s_{li_3},$
$h < jl < I-3, l_2 < j_2 < k_2, k_2 \in \{\text{J,i}-1\},\ j_2 < i_3, j_1+3 < i_3,$
$1 \leq m_n \leq n+1;$

(10) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3},$
$i_1 < j_1 < l-3, i_2 < j2 < I-2, ji+4 < j_3, 1 \leq m_n \leq n+1;$

(11) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{li_3},$
$\text{U} < \text{ji} < /-3, i2 < 32 < l-3,\ \text{ii}+4 < \text{j3},\ 1 \leq m_n \leq n+1;$

(12) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-3}s_{li_2}s_{l-2j_2}s_{l-1l-4}s_{li_3}s_{l-2j_3}s_{l-1k_3},$
$h < 3i < \text{i}-3, \text{»2} < \text{J2} < \text{Z}-4, i_3 < 33 < /C3\ \text{ji}+4 < j_3, l < m_n \leq \text{n}+1;$

(13) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{li_3},$
$i_1 < \text{j!} < \text{J}-3, \text{i}_2 < 32 < I-1\text{»}\ \text{Ji} < \text{^3}, 1 \leq rn_n \leq n+1;$

(14) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{li_3}s_{l-2j_3},$
$i_1 < \text{ji} < l\text{-}4, i_2 < 32 < \text{Z}-1, \text{i}_3 < \text{J3} < 2\text{-}1\ \text{ji} < \text{*3},\ 1 \leq \text{wi}_n \leq \text{n}+1;$

(15) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3}s_{l-2j_3},$
$\text{»i} < \text{Ji} < * \sim 4, \text{«}_2 < \text{J2} < \text{i}-2, \text{e}_3 < J3 < I-1\ J3 < 32 + 3,$
$1 \leq m_n \leq n+1;$

fltf,) $\text{Si}_{\text{mi}}\text{S2m}_2\text{...SJ_l}_{\text{m!}_1}\text{SH}_1\text{S(-2j}_1\text{Si-n-4Sii}_2\text{Si-2j}_2\text{S!-li-3Sli}_3{}^{\text{s}}\text{(-2j3'}$
$\text{ii} < \text{ji} < \text{i}-4, \text{i}_2 < \text{J2} < 2-3, \text{i}_3 < \text{J3} < /-1, 1 \leq \text{m}_\text{n} \leq \text{n}+1;$

^7^ $\text{5l}_{\text{mi}}\text{52m}_2\text{—•S/-lm,_iSKiS/-2j}_1\text{S/-l/-4S«}_2\text{S;_2i7}_2\text{S/-li-4S/t}_3\text{Sj-2j3}^{\text{S}}\text{J-lfe3.}$
$\text{ii} < \text{ji} < \text{J}-4, \text{i}_2 < \text{J2} < '-3,13 < \text{J3} < \text{J}-1, 1 \leq \text{m}_\text{n} \leq \text{n}+1;$

$s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1}s_{li_3}s_{l-2j_3}s_{li_4},$
$i_1 < j_1 < l-4, i_2 < j_2 < l-3, i_3 < j_3 < l-1, j_3 < i_4, j_2 < i_3,$
$1 \leq m_n \leq n+1;$

(19) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-2}s_{li_3}s_{l-2j_3}s_{li_4},$
$i_1 < j_1 < l-4, i_2 < j_2 < l-3, i_3 < j_3 < l-1, j_3 < i_4, j_2+1 < j_3,$
$1 \leq \text{m}_\text{n} \leq \text{n}+1;$

(20) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-3}s_{li_3}s_{l-2j_3}s_{li_4}$,
$h < h < \text{i-4,i2} < J2 < /\text{-}3, i_3 < j_3 < l\text{-}l, J3 < U,\ 1 \le m_n \le n+1;$

(21) $\mathrm{Sl_{mi}S2m_2\text{-}S(\text{-}lm,_iSlnS(\text{-}2j_1S(_li_4SH_2Si\text{-}2j_2S(_li\text{-}4Sli_3{}^{S}(\text{-}2J3^{S}H_4}$,
$H < h < 1\text{-}4, i2 < J2 < /\text{-}4, i_3 < j_3 < /\text{--}1, J3 < *4,\ 1 \le m_n \le n+1;$

(22) $s_{1m_1}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-4}s_{li_3}s_{l-2j_3}s_{l-1}s_{li_4}s_{lj_4}$,
$h < h < I - 4, 2 < to < {}^{l} - 4, 3 < h < I - 1.J3 < *4 < ji,$
$1 \le m_n \le n + 1;$

(23) $s_{1m_1}s_{2m_2}\cdots s_{l-1m_{l-1}}s_{li_1}s_{l-2j_1}s_{l-1l-4}s_{li_2}s_{l-2j_2}s_{l-1l-4}s_{li_3}s_{l-2l-5}s_{l-1}$
SU-4S1-21-3S1 (the longest word in E7).

Proof. As in Theorem 4.2, we use a computer, it considerably sped up the calculations of the reduced words. •

Theorem 5.3. *The Grobner—Shirshov bases of the Coxeter group* E7 *consists of the relations of Theorem 5.1. A normal form of elements in* E7 is *described in Theorem 5.2.*

Proof. The proof is the same as the proof of Theorem 5.3. But in this case we have (6+1)! possibilities for the word *v* and 576 possibilities for the word *w.* Thus we obtain that the number or reduced words relative to the relations in Theorem 5.1 is equal to the order of the group *E7:* 576 . (6 + l)! = 2903040 =I E_7 I . This proves the theorem. 0

Acknowledgement: The author expresses his deepest gratitude to Professor L. A. Bokut and Dr P. Kolesnikov for their kind guidance and useful discussions.

References

1. G. M. Bergman, *The diamond lemma for ring theory,* Adv. in Math. 29 (1978), 178-218.
2. L. A. Bokut, *Unsolvability of the word problem and subalgebras of finitely presented Lie algebras,* Izv. Akad.Nauk SSSR Ser. Mat. 36 (1972), 1173-1219 (in Russian).
3. L. A. Bokut, *Imbeddings into simple Lie algebras,* Algebra i Logika 15 (1976), 117-142.
4. L. A. Bokut, Shiao L.-S., *Grobner—Shirshov Bases for Coxeter Groups,* Communications in Algebra, 29(9) (2001), 4305-4319.
5. B. Buchberger, *An algorithm for finding a bases for the residue class ring of a zero-dimensional polynomial ideal,* (German). Ph. D. Thesis, University of Insbruck, Austria (1965).

6. W. Gröbner, (1939). *Über die algebraischen Eigenschaften der Integrale von linearen Differentialhleichungen mit konstanten Koeffizitienten*, Monatsh. der Math. 47, 247-284.
7. H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero, I, II* Ann. of Math. 79 (1964), 109-203; ibid. 79 (1964), 205-326.
8. A. I. Shirshov, *Some algorithmic problems for Lie algebras*, Sibirsk. Mat. Zh. 3 (1962), 292-296 (in Russian); translation appears in ACM SIGSAM Bull. 33 (1999), No. 2, 3-6.

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 257–273)

ON OVERGROUPS IN GL(n, F) OVER A SUBFIELD OF F

SHANGZHI LI

School of Science, Beihang University,
Beijing, 100083, China
E-mail: lisz@buaa.edu.cn

Let F be a field and K a subfield of F. We classify the overgroups in $GL(n, F)$ of an $Sp(n, K, f_K)$, $SU(n, K, f_K)$ or $\Omega(n, K, Q_K)$ provided that the index $[F : K]$ is not much bigger than the Witt index $\nu(f_K)$ or $\nu(Q_K)$.

Keywords: Classical groups; Maximal subgroups; Subfields; Overgroups.

1. Introduction

In classifying the maximal subgroups of finite classical groups, it is important to investigate the maximality of the subgroups in the classes $C_1 \sim C_8$ defined in Aschbacher in 1984. This research has been done by P. Kleidman-M. Liebeck [3]. The analysis depends heavily on the classification of finite simple groups. The author has done research on this topic as well; however, the results obtained are for the classical groups over arbitrary fields (not necessarily finite) or sometimes over arbitrary division rings. Our method is based on matrix techniques. In this paper, we concentrate mainly the treatment of Aschbacher class C_5. For a classical group G acting on an n-dimensional space V over a field F, the class C_5, by definition, is the collection of the normalizers of the classical groups acting on the n-dimensional spaces V_K over maximal subfields K of F so that $V = F \otimes_K V_K$.

2. Notations

We first denote by $Mat_{m\times n}R$ the set of all the $m \times n$ matrices $(a_{ij})_{m\times n}$ with entries a_{ij} in a ring R. We also write $Mat_m R$ instead of $Mat_{m\times m}R$. By $A^{(m\times n)}$ (resp.$A^{(m)}$), we mean $A \in Mat_{m\times n}R$ (resp. $A \in Mat_m R$) and we use A' to denote the transpose of A.

A matrix $A \in Mat_{m\times n}R$ can be identified with the R-linear mapping $\mathcal{A} : Mat_{1\times m}R \to Mat_{1\times n}R, x \mapsto xA$. We define $ImA = Im\mathcal{A}$ and

$KerA = Ker\mathcal{A}$. In this point of view, the general linear group $GL(n,R)$, i.e., the multiplication group of all the invertible matrices in Mat_nR, can be identified with the group of all the invertible R-linear transformations $x \mapsto xA$ (with $A \in GL(n,R)$) on $Mat_{1\times n}R$.

We write $\mathrm{diag}(A_1,\ldots,A_k)$ to denote the quasi-diagonal matrix with $A_1,\ldots,A_k$ as its diagonal blocks. We use I and O to denote the identity and the zero matrices, respectively. Also $E_{ij} \in Mat_{m\times n}R$ means that the matrix in $Mat_{m\times n}R$ having a singel 1 as its (i,j)-th entry, and all other entries are zero. For each $c \in R$ and $i \neq j$, we denote $T_{ij}(c) = I + cE_{ij}$, and define $SL(n,R) = \langle T_{ij}(c) \in Mat_nR \mid i \neq j, c \in R\rangle$. We denote $P_{ij} = I - E_{ii} - E_{jj} + E_{ij} - E_{ji} \in SL(n,R)$ when $i \neq j$.

An n-dimensional vector space V over a field F will be written as the space $Mat_{1\times n}F$ of all the n-dimensional rows relative to a basis $\{e_1,\ldots,e_n\}$ of V, by identifying a vector $\sum_{i=1}^n x_ie_i \in V$ with its coordinate $x = (x_1,\ldots,x_n) \in Mat_{1\times n}F$. A basis vector e_i becomes the row $(a_1,\ldots,a_n)$ with all $a_j = 0$ except $a_i = 1$. The linear group $GL(V)$ is written as the matrix group $GL(n,F)$. Let $\tau : a \mapsto \overline{a}$ be an involution of F, $\varepsilon = \pm 1 \in F$. Then an $H = \Delta + \varepsilon\overline{\Delta}' \in GL(n,F)$ (with $\Delta \in Mat_nF$) determines a unitary group $U(n,F,H) = \{A \in Mat_nF \mid AH\overline{A}' = H\}$ relative to τ, which is geometrically the group $U(V,f) = \{g \in GL(V) \mid f(xg,yg) = f(x,y)\ \forall x,y \in V\}$ determined by the nondegenerate trace-valued ε-hermitian form $f(x,y) = xH\overline{y}'$ on V. We admit the case $\tau = 1$, hence $U(n,F,H)$ may be a symplectic group $Sp(n,F,H)$ or an orthogonal group $O(n,F,H)$ (with $charF \neq 2$). When $U(n,F,H) \neq O(n,F,H)$, we may always assume that $\varepsilon = -1$, and in such case the $SU(n,F,H) = U(n,F,H) \cap SL(n,F)$ coincides with the commutator group of $U(n,F,H)$ with a few exceptions for small n and F. The normalizer of $SU(n,F,H)$ in $GL(n,K)$ is $GU(n,F,H) = \{A \in GL(n,F) \mid AH\overline{A}' = \lambda H$ for a $\lambda \in F^*\}$.

The orthogonal groups acting on V are defined to be an $O(V,Q,L) = \{g \in GL(V) \mid Q(xg) - Q(x) \in L, \forall x \in V\}$ which is determined by a quadratic form $Q(x) = x\Delta x'$ (with $\Delta \in Mat_nF$) and an $L = 0$ when $charF \neq 2$ or an F^2-subspace $L \neq F$ of F when $charF = 2$. Without loss of generality, we may assume that Q associates with a nondegenerate symmetric form $f(x,y) = Q(x+y) - Q(x) - Q(y) = xHy'$ with an invertible $H = \Delta + \Delta'$ and assume that $1 \in L$ when $L \neq 0$. We also define $O(V,Q,L) = Sp(V,f)$ for $L = F$. $O(V,Q,L)$ which can be identified with the matrix group $O(n,F,\Delta,L) = \{A \in GL(n,F) \mid A\Delta A' - \Delta \in K_L(n,F)\}$, where $K_L(n,F) = \{A = (a_{ij})_{n\times n} \in Mat_nF \mid A' = -A$ and all $a_{ii} \in L\}$. The commutator of $O(n,F,\Delta,L)$ is denoted by $\Omega(n,F,\Delta,L)$. The normalizer

of $\Omega(n,F,\Delta,L)$ in $GL(n,F)$ is $GO(n,K,\Delta,L) = \{A \in GL(n,F) \mid A\Delta A' - \lambda\Delta \in K_L(n,F)$ for a $\lambda \in F^*\}$, (with exceptions for small n and F). The $O(n,F,\Delta,0), \Omega(n,F,\Delta,0), GO(n,F,\Delta,0)$ are just the orthogonal groups $O(n,F,\Delta), \Omega(n,F,\Delta), GO(n,F,\Delta)$ as usually defined.

The Witt index ν of a $U(n,F,H)$ is defined as the dimension of a maximal totally isotropic subspace of the V equipped with the form $f(x,y) = xH\overline{y}'$. We shall always choose a suitable basis of V at the beginning, hence by replacing H by a $PH\overline{P}'$ $(P \in GL(n,F))$, we can reduce to the case in which H is in the standard form

$$\begin{pmatrix} O & I^{(\nu)} & \\ \varepsilon I_{(\nu)} & O & \\ & & H_0 \end{pmatrix},$$

where $H_0 \in Mat_{n-2\nu}K$ is diagonal and definite, where 'definite' means that $xH_0\overline{x}' \neq 0$ for all $0 \neq x \in Mat_{1\times(n-2\nu)}K$. The Witt index ν of an $O(n,F,\Delta,L)$ is defined as the dimension of a maximal totally L-singular subspace W of V (satisfying $xHy' = 0$ and $x\Delta x' \in L$ for all $x,y \in W$, where $H = \Delta + \Delta'$). We can reduce to the case in which Δ is in the standard form

$$\Delta = \begin{pmatrix} O & I_{(\nu)} & \\ O & & \\ & & \Delta_0 \end{pmatrix},$$

where $\Delta_0 \in Mat_{n-2\nu}K$ is L-definite that is, $x\Delta_0 x' \notin L$ for all $0 \neq x \in Mat_{1\times(n-2\nu)}K$, and $\Delta_0 + \Delta_0'$ is monomial.

3. Main Results

Let K be a subfield of a field F. Then $GL(n,K) = GL(n,F) \cap Mat_nK$ is a subgroup of $GL(n,F)$. Clearly, the underlying spaces V_K, V_F of $GL(n,K)$ resp. $GL(n,F)$ have the relation $V_F = F \otimes V_K$. When $\Delta \in Mat_nK$ and $L \subseteq K$, $O(n,F,\Delta,L)$ contains $O(n,K,\Delta,L)$ as its subgroup. If τ is an involution of F which extends an involution J of K, $H = -\overline{H}' \in GL(n,K)$, then a $U(n,F,H)$ relative to τ contains the $U(n,K,H)$ relative to J as its subgroup.

Let K be a maximal subfield of F and N_F a classical group acting on $Mat_{1\oplus n}F$, $N = SL(n,K)$, $SU(n,K,H)$ or $\Omega(n,K,\Delta,L)$. Then the normalizer M of N in N_F is a member of the class C_5 defined in [1]. We need to verify that M is a maximal subgroup of N_F. It suffices to determine all the overgroups of M in N_F. However, instead of just considering this case, we can do something more, that is, we want to classify the overgroups in

$GL(n,F)$ of an $N = SL(n,K), Sp(n,K,H), SU(n,K,H)$ or $\Omega(n,K,\Delta,L)$ for an arbitrary subfield (not necessarily maximal) K of F. When F is finite, then this has been done in [2] and [3]. Also we have discussed the overgroups of $SL(n,K)$ in $GL(n,F)$ in [4]. In this paper we will treat the other cases and classify the overgroups of $N = Sp(n,K,H), SU(n,K,H)$ or $\Omega(n,K,\Delta,L)$ in $G = GL(n,F)$ for an infinite field F. Our treatment applies to arbitrary fields F,K, but only exclude the case $K_J = F_2$ for avoiding lengthy discussion. For each field E between K and F, we have certain classical groups $N_E \trianglerighteq SL(n,E), Sp(n,E,H), SU(n,E,H), \Omega(n,E,\Delta)$ or $\Omega(n,E,\Delta,L)$ over E as overgroups of N in G. The question is : are there any groups between N and G other than those corresponding to the intermediate fields E ? We will settle the cases under certain assumption on the index $r = [F:K]$ and the Witt index ν of N. The following theorem is our main result.

Theorem 3.1. *Let K be a subfield of a field F with $[F:K] = r < \infty$ and $K_J \neq F_2$. Let $N \leq X \leq GL(n,F)$ of Witt index ν satisfying one of the following conditions:*

(a) $N = Sp(n,K,H)$, and $n > r$ or $n(n+2) > 8r$.
(b) $N = SU(n,K,H)$ and $\nu^2 > 2r$.
(c) $N = \Omega(n,K,\Delta)$ and $\nu(\nu-1) > 4r$.
(d) $N = \Omega(n,K,\Delta,L)$ with $L \neq 0$, and $\nu(\nu-1) \geq 2r$.

Then there exists a field E between K and F such that one of the following condition holds:

(i) $SL(n,E) \trianglelefteq X \leq GL(n,E) \cdot K^*$.
(ii) $SU(n,E,H) \trianglelefteq X \leq GU(n,E,H) \cdot F^*$, (including the case $SU(n,E,H) = Sp(n,E,H)$).
(iii) $X \trianglerighteq \Omega(n,E,\Delta,L) \geq N$ for an L, ($L \supsetneq K$ in case (a) when $charK = 2$).

When K is a maximal subfield, one can easily deduce a corollary on the maximality of the groups over K in the groups over F.

4. Some Preliminary Results

In our proof, we need the following known results.

Theorem 4.1. *Let E be a field. Let $N = SU(n,E,f)$ with Witt index $\nu(f) \geq 1$ with $N \leq X \leq GL(n,E)$. Then either $X \trianglerighteq SL(n,E)$ or $X \trianglerighteq SU(n,E,f)$. Let $N = \Omega(n,E,Q)$ with Witt index $\nu(Q) \geq 1$ and $N \leq X \leq GL(n,E)$. Then one of the following condition holds:*

(i) $X \trianglerighteq SL(n,E)$;
(ii) $X \trianglerighteq \Omega(n,E,Q)$, *or* $X \trianglerighteq \Omega(n,E,Q,L)$ *for an* L *when* $charE = 2$;
(iii) $GL(n,E) = SL(4,2) \cong A_8$, $N = \Omega^-(4,2) \cong A_5$ *and* $X = Sp(4,2)' \cong A_6$ *or* $X \cong A_7$, *or* $N = \Omega^+(4,2) \cong (S_3 \times S_3 \times S_2) \cap A_8$ *and* $X \cong (S_3 \times S_5) \cap A_8$.

Proof. See [5]. □

Theorem 4.2. *Let* K *be a subfield of a field* F *with* $[F:K] = r < \infty$. *Let* $SL(n,K) \leq X \leq GL(n,F)$ *with* $n > r$ *or* $\frac{1}{2}n(n-1) > r$. *Then* $SL(n,E) \trianglelefteq X \leq GL(n,E) \cdot K^*$ *for a field* E *between* K *and* F.

Proof. See [4]. □

5. Group Extension by an Element with Zero Entries

For each field E between K and F, we denote $G_E = GL(n,E) \cdot F^*$, and write $G_K = GL(n,K) \cdot F^*$. We also use N_E to denote an $SU(n,E,H)$ or $\Omega(n,E,\Delta,L_1)$ containing $N = SU(n,K,H)$ or $\Omega(n,K,\Delta,L)$. Recall that $SU(n,E,H)$ is a $Sp(n,E,H)$. We denote by G_E the normalizer of N_E in $GL(n,E)$ which is equal to $GU(n,E,H)$ or $GO(n,E,\Delta,L_1)$. Let τ be an involution of E relative to $N_E = SU(n,E,H)$. Then we denote $E_\tau = \{a \in E \mid a^\tau = a\}$, and we write $E_0 = E_\tau$ when $N_E = SU(n,E,H)$ or $charE \neq 2$, while $E_0 = E^2 = \{a^2 \mid a \in E\}$. When $charE = 2$, then $N_E = \Omega(n,E,\Delta,L)$. In particular, $K_0 = K_J$ or K^2.

Throughout this section, we suppose that $N_K = SU(n,K,H)$ of Witt index $\nu \geq 2$ or $N_K = \Omega(n,K,\Delta,L)$ of Witt index $\nu \geq 3$. We consider the extension group $Y = \langle N_K, g_0 \rangle$ of N_K by a $g_0 \in GL(n,F) \setminus G_K$ with certain entries zero. We prove that if g_0 has enough zero entries, then Y will contain an N_E over a field $E \supset K$.

Since the overgroups of $SL(n,K)$ in $GL(n,F)$ has been classified in Theorem 4.2, we need only to consider the case in which $Y \not\trianglerighteq SL(n,K)$. By Theorem 4.1, $X \cap GL(n,K) \trianglerighteq \widetilde{N}_K$ for a $\widetilde{N}_K = SU(n,K,H)$ or $\widetilde{N}_K = \Omega(n,K,\Delta,L_1)$. Replacing N_K by $\widetilde{N}_K$, we may assume that $Y \cap SL(n,K) \trianglerighteq N_K$. Namely, $Y \cap SL(n,K)$ lies in the normalizer $G_K = GU(n,K,H)$ or $GO(n,K,\Delta,L)$ of N_K in $GL(n,K)$.

We now prove the following theorem.

Lemma 5.1. *Let* $N_K = SU(n,K,H)$ *with Witt index* $\nu \geq 2$, *or* $N_K = \Omega(n,K,\Delta,L)$ *with Witt index* $\nu \geq 3$. *Suppose that* $g_0 = (d_{ij})_{n\times n} \in$

$GL(n,F) \setminus G_K$ *with* $d_{1,\nu+1} = 0$ *when* $N_K \neq \Omega(n,K,\Delta)$, *or* $d_{i,\nu+j} = 0$ *for* $1 \leq i,j \leq 2$ *when* $N_K = \Omega(n,K,\Delta)$. *If* $Y = \langle N_K, g_0 \rangle$ *with* $Y \cap GL(n,K) \leq G_K$, *then* $Y \geq N_E > N_K$ *for* $N_E = SU(n,E,H)$ *or* $\Omega(n,E,\Delta,L_1)$ *over a field* $E \supset K$.

This result leads to the following lemmas.

For an $S \in Mat_\nu F$, we shall denote $T_0(S) = \begin{pmatrix} I^{(\nu)} & O & \\ S & I^{(\nu)} & \\ & & I \end{pmatrix}$. We now write $T_0(S)$ for an $S \in Mat_k F$ with $k < \nu$. Denote $T_0(\mathrm{diag}(S, O^{(\nu-k)}))$. Note that $T_0(S) \in SU(n,K,H)$ when $\overline{S}' = S$, while $T_0(S) \in \Omega(n,K,\Delta,L)$ when $S' = S$ having the diagonal entries in L. We also denote $\widetilde{H} = \begin{pmatrix} O & I^{(\nu)} \\ \varepsilon I & O \end{pmatrix}$ and $\widetilde{\Delta} = \begin{pmatrix} O & I^{(\nu)} \\ O & O \end{pmatrix}$. These matrices are just the submatrices of H with respect to Δ which contains the (i,j)-th entries with $1 \leq i,j \leq 2\nu$.

For group elements a, b, we denote by $[a,b] = aba^{-1}b^{-1}$ as their commutator.

Lemma 5.2. *Let* $g_1 = (a_{ij})_{n\times n} \in GL(n,F) \setminus G_K$, *with all* $a_{1j} = 0$ $(j \geq 2)$ *when* $N_K = SU(n,K,H)$ *or* $\Omega(n,K,\Delta,L)$ $(L \neq 0)$, *or with all* $a_{1j} = a_{2j} = 0$ $(j \geq 3)$ *when* $N_K = \Omega(n,K,\Delta)$. *Let* $Y = \langle N_K, g_1 N_K g_1^{-1} \rangle$, *with* $Y \cap GL(n,K) \leq G_K$. *Then* Y *contains* $T = T_0 \begin{pmatrix} a & b \\ c & 0 \end{pmatrix}$ *with either* $b \notin K$ *or* $c \notin K$.

Proof. Write $g_1^{-1} = (\widetilde{a}_{ij})_{n\times n}$. Define $U_1 = \{(c_{ij})_{n\times n} \in GL(n,F) \mid a_{11} = 1, a_{\nu+1,\nu+1} \in K^*$, and $c_{1j} = c_{i,\nu+1} = 0 \; \forall j \neq 1, i \neq \nu+1\}$, $U_0 = \{(c_{ij})_{n\times n} \in U_1 \mid c_{\nu+1,\nu+1} = 1\}$. For each $g \in GL(n,F)$, we denote by $\phi(g)$ the matrix obtained by deleting the first and the $(\nu+1)$-st rows and columns of g. Then $\phi : U_1 \to GL(n-2,F)$ defines a homomorphism of groups.

Step I. We first replace g_1 by a $g \in U_1 \setminus \Gamma_K$ with $gN_Kg^{-1} < Y$. Then we reduce it to the case $g_1 \in U_1$.

In case $N_K = SU(n,K,H)$ or $\Omega(n,K,\Delta,L)$ $(L \neq 0)$, we have $T_0(1) \in N_K$, and hence $g_2 = g_1 T_0(1) g_1^{-1} = I + b_2 E_{21} + \cdots + b_n E_{n1} \in Y \cap U_0$, with $b_i = a_{i,\nu+1} a_{11}^{-1}$ $(i \geq 2)$. If $g_2 \notin \Gamma_K$, then we are done. Otherwise, for $g_2 \in SL(n,K) \cap Y \leq G_K$, we have $g_2 = T_{\nu+1,1}(b_{\nu+1})$. Thus, $b_i = 0$ and hence $a_{i,\nu+1} = 0$ for $i \neq \nu+1$, and $a_{\nu+1,\nu+1} a_{11}^{-1} = b_{\nu+1} \in K^*$. Now we replace g_1 by $g_1 a_{11}^{-1}$ so that we can make $a_{11} = 1$ and hence $g_1 \in U_1$, without changing Y.

In case if $N_K = \Omega(n,K,\Delta)$, then we take $g_2 = g_1 T_0 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = I +$

$\sum_{3\le i\le n,1\le j\le 2} b_{ij}E_{ij} \in Y\cap U_0$. If $g_2\notin \Gamma_K$, then we are done. Otherwise, we have $g_2\in SL(n,K)\cap Y\le GO(n,K,\Delta)$. Hence $g_2 = T_0\begin{pmatrix} 0 & s\\ -s & 0\end{pmatrix}\in O(n,H,\Delta)$, for an $s\in K^*$. This implies that $a_{i1}=a_{i2}=0$ for all $i\notin\{1,2,\nu+1,\nu+2\}$, and $\begin{pmatrix} a_{\nu+1,\nu+1} & a_{\nu+1,\nu+2}\\ a_{\nu+2,\nu+1} & a_{\nu+2,\nu+2}\end{pmatrix} = s\begin{pmatrix} a_{22} & -a_{21}\\ -a_{12} & a_{11}\end{pmatrix}$. Take $g_3 = g_1T_{3l}(1)T_{\nu+l,\nu+3}(-1)g_1^{-1} = (c_{ij})_{n\times n}\in Y\cap U_0$ for $l=1,2$. If $g_3\notin \Gamma_K$ for $l=1$ or 2, then we are done. Otherwise, if $g_3\in SL(n,K)$ for $l=1,2$, then $(c_{t1},c_{t2}) = a_{t3}(\widetilde{a}_{l1},\widetilde{a}_{l2})\in Mat_{1\times 2}K$ for $t\notin\{1,2,\nu+1,\nu+2\}$. Since g_1 is invertible, we can choose a $t\notin\{1,2,\nu+1,\nu+2\}$ with $a_{t3}=\lambda\neq 0$. Hence, $(\widetilde{a}_{ij})_{1\le i,j\le 2}=\lambda^{-1}A$ for an $A\in GL(2,K)$. Take $z=\mathrm{diag}(A^{-1},I^{(\nu-2)},A',I^{(n-\nu-2)})\in O(n,K,\Delta)$. Then $(\lambda z)^{-1}N_K(\lambda z)=N_K$. Replacing g_1 by $g_1(\lambda z)^{-1}$ and without changing Y, we can reduce it to the case $(\widetilde{a}_{ij})_{1\le i,j\le 2}=I^{(2)}$. Thus $(a_{ij})_{1\le i,j\le 2}=I^{(2)}$, $g_1\in U_1$, as desired.

Step II . To make $g_1\in U_0$ and $\phi(g_1)=I$.

Note that our standard H is a monomial matrix. Now in each k-th row of H, there is a unique nonzero entry at some k'-th column. In other words, for each e_k, there is a unique $e_{k'}$ such that $f(e_k,e_{k'})\neq 0$. For each $k\notin\{1,\nu+1\}$, there is an element $z_1 = I+E_{k1}+\alpha E_{\nu+1,1}+\beta E_{\nu+1,k'}\in N_K$ for suitable $\alpha\in K$ and $\beta\in K^*$. We consider $g_2 = g_1z_1g_1^{-1}\in Y$. Then $g_2=(b_{ij})_{n\times n}\in Y\cap U_0$ with $\phi(g_2)=I$. If we can choose k to obtain a $g_2\notin G_K$, then step II is finished. Otherwise we have $g_2\in SL(n,K)\cap Y\le G_K$ for all k, $b_{i1}=a_{ik}\in K$ for all $i\notin\{1,\nu+1\}$. Thus $\phi(g_1)\in GL(n-2,K)$.

Let $N_0=\phi(N_K\cap U_0)$. Then $N_0 = SU(n-2,K,\phi(H))$ when $N_K = SU(n,K,H)$, or $N_0=\Omega(n-2,K,\phi(\Delta),L)$ when $N_K=\Omega(n,K,\Delta,L)$. Consider the normalizer G_0 of N_0 in $GL(n-2,K)$ and suppose that $\phi(g)\notin G_0$, (which implies $N_0\neq Sp(2,K,\phi(H))$). If $Y_0=\langle N_0,\phi(g_1)N_0\phi(g_1)^{-1}\rangle\le \phi(Y\cap U_0)$, then by Theorem 4.1, either $Y_0=SL(n-2,K)$, or $Y_0=\Omega(n-2,K,\phi(\Delta),L_1)$ with $L_1\supset L$ when $charK=2$ and $N_K=\Omega(n,K,\Delta,L)$. For the case $Y_0=SL(n-2,K)$, we can take $g_3\in Y\cap U_0$ with $\phi(g_3)=T_{n-2,1}(1)\notin G_0$ when $n\ge 5$, or with $\phi(g_3)=T_{21}(\theta)\notin G_0$ $(\overline{\theta}\neq\theta)$ when $N_K=SU(4,K,H)\neq Sp(4,K,H)$. Then we obtain $g_2=g_3T_{21}(1)T_{\nu+1,\nu+2}(-1)g_3^{-1}\in Y\cap U_0$ with $\phi(g_2)=I$, and $g_2\notin G_K$. Hence, $g_2\notin\Gamma_K$. Replacing g_1 by g_2, then we are done. For the case $charK=2$, $N_K=\Omega(n,K,\Delta,L)$ and $N_0=\Omega(n-2,K,\Delta,L_1)$ (with $L_1\supset L$), we take $g_3\in Y\cap U_0$ with $\phi(g_3)=T_{\nu 1}(\theta)\notin G_0$ for a $\theta\in L_1\setminus L$, and consider $g_2=[z,g_3]\in Y$ for any $z=T_{21}(s)T_{\nu+1,\nu+2}(s)\in N_K$ with $s\in K^2$. We have

$g_2 = T_0 \begin{pmatrix} \beta(s) & s\theta \\ s\theta & 0 \end{pmatrix} U_0$, with $\beta(s) = s(b+c) + s^2\theta$ depending on s, where b, c are the $(\nu+1,2)$-nd,$(\nu+2,1)$-st entries of g_3 respectively. If $g_2 \notin \Gamma_K$ for an s then we are done. Otherwise, then for all the $s \in K^2$, we have $g_2 \in Y \cap G_K$ hence $\beta(s) \in L$. Taking a $1 \neq s \in K^2 \neq F_2$ we see that $\beta(s) + s\beta(1) = (s^2+s)\theta \in L$ and hence $\theta \in L$, a contradiction.

Now we suppose that $\phi(g_1) \in G_0 = GU(n-2, K, \phi(H))$ or $GO(n-2, K, \phi(\Delta), L)$. Take $z_0 = (z_{ij})_{n\times n} \in U_1 \cap GL(n, K)$ with $\phi(z_0) = \phi(g_1)$, $z_{i1} = z_{\nu+1,j} = 0$ for all $i \geq 2$ and $j \neq \nu+1$. Since $\phi(z_0) \in G_0$, we can choose which is $z_{\nu+1,\nu+1}$ suitable to make $z_0 \in G_K$. Hence $z_0^{-1} N_K z_0 = N_K$. Replacing g_1 by $g_1 z_0^{-1}$ we can reduce to the case $\phi(g_1) = I$, without changing Y. Moreover, from $g_1 T_{\nu+1,2}(1) T_{\nu+2,1}(-\varepsilon) g_1^{-1} \in SL(n, K) \cap Y \leq \Gamma_K$, we can see that $a_{\nu+1,\nu+1} = 1$ and hence $g_1 \in U_0$.

Step III . Y contains T as required by the lemma.

By step II, we can assume that $g_1 \in U_0$ and $\phi(g_1) = I$. First we suppose that $a_{21} \notin K$ or $a_{\nu+1,\nu+2} \notin K$. When $N_K \neq \Omega(n, K, \Delta)$, we have $T = [T_{\nu+2,2}(1), g_1] = T_0(aE_{\nu+1,1} - a_{\nu+1,\nu+2}E_{\nu+1,2} + a_{21}E_{\nu+2,1}) \in Y$ for some $a \in F$. That is what we need. When $N_K = \Omega(n, K, \Delta)$ (with $\nu \geq 3$), we can take $T_1 = [T_{\nu+2,3}(-1)T_{\nu+3,2}(1), g_1] \in Y$. Then $T = [T_{32}(-1)T_{\nu+2,\nu+3}(1), T_1] = T_0(a_{\nu+1,\nu+2}E_{\nu+1,2} + aE_{\nu+2,1})$ which is what we need.

Suppose that $a_{21}, a_{\nu+1,\nu+2} \in K$, but $a_{k1}, a_{\nu+1,k'} \notin K$ for some $k, k' \notin \{1, \nu+1\}$ with $f(e_k, e_{k'}) \neq 0$. Then by taking a $z = I + E_{2k} + \alpha E_{2,\nu+2} + \beta E_{k',\nu+2} \in N_K$ with $\beta \neq 0$, and replacing g_1 by $[z, g_1] \in Y$ we can reduce it to the known cases $a_{21} \notin K$ or $a_{\nu+1,\nu+2} \notin K$.

In the remaining case, g_1 has all the entries in K except $a_{\nu+1,1} \notin K$. Take $z_1 = P_{12}P_{\nu+1,\nu+2} \in N_K$ and $z_2 = T_{21}(-1)T_{\nu+1,\nu+2}(1) \in N_K$. Replacing g_1 by $[z_2, z_1 g_1 z_1^{-1}]$, then we can reduce it to a known case . □

Lemma 5.3. *Let $N_K = SU(n, K, H)$ or $\Omega(n, K, \Delta, L)$ of Witt index $\nu = 2$ or 3 respectively, $N_K < Y \leq GL(n, F)$. If Y contains a $T = T_0 \begin{pmatrix} a & b \\ c & 0 \end{pmatrix}$ with $b \notin K$ or $c \notin K$, then $Y \geq N_E > N_K$ for an $N_E = SU(n, E, H)$ or $\Omega(n, E, \Delta, L_1)$ over a field $E \supset K$.*

Proof. Step I. Y contains all the $T_{\nu+1,2}(\theta)T_{\nu+2,1}(-\varepsilon\overline{\theta})$ $(\theta \in E)$ for an $E \supset K$ endowed with a $\tau : a \mapsto \overline{a}$ which extends J.

The set $M = \{S \in Mat_\nu F \mid T_0(S) \in Y\}$ is an additive group. Moreover, if for $A, B \in GL(\nu, F)$, then we can find a $g = \mathrm{diag}(A, B, \ldots) \in Y$. Thus for each $S \in M$, we have $gT_0(S)g^{-1} = T_0(BSA^{-1})$, i.e., $BSA^{-1} \in M$, $BMA^{-1} = M$. In particular, $\overline{A}'MA = M$ holds for all $A \in SL(\nu, K)$,

and this equality holds for an $A \in GL(\nu, K)$ with suitable value of $\det A$, (e.g., $\det A \in K_J$ or $\det A \in K^{*2}$ when $N_K = SU(n, K, H), \Omega(n, K, \Delta, L)$ respectively).

We have $S_0 = aE_{11} + bE_{12} + cE_{21} \in M$. Thus we can deduce the following conclusions (a)$\sim$(c).

(a) $\widetilde{S}_0 = \overline{\Lambda(\lambda)} S_0 \Lambda(\lambda) = aE_{11} + b\lambda E_{12} + \overline{\lambda} c E_{21} \in Y$, where $\Lambda(\lambda) = \mathrm{diag}(1, \lambda, \ldots)$, with λ ranging over K_J^* when $N_K = SU(4, K, H)$, or over K^* otherwise.

(b) Taking $1 \neq \lambda \in K_J^*$ in (a), we see that $S_2 = \overline{\Lambda(\lambda - 1)^{-1}}(\widetilde{S}_0 - S_0)\Lambda(\lambda - 1)^{-1} = bE_{12} + cE_{21} \in M$, and thus $\widetilde{S}_1 = S_2 - S_0 = aE_{11} \in M$.

(c) Considering $PS_1P', PS_2P' \in M$ for $P = P_{ij} \in SL(\nu, K)$, we know that $bE_{ij} + cE_{ji}, aE_{ii} \in M$ for all $1 \leq i \neq j \leq \nu$. Moreover, considering the conjugates of $T_0(bE_{ij} + cE_{ji}) \in Y$ by suitable $z \in N_K$ fixing $e_j, e_{\nu+j}$ when interchanging $\langle e_i \rangle, \langle e_{\nu+i} \rangle$ we know that $T_{ij}(b)T_{\nu+j,\nu+i}(\varepsilon c) \in Y$. For any $\alpha E_{11} \in M$ and $b_i E_{12} - \varepsilon c_i E_{21} \in M$ $(i = 1, 2)$, we have the following conclusion (d)and (e).

(d) Since $\alpha E_{22} \in M$ and $T_{21}(-c_1)T_{\nu+1,\nu+2}(b_1) \in Y$, $T_{12}(b_1)\alpha E_{22} T_{21}(c_1) - \alpha E_{22} = b_1\alpha c_1 E_{11} + b_1\alpha E_{12} + \alpha c_1 E_{21} \in M$, and thus $b_1\alpha c_1 E_{11}, b_1\alpha E_{12} + \alpha c_1 E_{21} \in M$.

(e) When $\nu \geq 3$, since $\widetilde{S}_2 = b_2 E_{13} - \varepsilon c_2 E_{31} \in M$ and $T_{32}(-b_1)T_{\nu+2,\nu+3}(c_1) \in Y$ (by (c)), $T_{23}(c_1)\widetilde{S}_2 T_{32}(b_1) - \widetilde{S}_2 = b_2 b_1 E_{12} - \varepsilon c_1 c_2 E_{21} \in M$.

We now assume that $T \in Y$ can be chosen with $a \notin K$. Then $aE_{ii} \in M$ by (b) and (c). We prove by induction on k that $S(k) = \lambda a^k E_{12} - \varepsilon(-\varepsilon a)^k \overline{\lambda} E_{21} \in M$ holds for all integers $k \geq 0$ and $\lambda \in K$. $S(0) = \lambda E_{12} - \varepsilon\overline{\lambda} E_{21} \in M$ is known (since $T_0(S(0)) \in N_K < Y$). Assume that $S(k) \in M$. Then, by (d) we have $S(k) \in M$ and $aE_{11} \in M$. These imply that $(\lambda a^k)aE_{12} + a(-\varepsilon(-\varepsilon a)^k \overline{\lambda})E_{12} = S(k+1) \in M$. Hence, we have proved that $S(k) \in M$ for all the k. Now, it follows that $h(a)E_{12} - \varepsilon\overline{h}(-\varepsilon a)E_{21} \in M$ holds for all the polynomials $h(x) \in K[x]$, where $\overline{h}(x) \in K[x]$ is obtained by operating $J : \lambda \mapsto \overline{\lambda}$ on the coefficients of $h(x)$. Clearly, $\theta E_{12} - \varepsilon\overline{\theta} E_{21} \in M$ for all $\theta \in E$, where $E = K[a]$ is a proper extension field of K endowed with an involution $\tau : \theta \mapsto \overline{\theta}$. This extends J with $\overline{h(a)} = \overline{h}(-\varepsilon a)$. Now Y contains the 2-transvections $T_{\nu+1,2}(\theta)T_{\nu+2,1}(-\varepsilon\overline{\theta})$ $(\theta \in E)$ of $SU(n, E, H)$, as desired.

For the other cases, we can easily see that $a \in K$ holds for all the chosen T. Hence b, c are not both in K. By (b) we have $aE_{11} \in M$, i.e., $T_{\nu+1,1}(a) \in Y \cap SL(n, K) \leq G_K$. This implies that $a = 0$ when $N_K = \Omega(n, K, \Delta)$. By (a), we have that $\lambda c + b\overline{\lambda} \in K$ for all $\lambda \in K^*$. In particular, $b + c \in K$,

and hence neither of b, c lies in K. Clearly, $b = -c$ when $N_K = \Omega(n, K, \Delta)$. If $J \neq 1$, $\overline{\lambda} \neq \lambda$ for a $\lambda \in K^*$, then $\lambda(b+c) - (\lambda c + b\overline{\lambda}) = (\lambda - \overline{\lambda})c \in K$ and hence $c \in K$, a contradiction. Thus this case occurs only when $J = 1$, i.e., $N_K = Sp(n, K, H)$ or $\Omega(n, K, \Delta, L)$. When $N_K \neq \Omega(n, K, \Delta)$, since $E_{11} \in M$, $bcE_{11} \in M$ by (d), and hence $bc \in K$. From $b + c, bc \in K$, we conclude that $E = K[b] = K[c]$ is a quadratic extension of K and $\tau : h(b) \mapsto \overline{h(b)} = h(c)$ $(h(x) \in K[x])$ defines an involution of E which extends $J = 1$. $\theta E_{12} + \overline{\theta} E_{21} \in M$ holds for all $\theta = \lambda$ or $\theta = \lambda b$ (by (a)) with $\lambda \in K$, and hence it holds for all $\theta \in E = K \oplus Kb$. It is clear that, $T_{\nu+1,2}(\theta)T_{\nu+2,1}(\overline{\theta}) \in Y$ for all the $\theta \in E$. In the remaining case $N_K = \Omega(n, K, \Delta)$ (with $\nu \geq 3$), we have $b = -c$, $bE_{12} - bE_{21} \in M$. We now prove that $S(k) = b^k \lambda E_{12} - \lambda b^k E_{21} \in M$, for all integers $k \geq 0$. The case $S(0) \in M$ is known. Now we assume that $S(k) \in M$. Then $S(k+1) \in M$ follows from $bE_{12} - bE_{21} \in M$ and $S(k) \in M$ by (e). This proves that $S(k) \in M$ for all the k. Thus $\theta E_{12} - \theta E_{21} \in M$ for all the $\theta \in E = K[b]$, $T_{\nu+1,2}(\theta)T_{\nu+2,1}(-\theta) \in Y$, as desired.

Step II . Y contains $N(2\nu, E) \times I = \mathrm{diag}(A, I^{(n-2\nu)} \mid A \in N(2\nu, E) < SU(n, E, H)$, where $N(2\nu, E) = \Omega(2\nu, E, \widetilde{\Delta})$ when $N_K = \Omega(n, K, \Delta, L)$ or $N_K = Sp(n, K, H) = \Omega(n, K, \Delta, K)$ (in case $charK = 2$), or $N(2\nu, E) = SU(n, E, H)$ if otherwise.

We have shown that Y contains all the $T(\theta) = T_{\nu+1,2}(\theta)T_{\nu+2,1}(-\varepsilon\overline{\theta})$ $(\theta \in E)$. Now by onsidering the conjugates of these $T(\theta)$ with elements in N_K which are permutating $\{\pm e_i \mid 1 \leq i \leq 2\nu\}$, we know that Y contains all $T_{\nu+i,j}(\theta)T_{\nu+j,i}(-\varepsilon\overline{\theta}), T_{i,\nu+j}(\theta)T_{j,\nu+i}(-\varepsilon\overline{\theta})$, $T_{ij}(\theta)T_{\nu+i,\nu+j}(-\overline{\theta})$ with $\theta \in E$ and $1 \leq i \neq j \leq 2\nu$, these elements are enough to generate the group $N(2\nu, E) \times I$.

Step III . $Y \geq N_E > N_K$ for an $N_E = SU(n, E, H)$ or $\Omega(n, E, \Delta, L_1)$.

First we assume that $N_E = SU(n, E, H) = SU(V_E, f)$, where $V_E = Mat_{1\times n}E$, $f(x, y) = xH\overline{y}'$. It ia clear that $SU(V_E, f)$ can be generated by all the unitary transvections $\rho_{u,s} : x \mapsto x + f(x, u)su$ on V_E, with $0 \neq u \in V_E$, $f(u, u) = 0$, and $0 \neq s \in E_\tau$. For any $2\nu \leq k \leq n$, we denote $W_k = Ee_1 \oplus \cdots \oplus Ee_k$, and define $SU(W_k) = SU(W_k, f|_{W_k})@1_{W_k^\perp}$. We now prove by induction on k that $SU(W_k) \leq Y$ for all $k \geq 2\nu$. For $k = 2\nu$, we have $SU(W_{2\nu}) = N(2\nu, E) \times I \leq Y$. If $n = 2\nu$, then we are done. Suppose that $n \geq k \geq 2\nu + 1$. If $SU(W_{k-1}) \leq Y$ is known, then $\rho_{u,s} \in Y$ for all $u \in W_{k-1}$. Thus, for any $\rho_{u,s} \in SU(W_k) \setminus SU(W_{k-1})$ we can write $u = \lambda(w + e_k)$ with $0 \neq w_k \in W_{k-1}$ and $\lambda \in E^*$. We can take a $z_1 \in N_K$ which sends e_1 to a $v + e_k$ with $0 \neq v \in Mat_{1\times K}$. Note that $f(w + e_k, w + e_k) = f(v + e_k, v + e_k) = 0$ implies that $f(w, w) = f(v, v) = -f(e_k, e_k)$. By Witt's

extension theorem, we can take a $z_2 \in SU(W_{k-1}) \leq Y$ which sends v to w. Hence, sending $\lambda(v + e_k)$ to u. We have $\rho_{u,s} = (z_1 z_2)^{-1} \rho_{\lambda e_1, s}(z_1 z_2) \in Y$. Hence $SU(W_k) \leq Y$. This holds for all k. In particular, $N_E = SU(W_n) \in Y$.

Now we consider the case $N_E = \Omega(n, E, \Delta, L) = \Omega(V_E, Q, L)$, where $Q(x) = x\Delta x'$. We first prove that $Y \geq \Omega(V_E, Q)$. In this purpose, it suffices to prove that Y contains all the root elements $t_{u,w} : x \mapsto x + f(x,w)u - f(x,u)(w + Q(w)u)$ of $\Omega(V_E, Q)$, with $Q(u) = f(u,w) = 0$, where $f(x,y) = x(\Delta + \Delta')y' \ \forall x, y \in V_E$. Clearly, $N(2\nu, E) \times I \leq Y$ contains all the $t_{u,w}$ with $u, w \in W_{2\nu} = Ee_1 \oplus \cdots \oplus Ee_{2\nu}$ and N_K contains the elements z_i $(1 \leq i \leq n-1)$, which all fixing e_1 such that all the $e_2 z_i$ form an E-basis of $e_1^{\perp}$. Also, Y contains all the $z_i^{-1} t_{\lambda e_1, e_2} z_i = t_{\lambda e_1, e_2 z_i} = t_{e_1, \lambda e_2 z_i}$. Hence, it contains all the $\prod_i t_{e_1, \lambda_i e_2 z_i} = t_{e_1, w}$, where $w = \sum_i \lambda_i e_2 z_i$ ranges over $e_1^{\perp}$. For any $z \in Y$, we have $z^{-1} t_{e_1,w} z = t_{e_1 z, wz} \in Y$ with wz ranges over $(e_1 z)^{\perp}$. Hence, we need only to prove that $e_1 z$ $(z \in Y)$ which ranges over all the singular vectors in V_E. Clearly, $e_1 z$ $(z \in N(2\nu, E) \times I < Y)$ ranges over singular $u \in W_{2\nu}$ so that $t_{u,w} \in Y$ when $u \in W_{2\nu}$. Let $v \neq 0$ be any singular vector in V_E. Then we can find singular vectors u, u_1, v_1 with $u, u_1 \in W_{2\nu}$, with $f(u, u_1) = f(v, v_1) = 1$ and $(Eu \oplus Eu_1) \perp (Ev \oplus Ev_1)$. Thus $\langle t_{u, \lambda v_1}, t_{u_1, \lambda v} \mid \lambda \in E \rangle \leq Y$ induces $SL(2, E)$ on $Eu \oplus Ev$ which can send u to v. This proves that Y contains all the $t_{v,w}$, and clearly it contains $\Omega(V_E, Q)$. When $N_K = \Omega(n, K, \Delta)$, then $N_E = \Omega(n, E, \Delta)$ is what we need. Suppose that $N_K = \Omega(n, K, \Delta, L)$ with $L \neq 0$, including the case $N_K = Sp(n, K, H) = \Omega(n, K, \Delta, K)$ when $charK = 2$. We need to find in Y and $N_E = \Omega(n, E, \Delta, L_1) > N_K$. For each $1 \leq t \leq 2\nu$, N_K contains all the $\rho_{e_t, s}$ acting on V_E with $s \in L$. For any $\lambda \in E^*$, we can see that $N(2\nu, E) \times I < Y$ contains z sending e_t to λe_t, and hence Y contains $z^{-1} \rho_{e_t, s} z = \rho_{\lambda e_t, s} = \rho_{e_t, \lambda^2 s}$. This implies that Y contains $\rho_{e_t, s}$ for all s in the E^2-subspace L_1 of E spanned by L. We claim that $Y \geq N_E = \Omega(n, E, \Delta, L_1) > N_K$. To see this, it suffices to prove that $\rho_{u,s} \in Y$ for all $s \in L_1$ and $0 \neq u \in V_E$ with $Q(u) = u\Delta u' = c \in L_1$, since all these $\rho_{u,s}$ generate $\Omega(n, E, \Delta, L_1)$. Now $z_0 = \rho_{e_{\nu+1}, c} \in Y$ sends e_1 to $u_1 = e_1 + ce_{\nu+1}$ with $Q(u_1) = c = Q(u)$ and $\Omega(n, E, \Delta) \leq Y$ contains a z_1 sending u_1 to u. Hence $\rho_{u,s} = (z_1 z_0)^{-1} \rho_{e_1, s}(z_1 z_0) \in Y$ as desired. □

Lemma 5.4. *Let $g_0 \in GL(n, F) \setminus \Gamma_K$ and $Y = \langle N_K, g_0 \rangle \leq G_K$. Suppose that $0 \neq S_1 \in Mat_\nu K$ satisfying $T_0(\lambda S_1) \in N_K$ and $T_1(\lambda S_1) = g_0 T_0(\lambda S_1) g_0^{-1} \in \Gamma_K$, for all $\lambda \in K_0$. Then there exists a $g_1 \in Y \setminus \Gamma_K$ satisfying the condition described in Lemma 5.2 . Namely, g_1 has the $(1, j)$-th entries zero for all $j \geq 2$ when $N_K \neq \Omega(n, K, \Delta)$, or it has the (i, j)-th*

entries zero for $1 \leq i \leq 2, j \geq 3$ *when* $N_K = \Omega(n, K, \Delta)$.

Proof. We first prove that all $T_1(\lambda S_1)$ lie in $GL(n, K)$. Since $(T_1(S_1) - I)^2 = g_0(T_0(S_1) - I)^2 g_0^{-1} = 0$, $T_1(S_1)$ cannot be a diagonal matrix. Let b be a nonzero entry of $T_1(S_1)$ at some (k, l)-th position with $k \neq l$. Note that $T_0(\lambda S_1) = I + \lambda(T_0(S_1) - I) = \lambda T_0(S_1) + (1 - \lambda)I$. Hence $T_1(\lambda S_1) = \lambda T_1(S_1) + (1 - \lambda)I$. This leads to $T_1(\lambda S_1)$ contain the (k, l)-th entry which is equal to $\lambda b \neq 0$. Since $T_1(\lambda S_1)$ lies in $\Gamma_K = GL(n, K) \cdot F^*$, $b^{-1}T_1(\lambda S_1)$. But $b^{-1}T_1(\lambda S_1)$ has its (k, l)-th entry which is equal to $\lambda \in K_0^* \subseteq K^*$, we must have $b^{-1}T_1(\lambda S_1) \in GL(n, K)$. Write $b^{-1}\lambda T_1(S_1) + b^{-1}(1 - \lambda)I \in GL(n, K)$. Taking $\lambda = 1$, then we know that $b^{-1}T_1(S_1) \in GL(n, K)$, and thus $b^{-1}\lambda T_1(S_1) \in GL(n, K)$ for all $\lambda \in K_0^*$. This implies that $b^{-1}(1-\lambda)I \in Mat_n K$, $b^{-1}(1 - \lambda) \in K$. Since $|K_0| > 2$, we can choose $1 \neq \lambda_1 \in K_0^*$, $b^{-1}(1 - \lambda_1) \in K^*$ implies $b \in K^*$. This leads to $T_1(\lambda S_1) \in GL(n, K)$ for all $\lambda \in K_0^*$. In particular, we have $T_1(S_1) \in GL(n, K)$.

For each $\lambda \in K_0^*$, since $T_0(\lambda S_1) \in N_K$, $T_1(\lambda S_1) \in Y$ and $Y_K = \langle N_K, T_1(\lambda S_1) \mid \lambda \in K_0^* \rangle \leq Y \cap SL(n, K) \leq G_K$. Hence $N_K \trianglelefteq Y_K$. We take $z_1 \in N_K$ which sends $T_0(S_1)$ to $z_1^{-1}T_0(S_1)z_1 = T_0(S_0)$ by conjugacy with $S_0 = \mathrm{diag}(\Lambda_0, 0)$ for $\Lambda_0 \in GL(k, K)$, where $k = rank S_1 \leq \nu$. Moreover, for each $\lambda \in K_0^*$, $z_1^{-1}T_0(\lambda S_0)z_1 = I + z_1^{-1}(T_0(\lambda S_0) - I)z_1 = I + \lambda z_1^{-1}(T_0(S_0) - I)z_1 = T_0(\lambda S_0)$ hold. We have $(T_1(S_1) - I)^2 = 0$ since $(T_0(S) - I)^2 = 0$. Thus $Im(T_1(S_1) - I) \subseteq Ker(T_1(S_1) - I) = \{v \in V_K \mid vT_1(S_1) = v\}$. Hence we have $T_1(S_1) \in Y_K \leq G_K \leq GU(n, K, H)$, ($H = \Delta + \Delta'$ when $N_K = \Omega(n, K, \Delta, L)$). This proves that $T_1(S_1) \in U(n, K, H)$ and that $Im(T_1(S_1) - I)$ is a totally isotropic subspace relative to the form $f(x, y) = xH\overline{y}'$. Moreover, when $char K = 2$ and $N_K = \Omega(n, K, \Delta, L) \trianglelefteq Y_K$ we can also prove that $Im(T_1(S_1) - I)$ is totally L-singular. If $Im(T_1(S_1) - I)$ is not totally isotropic, then $f(u, v) \neq 0$ for some $u, v \in Im(T_1(S_1) - I) \subseteq Ker(T_1(S_1) - I)$. From $f(u, v) = f(uT_1(S_1), vT_1(S_1)) \neq 0$ and $T_1(S_1) \in GU(n, K, H)$, it follows that $T_1(S_1) \in U(n, K, H)$. This implies that $Im(T_1(S_1) - I) = Ker(T_1(S_1) - I)^{\perp} \subseteq Im(T_1(S_1) - I)^{\perp}$. Thus, $Im(T_1(S_1) - I)$ is totally isotropic, which is a contradiction. From this result, we conclude that $Im(T_1(S_1) - I)$ should be totally isotropic. Observe that $Im(T_1(\lambda S_1) - I) = Im(\lambda(T_1(S_1) - I)) = Im(T_1(S_1) - I)$ holds for each $\lambda \in K^*$, and that $dim(Im(T_1(S_1) - I)) = rank(T_1(S_1) - I) = rank S_1 = k$. Let $0 \neq u \in Im(T_1(S_1) - I)$. Since f is nondegenerate, there exists $v \in V_K$ with $f(u, v) = 1$. We have $vT_1(S_1) - v \in Im(T_1(S_1) - I)$, and hence $f(u, vT_1(S_1) - v) = 0$, $f(uT_1(S_1), vT_1(S_1)) = f(u, vT_1(S_1)) = f(u, v) = 1$. This implies that $T_1(S_1) \in U(n, K, H)$ since $T_1(S_1) \in GU(n, K, H)$, this im-

plies that $T_1(S_1) \in O(n,K,\Delta,L)$ when $N_K = \Omega(n,K,\Delta,L)$ since $T_1(S_1) \in GO(n,K,\Delta,L)$. Now we suppose $charK = 2$ and $N_K = \Omega(n,K,\Delta,L)$. We prove that $Q(x_1) \in L$ holds for all the $x_1 = x(T_1(S_1) - I) \in Im(T_1(S_1) - I)$ with $x \in V_K$. For each $\lambda \in K_0^*$, we have $\lambda x_1 = x(\lambda(T_1(S_1) - I)) = x(T_1(\lambda S_1) - I)$. Hence, $xT_1(\lambda S_1) = x + \lambda x_1$. From $T_1(\lambda S_1) \in O(n,K,\Delta,L)$ we know that $s_\lambda = Q(x + \lambda x_1) - Q(x) = \lambda^2 Q(x_1) - \lambda f(x,x_1) \in L$, and $s_1 = Q(x_1) - f(x,x_1) \in L$ in particular. Recall that L is a K_0-space in this case, and $|K_0| > 2$ by our assumption. Hence, we can take a $1 \neq \lambda \in K_0^*$, and $Q(x_1) = (\lambda^2 - \lambda)^{-1}(s_\lambda - \lambda s_1) \in L$ follows as desired.

Let $W_k = \{(a_1,\ldots,a_k,0,\ldots,0) \mid a_i \in F \text{ for } 1 \leq i \leq k\}$ be a k-dimensional totally isotropic subspace spanned by the first k vectors $e_1,\ldots,e_k$ in the standard basis $\{e_i \mid 1 \leq i \leq n\}$. Suppose that $Im(T_1(S_1) - I)$ and W_k are both totally isotropic and are totally L-singular when $N_K = \Omega(n,K,\Delta,L) \trianglelefteq Y_K$. Then by Witt's extension theorem, there exists $z \in Y_K$ which sends $Im(T_1(S_1) - I)$ onto W_k. Write $W_k = Im(T_1(S_1) - I)z = Im(T_2 - I)$ for the conjugate $T_2 = z^{-1}T_1(S_1)z$ of $T_1(S_1)$. Suppose that z sends $Ker(T_1(S_1) - I) = Im(T_1(S_1) - I)^\perp$ to $W_k^\perp = Ker(T_2 - I)$. Then we have $T_2 = T_0(S_2)$ for an $S_2 = \mathrm{diag}(\lambda, 0)$ with $\lambda \in GL(k,K)$. Let $g_2 = z^{-1}g_0z_1$. Then $g_2 \in Y \setminus \Gamma\Gamma_K$ and $T_0(S_2) = T_2 = g_2T_0(S_0)g_2^{-1}$. Comparing the both sides of the equality $(T_0(S_2) - I)g_2 = g_2(T_0(S_0) - I)$, we see that $g_2 = (b_{ij})_{n\times n}$ has the entries $b_{1j} = b_{2j} = b_{i,\nu+1} = b_{i,\nu+2} = 0$ for all the $j \geq k+1$ and all the $i \notin \{\nu+1,\ldots,\nu+k\}$. If $N_K \neq \Omega(n,K,\Delta)$, then N_K contains the unitary transvections $T_{\nu+1,1}(\lambda = T_0(\mathrm{diag}(\lambda,0,\ldots,0))$ for all the $0 \neq \lambda \in K_0$. Clearly, $T_3(\lambda) = g_2T_{\nu+1,1}(\lambda)g_2^{-1} \in Y$ have the $(1,j)$-th entries zero for all $j \geq 2$. If $T_3(\lambda_1) \notin \Gamma_K$ for a $\lambda_1 \in K_0^*$, then $g_1 = T_3(\lambda_1)$ is what we need. Otherwise, $T_3(\lambda) \in \Gamma_K$ for all the $\lambda \in K_j^*$. By the preceding treatment for the case $k = 1$, we can find a $z_2 \in N_K$ such that $g_1 = z_2^{-1}g_2 \in Y \setminus \Gamma_K$ has its $(1,j)$-th entries zero for all $j \geq 2$, as what we need. In case if $N_K = \Omega(n,K,\Delta)$, then $1 \neq T_0(S_1) \in N_K$ must have $rankS_1 = k \geq 2$. We take a 2-transvection $T_0(\widetilde{S}_0) \in N_K$ with $\widetilde{S}_0 = diag\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and consider $T_3(\lambda) = g_2T_0(\lambda S_0)g_2^{-1} \in Y$ for all the $\lambda \in K_0$. Then $T_3(\lambda)$ has all its (i,j)-th entries zero for $1 \leq i \leq 2$ and $j \geq 3$. If $T_3(\lambda_1) \notin \Gamma_K$ for a $\lambda_1 \in K_0^*$, then $g_1 = T_3(\lambda_1)$ works for it. Otherwise, $T_3(\lambda) \in \Gamma_K$ for all the $\lambda \in K_0^*$. By using the preceding treatment for the case $k = 2$, we know that the g_1 exists, as what we need. □

Proof of Lemma 5.1. We only need to show that Y contains a g_1 as

described in Lemma 5.2. If such g_1 exists, then we can apply Lemma 5.2 and 5.3 to conclude that Y contains an N_E, as Lemma 5.1 says.

In case if $N_K \neq \Omega(n, K, \Delta)$, then N_K contains $T_0(s) = T_{\nu+1,1}(s)$ for all $0 \neq s \in K_0$. Hence $g_2(s) = g_0 T_0(s) g_0^{-1} = (b_{ij})_{n\times n} \in Y$, with the entries $b_{1j} = 0$ for all $j \geq 2$. If $g_2(s_1) \notin \Gamma_K$ for an s_1, then $g_1 = g_2(s_1)$ is what we need. Or if otherwise, then $g_2(s) \in \Gamma_K$ for all the $s \in K_0^*$. Thus g_1 exists by Lemma 5.4. In the other case, we have $N_K = \Omega(n, K, \Delta)$, for each $s \in K^*$. Now we have a 2-transvection $T(s) = T_{\nu+1,2}(s) T_{\nu+2,1}(-s) \in N_K$. Take $g_2(s) = g_2 T(s) g_2^{-1} = (b_{ij})_{n\times n} \in X$. Then $b_{ij} = 0$ holds for all the $1 \leq i \leq 2$ and $3 \leq j \leq n$. If $g_2(s_1) \notin \Gamma_K$ for an s_1, then $g_1 = g_2(s_1)$ works. For if otherwise, we can apply Lemma 5.4 to show the existence of a needed g_1. □

6. Proof of Theorem 3.1

We consider the set Σ of all the fields R between K and F such that X contains an $N_E = SU(n, R, H)$ or $\Omega(n, R, \Delta, L)$. Suppose that Σ contains K which is non-empty, then Σ contains a maximal element E such that there is no subfield R of F proper containing E so that X contains an $N_R = SU(n, R, H)$ or $\Omega(n, R, \Delta, L)$. If $X \geq SL(n, E) \geq SL(n, K)$, then under the requirements on ν and r in Theorem 3.1 we have $n > r$ or $\frac{1}{2}n(n-1) > r$, and hence X is known by Theorem 4.2. We now assume that $X \not\geq SL(n, E)$. Then by Theorem 4.1, $X \cap GL(n, E) \trianglerighteq \widetilde{N}_E$ for some $\widetilde{N}_E = SU(n, E, H)$ or $\Omega(n, E, \Delta, L_1)$ containing N_E. Replacing N_E by $\widetilde{N}_E$, we may assume that $X \cap GL(n, E) \trianglerighteq N_E$. If $X \leq \Gamma_E = GL(n, E) \cdot K^*$, then $X \leq (X \cap GL(n, E)) \cdot K^*$ normalizes N_E, Theorem 3.1 holds . Now we assume that $X \not\leq \Gamma_E$. Then X contains some elements $g \notin \Gamma_E$. If X contains $g_0 \notin \Gamma_E$ as described in Lemma 5.1. Then by Lemma 5.1, X contains an $N_R = SU(n, R, H)$ or $\Omega(n, R, \Delta, L_1)$ for an $R \supsetneq E$. This contradicts the maximality of E. Thus in order to complete the proof of Theorem 3.1, it suffices to find a $g_0 \in X \setminus \Gamma_E$ with enough zero entries as in Lemma 5.1 . In fact, we need only to prove the following Lemma 6.1.

Lemma 6.1. *Under the requirements on ν and r given in Main theorem, we have a group X with $N < X \leq GL(n, F)$ and $X \not\leq \Gamma_E$ contains $g_0 = (d_{ij})_{n\times n} \notin \Gamma_E$, with $d_{1,\nu+1} = 0$ when $N \neq \Omega(n, K, \Delta)$, or with $d_{i,\nu+j} = 0$ for $1 \leq i, j \leq 2$ when $N = \Omega(n, K, \Delta)$.*

Proof. Take a $g = (a_{ij})_{n\times n} \in X \setminus \Gamma_E$. Write $g^{-1} = (\widetilde{a}_{ij})_{n\times n}$. Now we proceed to prove that the subgroup $Y = \langle N, g \rangle$ of X contains a $g_0 \notin \Gamma_E$,

as required.

Case 1. $N = Sp(n, K, H)$ *and* $n > r = [F : K]$. In this case, the n elements $a_{1j} \in F$ are linearly dependent over K, and so there exists a nonzero $(\lambda_1, \ldots, \lambda_n) \in Mat_{1\times n}K$ such that $\sum_{j=1}^{n} \lambda_j a_{1j} = 0$. We can take a matrix $P \in Sp(n, K, H)$ having λ_j as its $(j, \nu+1)$-st entry for $1 \leq j \leq n$. Then, we have $g_0 = gP = (b_{ij})_{n\times n} \in X\backslash\Gamma_K$, with $b_{1,\nu+1} = \sum_{j=1}^{n} \lambda_j a_{1j} = 0$, as desired.

In other cases, we can consider $T_1(S) = gT_0(S)g^{-1} \in Y$ with $T_0(S) \in N$. Denote by M the set of all the admitted S with $T_1(S)$ having the $(1, \nu+1)$-st entry zero when $N_E \neq \Omega(n, E, \Delta)$, or with the $(i, \nu+j)$-th entries zero for $1 \leq i, j \leq 2$ when $N_E = \Omega(n, E, \Delta)$. If we can find an $S_0 \in M$ with $T_1(S_0) \notin \Gamma_E$, then $g_0 = T_1(S_0)$ is what we need. Certainly, such $S_0 \neq 0$. Hence we first proves that $M \neq 0$. Observe that $T_0(S) \in N$ implies that $T_0(\lambda S) \in N$, for all the $\lambda \in K_0$, where $K_0 = K^2$ when $N = \Omega(n, K, \Delta, L)$ and $charK = 2$, while $K_0 = K_J$. Otherwise. Now we have $T_1(\lambda S) = I + \lambda(T_1(S) - I)$. Consequently, $S \in M$ implies $\lambda S \in M$ for all $\lambda \in K_0$. If we can prove $M \neq 0$, then we take a $0 \neq S_0 \in M$ and consider all the $T_0(\lambda S_0) \in Y$ $(\lambda \in K_0^*)$, which have zero entries at the needed positions. If $T_1(\lambda_0 S_0) \notin \Gamma_K$ for a chosen $\lambda = \lambda_0$, then $g_0 = T_1(\lambda S_0)$ is what we need. Otherwise, $T_1(\lambda S_0) \in \Gamma_K$ holds for all the $\lambda \in K_0^*$. Thus we can apply Lemma 5.4 to conclude that Y contains a g_1 as Lemma 5.2 requires. This certainly satisfies our requirement on g_0. Hence we only need to prove that $M \neq 0$. All we need is to find an $S = (s_{kl})_{\nu\times\nu} \neq 0$ with $T_1(S) = (b_{ij})_{n\times n}$ which have entries $b_{i,\nu+j} = \sum_{k,l=1}^{\nu} s_{kl} a_{i,\nu+l} \widetilde{a}_{k,\nu+j} = 0$ for $1 \leq i, j \leq t$, $t = 1$ or 2.

Case 2. $N = Sp(n, K, H)$ and $n(n+2) > 8r$. We have $n = 2\nu$, and hence $\nu(\nu+1)/2 > r$. Also $S' = S$, and hence $s_{kl} = s_{lk}$ for all $1 \leq k, l \leq \nu$. Thus $b_{1,\nu+1} = \sum_{1\leq k\leq l\leq \nu} s_{kl} c_{kl}$ for some $c_{kl} \in F, 1 \leq k \leq l \leq \nu$. Since $\nu(\nu+1)/2 > r = [F : K]$, the $\nu(\nu+1)/2$ elements $c_{kl} \in F$ $(1 \leq k \leq l \leq \nu)$ must be linearly dependent over K. We now choose $s_{kl} \in K$ not all zero so that we can make $\sum_{1\leq k\leq l\leq \nu} s_{kl} d_{kl} = 0$. These s_{kl} constitute a nonzero $S \in M$, as desired.

Case 3. $N = SU(n, K, H) \neq Sp(n, K, H)$ and $\nu^2 > 2r$. In this case, we have $[K : K_J] = 2$, $K = K_J \oplus K_J\theta$ for a given $\theta \in K \setminus K_J$. And $\overline{S'} = S$, i.e., $s_{lk} = \overline{s_{kl}}$ for all $1 \leq k, l \leq \nu$. Write $s_{kl} = \alpha_{kl} + \beta_{kl}\theta$ for all the $1 \leq k \leq l \leq \nu$, with $\alpha_{kl}, \beta_{kl} \in K_J$. Then all $\beta_{kk} = 0$ and $s_{lk} = \alpha_{kl} + \beta_{kl}\beta\theta$. We need to choose $S \neq 0$ so that we can make $b_{1,\nu+1} = \sum_{1\leq k,l\leq \nu} s_{kl} a_{1,\nu+k} \widetilde{a}_{l,\nu+1} = \sum_{1\leq k\leq \nu} \alpha_{kk} c_{kk} + \sum_{1\leq k<l\leq \nu} (\alpha_{kl} c_{kl} + \beta_{kl} c_{lk}) = 0$, where c_{kl} $(1 \leq k, l \leq \nu)$ are some elements in F. We have $\nu^2 > 2r = 2[F : K] = [F : K_J]$.

Consequently, the ν^2 elements $c_{kl} \in F$ $(1 \leq k, l \leq \nu)$ must be linearly dependent over K_J. Now we can choose α_{kl} $(1 \leq k \leq l \leq \nu)$ and β_{kl} $(1 \leq k < l \leq \nu)$not all zero, so that we can make $b_{1,\nu+1} = 0$. These α_{kl}, β_{kl} construct a $0 \neq S \in M$, as what we need.

Case 4. $N = \Omega(n, K, \Delta)$ and $\nu(\nu-1) > 4r$. S is alternating i.e., all $s_{kk} = 0$ $(1 \leq k \leq \nu)$ and all $s_{kl} = -s_{lk}$ $(1 \leq k < l \leq \nu)$. We need to choose $S \neq 0$ which makes $b_{i,\nu+j} = \sum_{1 \leq k < l \leq \nu} s_{kl} c_{kl}^{(ij)}$ for $1 \leq i, j \leq 2$, where $c_{kl}^{(ij)}$ are certain elements in F. First we choose an $S \neq 0$ to make $b_{1,\nu+1} = b_{1,\nu+2} = 0$, i.e., to annihilate the 2-dimensional vector $(b_{1,\nu+1}, b_{1,\nu+2}) = \sum_{1 \leq k < l \leq \nu} s_{kl}(c_{kl}^{(11)}, c_{kl}^{(12)})$ over F. Since $\nu(\nu-1) > 4r$, i.e., $\frac{1}{2}\nu(\nu-1) > 2r$ and $Mat_{1\times 2}F$ has dimension $2r$ over K, the $\frac{1}{2}\nu(\nu-1)$ vectors $(c_{kl}^{(11)}, c_{kl}^{(12)}) \in Mat_{1\times 2}F$ $(1 \leq k < l \leq \nu)$ must be linearly dependent over K. We now choose $s_{kl} \in K$ $(1 \leq k < l \leq \nu)$, not all zero which annihilate $(b_{1,\nu+1}, b_{1,\nu+2})$. These s_{kl} constitute an $S_1 \neq 0$ with $T_1(S_1)$ having the entries $b_{1,\nu+1}, b_{1,\nu+2}$ zero. Moreover, all the $T_1(\lambda S_1)$ $(\lambda \in K^*)$ have their $(1, \nu+1)$-st and $(1, \nu+2)$-nd entries zero. Hence, for each $s \in K^*$, the $T_2(\lambda, s) = T_1(\lambda S_1) T_0 \begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix} T_1(\lambda S_1)^{-1} \in X$ has its $(1, j)$-th entries zero for all $j \geq 2$. If $T_1(\lambda S_1) \in \Gamma_K$ for all $\lambda \in K^*$, then by Lemma 5.4, we can find a g_1 as required in Lemma 5.2. Otherwise, we have $T_1(\lambda_1 S_1) \notin \Gamma_K$ for a λ_1. If $T_2(\lambda_1, s) \in \Gamma_K$ for all the $s \in K^*$, then we can apply again Lemma 5.4 to obtain g_1 as in Lemma 5.2. Otherwise, we have $T_2(\lambda_1, s_1) \notin \Gamma_K$ for an s_1, and we can replace g_1 by $T_2(\lambda_1, s_1) \in X \setminus \Gamma_K$ at the very beginning, to reduce it to the case in which $g = (a_{ij})_{n\times n}$ has the entries $a_{1j} = 0$, for all $j \geq 2$. For such a g all the $T_1(S) = gT_0(S)g^{-1} = (b_{ij})_{n\times n}$ have their entries $b_{1j} = 0$ for all $j \geq 2$. We need only to choose an $S \neq 0$ which makes $b_{2,\nu+1} = b_{2,\nu+2} = 0$. Since $\frac{1}{2}\nu(\nu-1) > 2r = dim_K(Mat_{1\times 2}F)$, $(c_{kl}^{(21)}, c_{kl}^{(22)}) \in Mat_{1\times 2}F$ $(1 \leq k < l \leq \nu)$ should be linearly dependent over K. We now choose $s_{kl} \in K$ $(1 \leq k < l \leq \nu)$, not all zero which annihilates $(b_{1,\nu+1}, b_{1,\nu+2})$. These s_{kl} constitute an $S_1 \neq 0$ to annihilates the entries $b_{1,\nu+1}$ and $b_{1,\nu+2}$ of $T_1(S_1)$. Thus, $T_1(S_1)$ has its entries $b_{i,\nu+j} = 0$ for $1 \leq i, j \leq 2$ and $0 \neq S_1 \in M$, as desired.

Case 5. $N = \Omega(n, K, \Delta, L)$ with $L \neq 0$ (hence $1 \in L$), and $\nu(\nu-1) \geq 2r$. We have $s_{kl} = s_{lk}$ for all $1 \leq k < l \leq \nu$, and $s_{kk} \in L$ for all $1 \leq k \leq \nu$. We need to choose $S \neq 0$ which makes $b_{1,\nu+1} = \sum_{1 \leq k \leq l \leq \nu} s_{kl} c_{kl}$ for certain $c_{kl} \in F, 1 \leq k \leq l \leq \nu$. We have $\frac{1}{2}\nu(\nu-1) + 1 > r = [F : K]$. Hence, the $\frac{1}{2}\nu(\nu-1) + 1$ elements c_{kl} $(1 \leq k < l \leq \nu)$ and c_{11} in F must be linearly dependent over K. We can find $\lambda_{kl} \in K$ $(1 \leq k < l \leq \nu)$ and $\lambda_{11} \in K$,

not all zero, such that $\sum_{1\leq k<l\leq \nu} \lambda_{kl}c_{kl} + \lambda_{11}c_{11} = 0$. If $\lambda_{11} \notin L$ (hence $\lambda_{11} \neq 0$), then we can replace all the λ_{kl} by $\lambda_{kl}\lambda_{11}^{-1}$ ($1 \leq k < l \leq \nu$ or $k = l = 1$), so that reduce it to the case $\lambda_{11} = 1 \in L$. Hence we can always make $\lambda_{11} \in L$. Take $s_{kl} = \lambda_{kl}$ for $1 \leq k < l \leq \nu$ or $k = l = 1$, and take $s_{kk} = 0$ for all the $2 \leq k \leq \nu$. Then all these s_{kl} determine a $0 \neq S \in M$, as desired. This completes the proof of Lemma 6.1. And thus the proof of Theorem 3.1 is completed . □

References

1. M. Aschbacher, *On the maximal subgroups of finite classical groups*, Invent. Math., 76 (1984), 469-514.
2. N. Burgoyne, R. Griess and R. Lyons, *Maximal subgroups and automorphisms of Chevalley groups*, Pacific J. Math., 71 (1977), 365-403.
3. P. Kleidman and M. Liebeck, *The subgroup structure of the finite classical groups* (Cambridge Univ. Press, Cambridge, 1989).
4. S. Li, *Overgroups of $SL(n, K)$ in $GL(n, F)$ ($K \subset F$)*, Acta Math. Sinica, 33 (1990), No. 6, 774-778 (in Chinese).
5. S. Li, *Overgroups of $SU(n, K, f)$ or $\Omega(n, K, Q)$ in $GL(n, K)$*, Geometriae Dedicata 33 (1990), 241-250.
6. S. Li, *Subgroup Structure of Classical Groups*, (Shanghai Scientific and Technical Publishers, Shanghai, 1998).

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 275–291)

A SYMBOLIC CALCULUS ON DEFECT REVISIONS OF AXIOMATIC SYSTEMS

WEI LI

State Key Laboratory of Software Development Environment, Beihang University, Beijing, 100083, P.R.China
E-mail: liwei@nlsde.buaa.edu.cn

A symbolic calculus named as the **R**-calculus is built to revise the defects of axiomatic systems mechanically when some counterexamples are found. The **R**-calculus consists of the rules of logical connective symbols and logical quantifier symbols of first order languages. The concept of reachability, soundness and completeness of the **R**-calculus are introduced. The basic theorem of software testing based on the **R**-calculus is also introduced to help mechanizing model checking.

Keywords: Refutation; Axiomatic systems; Defects; Revision; Symbolic calculus; Model checking; Logical inference system; First order languages.

1. Introduction

In scientific investigations of a specific field, one always extracts the most fundamental and important properties from a large amount of knowledge of the field so as to establish axiomatic systems. Up to now, none of the axiomatic systems that are significant to mathematics or natural sciences are accomplished in one stroke. Usually, an axiomatic system describing the knowledge of a specific field might contain some axioms, laws or rules that are inconsistent with some counterexamples, researcher's observations or experiments. Under such circumstances, these axioms, laws or rules are called defects of the axiomatic system, and counterexamples, researcher's observations or experiments are called refutation by facts.

The aim of this paper is to build a symbolic calculus that can revise the defects mechanically when some refutations by facts of an axiomatic system are found. The idea of building **R**-calculus comes from the following

This work was supported by the National Basic Research Program of China (973 Grant No. 2005CB321901).

comparison:

The proof of a mathematical theorem is a procedure of making analysis and reasonings on the logical connectives and quantifiers inherent in the theorem. Most mathematical theorems can be described by first order languages. Symbolic calculi of various styles on logical connective symbols and quantifier symbols, which are usually called inference systems, were established within the frame of first order languages [1,2]. As a result, the proof of a mathematical theorem, which can be represented by a tree structure called a proof tree, can be accomplished through these symbolic calculi. Furthermore, some computer aided proof systems can be built to perform such a mathematical proof mechanically.

In contrast, the procedure of revising defects in an axiomatic system for some given refutation by facts is also a procedure of making analysis and reasoning on the logical connectives and quantifiers inherent in the axiomatic system and the refutation by facts. This kind of analysis and reasoning can also be described by first order languages. Thus a symbolic calculus, which is a defect-revision system on logical connective symbols and quantifier symbols, should be established within the frame of first order languages as well. The revision of defects in the process can be represented by a tree structure through this symbolic calculus. Similarly, some computer aided defect-revision systems should be built to perform defect revision mechanically.

The aim of this paper is to solve the problem of defect revisions based on the above strategy within the frame of the syntax and semantics of *classical* first order logic. Thus, Section 2 is a brief outline to the basic concepts of the syntax and semantics of first order languages. In Section 3, a symbolic inference system named as **G**-calculus is given to formalize and symbolize mathematical proofs. In Section 4, the concepts of refutation by facts and maximal contraction are introduced to describe the defects contained in an axiomatic system. In Section 5, **R**-calculus is introduced to perform defect-revisions "mechanically". Reachability, soundness and completeness of **R**-calculus are defined in Section 6. And finally, the basic theorem of testing is introduced in Section 7.

2. A formal language for symbolic calculus

First order languages are chosen to be the formal language to describe axiomatic systems and symbolic calculus [1] and [2]. Briefly, a first order language $\mathcal{L}$ has two sets of symbol strings. They are the set of terms and set of formulas. Terms are used to describe constants and functions used in

formulas. The set of terms is defined inductively as below

$$t ::= c \mid x \mid f(t_1, t_2, \ldots, t_n).$$

The formulas of $\mathcal{L}$ are used to describe the laws, rules, principles or non-logical axioms. The set of formulas of $\mathcal{L}$ is defined inductively as follows

$$\begin{aligned} A ::= & \ t_1 \doteq t_2 \mid P(t_1, t_2, \cdots, t_n) \\ & \mid \neg B \mid A \wedge B \mid A \vee B \mid A \supset B \mid \forall x A \mid \exists x A. \end{aligned}$$

A formula is called a sentence if there does not exist free variables in it. A formula is called atomic if it is a predicate ($t_1 \doteq t_2$ is treated as a predicate)[1].

Definition 2.1. A finite consistent set Γ of sentences is called an axiomatic system.

The sentences contained in Γ can be interpreted as non-logical axioms, rules, principles in mathematics and natural sciences.

A model $\mathbf{M}$ is a pair $\langle M, I \rangle$. M is called the domain of $\mathbf{M}$ which is a nonempty set. I is called the interpretation of $\mathbf{M}$ which is a mapping from $\mathcal{L}$ to M. The form $\mathbf{M} \models A$ means that for the given domain M and the interpretation I, A is true in M. $\mathbf{M} \models \Gamma$ means that for every $A \in \Gamma$, $\mathbf{M} \models A$.

Finally, A is called a logical consequence of Γ and is written as $\Gamma \models A$, if and only if for every $\mathbf{M}$, if $\mathbf{M} \models \Gamma$ holds, then $\mathbf{M} \models A$ holds.

3. G-calculus: A logical Inference System

The **G**-calculus is a Gentzen-style logical inference system which is a modified version of the inference system given in [1]. It is used to perform mathematical proofs symbolically. The **G**-calculus is built on sequents. A sequent is formed as $\Gamma \vdash \Delta$, where Γ and Δ could be any finite sets of formulas. Γ is called the antecedent of the sequent and Δ is called the succedent of the sequent, and $\vdash$ denotes the deductive relation [1,2]. The **G**-calculus consists of a logical axiom, the rules for logical connective symbols, rules for quantifier symbols, as well as the cut rule. Each rule of the **G**-calculus is a symbolic manipulation of the logical connective symbols and quantifier symbols **G** is given as below:

Definition 3.1. The logical axiom

Let Γ, Δ, Λ, Θ be sets of formulas, and A be a formula. The following sequent

$$\Gamma, A, \Delta \vdash \Lambda, A, \Theta$$

is called the logical axiom of **G**.

The above logical axiom means that if a formula appears in the succedent of the sequent as well as its antecedent, then the sequent holds without any proof.

Definition 3.2. $\vee$-rule

$$L: \frac{\Gamma, A, \Delta \vdash \Lambda \quad \Gamma, B, \Delta \vdash \Lambda}{\Gamma, A \vee B, \Delta \vdash \Lambda} \qquad R: \frac{\Gamma \vdash \Lambda, A, B, \Theta}{\Gamma \vdash \Lambda, A \vee B, \Theta}.$$

Definition 3.3. $\wedge$-rule

$$L: \frac{\Gamma, A, B, \Delta \vdash \Lambda}{\Gamma, A \wedge B, \Delta \vdash \Lambda} \qquad R: \frac{\Gamma \vdash \Lambda, A, \Theta \quad \Gamma \vdash \Lambda, B, \Theta}{\Gamma \vdash \Lambda, A \wedge B, \Theta}.$$

Definition 3.4. $\supset$-rule

$$L: \frac{\Gamma, \Delta \vdash A, \Lambda \quad B, \Gamma, \Delta \vdash \Lambda}{\Gamma, A \supset B, \Delta \vdash \Lambda} \qquad R: \frac{A, \Gamma \vdash B, \Lambda, \Theta}{\Gamma \vdash \Lambda, A \supset B, \Theta}.$$

Definition 3.5. $\neg$-rule

$$L: \frac{\Gamma, \Delta \vdash A, \Lambda}{\Gamma, \neg A, \Delta \vdash \Lambda} \qquad R: \frac{A, \Gamma \vdash \Lambda, \Theta}{\Gamma \vdash \Lambda, \neg A, \Theta}.$$

Definition 3.6. $\forall$-rule

$$L: \frac{\Gamma, A[t/x], \forall x A(x), \Delta \vdash \Lambda}{\Gamma, \forall x A(x), \Delta \vdash \Lambda} \qquad R: \frac{\Gamma \vdash \Lambda, A[y/x], \Theta}{\Gamma \vdash \Lambda, \forall x A(x), \Theta}.$$

Definition 3.7. $\exists$-rule

$$L: \frac{\Gamma, A[y/x], \Delta \vdash \Lambda}{\Gamma, \exists x A(x), \Delta \vdash \Lambda} \qquad R: \frac{\Gamma \vdash \Lambda, A[t/x], \exists x A(x), \Theta}{\Gamma \vdash \Lambda, \exists x A(x), \Theta}.$$

In the rules $\forall$-R and $\exists$-L, y is a "new" variable. Here "new" means y is different from the variables contained in Γ, Λ, Δ and A. y is formally called an eigen-variable. In the rules $\forall$-L and $\exists$-R, the term t can be any term, but the condition $x \notin FV(t)$ must be satisfied.

The formulas $A \wedge B$, $A \vee B$, $A \supset B$, $\neg A$, $\forall x A(x)$ and $\exists x A(x)$ in the denominators of the rules of the **G**-calculus are called **principal formulas** [1], and A, B, $A[t/x]$, $A[y/x]$ in the numerators of the corresponding rules are called **side formulas**.

Definition 3.8. Cut-rule

$$\frac{\Gamma \vdash \Delta, \alpha \quad \alpha, \Lambda \vdash \Theta}{\Gamma, \Lambda \vdash \Delta, \Theta}.$$

The above cut rule can be derived from other rules in the **G**-calculus [1]. Since in many cases it is convenient to construct formal proofs using the cut rule, it is adopted as a rule in the **G**-calculus.

It should be mentioned that the $\wedge$-L rule and the $\vee$-R rule tell us that in the sequent $A_1, \ldots, A_m \vdash B_1, \ldots, B_n$, the commas on the left hand side of $\vdash$ can be viewed as $\wedge$, and the commas on the right hand side of $\vdash$ can be viewed as $\vee$.

A proof of a given sequent $\Gamma \vdash A$ is a finite tree $\mathcal{T}$. Each node of $\mathcal{T}$ and its direct sons form an application of an inference rule of the **G**-calculus. The root of $\mathcal{T}$ is $\Gamma \vdash A$, and every leaf of $\mathcal{T}$ is an application of the axiom of **G**. The tree $\mathcal{T}$ is called a proof tree of $\Gamma \vdash A$. The sequent $\Gamma \vdash A$ is proved or holds if there exists a proof tree of the sequent. A is called a formal consequence of Γ. According to first order languages, the **G**-calculus is both sound and complete. Thus if A is a formal consequence of Γ, then A is a logical consequence of Γ, and vise versa.

Finally, it is assumed in this paper that two formulas A and B are treated as the same formulas if and only if $A \equiv B$, in other words, $(A \supset B) \wedge (B \supset A)$ is a tautology.

4. Necessary antecedents of formal consequences

In this section, we shall introduce the concept of necessary antecedent of a formal consequence A with respect to Γ.

Definition 4.1. Antecedent set of a proof tree.

Let Γ be a formula set, A be a formula and $\Gamma \vdash A$ be provable. Also let $\mathcal{T}$ be its proof tree with P, Q and R being formulas appearing in the proof tree $\mathcal{T}$.

(1) If Γ' is a formula set and $\Gamma', P \vdash P$ is a leaf of the proof tree $\mathcal{T}$, then the P on the left hand side of $\vdash$ is the antecedent of the P on its right hand side with respect to the proof tree $\mathcal{T}$.

(2) If a node of the proof tree $\mathcal{T}$ is an instance of a right rule of the **G**-calculus with Q appearing as the $B \wedge C$, $B \vee C$, $B \to C$, $\neg B$, $\forall x B(x)$ or $\exists x B(x)$ in the denominator of the rule, i.e. as the principal formula of the rule, and P appearing as the B, C, $B[t/x]$ or $B[y/x]$ in the numerator of the rule, i.e. as the side formula of the rule, then P is an antecedent of Q with respect to the proof tree $\mathcal{T}$.

(3) If a node of the proof tree is an instance of a left rule with P appearing as the $B \wedge C$, $B \vee C$, $B \to C$, $\neg B$, $\forall\, xB(x)$ or $\exists\, xB(x)$ in the denominator of the rule, i.e. as the principal formula of the rule, and Q appearing as the B, C, $B[t/x]$ or $B[y/x]$ in the numerator of the rule, i.e. as the side formula of the rule, then P is an antecedent of Q with respect to the proof tree $\mathcal{T}$. The side formula B, C, $B[t/x]$ or $B[y/x]$ is the antecedent of the side formula on the right hand side of $\vdash$ in the denominator of the rule with respect to the proof tree $\mathcal{T}$.

(4) If P is an antecedent of Q with respect to the proof tree $\mathcal{T}$ and Q is an antecedent of R with respect to the proof tree $\mathcal{T}$, then P is an antecedent of R with respect to the proof tree $\mathcal{T}$.

Let $\mathcal{P}(\Gamma, A, \mathcal{T})$ be the set consisting of all the antecedents of A with respect to the proof tree $\mathcal{T}$. The formula set $\Gamma \cap \mathcal{P}(\Gamma, A, \mathcal{T})$ is called the necessary antecedent set of the formula A with respect to Γ.

Example 4.1.

Consider the sequent $C, A, \forall x(A \to B(x)) \vdash \exists x B(x)$. According to the above definition, The necessary antecedent set of the formal consequence $\exists x B(x)$ of the sequent $C, A, \forall x(A \to B(x)) \vdash \exists x B(x)$ with respect to the proof tree $\mathcal{T}$ is:

$$\{C,\ A,\ \forall x(A \to B(x))\} \cap \{B[t/x], \forall x(A \to B(x)), A \to B[t/x], A\},$$

that is,

$$\{A, \forall x(A \to B(x))\}.$$

Lemma 4.1. *If the sequent $\Gamma \vdash A$ is provable and $\mathcal{T}$ is the proof tree of the sequent, then set $\mathcal{P}(\Gamma, A, \mathcal{T})$ is decidable.*

It should be mentioned that all the proofs of lemmas and theorems are omitted because of the limitation on the length of the paper. The interested readers can refer to [7] for the details.

Example 4.2. Given the sequent $A, A \to B, B \to C \vdash C$, consider its proof tree as follows([1]):

$$\dfrac{\dfrac{A \vdash A^{*2}, B, C \quad A, B^{*2} \vdash B^{*1}, C}{A, (A \to B)^{*2} \vdash B^{*1}, C}\ (2) \qquad C^{*1}, A \to B, B \to C \vdash C\ (3)}{A, A \to B, (B \to C)^{*1} \vdash C}\ (1)$$

we can verify that the necessary antecedent set of C with respect to the proof tree $\mathcal{T}$ is:

$$\{A, A \to B, B \to C\}.$$

It should be mentioned that a provable sequent might have several proof trees, the necessary antecedent sets of A do not have to be unique as well.

5. R-refutation and R-contraction

In this section, we shall discuss the concept of refutation by fact which formally describe counterexamples and researcher's observations and experiments in first order languages.

Definition 5.1. **R**-refutation.

Let Γ be an axiomatic system and Δ be an axiomatic system consisting of finite atomic sentences or their negations. If Δ and Γ are inconsistent, then we name Δ as an **R**-refutation of Γ.

Definition 5.2. Model of **R**-refutation.

Let Δ be an axiomatic system $\{A_1, \ldots, A_n\}$ consisting of finite atomic sentences or their negations such that $\Gamma \models \neg A_1 \vee \cdots \vee \neg A_n$ holds. If there exists a model $\mathbf{M}$ such that $\mathbf{M} \models \Delta$ holds, then we name the model $\mathbf{M}$ as the model of **R**-refutation of Γ with respect to Δ, or Γ is refuted by the model $\mathbf{M}$ with respect to Δ, under the condition that the following set of formulas

$$\Gamma_{\mathbf{M}(\Delta)} := \{\ B \mid B \in \Gamma,\ \mathbf{M} \models B,\ \mathbf{M} \models \Delta\ \}.$$

is maximal, that is, there does not exist another model $\mathbf{M}'$ satisfying $\Gamma_{\mathbf{M}(\Delta)} \subset \Gamma_{\mathbf{M}'(\Delta)}$.

Another form of Δ is $A_1 \wedge \cdots \wedge A_n$. Thus the model of **R**-refutation of Γ with respect to Δ is just the model of **R**-refutation of Γ with respect to $\{A_1 \wedge \cdots \wedge A_n\}$.

Definition 5.3. **R**-contraction.

We name an axiomatic system Λ as an **R**-contraction of an axiomatic system Γ with respect to an **R**-refutation Δ if Λ is a maximal subset of Γ that is consistent with Δ.

Example 5.1. **R**-contraction.

Let Γ be an axiomatic system consisting of the following set of sentences:

$$\{A, A \to B, B \to C, E \to F\}.$$

We can prove that $\Gamma \vdash C$ holds. There are three maximal subsets of Γ that are consistent with $\neg C$:

$$\{A, A \to B, E \to F\}, \quad \{A, B \to C, E \to F\} \quad \{A \to B, B \to C, E \to F\}.$$

They are all **R**-contractions of Γ with respect to $\neg C$. This example shows that the **R**-contraction of an axiomatic system with respect to its **R**-refutation is not unique.

6. R-calculus

In this section we shall define **R**-calculus on logical connective symbols and quantifier symbols. It can deduce all the **R**-contractions for a given **R**-refutation.

Definition 6.1. Let Γ be a finite set of sentences and Δ be a finite axiomatic system consisting of atomic sentences or their negations. We name

$$\Delta \mid \Gamma$$

as an **R**-configuration.

If Γ is an axiomatic system with Δ being the **R**-refutation of Γ, then we name $\Delta \mid \Gamma$ as an **R**-regular configuration.

For convenience, the axiomatic systems Δ and Γ on the two sides of "|" in the configuration can be regarded as either sets of sentences or sequences of sentences. They can be written in the form of A, B, Δ and A, B, Γ as well.

Lemma 6.1. *If* Δ *is* $\{A_1, \ldots, A_n\}$ *and* $\Delta \mid \Gamma$ *is an* **R**-*regular configuration, then*

$$\Gamma \vdash \neg(A_1 \wedge \cdots \wedge A_n)$$

is provable.

Definition 6.2.

$$\Delta \mid \Gamma \Longrightarrow \Delta' \mid \Gamma'$$

is named as an **R**-transition. It transforms the **R**-configuration $\Delta \mid \Gamma$ into the **R**-configuration $\Delta' \mid \Gamma'$.

In particular, the **R**-transition

$$\Delta \mid A, \Gamma \Longrightarrow \Delta \mid \Gamma$$

represents transforming the **R**-configuration $\Delta \mid A, \Gamma$ into $\Delta \mid \Gamma$ or equivalently speaking, the A in the sentence sequence A, Γ on the right hand side of "|" is deleted.

We shall introduce the **R**-calculus henceforth.

Definition 6.3. The **R**-calculus is a set of rules consisting of the **R**-axiom, **R**-logical connective symbol rules, **R**-quantifier symbol rules and **R**-cut rules.

Definition 6.4. **R**-axiom.

$$A, \Delta \mid \neg A, \Gamma \Longrightarrow A, \Delta \mid \Gamma$$

The **R**-axiom shows that if the axiomatic system $\{\neg A, \Gamma\}$ on the right hand side of the **R**-configuration contains $\neg A$, then it is inconsistent with A on the left hand side and hence the $\neg A$ on the right hand side of the **R**-configuration must be deleted.

Definition 6.5. **R**-$\wedge$ rule.

$$\frac{\Delta \mid A, \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid A \wedge B, \Gamma \Longrightarrow \Delta \mid \Gamma} \qquad \frac{\Delta \mid B, \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid A \wedge B, \Gamma \Longrightarrow \Delta \mid \Gamma}.$$

This rule shows that if A is deleted, then $A \wedge B$ must be deleted. Similarly if B is deleted, then $A \wedge B$ must be deleted.

Definition 6.6. **R**-$\vee$ rule.

$$\frac{\Delta \mid A, \Gamma \Longrightarrow \Delta \mid \Gamma \qquad \Delta \mid B, \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid A \vee B, \Gamma \Longrightarrow \Delta \mid \Gamma}.$$

This rule shows that if A and B can be deleted respectively, then $A \vee B$ must be deleted.

Definition 6.7. **R**-$\rightarrow$ rule.

$$\frac{\Delta \mid \neg A, \Gamma \Longrightarrow \Delta \mid \Gamma \qquad \Delta \mid B, \Gamma \Longrightarrow \Gamma}{\Delta \mid A \rightarrow B, \Gamma \Longrightarrow \Delta \mid \Gamma}.$$

The **R**-$\rightarrow$ rule is a special case of the **R**-$\vee$ rule.

Definition 6.8. **R**-$\forall$ rule.

$$\frac{\Delta \mid A[t/x], \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid \forall x A(x), \Gamma \Longrightarrow \Delta \mid \Gamma}.$$

t is a term and x is not a free variable in t.

The **R**-$\forall$ rule can be interpreted as: if there exists a term t such that $A[t/x]$ is inconsistent with Δ, then $\forall x A(x)$ cannot be consistent with Δ and hence it must be deleted.

Definition 6.9. **R**-$\exists$ rule.

$$\frac{\Delta \mid A[y/x], \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid \exists x A(x), \Gamma \Longrightarrow \Delta \mid \Gamma}$$

y is an arbitrary eigen-variable, i.e. the variable y cannot appear as a free variable in the denominator of the **R**-$\exists$ rule.

This rule can be interpreted as: for every eigen-variable y, if $A[y/x]$ is deleted, then $\exists x A(x)$ must be deleted as well.

Definition 6.10. **R**-cut rule.

$$\frac{\Gamma_1, A, \Gamma_2 \vdash C \qquad A \mapsto_{\mathcal{T}} C \qquad \Delta \mid C, \Gamma_2 \Longrightarrow \Delta \mid \Gamma_2}{\Delta \mid \Gamma_1, A, \Gamma_2 \Longrightarrow \Delta \mid \Gamma_1, \Gamma_2}.$$

The numerator of the **R**-cut rule specifies the following conditions:

(1) $\Gamma_1, A, \Gamma_2 \vdash C$ is provable and its proof tree is $\mathcal{T}$. This indicates that the formula C is a formal consequence of Γ_1, A, Γ_2;
(2) The condition $A \mapsto_{\mathcal{T}} C$ is true. This indicates that A is a necessary antecedent of C with respect to the proof tree $\mathcal{T}$;
(3) The **R**-transition $\Delta \mid C, \Gamma_2 \Longrightarrow \Delta \mid \Gamma_2$ is true. This indicates that the formal consequence C of Γ_1, A, Γ_2 is formally refuted by Δ and hence should be deleted.

The **R**-cut rule shows that when all the conditions of the numerator are satisfied, the A on the right hand side of the **R**-configuration $\Delta \mid \Gamma_1, A, \Gamma_2$ in the denominator should be deleted as well.

Lemma 6.2. *The* **R**-$\neg$ *derived rule*

$$\frac{\Delta \mid A', \Gamma \Longrightarrow \Delta \mid \Gamma}{\Delta \mid A, \Gamma \Longrightarrow \Delta \mid \Gamma}$$

is true with A and A' determined by the following table:

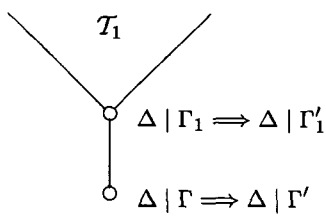
A	$\neg(B \wedge C)$	$\neg(B \vee C)$	$\neg\neg B$	$\neg(B \to C)$	$\neg\forall x B(x)$	$\neg\exists x B(x)$
A'	$\neg B \vee \neg C$	$\neg B \wedge \neg C$	B	$B \wedge \neg C$	$\exists x \neg B(x)$	$\forall x \neg B(x)$

Definition 6.11. **R**-proof tree.

Given an **R**-transition $\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$, the tree $\mathcal{T}$ is named as an **R**-proof tree of the **R**-transition if each node of $\mathcal{T}$ is an **R**-transition and satisfies:

(1) If the **R**-transition $\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$ is an instance of the **R**-axiom, then the **R**-proof tree $\mathcal{T}$ is a single node tree with the **R**-transition as its single node. In this case the single node is also the root of the tree $\mathcal{T}$.

(2) Suppose $\mathcal{T}_1$ is an **R**-proof tree whose root is the **R**-transition $\Delta \mid \Gamma_1 \Longrightarrow \Delta \mid \Gamma_1'$. If the following fraction (a) is an instance of a rule of the **R**-calculus, then the tree structure



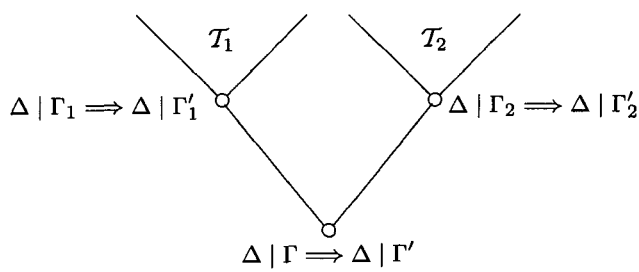
$$\frac{\Delta \mid \Gamma_1 \Longrightarrow \Delta \mid \Gamma_1'}{\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'} \quad (a)$$

is an **R**-proof tree of $\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$.

(3) Suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are **R**-proof trees whose roots are $\Delta \mid \Gamma_1 \Longrightarrow \Delta \mid \Gamma_1'$ and $\Delta \mid \Gamma_2 \Longrightarrow \Delta \mid \Gamma_2'$ respectively. If the following deduction (b)

$$\frac{\Delta \mid \Gamma_1 \Longrightarrow \Delta \mid \Gamma_1' \qquad \Delta \mid \Gamma_2 \Longrightarrow \Delta \mid \Gamma_2'}{\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'} \quad (b)$$

is an instance of a rule of the **R**-calculus, then the tree structure



is an **R**-proof tree of $\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$.

Definition 6.12. We name an **R**-transition provable if there exists an **R**-proof tree with the **R**-transition as its root. Otherwise we say that the **R**-transition $\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$ is unprovable.

We name

$$\Delta \mid \Gamma \Longrightarrow \cdots \Longrightarrow \Delta \mid \Gamma_n$$

as an **R**-transition sequence and denote it as

$$\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma_n$$

with $\Longrightarrow^n$ denoting finite transitions. We say that $\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma_n$ is provable if for every $i \leq n$, the **R**-transition sequence $\Delta \mid \Gamma \Longrightarrow^i \Delta \mid \Gamma_i$ is provable.

Definition 6.13. For a given **R**-configuration $\Delta \mid \Gamma$, if Δ and Γ are consistent, then we name $\Delta \mid \Gamma$ as an **R**-termination.

Example 6.1. Special theory of relativity.

In his famous book [5], Einstein explained how he discovered his special theory of relativity. Let us use the **R**-calculus to verify his informal reasoning. We use the sentence R to stand for the principle of relativity, N_1, N_2, N_3 for the three Newton laws of mechanics respectively, and E for the law of universal gravitation. The Galilean transformation can be described by the following sentence:

$$V: \ \forall x(B(x) \rightarrow A(x)),$$

where $B(x)$ stands for "x is a rigid body" and $A(x)$ stands for "If the velocity of x relative to a coordinate system K is $\vec{V}$ and the velocity of a coordinate system K' relative to K is $\vec{W}$, then the velocity of x relative to K' is $\vec{V} + \vec{W}$".

We can use the following axiomatic system to describe the laws in classical physics:

$$\{\forall x(B(x) \rightarrow A(x)), R, \ N_1, \ N_2, N_3, \ E\}.$$

Using the **G**-calculus, we know that

$$B[c], \forall x(B(x) \rightarrow A(x)) \vdash A[c] \qquad (**)$$

is provable. This sequent can be interpreted as: the classical physics predicts that "*the velocity of light in K' depends on both the relative velocity of K' to K and the velocity of light in K*".

Nonetheless all the scientific experiments and astronomical observations agree with its negation $\neg A[c]$, i.e. "*the light velocity does not depend on the velocity of the luminous body*". In this case, $A[c]$ is refuted by the above scientific experiments and astronomical observations. By his powerful logic intuitions, Einstein reached his conclusion that the Galilean transformation should be deleted.

We can now use the **R**-calculus to verify that Einstein's logic intuition is correct. Let

$$\Gamma := \{B[c], \forall x(B(x) \to A(x)), R,\ N_1,\ N_2, N_3,\ E\},$$
$$\Gamma' := \{R,\ N_1,\ N_2, N_3,\ E\}.$$

The $\neg A[c]$ is an **R**-refutation of Γ that has to be accepted since it is supported by experiments and observations and $\Gamma \vdash A[c]$ is provable. Since the photon can be treated as a particle, $\{B[c], \neg A[c]\}$ is an **R**-refutation of Γ. The verification by the **R**-calculus is as follows.

$$B[c], \neg A[c] \mid A[c], \Gamma' \Longrightarrow B[c], \neg A[c] \mid \Gamma'$$

is an instance of the **R**-axiom and

$$B[c], \neg A[c] \mid \neg B[c], \Gamma' \Longrightarrow B[c], \neg A[c] \mid \Gamma'$$

is an instance of the **R**-axiom as well. According to the **R**-$\to$ rule, we know that the following **R**-transition:

$$B[c], \neg A[c] \mid B[c] \to A[c], \Gamma' \Longrightarrow B[c], \neg A[c] \mid \Gamma'$$

is provable, that is, $B[c] \to A[c]$ is deleted. Finally, according to the **R**-$\forall$ rule we know that the following **R**-transition:

$$B[c], \neg A[c] \mid \forall x(B(x) \to A(x)), \Gamma' \Longrightarrow B[c], \neg A[c] \mid \Gamma'$$

is provable as well. This amounts to $\forall x(B(x) \to A(x))$ being deleted, i.e. we formally verified that the Galilean transformation should be deleted.

In the following example the **R**-calculus can deduce all the possible **R**-contractions.

Example 6.2. **R**-cut rule.

Let Γ be the axiomatic system given in Example 5.1 in Section 5:

$$\{A, A \to B, B \to C, E \to F\}.$$

According to the **G**-calculus, $\Gamma \vdash C$ is provable. Suppose that the logical consequence C of Γ is refuted by fact, that is, $\neg C$ is an **R**-refutation. There are three **R**-contractions of Γ with respect to $\neg C$:

$$\{A, A \to B, E \to F\},\ \{A, B \to C, E \to F\},\ \{A \to B, B \to C, E \to F\}.$$

In this example we will prove that, using the **R**-calculus, we can deduce every **R**-contraction as above. We first consider $\{A, A \to B, E \to F\}$. Let

$$\Gamma_1 := \{A, A \to B\}, \qquad \Gamma_2 := \{E \to F\}.$$

According to the **G**-calculus we know that:

$$\Gamma_1, B \to C, \Gamma_2 \vdash C$$

is provable. Since $B \to C$, which is an element of Γ, is an antecedent of C, $B \to C \mapsto C$ is provable. According to the **R**-axiom we also have:

$$\neg C \mid C, \Gamma_2 \Longrightarrow \neg C \mid \Gamma_2.$$

Thus we can apply the **R**-cut rule to obtain:

$$\neg C \mid \Gamma_1, B \to C, \Gamma_2 \Longrightarrow \neg C \mid \Gamma_1, \Gamma_2.$$

And Γ_1, Γ_2 is just $\{A, A \to B, E \to F\}$.

With the **R**-cut rule we can deduce another **R**-contraction $\{A, B \to C, E \to F\}$. In fact, let

$$\Gamma_1 := \{A\}, \qquad \Gamma_2 := \{B \to C, E \to F\}.$$

According to the **G**-calculus we know that:

$$\text{both } \Gamma_1, A \to B \vdash B \text{ and } B, \Gamma_2 \vdash C$$

are provable. The $A \to B$, which is an element of Γ, is an antecedent of B and hence $A \to B \mapsto B$ is provable. Further, the **R**-axiom indicates that:

$$\neg C \mid C, \Gamma_2 \Longrightarrow \neg C \mid \Gamma_2.$$

Using the **R**-cut rule we have:

$$\neg C \mid \Gamma_1, A \to B, \Gamma_2 \Longrightarrow \neg C \mid \Gamma_1, \Gamma_2$$

holds. And Γ_1, Γ_2 is just $\{A, B \to C, E \to F\}$. Finally, let

$$\Gamma_1 := \emptyset, \qquad \Gamma_2 := \{A \to B, B \to C, E \to F\}.$$

Using the **R**-cut rule we can obtain the third **R**-contraction $\{A \to B, B \to C, E \to F\}$.

7. Reachability, soundness and completeness

For a given **R**-refutation Δ, can the **R**-calculus deduce all the **R**-contractions of Γ with respect to Δ?

Definition 7.1. **R**-reachability.

Suppose that $\Delta \mid \Gamma$ is a given **R**-regular configuration. Γ' is an arbitrary **R**-contraction of Γ with respect to Δ. If there exists an **R**-transition sequence of length n such that:

$$\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma'$$

is provable with $\Delta \mid \Gamma'$ being an **R**-termination, then we say that the **R**-calculus is **R**-reachable.

We can prove the following theorem.

Theorem 7.1. **R**-*reachability.*
The **R**-*calculus is* **R**-*reachable.*

The converse theorem of Theorem 7.1 does not hold. In fact, for an arbitrary **R**-transition sequence of length n:

$$\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma'$$

with $\Delta \mid \Gamma'$ being an **R**-termination, Γ' is not necessarily an **R**-contraction of Γ with respect to Δ. Consider the following example:

Example 7.1. Let Γ be:

$$\{A, A \rightarrow B, B \rightarrow C, A \rightarrow E, E \rightarrow C\}.$$

According to the **G**-calculus, $\Gamma \vdash C$ is provable. Consider the formal refutation $\neg C$ of Γ. We can apply the **R**-cut rule to delete $A \rightarrow B$. Since

$$A, A \rightarrow E, E \rightarrow C \vdash C$$

is provable, we can apply the **R**-cut rule again to delete A and obtain:

$$\{B \rightarrow C, A \rightarrow E, E \rightarrow C\}.$$

Nonetheless the above formula is not a maximal subset of Γ with respect to $\neg C$ because the maximal contraction is:

$$\{A \rightarrow B, B \rightarrow C, A \rightarrow E, E \rightarrow C\}.$$

Lemma 7.1.
If $\Delta \mid \Gamma$ *is an* **R**-*termination, then there does not exist an axiomatic system* $\Gamma' \subset \Gamma$ *such that the* **R**-*transition*

$$\Delta \mid \Gamma \Longrightarrow \Delta \mid \Gamma'$$

is provable.

Definition 7.2. If for an arbitrary **R**-regular configuration $\Delta \mid \Gamma$ and an arbitrary **R**-contraction Γ' of Γ with respect to Δ, that is, for an arbitrary provable **R**-transition sequence of length n:

$$\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma',$$

there always exists a model **M** of **R**-refutation such that both $\mathbf{M} \models \Delta$ and $\Gamma_{\mathbf{M}(\Delta)} = \Gamma'$ hold, then we say that the **R**-calculus is **R**-sound.

Theorem 7.2. *The* **R**-*calculus is* **R**-*sound.*

Definition 7.3. If for an arbitrary **R**-regular configuration $\Delta \mid \Gamma$ and an arbitrary model **M** of **R**-refutation of Γ with respect to Δ, there always exists a provable **R**-transition sequence of length n:

$$\Delta \mid \Gamma \Longrightarrow^{n} \Delta \mid \Gamma_{\mathbf{M}(\Delta)},$$

then we say that the **R**-calculus is **R**-complete.

Theorem 7.3. *The* **R**-*calculus is* **R**-*complete.*

8. Basic theorem of testing

The following example shows that, even if Γ is inconsistent, the **R**-calculus can still deduce every maximal subset of Γ that is consistent with Δ.

Example 8.1. Inconsistent set of axioms.

Let

$$\Delta := \{x \doteq x\}, \qquad \Gamma := \{f(x) \doteq y, f(y) \doteq z, \neg(f(f(x)) \doteq z)\}.$$

Γ is not an axiomatic system. In fact, since $f(x) \doteq y$, we can substitute the variable y in $f(y) \doteq z$ by $f(x)$ to obtain $f(f(x)) \doteq z$. This formula is inconsistent with $\neg(f(f(x)) \doteq z)$.

Using the **R**-cut rule, we can obtain all the maximal subsets of Γ that are consistent with Δ. For instance, let $\Gamma_1 := \{f(x) \doteq y\}$ and $\Gamma_2 := \{\neg(f(f(x)) \doteq z)\}$. First, the transitivity of $\doteq$ indicates that:

$$\Gamma_1, f(y) \doteq z \vdash f(f(x)) \doteq z$$

is provable. The $\neg$-L rule and axiom rule further indicate that

$$f(f(x)) \doteq z, \Gamma_2 \vdash \neg(x \doteq x)$$

is provable. Hence according to the cut rule,

$$\Gamma_1, f(y) \doteq z, \Gamma_2 \vdash \neg(x \doteq x)$$

is provable. It is not difficult to prove that $f(y) \doteq z$ is a necessary antecedent of $\neg(x \doteq x)$. And by the **R**-axiom, the **R**-transition

$$x \doteq x \mid \neg(x \doteq x) \Longrightarrow x \doteq x \mid \emptyset$$

is provable. The **R**-cut rule indicates that $f(y) \doteq z$ should be deleted, i.e.

$$x \doteq x \mid \Gamma \Longrightarrow x \doteq x \mid \{f(x) \doteq y,\ \neg(f(f(x)) \doteq z)\}$$

is provable. And $\{f(x) \doteq y, \neg(f(f(x)) \doteq z)\}$ is a maximal subset of Γ that is consistent with $x \doteq x$. Similarly, we can deduce the other two maximal consistent subsets:

$$\{f(y) \doteq z, \neg(f(f(x)) \doteq z)\} \text{ and } \{f(x) \doteq y, f(y) \doteq z\}.$$

Theorem 8.1. *Basic theorem of testing.*

Let Δ be an arbitrary axiomatic system consisting of finite atomic sentences or their negations, and Γ be an arbitrary finite formula set. If Γ' is an arbitrary maximal subset of Γ that is consistent with Δ, then there exists an **R***-transition sequence of length n*

$$\Delta \mid \Gamma \Longrightarrow^n \Delta \mid \Gamma'$$

that is provable.

In summary, we have introduced the **R**-refutation and constructed the **R**-calculus and proved its **R**-reachability, **R**-soundness and **R**-completeness. Up to now we have two formal calculi and they are the **G**-calculus and **R**-calculus. They solved two different kinds of problems. The **G**-calculus solved the problem of converting mathematical proofs into symbolic calculus. However the **R**-calculus solved the problem of converting the defect-revisions of axiomatic systems into symbolic calculus. Finally, we also proved the basic theorem of testing that converted the defect-revisions of an inconsistent set of axioms into symbolic calculus as well. We believe it is a helpful approach to mechanizing software testings.

References

1. Gallier, J. H., *Logic for Computer Science: Foundations of Automatic Theorem Proving* (John Wiley and Sons, New York, 1987).
2. Paulson, L., *Logic and Computations: Interactive Proof with Cambridge LCF* (Cambridge University Press, Cambridge, 1987).
3. Li, W., *A logical framework for the evolution of specifications*, Proc. 5th European Symp, on Programming, LNCS 788, Edinburgh, 394-408. Springer-Verlag (Berlin, 1994).
4. Burgess, J. P., *Forcing.* In Barwise, J. (eds), Handbook of Mathematical Logic, North-Holland Publishing Company (Amsterdam, 1977).
5. Einstein, A., *Relativity: The Special and General Theory*, Translated by Robert Lawson, W. Henry Holt (New York 1920).
6. Li, W., *A Development Calculus for Specifications*, Science in China (Series F), 46, 390-400 (2003).
7. Li, W., **R**-*Calculus: An Inference System for Belief Revision*, The Computer Journal, 50(4), 378-390 (2007).

This page intentionally left blank

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 293–302)

SOME REMARKS ON THE BURNSIDE PROBLEM FOR LOOPS

PETER PLAUMANN *

Mathematisches Institut, Universität Erlangen-Nürnberg,
Bismarckstrasse 1 1/2, D-91054 Erlangen, Germany
E-mail: peter.plaumann@mi.uni-erlangen.de

LIUDMILA SABININA

Facultad de Ciencias, Universidad Autónoma del Estado de Morelos,
Avenida Universidad 1001, 62209 Cuernavaca, Morelos, México
E-mail: liudmila@buzon.uaem.mx

In a previous paper we have begun to study the Restricted Burnside Problem for some classes of loops reducing it to the positive solution for the class of groups. Here we continue this work, giving a positive answer for those Bruck loops which can be obtained by Glauberman's construction from a Moufang A–loops.

The restricted Burnside problem (see [14]) for a class $\mathcal{L}$ of loops is the question whether the following propositon holds:

There are natural numbers $\beta(d,n)$ such that $|L| \leq \beta(d,n)$ for every finite loop $L \in \mathcal{L}$ having d generators and exponent n.

If this proposition holds for $\mathcal{L}$, we call the class $\mathcal{L}$ an RB–class. In this case we call the numbers $\beta(d,n)$ Burnside coefficients for the class $\mathcal{L}$.

It should be mentioned that the loops considered by us do not have to be power associative. Consequently we say that a loop L has exponent n if for every word w in the free magna with one generator and for every $a \in L$ one has $w(a) = 1$.

By a celebrated result of Zelmanov (see [21]) the class of all groups

*The first author (Peter Plaumann) thanks to SRE, Cátedra especial at UAEM, for support and the second one (Liudmila Sabinina) to CONACyT, Grant C2–44100, to PROMEP and to the University of California.

is an RB–class. Reducing the restricted Burnside problem for the class of Moufang A–loops to the corresponding problem for groups, in [18] we have shown that the class of Moufang A–loops is an RB–class. We shall use this result and the methods applied in [18] to discuss the restricted Burnside problem for further classes of loops.

Amongst the loops treated here one finds the class $\mathcal{V}$ of all Bruck loops obtained by Glauberman's construction (see [9]) from Moufang A–loops and a large class of left Bol loops.

In a concluding section we add some remarks on the representation of a Moufang loop from a group with a triality which may be of interest with respect to our use of Moufang loops to obtain Bruck loops.

1.

In any loop L the associator (x, y, z) of elements $x, y, z \in L$ is defined by the equation

$$(xy)z = [x(yz)](x, y, z).$$

The subloop (L, L, L) generated by all associators in L is called the *associator subloop* of L. Obviously (L, L, L) is a *characteristic* subloop of L, i.e. every automorphism of L fixes (L, L, L). As usual, we define the nuclei of a loop L by

$$\begin{aligned} N_l(L) &= \{u \in L \mid [u, x, y] = 1 \text{ for all } x, y \in L\}, \\ N_m(L) &= \{u \in L \mid [x, u, y] = 1 \text{ for all } x, y \in L\}, \\ N_r(L) &= \{u \in L \mid [x, y, u] = 1 \text{ for all } x, y \in L\}, \\ N(L) &= N_l(L) \cap N_m(L) \cap N_n(L). \end{aligned}$$

Observe that, like the associator subloop, the nuclei of a loop are characteristic, but in general not normal. In [18] we have shown the following

Lemma 1.1. *Let L be a loop in which the nucleus $N(L)$ is normal. If $L/N(L)$ is finite, then (L, L, L) is generated by at most $|L/N(L)|^3$ elements.*

We define the associator series

$$L = A_0(L) \supseteq A_1(L) \supseteq \cdots \supseteq A_i(L) \supseteq A_{i+1}(L) \supseteq \cdots,$$

of a loop L recursively by

$$A_{i+1}(L) = \big(A_i(L), A_i(L), A_i(L)\big).$$

If $A_s(A) = 1 \neq A_{s-1}(A)$, we say that L is α–solvable of length n.

Lemma 1.2. *Let $\mathcal{L}$ be a class of loops such that in every loop in $\mathcal{L}$ the nucleus and all terms of the associator series are normal subloops. If all loops in $\mathcal{L}$ are α–solvable of bounded length s, then $\mathcal{L}$ is an RB–class.*

Proof. Let $\gamma(d,n)$ be Burnside coefficients for the class of all groups and let $L \in \mathcal{L}$ satisfy the hypothesis of the Lemma. By Lemma 1.1 the groups A_{i+1}/A_i are generated by d_i elements, where the numbers $d_0, \cdots, d_{s-1}$ are bounded by a function of d. Hence $|L| \leq \Pi_{i=0}^{s-1}\gamma(d_i,n)$. □

We will now indicate some situations in which Lemma 1.2 can be applied. For any loop Q the right and the left multiplications are defined by

$$xR_a = xa,\ xL_a = ax.$$

The group Mlt_Q generated by the right and left multiplications of Q is called the *multiplication group* of Q. The subgroup Inn_Q of Mlt_Q consisting the mappings fixing the identity of Q is called the group of *inner mappings* of Q.
It is known that the group Inn_Q is generated by the mappings

$$L_{x,y} = L_xL_yL_{yx}^{-1},\ R_{x,y} = R_xR_yR_{xy}^{-1},$$

$$T_a = R_aL_a^{-1}.$$

The loop Q is called an *A–loop* if Inn_Q consists of automorphisms of Q. If only the mappings $L_{x,y}$ are automorphisms of Q, then Q is called an *A_l–loop.*

Theorem 1.1. *Every class of A–loops of bounded solvable length is an RB–class.*

Proof. In an A–loop every characteristic subloop is normal ([2], Theorem 2.2, p. 309). Hence Lemma 1.2 applies. ■

Proposition 1.1. *Let L be a loop with nucleus $N(L)$ for which $L/N(L)$ is a group. Hence L is α–solvable of length at most 2.*

Proof. The associator (L,L,L) is contained in $N(L)$ and is a normal subloop of L (see [13], Lemma 2.5, p. 4). □

Recently Buchsteiner loops, that is loops satisfying the identity

$$x\backslash(xy \cdot z) = (y \cdot zx)/x$$

have found some interest (see [4]).

Theorem 1.2. *The variety of Buchsteiner loops is an RB–class.*

Proof. In a Buchsteiner loop L the condition $(L, L, L) \subseteq N(L)$ is satisfied (see [4], Theorem 7.17). The theorem follows from Proposition 1.1. □

A loop L is called a CC–loop if the the set of the of all left multiplications and the set of all right multiplications of L are closed under conjugation (see [7]). Already in [18] we have shown that the class of CC–loops is an RB–class. Since CC–loops are α-solvable of length ≤ 2, this also follows from our Proposition 1.1.

In [3] loops L considered for which the inner mapping group $\mathfrak{I}_L$ is abelian. Conditions on $\mathfrak{I}_L$ and its embedding in the multiplication loop $M(L)$ are given which imply that L is centrally nilpotent of class ≤ 2. By Proposition 1.1 the loops satisfying these conditions form an RB–class.

2.

A loop is called a *(left) Bruck loop* (see [19, 20]) if the left Bol identity

$$a(b \circ ac) = (a \circ ba)c$$

and the identity

$$(xy)^{-1} = x^{-1}y^{-1}$$

are satisfied. Glauberman ([8, 9]) has described a principle to construct a Bruck loop $G(1/2)$ from a given finite Moufang loop G of odd order (see also ([1], Chapter VII.5, p. 120ff) and ([12], Theorem 6.14, p. 77)). On the set G the multiplication in $G(1/2)$ is definded by

$$a \circ b = a^{\frac{1}{2}} b a^{\frac{1}{2}}.$$

For a class $\mathcal{V}$ of finite Moufang loops of odd order we denote by $\mathcal{V}(1/2)$ the class of all Bruck loops $G(1/2)$ with $G \in \mathcal{V}$.

For a finitely generated loop G the number $rk(G)$ is the minimal number of generators of G.

Proposition 2.1. *For a finite Moufang loop G of odd order the following statements are true:*

(i) G and $G(1/2)$ have the same exponent,
(ii) $rk(G) \leq rk(G(1/2))$.

Proof. Statement (i) follows from the fact that the n–th power of an element of the set G is the same in the loop G and the loop $G(1/2)$. If $\{g_1, \cdots, g_d\}$ is a set of generators for the Bruck loop $G(1/2)$, then the Moufang loop G is generated by $\{g_1^{\frac{1}{2}}, \cdots, g_d^{\frac{1}{2}}\}$. Hence (ii) holds. □

As an immediate consequence of Proposition 2.1 we obtain

Theorem 2.1. *If $\mathcal{L}$ is an RB–class of finite Moufang loops of odd order, then $\mathcal{L}(1/2)$ is an RB–class, too.*

Corollary 2.1. *If $\mathcal{L}$ is a class of finite Moufang A–loops, in particular a class of finite groups, of odd order, then $\mathcal{L}(1/2)$ is an RB–class.*

Proof. By Corollary 11 in [18] the class $\mathcal{A}$ of all Moufang A–loops is an RB–class. Hence $\mathcal{A}(1/2)$ is an RB–class by Theorem 2.1. □

Corollary 2.2. *If $\mathcal{L}$ is a class of finite Moufang loops of prime exponent $p > 2$, then $\mathcal{L}(1/2)$ is an RB–class.*

Proof. By [10] and [16] the class of finite Moufang loops of prime exponent $p > 2$ is an RB–class. □

Proposition 2.2. *For any loop L the following statements hold*

(i) $(ap, q, r) = (p, q, r)$ *for all* $p, q, r \in L$, $a \in N_l(L)$
(ii) If L is a Bol A_l–loop L, then $N_l(L) \triangleleft L$.

Proof. For $p, q, r \in L$, $a \in N_l(L)$ put $s = (ap, q, r)$. Then one has

$$(ap \cdot q)r = (ap \cdot qr)s.$$

On the other hand,

$$(ap \cdot q)r = (a \cdot pq)r = a(pq \cdot r)$$

and

$$(ap \cdot qr)s = \big(a \cdot (p \cdot qr)\big)s = a\big((p \cdot qr)s\big).$$

It follows that $s = (p, q, r)$. Hence (i) holds.

Assume now that L is a Bol A_l–loop. By hypothesis one has

$$N_l(L)L_{x,y} = N_l(L)$$

for all $x, y \in L$.

For $a \in N_l(L)$ and $x, y \in L$ one has $aR_{x,y} = a$. Thus $N_l(L)R_{x,y} = N_l(L)$.

For elements $a \in N_l(L)$, $t \in L$, we put $a^t = t^{-1} \cdot at$. By ([12], §6) one knows that that the element $(a^{-1}t^{-1})(at)$ lies in $N_l(L)$ and as a consequence of (ii) one has

$$(a^{-1}t^{-1})(at) = a^{-1}a^t,$$

since $a^{-1} \in N_l(L)$. Using (i) we obtain

$$(a^t, x, y) = (a^{-1}a^t, x, y) = 1$$

for all $x, y \in L$. Thus $a^t \in N_l(L)$ and we have shown that $N_l(L)$ is invariant under all inner mappings. □

In 1986 T. KRINKO, a student of L. Sabinin, showed in her diploma thesis the following

Theorem 2.2. *If B is a Bol A_l–loop, then the factor loop $B/N_l(B)$ is a Bruck loop.*

Proof. The theorem is easily deduced from ([12], Theorem 6.6, p. 72). □

Theorem 2.3. *Let $\mathcal{V}$ be an RB–class of Moufang loops and let $\mathcal{V}_1$ be the class of all Bol A_l–loops B satisfying the conditions*

(A) $B/N_l(B) \in \mathcal{V}(1/2)$,
(B) $(Q, Q, Q) \lhd Q$ *for all epimorphic images of* B.

Then $\mathcal{V}_1$ is an RB–class.

Proof. For a finite loop $S \in \mathcal{V}_1$ the Bruck loop $T = S/N_l(S) \in \mathcal{V}(1/2)$ has odd order. Hence it follows from ([9], Theorem 14, p. 412) that T is solvable. Thus the associator series

$$T = A_0(T) \supseteq A_1(T) \supseteq \cdots \supseteq A_i(T) \supseteq A_{i+1}(T) \supseteq \cdots,$$

satisfies $A_s(T) = 1$ for some natural number s.

Assume that S has rank d and exponent n. It follows from Theorem 2.1 that $|T| \leq \sigma_{n,d}(\mathcal{V}(1/2))$ where $\sigma_{n,d}(\mathcal{V}(1/2))$ denotes the Burnside constant for the class $\mathcal{V}(1/2)$. In particular the inequality $s \leq \sigma_{n,d}(\mathcal{V}(1/2))$ holds. Thus by Krinko's theorem one has $A_{\sigma_{n,d}(\mathcal{V}(1/2))+1}(S) = 1$. The proposition now follows from Lemma 1.2. □

3.

A group G is called a *group with a triality* if on G one has automorphisms σ, ρ satisfying

$$\sigma^2 = \rho^3 = 1, \quad \sigma\rho\sigma = \rho^{-1},$$

$$(g^{-1}g^{\sigma})(g^{-1}g^{\sigma})^{\rho}(g^{-1}g^{\sigma})^{\rho^2} = 1 \text{ for all } g \in G.$$

The group S generated by σ and ρ is an epimorphic image of the symmetric group S_3.

Glauberman [8] and Doro [6] have observed that groups with a triality can be used to define Moufang loops. Recently Grishkov and Zavarnitsine [11] have described the construction of a Moufang loop from a group (G, σ, ρ) with a triality in the following way. Put $H = \{g \in G \mid g^{\sigma} = g\}$ and $M = \{g^{-1}g^{\sigma} \mid g \in G\}$. Then the multiplication defined on M by

$$m \cdot n = m^{-\rho} n m^{-\rho^2}$$

turns the set M into a Moufang loop $\mathcal{M}(G)$. One observes that

$$m \cdot n = n^{-\rho^2} m n^{-\rho}$$

holds, too. Conversely, every Moufang loop can be obtained this way (see [11, 6, 15]).

In [8] one finds another way to obtain a Bruck loops from groups (Example 4, p. 379): Let G be a torsion group and σ be an automorphism of G satisfying $\sigma^2 = 1$. Assume that the set $G_{\sigma} = \{g^{-1}g^{\sigma} \mid g \in G\}$ does not contain elements of order 2. Then the multiplication $a \circ b = a^{\frac{1}{2}} b a^{\frac{1}{2}}$ defines the structure of a Bruck loop on G_{σ} and every Bruck loop can be obtained in this way([8], p. 382).

Proposition 3.1. *Let G be a group with triality $\langle \sigma, \rho \rangle$ such that the set $G_{\sigma} = \{g^{-1}g^{\sigma} \mid g \in G\}$ does not contain elements of order 2. Then the Bruck loops G_{σ} and $M(G)(1/2)$ are isomorphic.*

Proof. For $x, y \in M$ we denote by

$$x \circ y = x^{\frac{1}{2}} \cdot y \cdot x^{\frac{1}{2}}$$

the multiplication in $M(G)(1/2)$ and by

$$x * y = x^{\frac{1}{2}} y x^{\frac{1}{2}}$$

the multiplication in G_σ. Then

$$\begin{aligned} x \circ y &= (x^{\frac{1}{2}} \cdot y)^{-\rho} x^{\frac{1}{2}} (x^{\frac{1}{2}} \cdot y)^{-\rho^2} \\ &= \left((x^{\frac{1}{2}})^{-\rho} y (x^{\frac{1}{2}})^{-\rho^2}\right)^{-\rho} x^{\frac{1}{2}} \left((x^{\frac{1}{2}})^{-\rho} y (x^{\frac{1}{2}})^{-\rho^2}\right)^{-\rho^2} \\ &= \left(x^{\frac{1}{2}} y^{-\rho} (x^{\frac{1}{2}})^{\rho^2}\right) x^{\frac{1}{2}} \left((x^{\frac{1}{2}})^{\rho} y^{-\rho^2} x^{\frac{1}{2}}\right) \\ &= x^{\frac{1}{2}} y^{-\rho} \left((x^{\frac{1}{2}})^{\rho^2} x^{\frac{1}{2}} (x^{\frac{1}{2}})^{\rho}\right) y^{-\rho^2} x^{\frac{1}{2}} \\ &= x^{\frac{1}{2}} \left(y^{-\rho} y^{-\rho^2}\right) x^{\frac{1}{2}} \\ &= x^{\frac{1}{2}} y x^{\frac{1}{2}} \\ &= x * y, \end{aligned}$$

using for the last two transformations the triality condition. □

Remark 3.1. We observe that not every involutorial automorphism of a group G can be embedded into a triality $\langle \sigma, \rho \rangle$ as the following obvious example shows: Let C be the cyclic group of order 5 and take $\sigma = (x \mapsto x^{-1})$. Then C_σ is isomorphic to C. Since C does not admit an automorphism of order 3, for a triality $\langle \sigma, \rho \rangle$ on C the mapping ρ had to be trivial. But then the triality condition would imply $x^3 = 1$ for every element $x \in C$. This example, however, does not answer the question whether for a given Bruck loop B one always can find $(G, \langle \sigma, \rho \rangle)$ such that B can be obtained in the form $M(G)(1/2)$. Observe that for C, considered as a Bruck loop, this can be done. For this we choose the group $A = C \times C$ and the automorphisms α, τ given by

$$(x, y)^\alpha = (y, x)\,, \; (x, y)^\tau = (y, x^{-1} y^{-1}).$$

Then the relations $\alpha^2 = \tau^3 = 1, \alpha\tau\alpha = \tau^{-1}$ hold and $\langle \alpha, \tau \rangle$ is a triality on A such that $M(A)$ is isomorphic to C. Hence $M(A)(1/2) \cong C$.

In a group (G, ρ, σ) with a triality as before we use the notation $M = \{g^{-1} g^\sigma \mid g \in G\}$. For $x, y \in M$ we put

$$\gamma(x, y) = [y^\rho, x^{\rho^2}].$$

Using the triality relation and the definition of the multiplication $(\cdot)$ on M one shows that

$$[y^{-\rho^2}, x^{-\rho}] = [y^\rho, x^{\rho^2}]$$

holds for all $x, y \in M$.

Theorem 3.1. *Let G be a group with a triality. Then $u \in M$ belongs the nucleus of the Moufang loop $(M, \cdot)$ if and only if u commutes with all*

elements $\gamma(x,y)$, $x,y \in M$. *In particular, for a group* G *with a triality, which is nilpotent of class at most* 2, *the Moufang loop* $\mathcal{M}(G)$ *is a group.*

Proof. The element $u \in M$ lies in the nucleus of M if and only if

$$\text{(A)}\ x \cdot (u \cdot y) = (x \cdot u) \cdot y$$

for all $x,y \in M$ (see [1], Theorem 2.1, p. 114). Computing

$$x \cdot (u \cdot y) = x \cdot (y^{-\rho^2} u y^{-\rho}) = x^{-\rho}(y^{-\rho^2} u y^{-\rho}) x^{-\rho^2}$$

and

$$(x \cdot u) \cdot y = (x^{-\rho} u x^{-\rho^2}) \cdot y = y^{-\rho^2}(x^{-\rho} u x^{-\rho^2}) y^{-\rho},$$

the equation (A) takes the form

$$x^{-\rho}(y^{-\rho^2} u y^{-\rho}) x^{-\rho^2} = y^{-\rho^2} x^{-\rho} u x^{-\rho^2} y^{-\rho}.$$

But this equivalent to

$$u y^{-\rho} x^{-\rho^2} y^{\rho} x^{\rho^2} = y^{\rho^2} x^{\rho} y^{-\rho^2} x^{-\rho} u,$$

i.e. equivalent to

$$u[y^{\rho}, x^{\rho^2}] = [y^{-\rho^2}, x^{-\rho}] u. \qquad \square$$

References

1. R. B. Bruck, *A survey of binary systems*, (Ergebnisse der Mathematik und ihrer Grenzgebiete. Neue Folge. Heft 20.) (Springer-Verlag 1958).
2. R. B. Bruck, L. J. Page, *Loops whose inner mappings are automorphims*, Ann. Math. 63, 308-323 (1956).
3. P. Csörgö, *On connected transversals to abelian subgroups and loop theoretical consequences* (Manuscript).
4. P. Csörgö, A. Drápal, M. K. Kinyon, *Buchsteiner loops* (Submitted).
5. A. Drápal, M. Kinyon, *Buchsteiner loops: associators and constructions* (In preparation).
6. S. Doro, *Simple Moufang loops*, Math. Proc. Cambridge Philos. Soc. 83, no.3, 377–392 (1978).
7. A. Drapal, *Conjugacy closed loops and their multiplication groups*, J. Algebra 272, 838-850 (2004).
8. G. Glauberman, *On loops of odd order*, J. Algebra 1, 374-396 (1964).
9. G.Glauberman, *On loops of odd order*, II J. Algebra 8, 393-414 (1968).
10. A. N. Grishkov, *The weak Burnside problem for Moufang loops of prime period*, Sib. Math. J. 28, no.3, 401-405 (1987).
11. A. N. Grishkov, A. V. Zavarnitsine, *Groups with triality*, J. of Algebra and its Applications 5, 441-464 920060.

12. H. Kiechle, *Theory of K-loops*, Lecture Notes in Mathematics, 1778. (Springer-Verlag 2002).
13. M. K. Kinyon, K. Kunen, *Power–associative, conjugacy closed loops*, arXiv:math/0507278v3 [math.GR] 13 Jan 2006.
14. A. I. Kostrikin , *Around Burnside.* Transl. from the Russ. by James Wiegold. (Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, 20.) (Springer-Verlag 1990).
15. P. O. Mikheev, *Moufang loops and their enveloping groups*, Webs and quasigroups (1993).
16. G. P. Nagy, *Burnside problems for Moufang and Bol loops of small exponent*, Acta Sci. Math. 67, No.3-4, 687-696 (2001).
17. H. O. Pflugfelder, *Quasigroups and loops: introduction.* (Sigma Series in Pure Mathematics, 7) (Heldermann Verlag 1990).
18. P. Plaumann, L. Sabinina, *On nuclearly nilpotent loops of finite exponent*, To appear in Comm, in Algebra.
19. D. A. Robinson, *Bol loops*, Ph. D. Thesis, University of Wisconsin, Wis. (1964).
20. D. A. Robinson, *A loop–theoretic study of right–sided quasigroups*, Ann. Soc. Sci. Bruxelles, Sr. I 93, 7-16 (1979).
21. E. Zelmanov, *Nil rings and periodic groups*, KMS Lecture Notes in Mathematics (1992).
22. E. I. Zelmanov, *Solution of the restricted Burnside problem for groups of odd exponent*, (Russian). Izv. Akad. Nauk SSSR, Ser. Mat. 54, No.1, 42-59 (1990).
23. E. I. Zelmanov, *Solution of the restricted Burnside problem for 2-groups* (Russian). Mat. Sb. 182, No.4, 568-592 (1991).

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 303–334)

RPP SEMIGROUPS, ITS GENERALIZATIONS AND SPECIAL SUBCLASSES * † ‡

K. P. SHUM

Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong, China (SAR)

In this survey article, we will briefly introduce the recent development of rpp semigroups, its generalizations and some of its special subclasses. Some methods of constructions for such semigroups are introduced. Several structure theorems of such semigroups are described.

Keywords: Completely regular semigroups; Quasi-adequate semigroups; Rpp semigroups; Abundant semigroups; Ehresmann semigroups; Cyber groups.

1. Introduction

Recall that a semigroup S is regular if for any $a \in S$, there exists $x \in S$ such that $axa = a$. A regular semigroup S is called completely regular if every element has an inverse with which it commutes. We call a completely regular semigroup Clifford if the set of its idempotents is in its center. It is well known that a Clifford semigroups can always be expressed as a strong semilattice of groups (see [2,28]). Because of this remarkable result, people usually regard semigroups as generalized groups and the class of regular semigroups plays a most important role in the theory of semigroups. We call a regular semigroup S a left Clifford semigroup, for brevity, LC-semigroup [RC-semigroup] if $eS \subseteq Se$ [$Se \subseteq eS$] holds for all $e \in S$. A completely regular semigroup S is said to be an orthogroup if the set of its idempotents $E(S)$ of S forms a subsemigroup of S. A completely regular semigroup S is called inverse if every element of S has a regular inverse in S (see [2]). It was noticed by Yamada (see [50]) that the properties of $E(S)$

*AMS Mathematics subject classification : 20 M10

†This presentation is partially supported by a grant of Wu Jiehyee Charitable Foundation (HK)(2007/08.

‡to Professor L. A. Bokut on the occasion of his 70th anniversary.

of the regular semigroups S can be used to determine the structure of the regular semigroups,in particular, he considered the regular semigroups in which the idempotents satisfy certain identities. The structure of the above regular semigroups have been extensively investigated by many authors since 1950 (see [28,38,39,43,49,50] etc). For further generalization of regular semigroups, we can consider the quasi-regularity of the semigroups. We call a semigroup S quasi-regular if for any $a \in S$, there exists a positive integer n such that a^n is regular. In particular, all regular semigroups are special quasi-regular semigroups. The structure and the properties of quasi-regular semigroups and generalized orthogroups have been investigated by many authors, for instance, see the references in [46]. The structure of orthogroups was also studied. There are also some other ways to generalize orthogroups. Recently, Petrich and Reilly [39] have discussed some special orthogroups, that is, the $\mathcal{C}$ orthogroups. They called an orthogroup S a $\mathcal{C}$ orthogroup if its set of idempotents forms a $\mathcal{C}$ band. The most well known $\mathcal{C}$ bands were tabled in the text of Howie [28] as follows:

(i) regular band: the band satisfying the identity $efge = efege$,

(ii) left regular band: the band satisfying the identity $ef = efe$,

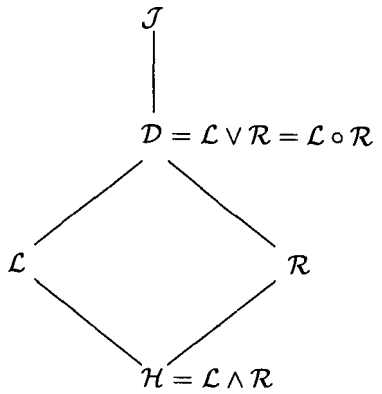
(iii) right regular band: the band satisfying the identity $fe = efe$,

(iv) normal band: the band satisfying the identity $efge = egfe$,

(v) left normal band: the band satisfying the identity $efg = egf$,

(vi) right normal band: the band satisfying the identity $gfe = fge$.

The most known $\mathcal{C}$-orthogroups are completely regular semigroups whose sets of idempotents form regular bands. In the literature, these semigroups are called crytogroups and have been studied by Reilly and Petrich [39] and Guo, Shum and Zhu called them quasi-C-semigroups [23]. Also, regular semigroups whose sets of idempotents form regular bands can be called cryptic groups. In studying the structure of a semigroup S, we usually use the well known Green's relations $\mathcal{L}$, $\mathcal{R}$, $\mathcal{D}$, $\mathcal{H}$ and $\mathcal{J}$. Their relationships are given in the following diagram:



We denote the set of all transformations of the semigroup S by $\mathcal{T}(S)$. For any $f \in \mathcal{T}(S)$, the image of f is denoted by Imf, and the kernel of f is denoted by $Kerf$. By definition,

$$Imf = \{f(x)|x \in S\}, \quad Kerf = \{(x,y) \in S \times S|f(x) = f(y)\},$$

respectively. For any $a \in S$, let $a_r(a_l) \in \mathcal{T}(S^1)$ be the inner right [left] translation on S^1 determined by a:

$$a_r : x \longmapsto xa \quad (a_l : x \longmapsto ax).$$

Then, $\mathcal{L}$, $\mathcal{R}$ and $\mathcal{J}$ can be defined as follows:

$$a\mathcal{L}b \Longleftrightarrow Ima_r = Imb_r,$$
$$a\mathcal{R}b \Longleftrightarrow Ima_l = Imb_l,$$
$$a\mathcal{J}b \Longleftrightarrow S^1aS^1 = S^1bS^1.$$

2. rpp Semigroups

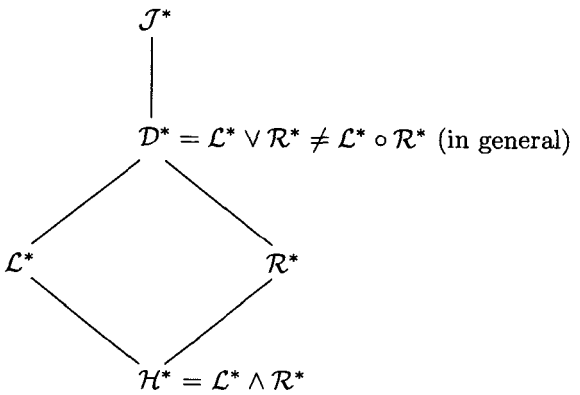
In generalizing regular semigroups, apart from generalizing the definition of regularity, another approach is to modify the Green's relations. In fact, by using this approach, many new results and new semigroups were obtained. Roughly speaking, if we extend the definition of regularity, then this is more or less equivalent to extend the landscape of regular semigroups to a bigger one. But if we modify the Green's relations, then we are going to build a huge pyramid on the ground regular semigroups. For rpp semigroups, we actually start to build, say the 60th floor, of the pyramid of regular semigroups. We call this floor the rpp floor of regular semigroups.

The concept of rpp semigroups was first inspired by rings. Recall that a ring R is a left *p.p.*-ring, for brevity *l.p.p.*-ring if every principal left ideal of

R, regarded as a left R-module, is projective. This concept was first introduced around 1960. Dually, we may define the right *p.p.*-rings. Naturally, we call a ring R a *p.p.*-ring if R is both an *l.p.p.*- and *r.p.p.*-ring. It can be observed that the class of *p.p.*-rings contains the classes of regular rings; hereditary rings; Baer rings; p-q Baer rings and semi-hereditary rings as its proper subclasses. In the literature, *p.p.*-rings have also been extensively studied by many authors. It is noteworthy that the definition of *p.p.*-rings can be extended to semigroups, in particular, Fountain [9] has introduced the concept of *rpp*-monoids in 1977. He called a semigroup abundant [11] if it is both an *lpp*- and *rpp*- semigroup. Similar to *p.p.*-rings, the class of abundant semigroups contains the class of regular semigroups as its proper subclass. The *rpp*-semigroup in which every idempotent is central is called *C-rpp*. It was proved by Fountain [9] that the *C-rpp* monoid is a strong semilattice of left cancellative monoids. Thus, by comparing the result of Fountain [9] with the well known Clifford theorem, we can immediately see that a *C-rpp*-semigroup is indeed a generalization of Clifford semigroup. Since a C-rpp semigroups, by definition, need not to be regular, we can regard C-rpp semigroups as a generalization of regular semigroups, from another approach which is different from just weakening the regularity of the semigroups. Perhaps we need to mention here that the theorem of Fountain on C-rpp monoids has been recently restricted to perfect abundant semigroups. It has been just proved that a perfect abundant semigroup can be expressed as a strong semilattice of cancellative planks. Thus, rpp semigroups can be further studied along this direction. Apart from abundant semigroups, there are still many other important subclasses within the class of rpp semigroups. If we regard the class of rpp semigroups as the 60th floor of the pyramid of regular semigroups , then the class of the abundant semigroups can be regarded as the 58th floor of this pyramid and the class of perfect abundant semigroups may be at the 56th floor.

In the theory of semigroups, it is well known that the idempotents of a semigroup play a crucial role and even more important than the idempotents in a ring. In particular, we notice that the Green's relation $\mathcal{H}$ is a congruence on a semigroups S and indeed it contains an unique idempotent of S. In studying rpp semigroups, Fountain has adopted the Green *-relations on semigroups introduced by Pastijn in [37]. In fact, the Green *-relations on the semigroup S are the usual Green relations on some over semigroup of a semigroup S. Thus, in this connection, the Green *- congruence $\mathcal{H}^*$ on S behaves as the usual Green $\mathcal{H}$ congruence which has a nonempty intersection with the set of idempotents of a semigroup S, in

particular, it contains an unique idempotent. Hence the Green *- relations are indeed the analogous Green's relations on the semigroup S and they are appropriate relations to study the structure of semigroups. The following diagram shows the relationships between the mutual Green *- relations on a semigroup S:



By definition, we have

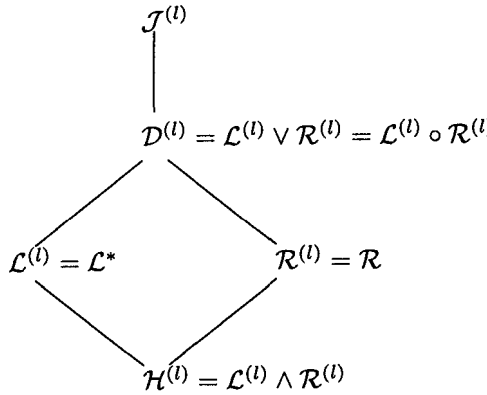
$$a\mathcal{L}^*b \Longleftrightarrow Kera_l = Kerb_l,$$
$$a\mathcal{R}^*b \Longleftrightarrow Kera_r = Kerb_r,$$
$$a\mathcal{J}^*b \Longleftrightarrow J^*(a) = J^*(b)$$

and $J^*(a)$ is the smallest ideal containing a saturated by $\mathcal{L}^*$ and $\mathcal{R}^*$.

Clearly, $\mathcal{L}^*[\mathcal{R}^*]$ is a right [left] congruence on S.

We observe that the Green *- relations $\mathcal{L}$ and $\mathcal{R}$ on a semigroups S can also be explicitly defined as follows: for any two elements $a, b \in S$,, $a\mathcal{L}^*b$ $[a\mathcal{R}^*b]$ if and only if $ax = ay$ $[xa = ya] \Longleftrightarrow bx = by$ $[xb = yb]$, for all $x, y \in S^1$ [37].

In order to amend that $\mathcal{L}^* \vee \mathcal{R}^* \neq \mathcal{L}^* \circ \mathcal{R}^*$ in general, Pastijn [37] has further modified the Green *-relations by using the new Green relations $\mathcal{D}^{(l)}$, $\mathcal{L}^{(l)}$, $\mathcal{R}^{(l)}$ and $\mathcal{H}^{(l)}$ and he has done some ground work for this kind of Green's relations. The above relations together with $\mathcal{J}^{(l)}$ form a new set of Green's relations which lie between the above two mentioned sets of the Green relations on S. We call them the Green (ℓ)-relations. Relationships of Green (ℓ)-relations are shown in the following diagram:



where $a\mathcal{J}^{(l)}b \Longrightarrow \mathcal{J}^{(l)}(a) = \mathcal{J}^{(l)}(b)$, and $\mathcal{J}^{(l)}(a)$ is the smallest ideal containing a which is saturated by $\mathcal{L}^{(l)}$. Clearly, $\mathcal{D}^{(l)}$ refines $\mathcal{J}^{(l)}$ since $\mathcal{J}^{(l)}(a)$ is also saturated by $\mathcal{R}$.

According Fountain, a semigroup S is rpp if all of its principal right ideals $aS^1(a \in S)$, regarded as the right S^1-systems, are projective [9].

Definition 2.1.

(i) A semigroup S is called rpp if and only if for any $a \in S$, the set

$$\mathcal{M}_a = \{e \in E(S) | S^1 a \subseteq Se \ \ (\forall x, y \in S^1)\ ax = ay \Longrightarrow ex = ey\}$$

is nonempty, where $E(S)$ is the set of all idempotents of S.

(ii) An rpp semigroup is said to be strongly rpp [21] if

$$(\forall a \in S)\ \ (\exists!\ e \in \mathcal{M}_a)\ \ ea = a.$$

We now denote such e by $a^\dagger$ where $a^\dagger \in R_a^* \cap E(S)$.

The concept of lpp semigroups is the left-right dual of rpp semigroups. A semigroup is said to be abundant [11] if it is both rpp and lpp. By the definition of $\mathcal{L}^*$ $[\mathcal{R}^*]$, S is rpp (ℓpp, abundant) if and only if S is $\mathcal{L}^*$-abundant, that is, $a^\dagger \in R_a^* \cap E(S)$ ($\mathcal{R}^*$-abundant, $\mathcal{L}^*$-abundant and $\mathcal{R}^*$-abundant). Obviously, $\mathcal{M}_a$ is the set $L_a^* \cap E(S)$, where L_a^* is the $\mathcal{L}^*$-class containing $a \in S$. For $\mathcal{R}^*$, we have similar properties.

It is clear that $\mathcal{L} \leq \mathcal{L}^*$ $[\mathcal{R} \leq \mathcal{R}^*]$ and for $a, b \in RegS$, the set of all regular elements of S, $a\mathcal{L}^*b$ $[a\mathcal{R}^*b]$ implies $a\mathcal{L}b$ $[a\mathcal{R}b]$. In fact, rpp [abundant] semigroups actually form a larger class of semigroups than the class of abundant [regular] semigroups, for example, a left cancellative [cancellative] monoid is clearly rpp [abundant] but it may not be abundant [regular].

By using the (ℓ)-Green relations, a completely regular semigroup S can be described as a $\mathcal{H}^{(l)}(=\mathcal{L}^*\cap\mathcal{R})$-abundant semigroup, in this case, $\mathcal{D}^{(l)}=\mathcal{D}$ is a semilattice congruence on S and every $\mathcal{H}^{(l)}(=\mathcal{H})$-class is a group. A semigroup S is said to be superabundant if S is $\mathcal{H}^*(=\mathcal{L}^*\cap\mathcal{R}^*)$-abundant [11]. For such a semigroup S, we can see that $\mathcal{D}^*=\mathcal{L}^*\circ\mathcal{R}^*$ on S and $\mathcal{D}^{(l)}=\mathcal{D}^*$ is a semilattice congruence on S. In this case, every $\mathcal{H}^*$-class is a cancellative monoid, for brevity, a c-monoid. For super-abundant semigroups, Ren, Kong and Shum have recently obtained some new results in [see 30,42].

We now call a super-abundant semigroup S an ortho-c-monoid if $E(S)\subseteq S$ forms a subsemigroup.

It can be easily seen that a cancellative monoid is not necessarily a group and hence,

$$\{\text{orthogroup}\}\subsetneqq\{\text{ortho-c-monoid}\}.$$

Obviously, super-abundant semigroups (ortho-c-monoid) play a similar role in the class of abundant semigroups to the case of completely regular semigroups (orthogroups) in the class of regular semigroups.

Fountain (see [9,13]) was the first to observe that the Green *- relations can be applied to study abundant semigroups, in particular, super-abundant semigroups. A series of papers have indicated that the Green *-relations are particularly appropriate for the study of abundant semigroups which play the same role as the usual Green's relations in regular semigroups. Recently, X. J. Guo, Shum and Ren have applied the Green (ℓ)- relations to study rpp semigroups, in particular, strongly rpp semigroups. Further applications of Green (ℓ)-relations to rpp semigroups were described in [45].

The relation $\mathcal{R}^*$ on a semigroup S has been generalized to the relation $\widetilde{\mathcal{R}}$ by El-Qallali [4] as following:

$$a\widetilde{\mathcal{R}}b\Longleftrightarrow(\forall e\in E(S))\text{“}ea=a\longleftrightarrow eb=b\text{”}.$$

The $\widetilde{\mathcal{L}}$ on S is the left-right dual of $\widetilde{\mathcal{R}}$ on S.

The following theorem describes the relationships between the Green (ℓ) relations and the Green *-relations.

Theorem 2.1.

(i) $\mathcal{L}\subsetneqq\mathcal{L}^{(l)}$, $\mathcal{D}\subsetneqq\mathcal{D}^{(l)}$, $\mathcal{J}\subsetneqq\mathcal{J}^{(l)}$ *and* $\mathcal{H}\subsetneqq\mathcal{H}^{(l)}$.
(ii) $\mathcal{R}^{(l)}\subsetneqq\mathcal{R}^*$, $\mathcal{D}^{(l)}\subsetneqq\mathcal{D}^*$, $\mathcal{J}^{(l)}\subsetneqq\mathcal{J}^*$ *and* $\mathcal{H}^{(l)}\subsetneqq\mathcal{H}^*$.

We introduce here an important condition on abundant semigroups,namely,the "idempotent connected" condition, in short, the IC condition which was first introduced by El-Qallali and Fountain [8] in 1981. They called a semigroup S an idempotent-connected (for brevity IC) semigroup if for every $a \in S$, with its corresponding idempotents $a^+ \in R_a^*(S)$ and $a^* \in L_a^*(S)$, respectively, there exists a bijection mapping $\alpha : \langle a^+ \rangle \to \langle a^* \rangle$ satisfying $xa = a(x\alpha)$, where $x \in \langle a^+ \rangle$. In particular,an adequate semigroups satisfying the IC condition is called type A and these type A semigroups were studied by Armstrong [1] and then by Lawson [31], respectively. As described by Armstrong and Lawson, we can easily see that the type A semigroup is precisely an analogue of an inverse semigroup in the class of abundant semigroups. On the other hand, El-Qallali and Fountain [7] also discovered that the quasi-adequate semigroups are abundant whose set of idempotents forms a subsemigroup. A special quasi-adequate semigroup is called type W. By studying the structure of a type W semigroup, one can see that the type W semigroup is closely related to the orthodox semigroup which was described by Hall [25]. Recently, X. J. Guo has shown that a semigroup S is a type W semigroup if and only if S is a quasi-adequate semigroup satisfying the IC condition (see [22]). By using the result of X. J. Guo, we can easily see that a type W semigroup is an analogue of orthodox semigroup in the class of regular semigroups. Furthermore, El-Qallali [6] discussed the $\mathcal{L}^*$-unipotent semigroup which is an abundant semigroup whose set of idempotents forms a left regular band. Inspired by his idea, Shum and Ren [40] have recently introduced the concept of $\mathcal{L}^*$–inverse semigroup. They call an IC abundant semigroup an $\mathcal{L}^*$–inverse semigroup if its idempotents form a left regular band. In fact, Ren and Shum [40] have proved that a semigroup S is an $\mathcal{L}^*$-inverse semigroup if and only if S is a left wreath product of a left regular band B and a type A semigroup Γ. Hence, a $\mathcal{L}^*$-inverse semigroup is a special kind of abundant semigroups sitting between the type A semigroup and the type W semigroup. Thus, the $\mathcal{L}^*$-inverse semigroup is an analogue of the left inverse semigroup. Since Yamada [50] also studied the quasi-inverse semigroups within the class of regular semigroups, it is natural to ask what will be the analogous semigroup of this kind of quasi-inverse semigroups within abundant semigroups? In answering this question, Ren and Shum have recently constructed such an analogue, namely the quasi $\mathcal{Q}^*$-inverse semigroups within the class of abundant/rpp semigroups. Their method of construction also used wreath product, in fact, we can see that such an analogue, namely the quasi $\mathcal{Q}^*$-inverse semigroup, is an IC abundant semi-

group whose idempotents form a regular band. For completely regular semigroups, their analogues in the class of abundant semigroups are naturally the superabundant semigroups. Thus, in the class of regular semigroups, the Clifford semigroups, inverse semigroups, left inverse semigroups, quasi-inverse semigroups and orthodox semigroups form a semigroup hierarchy. Their corresponding analogues in the range of rpp semigroups are therefore the C-a semigroups, abundant semigroups, adequate semigroups, type A semigroups, $\mathcal{L}^*$-inverse semigroups, quasi $\mathcal{Q}^*$-inverse semigroups and type W semigroups, respectively. This set of new generalization of semigroups forms a corresponding semigroup hierarchy within the class of rpp semigroups.

It was observed by El-Qallali in [4,5] that the class of abundant semigroups not only contains regular semigroups, but also contains cancellative monoids, semilattices of cancellative monoids, bands of cancellative monoids etc., as its special subclasses. Although many properties of abundant semigroups can be inherited from regular semigroups, but there are some remarkable differences between these two kinds of semigroups, both in structures and properties. We now list some of their main differences below:

(i) The homomorphic image of a regular semigroup is still regular, but the homomorphic image of an abundant semigroup is not necessarily abundant [5]. In other words, if ρ is a congruence on an abundant semigroup S then its quotient semigroup S/ρ is not necessarily abundant.

(ii) The semilattice of regular semigroups is obviously regular, however it is not true for semilattice of abundant semigroups. We cite here an interesting example, given by Fountain [11] for illustration.
Let A be a free monoid generated by elements x, y with identity 1. Let $B = \{ e, f \}$ be a left zero band and form $S = A \cup B$. Define the multiplication on S as follows: for elements in A and B, the multiplications in A and B are the same and we always let 1 be the identity element of S. For any $w \in A \setminus \{ 1 \}$and any $b \in B$, we define $wb = e$ if w is a word starting from x. Otherwise, we define $wb = f$. Also, we define $bw = b$. Then, we can verify that S is a non-commutative semigroup. Clearly A and B are both abundant semigroups and $S = A \cup B$ is a semilattice of the abundant semigroups A and B, however, S itself need not be abundant. This is because that $1, e, f$ are the only idempotents of S, but there does not exist any $\mathcal{L}^*$-relation between the element x and these three idempotents.

(iii) In regular semigroups, it is well-known that Lallement lemma holds,

that is, if ρ is a congruence on a regular semigroup S and $a\rho$ is an idempotent of S/ρ, then there exists an idempotent $e \in S$ such that $e\rho = a\rho$. However, in abundant semigroups or rpp semigroups, the Lallement lemma may not hold. The following is an example.

Let S be a free semigroup generated by x, y. Define a relation ρ on S by $(u, v) \in \rho$ if and only if the words u, v have the same alphabet. Clearly, such a semigroup is abundant and ρ is a congruence on S. We observe that there are only two ρ-classes on S, and consequently, S/ρ forms a left zero semigroup of two elements. Clearly, $x\rho$ is an idempotent but there does not exist an idempotent $e \in S$ such that $e\rho = x\rho$, since S does not contain any idempotent. Thus, Lallement lemma fails to be true in abundant semigroups and rpp semigroups.

From the above examples, we can see that although abundant semigroups and rpp semigroups are generalized regular semigroups, they have many particular properties which are not shared by regular semigroups. Because of these distinguished differences, we can not just simply say that rpp semigroups and abundant semigroups are analogues of regular semigroups.

In study rpp semigroups and abundant semigroups, El-Qallali and Fountain [8] introduced some useful concepts. We first let $L_a^*(S)$ and $R_a^*(S)$ be respectively the $\mathcal{L}^*$–class and the $\mathcal{R}^*$–class containing $a \in S$. Also, we let a^+ and a^* be the idempotents in $R_a^*(S)$ and $L_a^*(S)$, respectively. For the set E of idempotents of S, we let $B = \langle E \rangle$ be the core of S. For any $e \in E$, $\langle e \rangle$ is the subsemigroup generated by the idempotent in eBe. Clearly, $\langle e \rangle$ is generated by all the idempotents f with $f \leq e$.

Another important concept is the concept of good homomorphisms. We call a semigroup homomorphism $\varphi : S \to T$ good if for any $a, b \in S$, $a\mathcal{L}^*(S)b$ implies that $a\varphi\mathcal{L}^*(T)b\varphi$ and $a\mathcal{R}^*(S)b$ implies that $a\varphi\mathcal{R}^*(T)b\varphi$. Correspondingly, we call a semigroup congruence ρ good if its natural homomorphism $\varphi^\sharp : S \to S/\rho$ is good . Equipped with the above definitions, we can see immediately that the good homomorphic image of an rpp [abundant] semigroups is still rpp [abundant]. It was pointed out by El-Qallali [5] that every regular semigroup S is an IC semigroup. This is because for any $a \in S$ and any $a' \in V(a)$, if we define a mapping $\alpha : \langle aa' \rangle \to \langle a'a \rangle$ such that $x\alpha = a'xa$ for $x \in \langle aa' \rangle$, then we can easily observe that such mapping α is the required mapping that makes the regular semigroup S an IC semigroup.Furthermore, El-Qallali [6] introduced the concept of $\mathcal{L}^*$–unipotent semigroup. He called an rpp semigroup S $\mathcal{L}^*$-unipotent if its set of idempotents forms a subsemigroup and every $\mathcal{L}^*$-class of S contains a unique idempotent. As inspired by Yamada [50] and El-Qallali [4], we can define

the $\mathcal{L}^*$-inverse semigroup. We call an abundant semigroup S a $\mathcal{L}^*$-inverse semigroup if S is an IC semigroup whose idempotents form a left regular band [40]. Clearly, every left inverse semigroup is an $\mathcal{L}^*$- inverse semigroup because left inverse semigroup is obviously abundant and IC. To construct the $\mathcal{L}^*$- inverse semigroup, Ren and Shum have recently introduced the concept of "left wreath product" so that the structure theorem of left inverse semigroup obtained by Yamada [50] becomes a special case of their result. In fact, the $\mathcal{L}^*$- inverse semigroup is not only a generalization of left inverse semigroup in the class of rpp semigroups, but is also the most appropriate generalization of left inverse semigroup.

In constructing $\mathcal{L}^*$-inverse semigroups, we need some properties of the $\mathcal{L}^*$- unipotent semigroups.

The following is a crucial observation:

Let S be a quasi-adequate semigroup with band of idempotent E. Then the following conditions are equivalent :

(i) S is $\mathcal{L}^*$-unipotent ;
(ii) $eS \cap fS = efS$ for any $e, f \in E$;
(iii) $ef\mathcal{R}fe$ for any $e, f \in E$;
(ii) Green's relations $\mathcal{R}$ and $\mathcal{J}$ coincide on E ;
(iii) $a^+ea = ea$ for all $e \in E$, $a \in S$, $a^+ \in R_a^* \cap E$.

In closing this Section, we provide an example of $\mathcal{L}^*$-inverse semigroup and show that such semigroup contains the type A semigroup and the left inverse semigroup as its major components [40].

Example 2.1. Let $a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$. Put

$$T = \{2^n a,\ 2^n b \quad | \quad n \geqslant 0\}.$$

It is easy to see that T is a semigroup under the usual matrix multiplication.

We form a semigroup $S = \{a, b, c, d, e, f, g, h, u, v, w, x, y, z, a_n, b_m\}$ by the following Cayley table:

$*$	a	b	c	d	e	f	g	h	u	v	w	x	y	z	a_n	b_j
a	a	a	d	d	f	f	h	h	v	v	x	x	z	z	a_n	a_j
b	b	b	d	d	f	f	h	h	v	v	x	x	z	z	b_n	b_j
c	c	c	c	c	e	e	g	g	u	u	w	w	y	y	c	c
d	d	d	d	d	f	f	h	h	v	v	x	x	z	z	d	d
e	e	e	e	e	c	c	w	w	y	y	g	g	u	u	e	e
f	f	f	f	f	d	d	x	x	z	z	h	h	v	v	f	f
g	g	g	g	g	y	y	c	c	w	w	u	u	e	e	g	g
h	h	h	h	h	z	z	d	d	x	x	v	v	f	f	h	h
u	u	u	u	u	w	w	y	y	c	c	e	e	g	g	u	u
v	v	v	v	v	x	x	z	z	d	d	f	f	h	h	v	v
w	w	w	w	w	u	u	e	e	g	g	y	y	c	c	w	w
x	x	x	x	x	v	v	f	f	h	h	z	z	d	d	x	x
y	y	y	y	y	g	g	u	u	e	e	c	c	w	w	y	y
z	z	z	z	z	h	h	v	v	f	f	d	d	x	x	z	z
a_m	a_m	a_m	d	d	f	f	h	h	v	v	x	x	z	z	a_{m+n}	a_{m+j}
b_i	b_i	b_i	d	d	f	f	h	h	v	v	x	x	z	z	b_{i+n}	b_{i+j}

where $a_n = 2^n a, a_m = 2^m a, b_i = 2^i b$ and $b_j = 2^j b$ for any $n, m, i, j \geqslant 1$.

In fact, the multiplication on S is defined by extending the multiplication on the matrix semigroup T so that S becomes an infinite semigroup, where $E(S) = \{a, b, c, d\}$ is the set of all idempotents of S.We can check that the $\mathcal{L}^*$-classes of S are $\{c, d, e, f, g, h, u, v, w, x, y, z\}$; $\{a, b, a_n, b_i |\ n, i \geqslant 1\}$ and the $\mathcal{R}^*$-classes of S are $\{a, a_n \mid n \geqslant 1\}$; $\{b, b_i \mid i \geqslant 1\}$; $\{c, e, g, u, w, y\}$ and $\{d, f, h, v, x, z\}$, respectively. Thus each $\mathcal{R}^*$-class and each $\mathcal{L}^*$-class of S contains an idempotent and hence by definition, S is an abundant semigroup.

Furthermore, we can easily see that $S_1 = \{a, b, c, d,\ e, f, g, h, u, v, w, x, y, z\}$ is a left inverse subsemigroup of S and $E(S)$ forms a left regular band. Since $\langle a \rangle = \{a, d\}, \langle b \rangle = \{b, d\}, \langle c \rangle = c$ and $\langle d \rangle = d$ in S, where $\langle a \rangle$ denotes the subsemigroup of S generated by all $x \in E(S)$ with $x \leqslant a$, it can be checked that S is an IC abundant semigroup and by definition, S is an $\mathcal{L}^*$-inverse semigroup. Because every element of $S \setminus S_1$ is non-regular, S is not a left inverse semigroup. Also, since the idempotents of S do not commute each other, S is not type A. This example illustrates that the class of type A semigroups and the class of left inverse semigroups are two proper subclasses of the class of $\mathcal{L}^*$-inverse semigroups.

3. Quasi-adequate semigroups and cyber groups

We now introduce some special rpp semigroups, namely the good quasi-adequate semigroups and cyber groups. As mentioned above, the homo-

morphic image of an abundant [quasi-adequate] semigroup is not necessarily abundant [quasi-adequate], El-Qallali and Fountain in [7] therefore considered the good homomorphism again on quasi-adequate semigroups. It was shown in [7] that the minimum adequate good congruence γ exists in a quasi-adequate semigroup. Thus if S is an orthodox semigroup, then γ is just the minimum inverse semigroup congruence on S (see [43]). It is noted that the quasi-adequate semigroups within the class of abundant semigroups are a kind of semigroups corresponding to the class of orthodox semigroups within the class of regular semigroups.

For an abundant semigroup S, since $L_a^* \cap E(S) \neq \emptyset$ and $R_a^* \cap E(S) \neq \emptyset$, for all $a \in S$, we use a^* to denote any idempotent in L_a^*, and use $a^\dagger$ to denote any idempotent in R_a^*. In order to locate the minimum adequate good congruence on a quasi-adequate semigroup S, Fountain [10] defined a relation δ on S by

$$a\delta b \Longleftrightarrow E(a^\dagger)aE(a^*) = E(b^\dagger)bE(b^*)$$

for some $a^\dagger, a^*, b^\dagger, b^*$, where $E(e)$ is the $\mathcal{J}$-class containing e in the band $E(S), e \in E(S)$. Because $E(S) = \bigcup_{\alpha \in Y} E_\alpha$, where each E_α is a rectangular band and Y is a semilattice, we see that $E(e)$ is the one of the rectangular bands E_α . Moreover, if a semigroup S is orthodox , then δ is the minimum inverse congruence. The following result was proved in [7]:

Lemma 3.1. *The following statements hold on a quasi-adequate semigroup S.*

(i) The relation δ is a congruence on S if and only if for all $a, b \in S$, $aE(a^*)E(b^\dagger)b \subseteq E((ab)^\dagger)abE((ab)^*)$,
(ii) $\mathcal{H}^* \cap \delta = \iota$, where $\mathcal{H}^* = \mathcal{L}^* \cap \mathcal{R}^*$.

We note here that an IC-adequate semigroup is type- A and it was proved in [1] that an adequate semigroup S is type- A if and only if for all $e \in E(S), a \in S$, $ae = (ae)^\dagger a$ *and* $ea = a(ea)^*$.

Lawson [31] has shown that the type -A semigroups play a crucial role in the structure of adequate semigroups. Naturally, one would ask: Can we find some special adequate semigroups which correspond to some type A semigroups among the adequate semigroups ? For this question , we give the following weak type A conditions.

Definition 3.1. *Let S be a quasi-adequate semigroup.* Then we define the the weak type- A condition if for any $a \in S$ and any $e, f \in E(S)$,

(i) $aef\mathcal{R}^*(ae)^\dagger af$,
(ii) $fea\mathcal{L}^*fa(ea)^*$.

We formulate a definition:

Definition 3.2. A quasi-adequate semigroup S is called good if the following conditions hold:

(i) the relation δ on S is a congruence,
(ii) the weak type A condition holds on S.

We can easily see that δ is the minimum adequate good congruence on a good quasi-adequate semigroup. We now formulate the following theorem:

Theorem 3.1. *A regular band E is a good quasi-adequate semigroup. Conversely, if a good quasi-adequate semigroup is a band, then it is regular.*

In the above theorem, we observe that the idempotents set of a good quasi-adequate semigroup forms a regular band. Also, an adequate semigroup S is good if S is good as a quasi-adequate semigroup, that is, the weak type-A condition holds on S (since $\delta = \gamma = \iota$ in an adequate semigroup). According to the above observation, it is clear that a type-A semigroup is always good adequate but not conversely. The following is an example.

Example 3.1. (see [30]) Let $S = \{e, f, g, h, z, a, b, c\}$ having the following Cayley table

	e	f	g	h	z	a	b	c
e	e	f	g	z	z	a	b	c
f	f	f	z	z	z	b	b	z
g	g	z	g	z	z	c	z	c
h	z	z	z	h	z	z	z	z
z	z	z	z	z	z	z	z	z
a	z	z	z	a	z	z	z	z
b	z	z	z	b	z	z	z	z
c	z	z	z	c	z	z	z	z

In the above table, $\mathcal{L}^*$-classes of S are $\{a, b, c, h\}$, $\{e\}$, $\{f\}$, $\{g\}$, and $\{z\}$. And we can see that the $\mathcal{R}^*$-classes of S are $\{a, e\}$, $\{b, f\}$, $\{c, g\}$, $\{h\}$, $\{z\}$. Thus S is an abundant semigroup. Also we observe that the idempotents e, f, g, h, z commute with each other. Obviously $\delta = \gamma = \iota$ on S. We can check that the weak type A condition holds on S. Hence, S is a good adequate semigroup. However, since $fa = b \neq a = ah = a(fa)^*$, this shows

that the type-A conditions does not hold on S. Hence S is a good adequate semigroup but not type A.

Inspired by the lpp and rpp conditions for good quasi-adequate semigroups, Y. Q. Guo X. J. Guo and Shum introduced the concept of left GC-lpp semigroup in [16]. They called an lpp semigroup S left GC-lpp if $E(S)$ forms a left regular band and $ae = (ae)^{\dagger}a$ holds for all elements $a \in S$ and $e \in E(S)$, where $E(S)$ is the set of idempotents of S and $(ae)^{\dagger}$ is a corresponding idempotent of ae lying in the $\mathcal{R}^*$-class R^*_{ae}. The left GC-lpp semigroups can be described by using McAlister triples and weak semidirect products in [18]. Some construction theorems of lpp semigroups were also established in [14,16]. It can be easily seen that the class of lpp semigroups is a common generalization of the classes of left type-A semigroups and right inverse semigroups. More precisely, a left type-A semigroup is just a left GC-lpp semigroup whose band of idempotents forms a semilattice, and conversely [16].

We call an abundant semigroup S a left cyber group [16] if the set of idempotents $E(S)$ is a left regular band (that is, a band satisfying the identity $xyx = xy$ and also $ae = (ae)^{\dagger}a$ holds for all $a \in S$ and $e \in E(S)$). Clearly, left cyber groups are special left GC-lpp semigroups. In fact, a cyber group is a natural generalization of a regular cryptogroups within the class of abundant/rpp semigroups.

A lpp semigroup S is called strongly lpp if for all $x \in S$, there exists a unique idempotent x^*, which is $\mathcal{R}^*$-related to x, such that $x = xx^*$. Dually, we may define strongly rpp semigroup. In the literature, Y. Q. Guo, Shum and Zhu [22] were the first who investigated the so called right C-rpp semigroup. The structure of right C-rpp semigroups have recently been investigated by Shum and Ren [49]. Dually, a left C-lpp semigroup S can be defined as an lpp semigroup in which $\mathcal{D}^{(r)}$ is a congruence on S and $eS \subseteq Se$ for all $e \in E(S)$, where $\mathcal{D}^{(r)}$ is the smallest equivalence relation on the semigroup S containing $\mathcal{R}^*$ and $\mathcal{L}$. According to [14], an lpp semigroup S is left C-lpp [22] if and only if $\mathcal{D}^{(r)}$ is a semilattice congruence on S such that $\mathcal{D}|_{E(S)} = \mathcal{L}|_{E(S)}$. In particular, it is also known that $E(S)$ forms a left regular band and $a^{\dagger} = a^*$, for all $a \in S$. In view of the results in [41], a super abundant semigroup S is a left super cyber group if the set $E(S)$ of idempotents of the super abundant semigroup S forms a left regular band. In addition, new structure theorems of generalized superabundant semigroups have been given in [30,42].

The structure of left cyber groups can be described by using McAlister triple and weak semi-direct product (see [18,43]). We have the following

properties of the left GC-lpp semigroups [16]:

(i) Let S be an lpp semigroup whose set of idempotents forms a left regular band. Then S is a left GC-lpp semigroup if and only if for all $a \in S$ and $e \in E(S)$, there exists $b \in S$ such that $ae = ba$.

(ii) Let S be a left GC-lpp semigroup. Then the following conditions hold:

(a) Define a relation

$$\gamma = \{(x, y) \in S \times S : x = ey, e \in E(y^\dagger)\} on S,$$

where $E(y^\dagger)$ is the $\mathcal{D}$-class of $E(S)$ containing the idempotent $y^\dagger \in R_y^* \bigcap E(S)$. Then γ is the minimum $\mathcal{R}^*$-congruence (a $\mathcal{R}^*$-congruence is a congruence preserving the relation $\mathcal{R}^*$) such that S/γ is a left type-A semigroup;

(b) $\gamma \bigcap \mathcal{R}^* = \iota_S$ (the identity relation on S).

We call a non-empty set M a partial groupoid if there is a partial binary operation defined on M. A partial groupoid $(M, \circ)$ is called a partial band groupoid if for all $a \in M$, $a \circ a$ is defined, then $a \circ a = a$. Moreover, $(M, \circ)$ is called a strong band groupoid provided that for all $x, y, z \in M$, if $y \circ z$ and $x \circ (y \circ z)$ are defined, then $x \circ y$ is also defined. A nonempty subset N of $(M, \circ)$ is called a left regular subband of M if $(N, \circ)$ is a left regular band.

We now construct a left GC-lpp semigroup by using McAlister triple:

(McAlister triple) Let $\mathcal{X}$ be a strong band groupoid, $\mathcal{Y}$ a left regular subband of $\mathcal{X}$ with structure semilattice Y in which there is $i \in \mathcal{X}$ but

$$i \notin \mathcal{Y}, \text{ and } \mathcal{Y} = \bigcup_{\alpha \in Y} E_\alpha$$

the semilattice decomposition of $\mathcal{Y}$ over left zero bands E_α ($\alpha \in Y$). Let T be a left type-A semigroup with semilattice Y of idempotents such that $\mathcal{X}$ is a left T^1-system with $1x = x$ for all $x \in \mathcal{X}$, but $vy \notin \mathcal{Y} \bigcup \{i\}$ for all $y \in \mathcal{Y} \bigcup \{i\}$, $v \in T$.

Suppose that the following conditions hold:

(M1) $T^1 \mathcal{Y} \bigcup \{i\} = \mathcal{X}$;

(M2) for all $x \in \mathcal{Y}$ and $t \in T$, $x \circ (ti) = x$ if and only if $x \in E_{t^\dagger}$;

(M3) for all $x \in \mathcal{Y}$ and $t \in T$, if $x \circ (ti) = x$, then $x \circ (ty) \in \mathcal{Y}$;

(M4) for all $x, y \in \mathcal{Y}, Z \in \mathcal{X} \backslash \mathcal{Y} \bigcup \{i\}$ and $t \in T$, if $y \circ z$ and $x \circ (t(y \circ z))$ are defined, then $(x \circ ty) \circ (tz)$ is also defined and $x \circ (t(y \circ z)) = (x \circ (ty)) \circ (tz)$;

(M5) for all $x, y \in \mathcal{Y}$ and $\alpha \in Y$, if $x \circ (\alpha i) = x$, then $x \circ (\alpha y) = x \circ y$.

Put

$$\mathcal{M} = \mathcal{M}(T, \mathcal{X}, \mathcal{Y}) = \{(y,t) \in \mathcal{Y} \times T : y \circ (ti) = y\}$$

and define a multiplication on $\mathcal{M}$ by

$$(x,t) \star (y,u) = (x \circ (ty), tu).$$

Then $(\mathcal{M}, \star)$ is a left GC-lpp semigroup such that $\mathcal{M}/\gamma \cong T$. Conversely, any such semigroup can be so constructed. In this case, the triple $(T, \mathcal{X}, \mathcal{Y})$ is called a McAlister triple.

The following is an important tool for constructing the left GC-lpp semigroups:

(Weak semidirect product) Let T be a left type-A semigroup with semilattice Y of idempotents and $L = (Y, L_\alpha)$ the semilattice decomposition of the left regular band L into left zero bands L_α ($\alpha \in Y$). Denote by $End(L)$ the semigroup of endomorphisms of L. Suppose that φ is a mapping of T into $End(L)$ satisfying the following conditions:

(WS1) for all $u \in T$ and $x \in L_\alpha$, $\varphi_u x \in L_{(u\alpha)^\dagger}$;

(WS2) for all $u, v \in T$, there exists $(uv)^\circ \in L_{(uv)^\dagger}$ such that $\varphi_u \varphi_v = \lambda_{(uv)^\circ} \varphi_{uv}$, where φ_u is the image of $u \in T$ and λ_x is the inner left translation determined by $x \in L$.

Now, form the following set WS by

$$WS = WS(L, T; \varphi) = \{(x,t) \in L \times T : x \in L_{t^\dagger}\}$$

and define a multiplication $\bullet$ on WS by

$$(x,u) \bullet (y,v) = (x(\varphi_u y), uv).$$

Then $(WS, \bullet)$ is a left GC-lpp semigroup with $WS/\gamma = T$. Conversely, any left GC-lpp semigroup can be constructed in the above manner.

We call $(WS, \bullet)$ the weak semidirect product of the left regular band L together with the left type-A semigroup T over the mapping $\varphi : T \to End(L)$.

In the above construction, $(WS, \bullet)$ is named the weak semidirect product of a left regular band L, together with a left type-A semigroup T with respect to the mapping $\varphi : T \to End(L)$.

By using McAlister triples and weak semidirect products of semigroups, we obtain a construction theorem for right inverse semigroups.

Theorem 3.2. *Let $(T, \mathcal{X}, \mathcal{Y})$ be a McAlister triple in which T is an inverse semigroup. Then $\mathcal{M}(T, \mathcal{X}, \mathcal{Y})$ is a right inverse semigroup. Conversely, every right inverse semigroup can be so constructed in the above manner.*

In the following, we characterize the left cyber groups by considering the quasi-adequate semigroups [20].

Theorem 3.3. *Let S be a quasi-adequate semigroup. Then S is a left cyber group if and only if the relation γ is a congruence on S such that S/γ is a left type-A semigroup.*

Another important subclass of rpp semigroups is the perfect rectangular band of adequate semigroups [15]. For this topic, we have the following theorem.

Theorem 3.4. *Let S be an abundant rectangular band $I \times \Lambda$ of adequate semigroups $S_{i\lambda}$ for $(i, \lambda) \in I \times \Lambda$. Then S is a perfect rectangular band of adequate semigroups $S_{i\lambda}$ for $(i, \lambda) \in I \times \Lambda$ if and only if $RegS$ is a perfect rectangular band $I \times \Lambda$ of inverse semigroups $RegS_{i\lambda}$ for $(i, \lambda) \in I \times \Lambda$.*

In view of the above theorem , we establish the following theorem.

Theorem 3.5. *Let S be an abundant perfect rectangular band $I \times \Lambda$ of adequate semigroups $S_{i\lambda}$ for $(i, \lambda) \in I \times \Lambda$. Then*

(1) every $S_{i\lambda}$ is a quasi-ideal adequate transversal of S;

(2) S is a locally adequate semigroup, that is, the local submonoid eSe of S is an adequate semigroup for all $e \in E(S)$.

Because an IC abundant perfect rectangular band of type-A semigroups can be described as an IC abundant semigroup which can be expressed as a disjoint union of quasi-ideal adequate transversals, it is natural to ask whether an abundant semigroup which is the disjoint union of quasi-ideal adequate transversals can be expressible as a perfect rectangular band of adequate semigroups or not ? We do not know whether the general case holds or not, however, if the semigroup satisfies the regularity condition, then the answer to this problem is positive.

Theorem 3.6. *(see [15]) Let S be an abundant semigroup satisfying the regularity condition, that is, $RegS$ is a subsemigroup of S. Then S can be expressed as a perfect rectangular band of adequate semigroups if and only if S is a disjoint union of quasi-ideal adequate transversals [21].*

We also have the following construction theorem for abundant perfect rectangular band of adequate semigroups.

Theorem 3.7. *(see [33,35]) Let I, Λ be indexed sets and T an adequate semigroup. Then the Rees matrix semigroup $\mathcal{M}(T, I, \Lambda; P)$ is expressible as an abundant perfect rectangular band of adequate semigroups and each of which is isomorphic to T.*

Conversely, an abundant perfect rectangular band of adequate semigroups can be constructed in the above manner.

4. Quasi-C-Ehresmann semigroups and their subclasses

We consider a generalization of rpp semigroups, namely, the quasi-C-Ehresmann semigroups which are special U-semiabundant semigroups that are orthodox. In generalizing the regular semigroups, we also use the Green $*$- relations. Since $E(S)$ of the semigroup S plays an essential role in the structure of abundant semigroup so that the semigroup S itself can be more or less determined by the properties of $E(S)$. It was noticed by Lawson [35] that we only need to pay special attention to a subset U of $E(S)$ instead of the whole set $E(S)$. In particular, if U is a subset of $E(S)$ of a semiabundant semigroup S then U is perhaps good enough to provide sufficient information for the whole semigroup S. In order to emphasize that U is a subset of $E(S)$ of a semigroup S, we usually denote this semigroup by $S(U)$. Then, he introduced the concept of U-semiabundant semigroup. Unlike the usual Green's relations, Lawson [35] observed that $\widetilde{\mathcal{L}}^U$ is not necessarily a right congruence on S and $\widetilde{\mathcal{R}}^U$ is not necessarily a left congruence on S. Due to the above observation, Fountain, Gomes and Gould in [12] called a semigroup $S(U)$ satisfied the condition C [(CR) , (CL)] if $\widetilde{\mathcal{L}}^U$ is a right congruence on S [$\widetilde{\mathcal{R}}^U$ is a left congruence]. They also called a U-semiabundant semigroup $S(U)$ an Ehresmann semigroup if U is a subsemilattice of S and $S(U)$ satisfies both conditions (CR) and (CL). An Ehresmann semigroup $S(U)$ in which the idempotents of U are central is called a C-Ehresmann semigroup. The structure of fundamental C-Ehresmann semigroups have been described in [13]. In fact, this kind of semigroups is a special generalization of Clifford semigroups within the class of U-semiabundant semigroups.

On the other hand, for any element a of a semigroup S, Y. He [27] defined the orbit of a in $U \subseteq E(S)$ by $U_a = \{u \in U |\ ua = a = au\}$ and defined $\widetilde{\mathcal{Q}}^U = \{(a, b) \in S \times S |\ U_a = U_b\}$. Clearly, $\widetilde{Q}_a^U$ is an equivalence class of the semigroup S containing $a \in S$. The semigroup $S(U)$ is called U-

liberal in [26] if every $\tilde{Q}$^-class of 5 contains an element in *U.* It is routine to check that a $\tilde{Q}$^-class contains at most one element of *U.* Denote the unique element in $\tilde{Q}$^*C**U,* if it exists, by *afj.* The structure of [/-liberal semigroups and left C-Ehresmann semigroups have also been recently investigated by Y. He [26], G. Li-Shum [35] and others.

For a [/-liberal semigroup *S(U),* we call the following condition the ET-condition:

$$(\mathrm{Va},6\text{€}5)\ (ab)b\ \tilde{V}(U)\ albfr,$$

where

$$\tilde{\mathcal{D}}(U) = \{(\mathrm{e}, f)\text{€}UxU\backslash\ (3g\ £\ U)\ e\tilde{U}gZf\}.$$

Now, an [/-liberal semigroup *S{U)* is called orthodox if *U* is a subsemigroup of *S(U)* and the ET-condition holds on *S(U)* [36]. We observe that *U*-liberal semigroups are not always orthodox. For example, if we let *S* = *{l,e,a,b,* /, 0} be a semigroup with the following Cayley table,

	1	e	*a*	1	*b*	O
1	**1**	***e***	***a***	1	***b***	O
e	e	***e***	*a*	O	**O**	O
a	a	**O**	O	a	e	O
/	/	0	0	/	*b*	0
b	*b*	*b*	/	0	0	0
0	0	0	0	0	0	0

then *S* is a finite inverse monoid and hence *S(E(S))* is a ^(S'j-liberal. Clearly, *E(S)* is a subsemigroup of *S{E(S)).* Since $a°_{E,S}$^ = 1, $e\%_{S}$-, — *e* and $(ae)°_{E(S)}$ = 0, we see that the ET-condition does not always hold in *S(E(S)).* Thus, the above semigroup is an [/-liberal semigroup but if is not orthodox.

We note that the orthodox [/-liberal semigroups are generalization of orthogroups within the class of [/-semiabundant semigroups.

As an analogue of regular orthogroups , we call an orthodox [/-liberal semigroup *S(U)* quasi-C-Ehresmann if *U* forms a regular band. Clearly, quasi-C-Ehresmann semigroup is a generalized rpp-semigroup. The structure of quasi-C-Ehresmann semigroups has been recently described by G. Li and Shum in [24] and they have provided a construction method for such kind of semigroups. Consequently, many important results on quasi-C-semigroups within the class of regular semigroups are extended to quasi-C-Ehresmann semigroups within the class of rpp semigroups, in particular,

within the class of U-semiabundant semigroups. We cite some useful definitions given by Lawson [32] and Y. He [26].

Notation 4.1. Let $U \neq \emptyset$ be a subset of $E(S)$. Then the following set

$$Reg_U(S) = \{a \in S | (\exists e, f \in U) e \ \widetilde{\mathcal{L}}\ a\ \widetilde{\mathcal{R}}\ f\}$$

is called the set of U-regular elements of S.

Remark 4.1. Clearly $U \subseteq Reg_U(S)$. Moreover, it is clear that $a \in Reg_U(S)$ if and only if $a \in Reg(S)$ and $V_U(a) = \{a' \in V(a) | \ aa', a'a \in U\} \neq \emptyset$. We now call the elements in $V_U(a)$ the U-inverses of $a \in S$.

For any $a \in S$. Define $U_a^l = \{u \in U | \ ua = a\}$ and $U_a^r = \{u \in U | \ au = a\}$. Let $U_a = U_a^l \cap U_a^r$. Then $U_a = \{u \in U | \ ua = a = au\}$. We call U_a the orbit of a in U.

Notation 4.2. Define the following relations on the semigroup S with respect to $U \subseteq E(S)$:

$$\begin{aligned}
\widetilde{\mathcal{L}}^U &= \{(a,b) \in S \times S | \ U_a^r = U_b^r\}, \\
\widetilde{\mathcal{R}}^U &= \{(a,b) \in S \times S | \ U_a^l = U_b^l\}, \\
\widetilde{\mathcal{H}}^U &= \widetilde{\mathcal{L}}^U \cap \widetilde{\mathcal{R}}^U, \\
\widetilde{\mathcal{Q}}^U &= \{(a,b) \in S \times S | \ U_a = U_b\}.
\end{aligned}$$

As usual, we denote the $\widetilde{\mathcal{L}}^U$, $\widetilde{\mathcal{R}}^U$, $\widetilde{\mathcal{H}}^U$ and $\widetilde{\mathcal{Q}}^U$-classes containing a $\in S$ by $\widetilde{L}_a^U$, $\widetilde{R}_a^U$, $\widetilde{H}_a^U$ and $\widetilde{Q}_a^U$, respectively. We state here repeatedly that $\widetilde{\mathcal{L}}^U$ is not necessarily a right congruence on S and $\widetilde{\mathcal{R}}^U$ is not necessarily a left congruence on S.

Notation 4.3. Let S be a semigroup. Denote the semilattice of the semigroups S_α by $S = (Y; S_\alpha)$. In particular, if S is a band, then $(Y; S_\alpha)$ is the greatest semilattice decomposition of S. By $S = [Y; S_\alpha; \phi_{\alpha,\beta}]$, we mean S is a strong semilattice of the semigroups S_α with structure homomorphisms $\phi_{\alpha,\beta}$.

Definition 4.1.

(i) A semigroup $S(U)$ is called an U-semiabundant if every $\widetilde{\mathcal{L}}^U$- and every $\widetilde{\mathcal{R}}^U$-class of S contains at least one element in U, that is, $\widetilde{L}_a^U \cap U \neq \emptyset$ and $\widetilde{R}_a^U \cap U \neq \emptyset$ for all $a \in S$.

(ii) A semigroup $S(U)$ is called an U-semi-superabundant if every $\widetilde{\mathcal{H}}^U$-class of S contains an element of U, that is, $\widetilde{H}_a^U \cap U \neq \emptyset$ for all $a \in S$. In this case, the unique element in $\widetilde{H}_a^U \cap U$ is denoted by a_U°.

The following results were given by Y. He in [26].

Lemma 4.1. *Let $\mathcal{F}$ be one of the Green relations $\mathcal{L}$, $\mathcal{R}$ or $\mathcal{H}$ and $\widetilde{\mathcal{F}}^U$ its corresponding $\widetilde{Green}$ relations on the semigroup S. Then, for any $a, b \in S$, we have*

(i) $\mathcal{F} \subseteq \widetilde{\mathcal{F}}^U$ and for $a, b \in Reg_U(S)$, $(a, b) \in \widetilde{\mathcal{F}}^U$ if and only if $(a, b) \in \mathcal{F}$.
(ii) $\widetilde{\mathcal{H}}^U \subseteq \widetilde{\mathcal{Q}}^U$ and $\widetilde{\mathcal{Q}}_a^U$ contains at most one element in U.
(iii) If $S(U)$ is a U-semi-superabundant semigroup then $S(U)$ is U- liberal with $\widetilde{\mathcal{Q}}^U = \widetilde{\mathcal{H}}^U$.

Definition 4.2. Let $T = (Y; T_\alpha)$ be a semilattice of monoids T_α whose identity is denoted by 1_{T_α}. If $T = [Y; T_\alpha, \phi_{\alpha,\beta}]$ is a strong semilattice with structure homomorphisms $\phi_{\alpha,\beta}$ such that $\phi_{\alpha,\beta} : T_\alpha \longrightarrow T_\beta$ is defined by $x \mapsto x1_{T_\beta}$ for any $\beta \leq \alpha$, then T is called an $E-$ semilattice of monoids T_α, where $E = \{1_{T_\alpha} | \alpha \in Y\}$.

The following identification result of C-Ehreshmann semigroup was given in [24].

Lemma 4.2. *Let T be a semigroup with $E \subseteq E(T)$. Then $T(E)$ is a C-Ehresmann semigroup if and only if $T(E)$ is a E-semilattice of monoids.*

By the above Lemma, we call $T = (Y; T_\alpha)$ a C-Ehresmann semigroup if $T(E)$ is a E-semilattice of monoids T_α, where $E = \{1_{T_\alpha} | \alpha \in Y\}$.

For brevity, we call $I \times T \times \Lambda$ the direct product of a rectangular band $I \times \Lambda$ and a monoid T, a rectangular monoid. The direct product $I \times T$ of a left zero semigroup I and a monoid T is called a "left monoid".

Now, let $I = (Y; I_\alpha)$ and $T = (Y; T_\alpha)$ be two semigroups. Form the set product $S_\alpha = I_\alpha \times T_\alpha$ and the set union $S = \bigcup_{\alpha \in Y} S_\alpha$. Define the following mapping $\eta : S \to \mathcal{T}_l(I)$ by $(i, a) \mapsto (i, a)^\sharp$, where $\mathcal{T}_l(I)$ is the left transformation semigroup acting on I.

If the mapping η satisfies the following conditions: for any $(i, a) \in S_\alpha$ and $(j, b) \in S_\beta$,

(i) $(i, a)^\sharp j \in I_{\alpha\beta}$, in particular, $(i, a)^\sharp j = ij$ when $\alpha \leq \beta$;
(ii) $(i, a)^\sharp (j, b)^\sharp = ((i, a)^\sharp j, ab)^\sharp$,

then we can verify that S becomes a semigroup under the multiplication

$$(i, a)(j, b) = ((i, a)^\sharp j, ab).$$

According to [19], the above semigroup S is called a left semi-spined product of I and T with respect to Y and η, denoted by $I \times_{Y,\eta} T$.

An U-liberal semigroup $S(U)$ is called left C-Ehresmann if U is a subsemigroup of S such that $S(U)$ satisfies the ET-condition and $uS \subseteq Su$ for all $u \in U$.

The following is a construction theorem of left C-Ehresmann semigroups.

Theorem 4.1. *Let $T = (Y; T_\alpha)$ be a C-Ehresmann semigroup and $S = \bigcup_{\alpha \in Y} S_\alpha$, where $S_\alpha = I_\alpha \times T_\alpha$, I_α is a non-empty set. Let $U = \{(i, 1_{T_\alpha}) | i \in I_\alpha, \alpha \in Y\}$. For any $\gamma \leq \alpha$ in Y, let*

$$\Psi_{\alpha,\gamma} : S_\alpha \to \mathcal{T}_l(I_\gamma) \text{ by } (i,a) \mapsto \psi_{\alpha,\gamma}^{(i,a)}$$

satisfies the following conditions for any $\alpha, \beta \in Y$:

(i) for any $(i,a) \in S_\alpha$, $\psi_{\alpha,\alpha}^{(i,a)}$ is the constant mapping on I_α, whose value is i;

(ii) for any $(i,a) \in S_\alpha$ and $(j,b) \in S_\beta$, $\psi_{\alpha,\alpha\beta}^{(i,a)} \psi_{\beta,\alpha\beta}^{(j,b)}$ is the constant mapping on $I_{\alpha\beta}$, whose value is denoted by $< \psi_{\alpha,\alpha\beta}^{(i,a)} \psi_{\beta,\alpha\beta}^{(j,b)} >$;

(iii) for any $\delta \leq \alpha\beta \in Y$, $\psi_{\alpha\beta,\delta}^{(k,ab)} = \psi_{\alpha,\delta}^{(i,a)} \psi_{\beta,\delta}^{(j,b)}$, where $k = < \psi_{\alpha,\alpha\beta}^{(i,a)} \psi_{\beta,\alpha\beta}^{(j,b)} >$.

Then $S(U)$ forms a left C-Ehresmann semigroup under the multiplication

$$(i,a)(j,b) = (< \psi_{\alpha,\alpha\beta}^{(i,a)} \psi_{\beta,\alpha\beta}^{(j,b)} >, ab),$$

for all $(i,a) \in S_\alpha$ and $(j,b) \in S_\beta$.

Conversely, every left C-Ehresmann semigroup can be constructed in the above manner.

We give here an example to illustrate that not all quasi-adequate semigroups are good quasi-adequate.

Example 4.1. Let $\bar{S}$ be a semigroup with the following Caylay table

·	e	*mn*	*l*	*fghabctwqrso*
e	*e*	*mn*	*1*	0 0 0 *abctwqrsO*
m	*mm*	*I*	*1*	0 0 0 *t b s t q qr sO*
n	*n*	*1*		*nlOOOwrcqwqrsO*
l	*1*	*1*		*11oooqrsqqqrso*
f	0	O	*OO*	*f g h O* OOO O OOOO
g	0	o o o	g g g	o ooo o o o o o
h	0	0 0 0	*h h h*	0 000 0 0 0 0
a	0	0 0 0	a b c	0 000 0 0 0 0.
b	0	O	OObbb	OOOOOOOOO
c	0	o o o	c c c	o ooo o o o o o
t	0	0	00 t b s	O OOO O OOOO
w	0	O	OOwrc	OOOOOOOOO
q	0	0 o o	oq r s	o ooo o o o o o
r	0	O O O	r r r	O O O O O O O O O
s	0	o o o	s s s	o o o o o o o o o
0	0	0	000000000	00000

By direct computation, $E(\bar{S})$ = {e, m, n, *I, f, g, h,* 0} is a band and

$$\bar{S}/\mathcal{L}^* = \{\{e\},\ \{m\},\ \{n\},\ \{I\},\ \{f,a,t,w,q\},\ \{g,h,b,c,r,s\},\ \{0\}\},$$

$$\bar{S}/\mathcal{R}^* = \{\{e,«\},\ \{m,b,t\},\ \{n,c,w\},\ \{l,q,r,s\},\ \{/\},\ \{\#\},\ \{h\},\ \{0\}\}.$$

Thus $\bar{5}$ is a quasi-adequate semigroup. Since $6 = t\S \cup \{(g, ft), (ft,3)\}$ but $(a \leqslant 7, aft) = (b, c)\ \&\ 5$, $\bar{S}$ is not good quasi-adequate.

We provide here a method of construction for quasi-C-Ehresmann semigroups.

Theorem 4.2. *Let $T = [Y; T_a]$ be a C-Ehresmann semigroup, $I = [Y\backslash I_a]$ a left regular band and $A = [Y; A_a)$ a right regular band, let %{A) be the right transformation semigroup acting on* A *and %(I) the left transformation semigroup acting on I. If the following mappings*

$$\pounds \bullet \mathop{\mathrm{I\,I}}_{\alpha\in Y} (I \times T) \to \mathcal{T}_I(I), (i,x) \mapsto (i,x)^\sharp,$$

$$n: \mathop{(\mathrm{J}}_{\alpha\in Y} (T_a \times \mathrm{A_a}) \to 7;(\mathrm{A}), (\mathrm{I}, \mathrm{A}) \mapsto (x,\lambda)^*$$

satisfy the following conditions and their duals

(LI) *if(i,z) E I,xT, andj* E $\mathrm{I}_{,}$, *then (i,z)~j* E $\mathrm{I}_{,\mathrm{p}}$;
(L_z) *if* $a \leq_{IP}$ *in* (L1) for a, P E Y, *then (i, z)ffj = i;*
(L_3) if (i, x) E *I,* x T' and *(j, y) E Ip x Tp, then* $(i,x)^\sharp(j,y)^\sharp = ((i,x)^\sharp j, xy)^\sharp$.

Then $S(U) = \bigcup_{\alpha \in Y} (I_\alpha \times T_\alpha \times \Lambda_\alpha)$ *forms a quasi-C-Ehresmann semigroup, where* $U = \bigcup_{\alpha \in Y} (I_\alpha \times \{1_{T_\alpha}\} \times \Lambda_\alpha)$ *, under the following multiplication:*

$$(i, x, \lambda)(j, y, \mu) = ((i, x)^\sharp j, xy, \lambda(y, \mu)^*).$$

Conversely, every quasi-C-Ehresmann semigroup can be constructed by the above manner.

5. Some applications

Let S be a semigroup. Denote the equivalent relations $\widetilde{\mathcal{Q}}^{E(S)}, \widetilde{\mathcal{L}}^{E(S)}, \widetilde{\mathcal{R}}^{E(S)}$ and $\widetilde{\mathcal{H}}^{E(S)}$ by $\widetilde{\mathcal{Q}}, \widetilde{\mathcal{L}}, \widetilde{\mathcal{R}}$ and $\widetilde{\mathcal{H}}$, respectively. In fact, the notations of $\widetilde{\mathcal{L}}, \widetilde{\mathcal{R}}$ and $\widetilde{\mathcal{H}}$ were first mentioned by El-Qallali [4] . If $S(E(S))$ is a $E(S)$-liberal semigroup, that is, every $\widetilde{\mathcal{Q}}$-class of S contains an idempotent of S, then S is called liberal and in this case, we simply denote $a^\circ_{E(S)}$ by a°. Similarly, if $S(E(S))$ is an orthodox $E(S)$-liberal semigroup, that is, $U = E(S)$, then S is said to be a full orthodox liberal semigroup. If $S(E(S))$ is an C-Ehresmann semigroup, then we simply call the semigroup S a full Ehresmann semigroup with central idempotents (see [36]).

Now, we let T be an unipotent semigroup. Then the direct product $S = I \times T \times \Lambda$ of a rectangular band $I \times \Lambda$ with T is called the rectangular unipotent semigroup. In particular, the direct product $S = I \times T$ of a left zero band I and a unipotent semigroup T is called a left unipotent semigroup. Right unipotent semigroups can be dually defined [6]. It can be easily proved that S is a C- full Ehresmann semigroup if and only if S is a semilattice of unipotent semigroups and that S is a left C-full Egresmann semigroup if and only if S is a semilattice of left unipotent semigroups.

Hereafter, a semigroup $T = [Y; T_\alpha]$ is called a C-full Ehresmann semigroup if T itself is a C-full Ehresmann semigroup which can be expressed as a semilattice of unipotent semigroups $T_\alpha (\alpha \in Y)$, and the semigroup $S = [Y; I_\alpha \times T_\alpha]$ is said to be left C-full Ehresmann when S itself is a left C-full Ehresmann semigroup which can be expressed as a semilattice of left unipotent semigroups $S_\alpha = I_\alpha \times T_\alpha (\alpha \in Y)$.

The structure of orthodox liberal semigroups and quasi-C-full Ehresmann semigroups can be described by the following theorem:

Theorem 5.1. *The following statements are equivalent for a semigroup* S*:*

(i) S *is an orthodox liberal semigroup;*

(ii) S *is a semilattice of rectangular unipotent semigroups* $S_\alpha (\alpha \in Y)$ *and* $E(S)$ *is a subsemigroup of* S*;*

(iii) S is a semi-superabundant semigroup satisfying the (C) condition and $E(S)$ is a subsemigroup of S.

Theorem 5.2. *A semigroup S is a quasi-C-full Ehresmann semigroup if and only if S is isomorphic to a spined product $S_1 \times_T S_2$ of a left C-full Ehresmann semigroup $S_1 = [Y; I_\alpha \times T_\alpha]$ and a right C-full Ehresmann semigroup $S_2 = [Y; T_\alpha \times \Lambda_\alpha]$ having a common C-full Ehresmann semigroup component $T = [Y; T_\alpha]$ with respect to the semigroup homomorphisms $\phi : (i, x) \mapsto x$ for $(i, x) \in S_1$ and $\psi : (x, \lambda) \mapsto x$ for $(x, \lambda) \in S_2$.*

Following Fountain [11], a semigroup S is said to be superabundant if every H^*-class of S contains at least one idempotent of S. He observed that if S is a superabundant semigroup then $\mathcal{H}^* = \widetilde{\mathcal{H}} = \widetilde{\mathcal{Q}}$, and, for any $a \in S$, $a^\circ = a^\diamond$.

Now, we give the following definition.

Definition 5.1. A superabundant semigroup S is called orthodox superabundant if $E(S)$ forms a subsemigroup of S. In particular, a superabundant semigroup with central idempotents is called a C-superabundant semigroup. An orthodox superabundant semigroup S is said to be left C-superabundant; right C-superabundant and quasi-C- superabundant,respectively when $E(S)$ forms a left regular band; a right regular band and a regular band, respectively.

We have the following lemma.

Lemma 5.1. *Let S be an orthodox superabundant semigroup. Then S is an orthodox liberal semigroup which can be expressed as a semilattice of direct products of cancellative monoids and rectangular bands. In particular, if S is a left (resp., right) C-superabundant semigroup, then S is a left (resp., right) C-full Ehresmann semigroup which is expressible as a semilattice of direct products of cancellative monoids and left (resp., right) zero bands.*

Recall that a semigroup $S = [Y; I_\alpha \times T_\alpha]$ [resp., $S = [Y; T_\alpha \times \Lambda_\alpha]$ is said to be a left [resp., right] C- superabundant if S is a left [resp., right] C-superabundant which can be expressed as a semilattice of direct products of left zero bands and cancellative monoids $I_\alpha \times T_\alpha (\alpha \in Y)$ (resp., cancellative monoids and right zero bands $(T_\alpha \times \Lambda_\alpha$). Then, by Fountain [11], one can easily show that a semigroup S is C-superabundant if and only if S is a semilattice of cancellative monoids. Thus, a C-superabundant semigroup $S = [Y; T_\alpha]$ is C-superabundant which can be expressed as a semilattice of cancellative monoids $T_\alpha (\alpha \in Y)$.

The following theorem characterizes the quasi-C-superabundant semigroups. In fact, it also gives a method of construction of such semigroups.

Theorem 5.3. *Let S be a semigroup. Then S is a quasi-C-superabundant semigroup if and only if S is isomorphic to a spined product $S_1 \times_T S_2$ of a left C-superabundant semigroup $S_1 = [Y; I_\alpha \times T_\alpha]$ and a right C-superabundant semigroup $S_2 = [Y; T_\alpha \times \Lambda_\alpha]$ having a common C-superabundant semigroup component $T = [Y; T_\alpha]$ with respect to the semigroup homomorphisms $\phi : (i, x) \mapsto x$ for $(i, x) \in S_1$ and $\psi : (x, \lambda) \mapsto x$ for $(x, \lambda) \in S_2$.*

It is well known that a regular orthogroup S is a completely regular semigroup whose set of idempotents $E(S)$ forms a regular band. Such regular orthogroups were called quasi-C-semigroups. By observing that regular orthogroups are quasi-C-superabundant regular semigroups and cancellative regular monoids are groups, we re-obtain the following known theorem in [36].

Theorem 5.4. *The following statements are equivalent :*

(i) A semigroup S is a quasi-C-semigroup if and only if S is isomorphic to a spined product $S_1 \times_T S_2$ of a left C-semigroup $S_1 = [Y; I_\alpha \times T_\alpha]$ and a right C-semigroup $S_2 = [Y; T_\alpha \times \Lambda_\alpha]$ having a common Clifford semigroup component $T = [Y; T_\alpha]$ with respect to the semigroup homomorphisms $\phi : (i, x) \mapsto x$ for $(i, x) \in S_1$ and $\psi : (x, \lambda) \mapsto x$ for $(x, \lambda) \in S_2$.

(ii) A semigroup S is a quasi-C-semigroup if and only if S can be expressible as the semi-spined product of a Clifford semigroup, a left regular band and a right regular band.

(iii) A semigroup S is a quasi-C-semigroup if and only if S can be expressible as a $\triangle$-product of a Clifford semigroup, a left regular band and a right regular band.

6. Conclusion

According to Fountain [11] and Lawson [32], the relations $\mathcal{L}^*$ and $\widetilde{\mathcal{L}}^U$ on S are defined by

$$\mathcal{L}^* = \{(a, b) \in S \times S | (\forall x, y \in S^1)\ ax = ay \Leftrightarrow bx = by\},$$
$$\widetilde{\mathcal{L}}^U = \{(a, b) \in S \times S | U_a^r = U_b^r\}.$$

The equivalent relation $\widetilde{\mathcal{L}}$ [$= \widetilde{\mathcal{L}}^{E_S}$] on S was first discussed by El-Qallali in [4]. The left congruence $\mathcal{R}^*$ and the equivalent relations $\widetilde{\mathcal{R}}$, $\widetilde{\mathcal{R}}^U$ on S can

be defined dually. The equivalent relations $\mathcal{H}^*$, $\widetilde{\mathcal{H}}$ and $\widetilde{\mathcal{H}}^U$ are defined by

$$\mathcal{H}^* = \mathcal{L}^* \cap \mathcal{R}^*, \quad \widetilde{\mathcal{H}} = \widetilde{\mathcal{L}} \cap \widetilde{\mathcal{R}}, \quad \widetilde{\mathcal{H}}^U = \widetilde{\mathcal{L}}^U \cap \widetilde{\mathcal{R}}^U.$$

It is clear that the results of $\mathcal{L}$ [$\mathcal{L}^*$, $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{L}}^U$, resp.] are duals of $\mathcal{R}$ [$\mathcal{R}^*$, $\widetilde{\mathcal{R}}$ and $\widetilde{\mathcal{R}}^U$, resp.].

Let ρ and σ be equivalent relations on the semigroup S. Then, we have the following statements:

(1) (ρ, U)-*surjective* if for any $a \in S$, $|\rho_a \cap U| \geq 1$;
(2) *strongly* (ρ, U)-*surjective* if for any $a \in S$, $|\rho_a \cap U_a| = 1$;
(3) *super* $(\rho, U; \sigma)$-*surjective* if S is strongly (ρ, U)-surjective and for any $a \in S$, the unique element in $\rho_a \cap U_a$ is σ-equivalent to a.

It is evident that we have the following properties for ρ and σ) on S:

$$\text{super}(\rho, U; \sigma)\text{-surjective} \Longrightarrow \text{strongly } (\rho, U)\text{-surjective} \Longrightarrow (\rho, U)\text{-surjective}.$$

By using the above notations, we call a semigroup S a completely regular if S is $(\mathcal{H}, E_S)$-surjective. Denote the class of completely regular semigroups by **CReg**. The structure of completely regular semigroups has been fully described in [39].

The following results can be used to describe various types of generalized regular semigroups and generalized rpp semigroups.

Lemma 6.1. *Let $a \in S$ and $e \in E_S$. Then the following statements hold:*

(i) $\mathcal{L} \subseteq \mathcal{L}^* \subseteq \widetilde{\mathcal{L}} \subseteq \widetilde{\mathcal{L}}^U$;
(ii) $\mathcal{L}|_{E_S} = \mathcal{L}^*|_{E_S} = \widetilde{\mathcal{L}}|_{E_S}$ *and* $\widetilde{\mathcal{L}}^U|_U = \mathcal{L}|_U$;
(iii) if $S \in \mathbf{Rpp}$, *then* $\widetilde{\mathcal{L}} = \mathcal{L}^*$;
(iv) if $S \in \mathbf{Reg}$, *then* $\mathcal{L} = \mathcal{L}^* = \widetilde{\mathcal{L}}$;
(v) $(a, e) \in \mathcal{L}^*$ *if and only if* $e \in I_a^r$ *and, for all* $x, y \in S^1$, $ax = ay$ *implies* $ex = ey$;
(vi) each $\widetilde{\mathcal{H}}^U$*-class in* S *contains at most one element of* U.

By applying Lemma 6.1, we can easily see that, for a semigroup S, we have the following terminologies:

$$(\widetilde{\mathcal{H}}^U, U)\text{-surjective} = \text{strongly } (\widetilde{\mathcal{H}}^U, U)\text{-surjective},$$

$$(\widetilde{\mathcal{H}}, E_S)\text{-surjective} = \text{strongly } (\widetilde{\mathcal{H}}, E_S)\text{-surjective},$$

$$(\mathcal{H}^*, E_S)\text{-surjective} = \text{strongly } (\mathcal{H}^*, E_S)\text{-surjective},$$

$(\mathcal{H}, E_S)$-surjective = strongly $(\mathcal{H}, E_S)$-surjective.

In closing, we give a table which provides useful information for several known classes of generalized regular semigroups and generalized rpp semigroups.

Class	Definition	Notation
completely regular semigroups	$(\mathcal{H}, E_S)$-surjective	**CReg**
regular semigroups	$(\mathcal{L}, E_S)$-surjective or, alternatively, $(\mathcal{R}, E_S)$-surjective	**Reg**
superabundant semigroups	$(\mathcal{H}^*, E_S)$-surjective	**SA**
abundant semigroups	$(\mathcal{L}^*, E_S)$ and $(\mathcal{R}^*, E_S)$-surjective	**Ab&**
super rpp semigroups	super $(\mathcal{L}^*, E_S; \widetilde{\mathcal{R}})$-surjective	**SuR**
strongly rpp semigroups	strongly $(\mathcal{L}^*, E_S)$-surjective	**SR**
rpp semigroups	$(\mathcal{L}^*, E_S)$-surjective	**Rpp**
semi-superabundant semigroups	$(\widetilde{\mathcal{H}}, E_S)$-surjective	**SSA**
semiabundant semigroups	$(\widetilde{\mathcal{L}}, E_S)$ and $(\widetilde{\mathcal{R}}, E_S)$-surjective	**AbS**
strongly semi-rpp semigroups	strongly $(\widetilde{\mathcal{L}}, E_S)$-surjective	
semi-rpp semigroups	$(\widetilde{\mathcal{L}}, E_S)$-surjective	
$\mathcal{P}$-semi-superabundant semigroups	$(\widetilde{\mathcal{H}}^U, U)$-surjective	
$\mathcal{P}$-semiabundant semigroups	$(\mathcal{L}^U, U)$ and $(\widetilde{\mathcal{R}}^U, U)$-surjective	$\mathcal{P}$**AbS**
strongly $\mathcal{P}$-semi-rpp semigroups	strongly $(\widetilde{\mathcal{L}}^U, U)$-surjective	$S\mathcal{P}$**Srrp**
$\mathcal{P}$-semi-rpp semigroups	$(\widetilde{\mathcal{L}}^U, U)$-surjective	**calPSrrp**

(i) In the above table, the dual of a " rpp " semigroups is a " lpp " semigroup. Consequently,the dual classes of the classes of semigroups **Rpp**, **SR**, **SuR**, **SSR** and $S\widetilde{\mathcal{P}}$**Srrp** are the classes of the semigroups **Lpp**, **SL**, **SuL**, **SSL** and $S\widetilde{\mathcal{P}}$**Slpp**, respectively.

(ii) In the last four cases in the table, the set U is *the set of all projections of* S, and S with such a subset U of $(E(S)$ is denoted by $S(U)$. The $\mathcal{P}$-semiabundant semigroup $S(U)$ was called a U-semiabundant semigroup by Lawson in [33]. The Rees matrix semigroups investigated by Lawson in [33,34] are in fact the $\mathcal{P}$-semi-superabundant semigroups.

Furthermore, we can also see that **SA**, **SSA** and **PPSSA** are the generalizations of **CReg** in the categories of abundant semigroup, semiabundant semigroup and $\mathcal{P}$-semiabundant semigroup, respectively. In a recent paper [27], it has been shown by Y. He, Chen and Shum that **SR**, **SSR** and

SPSR are generalizations of **CReg** in the categories rpp semigroups, semi-rpp semigroups and $\mathcal{P}$-semi-rpp semigroups, respectively. Hence, we can call the members of **SA**, **SSA**, **PSSA**, **SuR**, **SR**, **SSR**, **S$\mathcal{P}$SR**, **SuL**, **SL**, **SSL** and **S$\mathcal{P}$SR** the *generalized completely regular semigroups*. We notice that many results in **CReg** have already been generalized into **SA**, **SuR** and **SR** (see [12,14,15,21,24,29,41,44,46,48,51]). Thus, by using the above table as an aid, we can establish a platform for various kinds of generalizations of rpp semigroups and we can therefore further study their mutual relationships systematically.

Finally, we propose a problem: By using the techniques proposed by Lawson (see [33-35]), can we establish a Rees Matrix cover theorem for a quasi-adequate semigroup or a type A-semigroup ?

Acknowledgement

The author would like to thank Professor Mario Petrich for giving him many valuable comments and helpful suggestions concerning the presentation of this paper.

References

1. Armstrong S., The structure of type A semigroups, Semigroup Forum 29 (1984), no. 3, 319–336.
2. Clifford A. H. and G. B. Preston, The algebraic theory of semigroups, Vol. I & Vol. II, Math. Surveys of the American Math. Soc., Providence R.I., (1961) & (1967)
3. Du L. and K. P. Shum , On left C-wrpp semigroups, Semigroup Forum 67 (2003), no. 3, 373–387.
4. El-Qallali A., On the construction of a class of abundant semigroups, Acta Math. Hung., 1990, 56(1-2), 77-91.
5. El-Qallali A., Abundant semigroups with a multiplicative type-A transversal, Semigroup Forum ,1993, 47, 327-340.
6. El-Qallali A., $\mathcal{L}^*$-unipotent semigroups, J. Pure and Applied Algebra,(1989),19-23.
7. El-Qallali A. and J. B. Fountain , Quasi-adequate semigroups, Proc. Roy. Soc. Edinburgh Sec. 91A (1981), 91-99.
8. El-Qallali A. and J. B. Fountain , Idempotent-connected semigroups, Proc. Roy. Soc. Edinburgh Sec.,91A (1981), 79-90.
9. Fountain J. B., A class of right PP-monoids. Quart. J. Math., Oxford Ser., (2) 28 (1977), no. 111, 285–300.
10. Fountain J. B., Adequate semigroups, Proc. Edinburgh Math. Soc., 22 (1979), 113-125.
11. Fountain J. B., Abundant semigroups, Proc. London Math. Soc. (3) 44 (1982), 103-129.

12. Fountain J. B., G. M. S. Gomes and V. A. Gould , A Munn type representation for a class of E−semiadequate semigroups, J. Algebra, 218(1999): 693−714.
13. Gomes G. M. S. and V. A. Gould , Fundamental Ehresmann semigroups, Semigroup Forum, 63(2001): 11−33.
14. Guo X. J., Abundant left C-lpp proper semigroups, Southeast Asian Bull. Math., 2000, 23, 41-50.
15. Guo X. J., K. P. Shum and Y. Q. Guo , Perfect rpp semigroups, Comm. Algebra ,29 (2001), no. 6, 2447–2459.
16. Guo X. J, Y. Q. Guo and K. P. Shum , On left abundant semigroups, Comm. in Algebra, Vol 32, no 6,2061-2085, 2004.
17. Guo X. J. and K. P. Shum , On translational hulls of type-A semigroups. J. Algebra, 269 (2003), no. 1, 240–249.
18. Guo X. J., K. P. Shum and L. Zhang , Regular F-abundant semigroups, Comm. Algebra 33 (2005), no. 12, 4383–4401.
19. Guo X. J., Y. Q. Guo and K. P. Shum , Semispined product structure of left C-a semigroups, In: Semigroups edited by K.P. Shum, Y. Q. Guo, M. Ito and Y. Fong , Springer-verlag, Singapore,1995., p.157-166.
20. Guo X. J. and K. P. Shum , On left cyber groups. Int. Math. J., 5 (2004), no. 8, 705–717.
21. Guo X. J. and L. M. Wang , Idempotent-connected abundant semigroups which are disjoint unions of quasi-ideal adequate transversals, Comm. in Algebra, 2002, 30, 1779-1800.
22. Guo, X. J. and K. P. Shum , Left semi-perfect abundant semi-groups of type W, Proc. Roy. Soc. Edinburgh Sect. A 135 (2005), no. 3, 603–613.
23. Guo Y. Q., K. P. Shum and Zhu P.Y. , The structure of left C-rpp semigroups, Semigroup Forum , 1995 , 50(1):9-23.
24. Guo Y. Q., K. P. Shum and P. Y. Zhu , On quasi-C-semigroups and some special subclasses, Algebra Colloquium, V.6, N.1 (1999), 105–120.
25. Hall T. E. , On regular semigroups. J. Algebra, 24 (1973), 1–24.
26. He Y., F. Shao and Shi-qun Li and W. Gao , On left C-$\mathcal{U}$-liberal semigroups, Czechoslovak Math. J. 56(131), (2006), no. 4, 1085–1108.
27. He Y., Chen Y. Q. and K. P. Shum , Projectively condensed Semigroups, (to appear in Acta Math Hung. Sci., 2008).
28. Howie J. M., Introduction to Semigroups,Academic Press,1976.
29. Kong X. Z. and K. P. Shum , On the structure of regular crypto semigroups , Comm. Algebra, 29 (2001), no. 6, 2461–2479.
30. Kong X. Z. and K. P. Shum , A structure theorem of generalized superabundant semigroups , (to appear in Algebra Colloqium 2008).
31. Lawson M. V. , The structure of type-A semigroups, Quart. J. Math. Oxford (2), 1986, 37, 279-298.
32. Lawson M. V. , The natural partial orders on an abundant semigroups , Proc. Edinburgh Math. Soc. 1987, 30, 169-186.
33. Lawson M. V. , Abundant Rees matrix semigroups , J. Austral. Math. Soc. (Series A) 1987, 42, 132-142.
34. Lawson M. V. , Rees matrix semigroups , Proc. Edinburgh Math. Soc., 1990,

33: 23–37.
35. Li G. and K. P. Shum , A Rees matrix representation theorem for U-semi-superabundant semigroups , Int. Math. J., 3 (2003), no. 3, 353–363.
36. Li G., Y. Q. Guo and K. P. Shum , Quasi-C-Ehresmann semigroups and their subclasses, Semigroup Forum , 70 (2005), no. 3, 369–390.
37. Pastijn F. , A generalization of Green's equivalence relations for halfgroupoids , Simon Stevin 49 (1975/76), no. 4, 165–175.
38. Petrich M. , The structure of completely regular semigroups , Trans. Amer. Math. Soc. 1974(189):211–236.
39. Petrich M. and N. R. Reilly , Completely Regular Semigroups , A Wiley & Sons, Intersience Publication, (1999).
40. Ren X. M. and K. P. Shum , The structure of $\mathcal{L}^*$-inverse semigroups , Sci. China Ser. A 49 (2006), no. 8, 1065–1081.
41. Ren X. M. and K. P. Shum , Supper abundant semigroups ,Science in China,Series A,(2004) No.1,1-15.
42. Ren X. M. and K. P. Shum , The structure of superabundant semigroups , *Sci. China Ser. A* 47(5), 2004, 756–771.
43. Shum K. P. and Y. Q. Guo , Regular semigroups and their generalizations , "Rings, groups, and algebras", 181–226, Lecture Notes in Pure and Appl. Math., 181, Dekker, New York,1996.
44. Shum K. P. and X. M. Ren , Abundant semigroups with left central idempotents , Pure Math. Appl., 10 (1999), no. 1, 109–113.
45. Shum K. P., X. J. Guo and X. M. Ren , (l)-Green's relations and perfect rpp semigroups , Proceedings of the Third Asian Mathematical Conference, 2000 (Diliman), 604–613, World Sceintific Inc (2002).
46. Shum K. P., X. M. Ren and Y. Q. Guo , On C^*-quasiregular semigroups , Comm. Algebra,1999,27 (19),4251-4274.
47. Shum K. P. and X. M. Ren , Abundant semigroups and their special subclasses , Proceedings of the International Conference on Algebra and its Applications ,(ICAA 2002) (Bangkok),66–86,Chulalongkorn Univ., Bangkok, 2002.
48. Shum K. P. and X. M. Ren , On generalized orthogroups , Comm. Algebra 29,2001,no. 6, 2341–2361.
49. Shum K. P. and X. M. Ren , The structure of right C-rpp semigroups , Semigroup Forum ,68 (2004), no. 2, 280–292.
50. Yamada M. , Orthodox semigroups whose idempotents satisfy a certain identity , Semigruup Forum, 6 (1973), 113-128.
51. Zhu P.Y., Y. Q. Guo and K. P. Shum , Struture and chacterizations of left C^* semigroups , Science in China, Series A, 6(1992):791–805.

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 335–352)

CONFORMAL FIELD THEORY AND MODULAR FUNCTOR

KENJI UENO

Department of Mathematics, Graduate School of Science,
Kyoto University, Kyoto, 606-8502, Japan
E-mail: ueno@math.kyoto-u.ac.jp

1. Introduction

In the preset paper we shall report our recent results with J. E. Andersen ([1–3]) on Modular Functor and Topological Quantum Field Theory .

Conformal field theory (CFT) is a functor from the category of pointed Riemann surfaces with coordinates to the category of finite dimensional complex vector spaces, which satisfies several basic properties (see §3 below). Modular functor (MF) is a functor from the category of pointed oriented surfaces with tangent vectors and Lagrangian subspace to the category of finite dimensional complex vector spaces, which satisfies similar properties to those of conformal field theory (see §2). The difference is that CFT does depend on a complex structure of a surface but MF only depends on a differentiable structure of a surface. Nevertheless we can construct a modular functor by using non-abelian and abelian conformal filed theories ([1,2]), and the naturally defined projectively flat connections.

Modular functor defines Topological Quantum Field Theory (TQFT) ([6,15]), which gives invariants of three dimensional manifolds. Thus, to construct TQFT is is enough to construct MF.

In the following, we shall discuss how to construct MF form CFT. Details can be found in our paper [2].

2. Modular Functor

In this section, we shall give the axioms for a modular functor. These are due to G. Segal ([9]). We present them here in a topological form which is due to K. Walker [15] (see also [6]).

By a closed surface, we mean a smooth real two-dimensional manifold.

For a closed oriented surface Σ of genus g we have the non-degenerate skew-symmetric intersection pairing

$$(\cdot,\cdot) : H_1(\Sigma,\mathbb{Z}) \times H_1(\Sigma,\mathbb{Z}) \to \mathbb{Z}.$$

Suppose Σ is connected. In this case a Lagrangian subspace $L \subset H_1(\Sigma,\mathbb{Z})$ is by definition a subspace, which is maximally isotropic with respect to the intersection pairing. A $\mathbb{Z}$-basis $(\vec{\alpha},\vec{\beta}) = (\alpha_1,\ldots,\alpha_g,\beta_1,\ldots\beta_g)$ for $H_1(\Sigma,\mathbb{Z})$ is called a symplectic basis if

$$(\alpha_i,\beta_j) = \delta_{ij}, \quad (\alpha_i,\alpha_j) = (\beta_i,\beta_j) = 0,$$

for all $i,j = 1,\ldots,g$.

For any real vector space V, we define $PV = (V \setminus \{0\})/\mathbb{R}_+$.

A *pointed surface* (Σ,P) is an oriented closed surface Σ with a finite set $P \subset \Sigma$ of points. A. A *morphism of pointed surfaces* $f : (\Sigma_1,P_1) \to (\Sigma_2,P_2)$ is an isotopy class of orientation preserving diffeomorphisms which maps P_1 to P_2.

A *marked surface* $\mathbf{\Sigma} = (\Sigma,P,V,L)$ is an oriented closed smooth surface Σ with a finite subset $P \subset \Sigma$ of points with projective tangent vectors $V \in \sqcup_{p\in P} PT_p\Sigma$ and a Lagrangian subspace $L \subset H_1(\Sigma,\mathbb{Z})$.

A *morphism* $\mathbf{f} : \mathbf{\Sigma}_1 \to \mathbf{\Sigma}_2$ of marked surfaces $\mathbf{\Sigma}_i = (\Sigma_i,P_i,V_i,L_i)$ is an isotopy class of orientation preserving diffeomorphisms $f : \Sigma_1 \to \Sigma_2$ that maps (P_1,V_1) to (P_2,V_2) together with an integer s. Hence we write $\mathbf{f} = (f,s)$.

Let $(\Sigma,\{p_-,p_+\} \sqcup P,\{v_-,v_+\} \sqcup V,L)$ be a marked surface, where we have selected an ordered pair of marked points with projective tangent vectors $((p_-,v_-),(p_+,v_+))$, at which we will perform the glueing. Let $c : P(T_{p_-}\Sigma) \to P(T_{p_+}\Sigma)$ be an orientation reversing projective linear isomorphism such that $c(v_-) = v_+$. Such a c is called a *glueing map* for Σ. Let $\tilde{\Sigma}$ be the oriented surface with boundary obtained from Σ by blowing up p_- and p_+, i.e.

$$\tilde{\Sigma} = (\Sigma \setminus \{p_-,p_+\}) \sqcup P(T_{p_-}\Sigma) \sqcup P(T_{p_+}\Sigma),$$

with the natural smooth structure induced from Σ. Let now Σ_c be the closed oriented surface obtained from $\tilde{\Sigma}$ by using c to glue the boundary components of $\tilde{\Sigma}$. We call Σ_c the glueing of Σ at the ordered pair $((p_-,v_-),(p_+,v_+))$ with respect to c.

Let now Σ' be the topological space obtained from Σ by identifying p_- and p_+. We then have natural continuous maps $q : \Sigma_c \to \Sigma'$ and $n : \Sigma \to \Sigma'$. On the first homology group n induces an injection and q a surjection, so we can define a Lagrangian subspace $L_c \subset H_1(\Sigma_c,\mathbb{Z})$ by $L_c = q_*^{-1}(n_*(L))$.

We note that the image of $P(T_{p_-}\Sigma)$ (with the orientation induced from $\tilde{\Sigma}$) induces naturally an element in $H_1(\Sigma_c, \mathbb{Z})$ and as such it is contained in L_c.

Let $\tilde{\Sigma} = (\Sigma, \{p_-, p_+\} \sqcup P, \{v_-, v_+\} \sqcup V, L)$ be a marked surface. Let

$$c : P(T_{p_-}\Sigma) \to P(T_{p_+}\Sigma)$$

be a glueing map and Σ_c the glueing of Σ at the ordered pair $((p_-, v_-), (p_+, v_+))$ with respect to c. Let $L_c \subset H_1(\Sigma_c, \mathbb{Z})$ be the Lagrangian subspace constructed above from L. Then the marked surface $\tilde{\Sigma}_c = (\Sigma_c, P, V, L_c)$ is defined to be the *glueing* of $\tilde{\Sigma}$ at the ordered pair $((p_-, v_-), (p_+, v_+))$ with respect to c.

To define Modular Functor we need a label set. A *label set* Λ is a finite set furnished with an involution $\lambda \mapsto \lambda^\dagger$ and a trivial element 0 such that $0^\dagger = 0$. Let Λ be a label set. The category of Λ-*labeled marked surfaces* consists of marked surfaces with an element of Λ assigned to each of the marked point and morphisms of labeled marked surfaces are required to preserve the labelings. An assignment of elements of Λ to the marked points of $\tilde{\Sigma}$ is called a labeling of $\tilde{\Sigma}$ and we denote the labeled marked surface by $(\tilde{\Sigma}, \lambda)$, where λ is the labeling.

Definition 2.1. A *modular functor* based on the label set Λ is a functor V from the category of labeled marked surfaces to the category of finite dimensional complex vector spaces satisfying the axioms MF1 to MF5 below.

MF1(*Disjoint union axiom*): For any pair of labeled marked surfaces, there is an isomorphism

$$V((\tilde{\Sigma}_1, \lambda_1) \sqcup (\tilde{\Sigma}_2, \lambda_2))) \cong V(\tilde{\Sigma}_1, \lambda_1) \otimes V(\tilde{\Sigma}_2, \lambda_2).$$

The identification is associative.

MF2 (*Glueing axiom*): Let $\tilde{\Sigma}$ and $\tilde{\Sigma}_c$ be marked surfaces such that $\tilde{\Sigma}_c$ is obtained from $\tilde{\Sigma}$ by glueing at an ordered pair of points and projective tangent vectors with respect to a glueing map c. Then there is an isomorphism

$$V(\tilde{\Sigma}_c, \lambda) \cong \bigoplus_{\mu \in \Lambda} V(\tilde{\Sigma}, \mu, \hat{\mu}, \lambda),$$

which is associative, compatible with glueing of morphisms, disjoint unions and it is independent of the choice of the glueing map in the obvious way.

MF3 (*Empty surface axiom*): Let $\emptyset$ denote the empty labeled marked surface. Then

$$dim_{\mathbb{C}} V(\emptyset) = 1.$$

MF4 (*Once punctured sphere axiom*): Let $\tilde{\Sigma} = (S^2, \{p\}, \{v\}, 0)$ be a marked sphere with one marked point. Then

$$dim_{\mathbb{C}} V(\tilde{\Sigma}, \lambda) = \begin{cases} 1, \lambda = 0 \\ 0, \lambda \neq 0. \end{cases}$$

MF5 (*Twice punctured sphere axiom*): Let $\tilde{\Sigma} = (S^2, \{p_1, p_2\}, \{v_1, v_2\}, \{0\})$ be a marked sphere with two marked points. Then

$$dim_{\mathbb{C}} V(\tilde{\Sigma}, (\lambda, \mu)) = \begin{cases} 1, \lambda = \mu^\dagger \\ 0, \lambda \neq \mu^\dagger. \end{cases}$$

3. Conformal Field Theory

A pointed Riemann surface with formal neighbourhoods

$$\mathfrak{X} = (C; q_1, \ldots, q_N; \eta_1, \ldots, \eta_N)$$

consists of a Riemann surface C and an ordered N-tuple of N distinct points $(q_1, \ldots, q_N)$ on C together with formal neighbourhoods

$$\eta_j : \hat{\mathcal{O}}_{C,q_j} \cong \mathbb{C}[[\xi_j]]$$

for $j = 1, \ldots, N$.

3.1. *Non-Abelian Conformal Field Theory*

Let us first consider non-abelian conformal field theory (WSWN model).

Let $\mathfrak{g}$ be a simple Lie algebra over the complex numbers $\mathbb{C}$, which we fix throughout the paper. Let $\mathfrak{h}$ be its Cartan subalgebra.

Let $(,)$ be a constant multiple of the Cartan-Killing form of the simple Lie algebra $\mathfrak{g}$. For each element of $\lambda \in \mathfrak{h}$, there exists a unique element $H_\lambda \in \mathfrak{h}^*$ such that

$$\lambda(H) = (H_\lambda, H)$$

for all $H \in \mathfrak{h}$. For $\alpha \in \Delta$, H_α is called the *root vector* corresponding to the root α.

On $\mathfrak{h}^*$ we introduce an inner product by

$$(\lambda, \mu) = (H_\lambda, H_\mu).$$

Let us normalize the inner product $(,)$ by requiring that θ, the highest (or longest) root, has length squared

$$(\theta, \theta) = 2.$$

A weight $\lambda \in \mathfrak{h}_{\mathbf{R}}^*$ is called an integral weight if

$$2(\lambda, \alpha)/(\alpha, \alpha) \in \mathbf{Z}$$

for any $\alpha \in \Delta$. A weight $\lambda \in \mathfrak{h}_{\mathbf{R}}^*$ is called a dominant weight if

$$w(\lambda) \leq \lambda$$

for any element w of the Weyl group W of $\mathfrak{g}$. By P_+ we denote the set of dominant integral weights of $\mathfrak{g}$. A weight λ is the highest weight of an irreducible left $\mathfrak{g}$-module if and only if $\lambda \in P_+$. Let V_λ be the irreducible left $\mathfrak{g}$-module of highest weight λ. It is well-known that a finite dimensional irreducible left $\mathfrak{g}$-module is a highest weight module and two irreducible left $\mathfrak{g}$-modules are isomorphic if and only if they have the same highest weight.

Let w be the longest element of W. Then we define an involution $\dagger$ on P_+ by

$$\lambda^\dagger = -w(\lambda).$$

One has that the opposite of dual of the left-$\mathfrak{g}$-module V_λ is isomorphic to left-$\mathfrak{g}$-module $V_{\lambda^\dagger}$, meaning there exists a non-degenerate $\mathfrak{g}$-invariant perfect pairing

$$(,) : V_\lambda \otimes V_{\lambda^\dagger} \to \mathbb{C}.$$

The affine Lie algebra $\widehat{\mathfrak{g}}$ over $\mathbb{C}((\xi))$ associated with $\mathfrak{g}$ is defined to be

$$\widehat{\mathfrak{g}} = \mathfrak{g} \otimes \mathbb{C}((\xi)) \oplus \mathbb{C}c,$$

where c is an element of the center of $\widehat{\mathfrak{g}}$ and the Lie algebra structure is given by

$$[X \otimes f(\xi), Y \otimes g(\xi)] = [X, Y] \otimes f(\xi)g(\xi) + c \cdot (X, Y) \operatorname*{Res}_{\xi=0}(g(\xi)df(\xi)),$$

for

$$X, Y \in \mathfrak{g},\ f(\xi),\ g(\xi) \in \mathbb{C}((\xi)).$$

Let us fix a positive integer ℓ (called the *level*) and put

$$P_\ell = \{\, \lambda \in P_+ \,|\, 0 \leq (\theta, \lambda) \leq \ell \,\}.$$

For all levels ℓ we observe that $\dagger$ takes P_ℓ to itself.

Theorem 3.1. *For each $\lambda \in P_\ell$, there exists a unique left $\widehat{\mathfrak{g}}$-module $\mathcal{H}_\lambda$ (called the integrable highest weight $\widehat{\mathfrak{g}}$-module of level ℓ) satisfying the following properties.*

(1) $V_\lambda = \{\, |v\rangle \in \mathcal{H}_\lambda \mid \widehat{\mathfrak{g}}_+|v\rangle = 0 \,\}$ *is the irreducible left* $\mathfrak{g}$*-module with highest weight* λ.

(2) The central element c *acts on* $\mathcal{H}_\lambda$ *as* $\ell \cdot id$.

(3) $\mathcal{H}_\lambda$ *is generated by* V_λ *over* $\widehat{\mathfrak{g}}_-$ *with only one relation*

$$(X_\theta \otimes \xi^{-1})^{\ell-(\theta,\lambda)+1}|\lambda\rangle = 0$$

where $X_\theta \in \mathfrak{g}$ *is the element corresponding to the maximal root* θ *and* $|\lambda\rangle \in V_\lambda$ *is a highest weight vector.*

We use the following notation

$$X(n) = X \otimes \xi^n, \quad X \in \mathfrak{g}$$
$$X(z) = \sum_{n\in\mathbf{Z}} X(n) z^{-n-1}$$

where z is a variable. The normal ordering $\circ\ \circ$ is defined by

$$\circ X(n)Y(m) \circ = \begin{cases} X(n)Y(m), & n < m, \\ \frac{1}{2}(X(n)Y(m) + Y(m)X(n)) & n = m, \\ Y(m)X(n) & n > m. \end{cases}$$

Note that, if $n > m$ and $X = Y$, we have

$$\circ X(n)X(m) \circ = X(n)X(m) - n\delta_{n+m,0}(X,X) \cdot c.$$

The *energy-momentum* tensor $T(z)$ of level ℓ is defined by

$$T(z) = \frac{1}{2(g^* + \ell)} \sum_{a=1}^{\dim \mathfrak{g}} \circ J^a(z)J^a(z) \circ ,$$

where $\{J^1, J^2, \ldots, J^{\dim \mathfrak{g}}\}$ is an orthonormal basis of $\mathfrak{g}$ with respect to the Cartan-Killing form $(,)$ and g^* is the dual Coxeter number of $\mathfrak{g}$.

Put

$$L_n = \frac{1}{2(g^* + \ell)} \sum_{m\in\mathbf{Z}} \sum_{a=1}^{\dim \mathfrak{g}} \circ J^a(m)J^a(n-m) \circ .$$

Then we have the expansion

$$T(z) = \sum_{n\in\mathbf{Z}} L_n z^{-n-2}.$$

The operator L_n is called the n'th Virasoro operator and it acts on $\mathcal{H}_\lambda$.

For $X \in \mathfrak{g}$, $f = f(z) \in \mathbb{C}((z))$ and $\underline{\ell} = \ell(z)\frac{d}{dz} \in \mathbb{C}((z))\frac{d}{dz}$, we use the following notation.

$$X[f] = \operatorname*{Res}_{z=0}(X(z)f(z)dz)$$
$$T[\underline{\ell}] = \operatorname*{Res}_{z=0}(T(z)\ell(z)dz).$$

In particular, we have

$$L_0 = T[\xi\frac{d}{d\xi}].$$

To define a filtration $\{F_\bullet\}$ on $\mathcal{H}_\lambda$, we first define the subspace $\mathcal{H}_\lambda(d)$ of $\mathcal{H}_\lambda$ for a non-negative integer d by

$$\mathcal{H}_\lambda(d) = \{\, |v\rangle \in \mathcal{H}_\lambda \mid \ L_0|v\rangle = (d + \Delta_\lambda)|v\rangle \,\},$$

where

$$\Delta_\lambda = \frac{(\lambda,\lambda) + 2(\lambda,\rho)}{2(g^* + \ell)}, \quad \rho = \frac{1}{2}\sum_{\alpha\in\Delta_+}\alpha.$$

The subspaces $\mathcal{H}_\lambda(d)$ are finite dimensional vector space and one has that

$$\mathcal{H}_\lambda = \bigoplus_{d=0}^{\infty}\mathcal{H}_\lambda(d).$$

Now we define the filtration $\{F_p\mathcal{H}_\lambda\}$ by

$$F_p\mathcal{H}_\lambda = \sum_{d=0}^{p}\mathcal{H}_\lambda(d).$$

Put

$$\mathcal{H}_\lambda^\dagger(d) = Hom_{\mathbb{C}}(\mathcal{H}_\lambda(d), \mathbb{C}).$$

Then the dual space $\mathcal{H}_\lambda^\dagger$ of $\mathcal{H}_\lambda$ is defined to be

$$\mathcal{H}_\lambda^\dagger = Hom_{\mathbb{C}}(\mathcal{H}_\lambda, \mathbb{C}) = \prod_{d=0}^{\infty}\mathcal{H}_\lambda^\dagger(d).$$

By our definition $\mathcal{H}_\lambda^\dagger$ is a right $\widehat{\mathfrak{g}}$-module. A decreasing filtration $\{F^p\mathcal{H}_\lambda^\dagger\}$ is defined by

$$F^p\mathcal{H}_\lambda^\dagger = \prod_{d\geq p}\mathcal{H}_\lambda^\dagger(d).$$

There is a unique canonical perfect bilinear pairing

$$\langle \ | \ \rangle : \mathcal{H}_\lambda^\dagger \times \mathcal{H}_\lambda \longrightarrow \mathbb{C}. \tag{1}$$

The Lie algebra $\widehat{\mathfrak{g}}_N$ is defined as

$$\widehat{\mathfrak{g}}_N = \bigoplus_{j=1}^{N} \mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}((\xi_j)) \oplus \mathbb{C}c$$

with the following commutation relations.

$$[(X_j \otimes f_j), (Y_j \otimes g_j)] = ([X_j, Y_j] \otimes f_j g_j) + c \sum_{j=1}^{N} (X_j, Y_j) \operatorname*{Res}_{\xi_j=0} (g_j df_j)$$

where (a_j) means $(a_1, a_2, \ldots, a_N)$ and c belongs to the center of $\widehat{\mathfrak{g}}_N$.

Let $\mathfrak{X} = (C; q_1, q_2, \ldots, q_N; \eta_1, \eta_2, \ldots, \eta_N)$ be a pointed Riemann Surface with formal neighbourhoods and define

$$\widehat{\mathfrak{g}}(\mathfrak{X}) = \mathfrak{g} \otimes_{\mathbb{C}} H^0(C, \mathcal{O}_C(* \sum_{j=1}^{N} q_j)).$$

We have the natural embedding

$$t = \oplus t_i : H^0(C, \mathcal{O}_C(* \sum_{j=1}^{N} q_j)) \hookrightarrow \bigoplus_{j=1}^{N} \mathbb{C}((\xi_j))$$

given by Laurent expansion using the formal neighbourhoods. In the following we often regard $H^0(C, \mathcal{O}_C(* \sum_{j=1}^{N} q_j))$ as a subspace of $\bigoplus_{j=1}^{N} \mathbb{C}((\xi_j))$. One has by lemma 1.1.15 in [14], that $\widehat{\mathfrak{g}}(\mathfrak{X})$ is a Lie subalgebra of $\widehat{\mathfrak{g}}_N$.

Let us fix a non-negative integer ℓ. For each $\vec{\lambda} = (\lambda_1, \ldots, \lambda_N) \in (P_\ell)^N$, the left $\widehat{\mathfrak{g}}_N$-module $\mathcal{H}_{\vec{\lambda}}$ and a right $\widehat{\mathfrak{g}}_N$-module $\mathcal{H}_{\vec{\lambda}}$ are defined by

$$\mathcal{H}_{\vec{\lambda}} = \mathcal{H}_{\lambda_1} \otimes_{\mathbb{C}} \cdots \otimes_{\mathbb{C}} \mathcal{H}_{\lambda_N}$$
$$\mathcal{H}^{\dagger}_{\vec{\lambda}} = \mathcal{H}^{\dagger}_{\lambda_1} \widehat{\otimes}_{\mathbb{C}} \cdots \widehat{\otimes}_{\mathbb{C}} \mathcal{H}^{\dagger}_{\lambda_N}.$$

The hats over the tensor product means that the algebraic tensor product has been completed with respect to the induced filtration.

For each element $X_j \in \mathfrak{g}$, $f(\xi_j) \in \mathbb{C}((\xi_j))$, the action ρ_j of $X_j[f_j]$ on $\mathcal{H}_{\vec{\lambda}}$ is given by

$$\rho_j(X_j[f_j])|v_1 \otimes \cdots \otimes v_N\rangle = |v_1 \otimes \cdots \otimes v_{j-1} \otimes (X_j[f_j])v_j \otimes v_{j+1} \otimes \cdots v_N\rangle$$

where $|v_1 \otimes \cdots \otimes v_N\rangle$ means $|v_1\rangle \otimes \cdots \otimes |v_N\rangle$, $|v_j\rangle \in \mathcal{H}_{\lambda_j}$. The left $\widehat{\mathfrak{g}}_N$-action is given by

$$(X_1 \otimes f_1, \ldots, X_N \otimes f_N)|v_1 \otimes \cdots v_N\rangle = \sum_{j=1}^{N} \rho_j(X_j[f_j])|v_1 \otimes \cdots v_N\rangle.$$

Similarly, the right $\widehat{\mathfrak{g}}_N$-action on $\mathcal{H}^\dagger_{\vec{\lambda}}$ is defined by

$$\langle u_1 \otimes \cdots u_N | (X_1 \otimes f_1, \ldots, X_N \otimes f_N) = \sum_{j=1}^{N} \langle u_1 \otimes \cdots u_N | \rho_j(X_j[f_j]).$$

As a Lie subalgebra, $\widehat{\mathfrak{g}}(\mathfrak{X})$ operates on $\mathcal{H}_{\vec{\lambda}}$ and $\mathcal{H}^\dagger_{\vec{\lambda}}$ as

$$(X \otimes f)|v_1 \otimes \cdots \otimes v_N\rangle = \sum_{j=1}^{N} \rho_j(X \otimes t_j(f))|v_1 \otimes \cdots v_N\rangle$$

and as

$$\langle u_1 \otimes \cdots \otimes u_N|(X \otimes f) = \sum_{j=1}^{N} \langle u_1 \otimes \cdots \otimes u_N|\rho_j(X \otimes t_j(f)).$$

The pairing $\langle \ | \ \rangle$ introduced in (1) induces a perfect bilinear pairing

$$\langle \ | \ \rangle : \mathcal{H}^\dagger_{\vec{\lambda}} \times \mathcal{H}_{\vec{\lambda}} \to \mathbb{C} \tag{2}$$

given by

$$(\langle u_1 \otimes \ldots \otimes u_N|, |v_1 \otimes \ldots \otimes u_N\rangle) \to \langle u_1|v_1\rangle\langle u_2|v_2\rangle \cdots \langle u_N|v_N\rangle$$

which is $\widehat{\mathfrak{g}}_N$-invariant:

$$\langle \Psi(X_j \otimes f_j)|\Phi\rangle = \langle \Psi|(X_j \otimes f_j)\Phi\rangle.$$

Now we are ready to define the space of abelian conformal block attached to $\mathfrak{X}$.

Definition 3.1. Assume that $\mathfrak{X}$ is saturated. Put

$$\mathcal{V}_{\vec{\lambda}}(\mathfrak{X}) = \mathcal{H}_{\vec{\lambda}}\mathcal{H}_{\vec{\lambda}}/\widehat{\mathfrak{g}}(\mathfrak{X})\mathcal{H}_{\vec{\lambda}}.$$

The vector space $\mathcal{V}_{\vec{\lambda}}(\mathfrak{X})$ is called the *space of covacua* attached to $\mathfrak{X}$. The *space of conformal blocks* attached to $\mathfrak{X}$ is defined as

$$\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{X}) = Hom_{\mathbb{C}}(\mathcal{V}_{\vec{\lambda}}(\mathfrak{X}), \mathbb{C}).$$

One gets that

$$\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{X}) = \{ \langle \Psi| \in \mathcal{V}^\dagger_{\vec{\lambda}} \,|\langle \Psi|\, \widehat{\mathfrak{g}}(\mathfrak{x}) = 0 \}. \tag{3}$$

Moreover, the pairing (2) induces a perfect pairing

$$\langle \ | \ \rangle : \mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{x}) \times \mathcal{V}_{\vec{\lambda}}(\mathfrak{x}) \to \mathbb{C}.$$

Note that the above definitions can be extended to the case in which the curve C may have singularities but the points q_j's are non-singular point of C.

The following theorem is proved in [11].

Theorem 3.2. *The vector spaces* $\mathcal{V}_{\vec{\lambda}}(\mathfrak{x})$ *and* $\mathcal{V}^{\dagger}_{\vec{\lambda}}(\mathfrak{x})$ *are finite-dimensional.*

We let $\mathcal{D}$ be the automorphism group $Aut\mathbb{C}((\xi))$ of the field $\mathbb{C}((\xi))$ of formal Laurent series as a $\mathbb{C}$-algebra. There is a natural isomorphism

$$\begin{aligned} \mathcal{D} &\simeq \{ \sum_{n=0}^{\infty} a_n \xi^{n+1} \mid a_0 \neq 0 \} \\ h &\mapsto h(\xi) \end{aligned}$$

where the composition $h \circ g$ of h, $g \in \mathcal{D}$ corresponds to the formal power series $h(g(\xi))$.

Put

$$\mathcal{D}^p = \{ h \in \mathcal{D} \mid h(\xi) = \xi + a_p \xi^{p+1} + \cdots \}$$

for a positive integer p. Also let

$$\underline{d} = \mathbb{C}[[\xi]]\xi \frac{d}{d\xi}$$

$$\underline{d}^p = \mathbb{C}[[\xi]]\xi^{p+1} \frac{d}{d\xi} \quad p = 0, 1, 2, \ldots$$

For any $\underline{l} \in \underline{d}$ and $f(\xi) \in \mathbb{C}[[\xi]]$ define $\exp(\underline{l})(f(\xi))$ by

$$\exp(\underline{l})(f(\xi)) = \sum_{k=0}^{\infty} \frac{1}{k!} (\underline{l}^k f(\xi)).$$

Set

$$\mathcal{D}^0_+ = \{ h \in \mathcal{D} \mid h(\xi) = a\xi + a_1 \xi^2 + \cdots, \quad a > 0 \}$$

$$\underline{d}^0_+ = \{ l(\xi) \frac{d}{d\xi} \mid l(\xi) = \alpha\xi + \alpha_1 \xi^2 + \cdots, \quad \alpha \in \mathbb{R} \}$$

Then, we have the following result.

Lemma 3.1. *The exponential map*

$$\begin{aligned} \exp \ : \ \underline{d} &\to \mathcal{D} \\ \underline{l} &\mapsto \exp(\underline{l}) \end{aligned}$$

is surjective. Moreover, the exponential map induces an isomorphism

$$\exp \ : \ \underline{d}^0_+ \simeq \mathcal{D}^0_+.$$

For any element $\underline{l} \in \mathcal{D}_+^0$ we define $\exp(T[\underline{l}])$ by

$$\exp(T[\underline{l}]) = \sum_{k=0}^{\infty} \frac{1}{k!} T[\underline{l}]^k.$$

Then, $\exp(T[\underline{l}])$ operates on $\mathcal{H}_\lambda$ from the left and on $\mathcal{H}_\lambda^\dagger$ from the right. By Lemma 3.1, for any automorphism $h \in \mathcal{D}_+^0$, there exist a unique $\underline{l} \in \underline{d}_+^0$ with $\exp(\underline{l}) = h$. Now for $h \in \mathcal{D}_+^0$ define the operator $G[h]$ by

$$G[h] = \exp(-T[\underline{l}]) \tag{4}$$

where $\exp(\underline{l}) = h$. Then, $G[h]$ operates on $\mathcal{H}_\lambda$ from the left and on $\mathcal{H}_\lambda^\dagger$ from the right.

Theorem 3.3 ([14], Theorem 3.2.5). *For any $h_j \in \mathcal{D}_+^0$, $j = 1, 2, \ldots, N$ and a pointed Riemann surface with formal neighbourhoods*

$$\mathfrak{X} = (C; q_1, q_2, \ldots, q_N; \xi_1, \xi_2, \ldots, \xi_N)$$

put

$$\mathfrak{X}_{(h)} = (C; q_1, q_2, \ldots, q_N; h_1(\xi_1), h_2(\xi_2), \ldots, h_N(\xi_N)).$$

Then, the isomorphism $G[h_1]\widehat{\otimes}\cdots\widehat{\otimes}G[h_N]$

$$\mathcal{H}_{\vec{\lambda}}^\dagger \to \mathcal{H}_{\vec{\lambda}}^\dagger$$
$$\langle \phi_1 \widehat{\otimes} \cdots \widehat{\otimes} \phi_N | \mapsto \langle \phi_1 G[h_1] \widehat{\otimes} \cdots \widehat{\otimes} \phi_N G[h_N] |$$

induces the canonical isomorphism

$$G[\vec{h}] = G[h_1]\widehat{\otimes}\cdots\widehat{\otimes}G[h_N] : \mathcal{V}_{\vec{\lambda}}^\dagger(\mathfrak{X}) \to \mathcal{V}_{\vec{\lambda}}^\dagger(\mathfrak{X}_{(h)}).$$

Now let us compare the axioms of MF and properties of the space of conformal blocks.

Corresponding to **MF1** by (3) we have

$$\mathcal{V}_{\vec{\lambda}_1, \vec{\lambda}_2}^\dagger(\mathfrak{X}_1 \sqcup \mathfrak{X}_2) = \mathcal{V}_{\vec{\lambda}_1}(\mathfrak{X}_1) \otimes \mathcal{V}_{\vec{\lambda}_2}(\mathfrak{X}_2).$$

Corresponding to **MF2** we have the following theorem.

Theorem 3.4. *Let $\mathfrak{X} = (C; q_1, q_2, \ldots, q_N; \eta_1, \eta_2, \ldots, \eta_N)$ be a pointed nodal curve with formal neighbourhoods. We let $\nu : \widetilde{C} \to C$ be the desingularization at the node p. Put*

$$\widetilde{\mathfrak{X}} = (\widetilde{C}; q_1, q_2, \ldots, q_N, p_+, p_-; \eta_1, \eta_2, \ldots, \eta_N, \xi_+, \xi_-)$$

where $\nu^{-1}(p) = \{p_+, p_-\}$ *and* ξ_+ *(resp.* ξ_-*) is a coordinate with center* p_+ *(resp.* p_-*.) Then there exists a canonical isomorphism*

$$\mathcal{V}_{\vec{\lambda}}^{\dagger}(\mathfrak{X}) = \bigoplus_{\mu \in P_l} \mathcal{V}_{\vec{\lambda},\mu,\mu^{\dagger}}(\widetilde{\mathfrak{X}}).$$

Corresponding to **MF3** by Theorem 3.1 and (3) we have

$$\mathcal{V}(\emptyset) = \mathbb{C},$$

since we may regard $\widehat{\mathfrak{g}}(\emptyset) = \widehat{\mathfrak{g}}$.

Finally corresponding to **MF4** and **MF5** we have

$$\dim_{\mathbb{C}} \mathcal{V}_{\lambda}(\mathbb{P}^1; q; \eta) = \begin{cases} 1 & \lambda = 0 \\ 0 & \lambda \neq 0, \end{cases}$$

$$\dim_{\mathbb{C}} \mathcal{V}_{\lambda,\mu}(\mathbb{P}^1; q_1, q_2; \eta_1, \eta_2) = \begin{cases} 1 & \lambda = \mu^{\dagger} \\ 0 & \lambda \neq \mu^{\dagger}. \end{cases}$$

Thus our CFT has similar properties to the axioms of MF. But CFT does depend on complex structures of Riemann surfaces, namely for pointed Riemann surfaces with coordinates

$$\mathfrak{X} = (C; q_1, q_2, \ldots, q_N; \eta_1, \eta_2, \ldots, \eta_N)$$
$$\mathfrak{X}' = (C; q_1', q_2', \ldots, q_N'; \eta_1', \eta_2', \ldots, \eta_N')$$

there is an isomorphism

$$\mathcal{V}_{\vec{\lambda}}^{\dagger}(\mathfrak{X}) \simeq \mathcal{V}_{\vec{\lambda}}^{\dagger}(\mathfrak{X}')$$

but the isomorphism is not canonical. To study the isomorphism let us consider a family of pointed Riemann surfaces with coordinates.

A family of pointed Riemann surfaces with formal neighbourhoods

$$\mathfrak{F} = (\pi \mathcal{C} \to \mathcal{B}; \vec{s}; \vec{\eta})$$

is the following date:

(1) Connected complex manifolds $\mathcal{C}$ and $\mathcal{B}$, such that $dim_{\mathbb{C}} C = dim_{\mathbb{C}} \mathcal{B} + 1$.
(2) A holomorphic submersion $\pi : \mathcal{C} \to \mathcal{B}$.
(3) Holomorphic sections s_j, $j = 1, \ldots, N$ of π.
(4) Filtered $\mathcal{O}_{\mathcal{B}}$-algebra isomorphisms

$$\eta_j :: \widehat{\mathcal{O}}_{/s_j} = \varprojlim_{n \to \infty} \mathcal{O}_Y / I_j^n \simeq \mathcal{O}_{\mathcal{B}}[[\xi]],$$

where I_j is the defining ideal of $s_j(\mathcal{B})$ in $\mathcal{C}$, $j = 1, \ldots, N$.

Let $\mathfrak{F} = (\pi : \mathcal{C} \to \mathcal{B}; s_1, \ldots, s_N; \eta_1, \ldots, \eta_N)$ be a family of pointed saturated Riemann Surfaces of genus g with formal neighbourhoods.

The sheaf $\widehat{\mathfrak{g}}_N(\mathcal{B})$ of affine Lie algebra over $\mathcal{B}$ is the sheaf of $\mathcal{O}_\mathcal{B}$-module

$$\widehat{\mathfrak{g}}_N(\mathcal{B}) = \mathfrak{g} \otimes_\mathbf{C} (\bigoplus_{j=1}^{N} \mathcal{O}_\mathcal{B}((\xi_j))) \oplus \mathcal{O}_\mathcal{B} \cdot c$$

with the following commutation relation, which is $\mathcal{O}_\mathcal{B}$-bilinear.

$$[(X_1 \otimes f_1, \ldots, X_N \otimes f_N), (Y_1 \otimes g_1, \ldots, Y_N \otimes g_N)]$$
$$= ([X_1, Y_1] \otimes (f_1 g_1), \ldots, [X_N, Y_N] \otimes (f_N g_N)) \oplus c \cdot \sum_{j=1}^{N} (X_j, Y_j) \operatorname*{Res}_{\xi_j = 0} (g_j df_j)$$

where $X_j, Y_j \in \mathfrak{g}$, $\quad f_j, g_j \in \mathcal{O}_\mathcal{B}((\xi_j))$ and we require c to be central. Put

$$\widehat{\mathfrak{g}}(\mathfrak{F}) = \mathfrak{g} \otimes_\mathbf{C} \pi_*(\mathcal{O}_\mathcal{C}(*S)),$$

where we define

$$S = \sum_{j=1}^{N} s_j(\mathcal{B})$$
$$\pi_*(\mathcal{O}_\mathcal{C}(*S)) = \varinjlim_k \pi_*(\mathcal{O}_\mathcal{C}(kS)).$$

Laurent expansion using the formal neighbourhoods η_j's gives an inclusion:

$$\tilde{t} : \pi_*(\mathcal{O}_\mathcal{B}(*S)) \to \bigoplus_{j=1}^{N} \mathcal{O}_\mathcal{B}((\xi_j))$$

and we may regard $\widehat{\mathfrak{g}}(\mathfrak{F})$ as a Lie subalgebra of $\widehat{\mathfrak{g}}_N(\mathcal{B})$. For any $\vec{\lambda} = (\lambda_1, \ldots, \lambda_N) \in (P_\ell)^N$, put

The pairing (2) induces an $\mathcal{O}_\mathcal{B}$-bilinear pairing

$$\langle \ | \ \rangle : \mathcal{H}^\dagger_{\vec{\lambda}}(\mathcal{B}) \times \mathcal{H}_{\vec{\lambda}}(\mathcal{B}) \to \mathcal{O}_\mathcal{B}.$$

The sheaf of affine Lie algebra $\widehat{\mathfrak{g}}_N(\mathcal{B})$ acts on $\mathcal{H}_{\vec{\lambda}}(\mathcal{B})$ and $\mathcal{H}^\dagger_{\vec{\lambda}}(\mathcal{B})$ by

$$((X_1 \otimes \sum_{n \in \mathbf{Z}} a_n^{(1)} \xi_1^n), \ldots, (X_N \otimes \sum_{n \in \mathbf{Z}} a_n^{(N)} \xi_N^n))(F \otimes |\Psi\rangle)$$

$$= \sum_{j=1}^{N} \sum_{n \in \mathbf{Z}} (a_n^{(j)} F) \otimes \rho_j(X_j(n)) |\Psi\rangle$$

The action of $\widehat{\mathfrak{g}}_N(\mathcal{B})$ on $\mathcal{H}^\dagger_{\vec{\lambda}}(\mathcal{B})$ is the dual action of $\mathcal{H}_{\vec{\lambda}}(\mathcal{B})$, that is,

$$\langle \Psi a | \Phi \rangle = \langle \Psi | a\Phi \rangle \quad \text{for any } \ a \in \widehat{\mathfrak{g}}_N.$$

Now we can define a sheaf $\mathcal{V}_{\vec{\lambda}}(\mathfrak{F})$ of conformal blocks and a sheaf $\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F})$ of covacua similarly as above.

Definition 3.2. For the family $\mathfrak{F}$ of pointed Riemann surfaces with formal neighbourhoods, we define the sheaves of $\mathcal{O}_{\mathcal{B}}$-modules on $\mathcal{B}$

$$\mathcal{V}_{\vec{\lambda}}(\mathfrak{F}) = \mathcal{H}_{\vec{\lambda}}(\mathcal{B}) / \widehat{\mathfrak{g}}(\mathfrak{F}) \mathcal{H}_{\vec{\lambda}}(\mathcal{B})$$

$$\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F}) = \underline{\mathrm{Hom}}_{\mathcal{O}_{\mathcal{B}}}(\mathcal{V}_{\vec{\lambda}}(\mathcal{B}), \mathcal{O}_{\mathcal{B}}).$$

These are the *sheaf of covacua* and the *sheaf of conformal blocks* attached to the family $\mathfrak{F}$.

Note that we have

$$\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F}) = \{ \langle \Psi | \in \mathcal{H}^\dagger_{\vec{\lambda}}(\mathfrak{F}) \mid \ \langle \Psi | a = 0 \quad \text{for any } \ a \in \widehat{\mathfrak{g}}(\mathfrak{F}) \}.$$

For a point $b \in \mathcal{B}$ put

$$\mathfrak{F}_b = (\pi^{-1}(b); s_1(b), \ldots, s_N(b); \eta_1(b), \ldots, \eta_N(b)).$$

Theorem 3.5 ([11]). *The sheaves $\mathcal{V}_{\vec{\lambda}}(\mathfrak{F})$ and $\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F})$ of conformal blocks and covacua are locally free $\mathcal{O}_{\mathcal{B}}$-modules of finite rank and carries projectively flat connections. In particular we have have*

$$\mathcal{V}_{\vec{\lambda}}(\mathfrak{F}) \otimes_{\mathcal{O}_{\mathcal{B}}} \mathbb{C}_b = \mathcal{V}_{\vec{\lambda}}(\mathfrak{F}_b)$$

$$\mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F}) \otimes_{\mathcal{O}_{\mathcal{B}}} \mathbb{C}_b = \mathcal{V}^\dagger_{\vec{\lambda}}(\mathfrak{F}_b)$$

where $\mathcal{V}_{\vec{\lambda}}(\mathfrak{F}_b) = \mathcal{O}_{\mathcal{B},b}/\mathfrak{m}_b$.

3.2. *Abelian Conformal Field Theory*

By using the fermion Fock space we can construct abelian conformal field theory. Details can be found in [1].

In abelian conformal field theory, to each pointed Riemann surface with formal neighbourhoods

$$\mathfrak{X} = (C; q_1, \ldots, q_N; \eta_1, \ldots, \eta_N)$$

we associate a vector space $\mathcal{V}^\dagger_{ab}(\mathfrak{X})$ called abelian conformal block. We can prove the following theorems.

Theorem 3.6 ([1], Theorem 3.2). *For any pointed Riemann surface* $\mathfrak{X} = (C; Q_1, \ldots, Q_N; \xi_1, \ldots \xi_N)$ *with formal neighbourhoods we have*

$$\dim_{\mathbf{C}} \mathcal{V}^{\dagger}_{ab}(\mathfrak{X}) = 1.$$

Theorem 3.7 ([1], Theorem 3.2.). *The space of abelian conformal blocks* $\mathcal{V}^{\dagger}_{ab}(\mathfrak{X})$ *is isomorphic to the determinant of the canonical bundle* ω_C.

Next let us consider a nodal curve C with node P. Let $\widetilde{C}$ be the Riemann surface obtained by resolving the singularity at P and let $\pi : \widetilde{C} \to C$ be the natural holomorphic mapping. Then $\pi^{-1}(P)$ consists of two points P_+ and P_-. Let

$$\mathfrak{X} = (C; q_1, \ldots, q_N; \xi_1, \ldots, \xi_N)$$

be a pointed nodal curve with formal neighbourhoods and we let

$$\widetilde{\mathfrak{X}} = (\widetilde{C}; P_+, P_-, q_1, \ldots, q_N; z, w, \xi_1, \ldots, \xi_N)$$

be the associated pointed Riemann surface with formal neighbourhoods.

Theorem 3.8 ([1], Theorem 3.5). *There exists a natural isomorphism*

$$\iota^*_{+,-} : \mathcal{V}^{\dagger}_{ab}(\widetilde{\mathfrak{X}}) \cong \mathcal{V}^{\dagger}_{ab}(\mathfrak{X}).$$

About the coordinate change we have the following proposition.

Proposition 3.1. *For any* $h_j \in \mathcal{D}^0_+$, $j = 1, 2, \ldots, N$ *and* N*-pointed curve*

$$\mathfrak{X} = (C; Q_1, Q_2, \ldots, Q_n; \xi_1, \xi_2, \ldots, \xi_N)$$

with formal neighbourhoods, put

$$\mathfrak{X}_{(h)} = (C; Q_1, Q_2, \ldots, Q_N; h_1(\xi_1), h_2(\xi_2), \ldots, h_N(\xi_N)).$$

Then there exists a canonical isomorphism

$$\mathcal{V}^{\dagger}_{ab}(\mathfrak{X}) \simeq \mathcal{V}^{\dagger}_{ab}(\mathfrak{X}_{(h)})$$

Now similar to the case of non-abelian conformal field theory we can define a sheaf of abelian conformal blocks $\mathcal{V}^{\dagger}_{ab}(\mathfrak{X})$ for a family $\mathfrak{X} = (\pi\mathcal{C} \to \mathcal{B}; s_1, \ldots, s_N; \eta_1, \ldots, \eta_N)$ of pointed curves surfaces of formal coordinates.

Theorem 3.9. *The sheaf* $\mathcal{V}^{\dagger}_{ab}(\mathfrak{F})$ *is an invertible* $\mathcal{O}_{\mathcal{B}}$*-module and carries a projectively flat connection.*

4. Modular Functor

4.1. *Teichmüller space*

The *Teichmüller space* $\mathcal{T}_{(\Sigma,P)}$ of the pointed surface (Σ, P) is by definition the set of equivalence classes of complex structures on (Σ, P).

Theorem 4.1 (Bers). *There is a natural structure of a finite dimensional complex analytic manifold on Teichmüller space $\mathcal{T}_{(\Sigma,P)}$. Associated to any morphism of pointed surfaces $f : (\Sigma_1, P_1) \to (\Sigma_2, P_2)$ there is a biholomorphism $f^* : \mathcal{T}_{(\Sigma_1,P_1)} \to \mathcal{T}_{(\Sigma_2,P_2)}$ which is induced by mapping a complex structure $\mathbf{C} = (C, Q, W)$, $\phi : \Sigma_1 \to C$ to $\phi \circ f^{-1} : \Sigma_2 \to C$. Moreover, compositions of morphisms go to compositions of induced biholomorphisms.*

There is an action of $\mathbb{R}$ on $\mathcal{T}_{(\Sigma,P)}$ given by scaling the tangent vectors. This action is free and the quotient $\mathcal{T}^{(r)}_{(\Sigma,P)} = \mathcal{T}_{(\Sigma,P)}/\mathbb{R}$ is a smooth manifold, which we call the *reduced* Teichmüller space of the pointed surface (Σ, P). Moreover the projection map π_P descend to a smooth projection map from $\mathcal{T}^{(r)}_{(\Sigma,P)}$ to $\sqcup_{p\in P} P(T_p\Sigma)$, which we denote $\pi^{(r)}_P$. We denote the fiber of this map over $V \in \sqcup_{p\in P} P(T_p\Sigma)$ by $\mathcal{T}_{(\Sigma,P,V)}$. Teichmüller space of a marked surface $\mathbf{\Sigma} = (\Sigma, P, V, L)$ is by definition $\mathcal{T}_{\mathbf{\Sigma}} = \mathcal{T}_{(\Sigma,P,V)}$, which we call the Teichmüller space of the marked surface. Morphisms of marked surfaces induce diffeomorphism of the corresponding Teichmüller spaces of marked surfaces, which of course also behaves well under composition. We observe that the self-morphism (id, s) of a marked surface acts trivially on the associated Teichmüller space for all integers s. General Teichmüller theory implies that

Theorem 4.2. *The Teichmüller space $\mathcal{T}_{\mathbf{\Sigma}}$ of any marked surface $\mathbf{\Sigma}$ is contractible.*

We note that we have a canonical isomorphism

$$\mathcal{O}_{C,q}/\mathfrak{m}_q^2 \simeq \mathbb{C} \oplus T_q^*C,$$
$$f \mapsto (f(q), df_q).$$

Hence a firsts order neighbourhood determines and is also determined by a non-zero vector in T_q^*C, specified by the property that it maps to $1 \in \mathbb{C}$ or equivalently a vector in T_qC pairing to unity with this vector.

4.2. *Construction of Modular Functor*

Let $\mathfrak{F} = (\pi\mathcal{C} \to \mathcal{B}; \vec{s}; \vec{\eta})$ be a family of pointed Riemann surfaces with formal neighbourhoods and assume we have a smooth fiber preserving diffeomor-

phism $\Phi_{\mathfrak{F}}$ from $Y = \Sigma \times \mathcal{B}$ to $\mathcal{C}$ taking the marked points to the sections $\vec{s}$ and inducing the identity on $\mathcal{B}$. This is always possible if $\mathcal{B}$ is small enough. This data induces unique holomorphic map $\Psi_{\mathfrak{F}}$ from $\mathcal{B}$ to the Teichmüller space $\mathcal{T}_{(\Sigma,P)}$ of the surface (Σ, P) by the universal property of Teichmüller space. Note that the Teichmüller space $\mathcal{T}_{(\Sigma,P)}$ is covered by the image $\Psi_{\mathfrak{F}}$ of families $\mathfrak{F}$ of pointed Riemann surfaces with coordinates.

As is shown in [2] §5, there exists a holomorphic vector bundle $\mathcal{V}^{\dagger}_{\vec{\lambda}} = \mathcal{V}^{\dagger}_{\vec{\lambda}}(\Sigma, P)$ over Teichmüller space $\mathcal{T}_{(\Sigma,P)}$ is specify to be the bundle $(\Psi^{-1}_{\mathfrak{F}})^* \mathcal{V}^{\dagger}_{\vec{\lambda}} \mathfrak{F})$ over $\Psi_{\mathfrak{F}}(\mathcal{B})$ for any good families of complex structures on (Σ, P). The holomorphic bundle $\mathcal{V}^{\dagger}_{\vec{\lambda}}$ is called the bundle of conformal blocks over the Teichmüller space.

Now by the above Theorem 3.5 the holomorphic bundle $\mathcal{V}^{\dagger}_{\vec{\lambda}} \to \mathcal{T}_{(\Sigma,P)}$ carries a projectively flat connection. If the connection were flat, then then the space of flat sections $H^0_{flat}(\mathcal{T}_{(\Sigma,P)}, \mathcal{V}^{\dagger}_{\vec{\lambda}})$ would define Modular Functor. To define a flat connection one possibility is to find a holomorphic line bundle $\mathcal{L}$ on the Teichmüller space $\mathcal{T}_{(\Sigma,P)}$ such that $\mathcal{V}^{\dagger}_{\vec{\lambda}} \otimes_{\mathcal{O}_{\mathcal{B}}} \mathcal{L}$ caries a flat connection. To do this we use the above abelian conformal field theory.

First note that similar to above we have the sheaf $\mathcal{V}^{\dagger}_{ab}$ of abelian conformal blocks over the Teichmüller space $\mathcal{T}_{(\Sigma,P)}$. Then we have the following result on curvatures.

Theorem 4.3. *The curvature form ω of $\mathcal{V}^{\dagger}_{\vec{\lambda}} \to \mathcal{T}_{(\Sigma,P)}$ and the curvature form ω_{ab} are related by*

$$\omega = \frac{c_v}{2} \omega_{ab} \otimes id$$

where

$$c_v = \frac{\ell \cdot \dim_{\mathbb{C}} \mathfrak{g}}{g^* + \ell}.$$

Here g^* is the dual Coxeter number of the Lie algebra $\mathfrak{g}$.

Thus the desired line bundle $\mathcal{L}$ would be given by

$$\mathcal{L} = \left(\mathcal{V}^{\dagger}_{ab}\right)^{-c_v/2}. \tag{5}$$

But since $c_v/2$ is a rational number, the definition (5) has a meaning if and only if there exists nowhere vanishing holomorphic section of $\mathcal{V}^{\dagger}_{ab}$ over the Teichmüller space. In our paper [1] it is shown that such a section indeed exists, if we fix a Lagrangian subspace of each pointed Riemann surface (for the details see [1]). Thus we can define the line bundle $\mathcal{L}$ as in (5). Moreover

in our paper [2] we show that the vector space

$$H^0_{flat}(\mathcal{T}_{(\Sigma,P)}, \mathcal{V}^\dagger_\lambda \otimes_{\mathcal{O}_\mathcal{B}} \mathcal{L})$$

defines a modular functor.

Finally, in our paper [3] it is shown that in case $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$, TQFT obtained form our MF is equivalent to the Reshetikhin-Turaev TQFT developed in [4] and [5].

References

1. J. E. Andersen and K. Ueno, *Abelian Conformal Field Theory and Determinant Bundles*, Internat. J. Math. 18 (2007), 919–993.
2. J. E. Andersen and K. Ueno, *Geometric construction of modular functors from conformal field theory*, J. Knot Theory Ramifications 16 (2007), 127–202.
3. J. E. Andersen and K. Ueno, *Construction of the Reshetikhin-Turaev TQFT from conformal field theory*, Preprint in preparation.
4. C. Blanchet, N. Habegger, G. Masbaum and P. Vogel, *Three-manifold invariants derived from the Kauffman bracket*, Topology 31 no. 4 (1992), 685–699.
5. C. Blanchet, N. Habegger, G. Masbaum and P. Vogel, *Topological quantum field theories derived from the Kauffman bracket*, Topology, 34 no. 4 (1995), 883–927.
6. J. Grove, *Constructing TQFTs from modular functors*, J. Knot Theory Ramifications 10 no. 8 (2001), 1085–1131.
7. N. Kawamoto, Y. Namikawa, A. Tsuchiya and Y. Yamada, *Geometric realization of conformal field theory on Riemann surfaces*, Comm. Math. Phys., 116 (1988), 247–308.
8. V. G. Knizhnik and A. B. Zamolodchikov, *Current algebra and Wess-Zumino model in two dimensions*, Nucl. Phys. B, 247 (1984), 83–103.
9. G. Segal, *The definition of conformal field theory*, Topology, geometry and quantum field theory, 421–577, London Math. Soc. Lecture Note Ser., 308 (2004).
10. A. Tsuchiya and Y. Kanie, *Vertex Operators in Conformal Field Theory on $\mathbb{P}^1$ and Monodromy Representations of Braid Group*, Advanced Studies in Pure Mathematics 16 (1988), 297–326.
11. A. Tsuchiya, K. Ueno and Y. Yamada, *Conformal Field Theory on Universal Family of Stable Curves with Gauge Symmetries*, Advanced Studies in Pure Mathematics 19 (1989), 459–566.
12. V. Turaev, *Quantum invariants of knots and 3-manifolds*, (W. de Gruyter, Berlin, 1994).
13. K. Ueno, *On conformal field theory*, London Math. Soc. Lecture Note 208 (1995), 283–345.
14. K. Ueno, *Introduction to conformal field theory with gauge symmetries*, Geometry and physics (Aarhus, 1995), Lecture Notes in Pure and Appl. Math. 184 (1997), 603–745, Dekker, New York.
15. K. Walker, *On Witten's 3-manifold invariants*, Preliminary version 2 (Preprint, 1991).

Advances in Algebra and Combinatorics
edited by K. P. Shum et al.
© 2008 World Scientific Publishing Co. (pp. 353–371)

CLASSIFICATION OF 5-DIMENSIONAL MD-ALGEBRAS HAVING COMMUTATIVE DERIVED IDEALS *

LE ANH VU and KAR PING SHUM

Department of Mathematics and Informatics,
University of Pedagogy, Ho Chi Minh City, Viet nam
E-mail: vula@math.hcmup.edu.vn
Department of Mathematics,
The University of Hong Kong, Pokfulam Road
E-mail: kpshum@maths.hku.hk

In this paper, we study a subclass of the class of MD-algebras, i.e., the class of solvable real Lie algebras such that the K-orbits of its corresponding connected and simply connected Lie groups are either orbits of dimension zero or orbits with maximal dimensions. Our main result is to classify, up to isomorphism, all the 5-dimensional MD-algebras having commutative derived ideals.

Keywords: Lie group; Lie algebra; $MD5$-group; $MD5$-algebra; C^*-algebra; K-orbits.

1. Introduction

The method of describing the structure of C^*-algebras by using K-functors was first suggested by D. N. Diep ([2]) in 1974. By applying the K-homology functors proposed by Brown - Douglas - Fillmore (for brevity, the BDF K-functors), Diep gave a description for the $C^*(\mathrm{Aff}\mathbb{R})$ of the group $\mathrm{Aff}\mathbb{R}$ of the affine transformations of the real line. In 1975, by using the method of Diep, J. Rosenberg ([7, 8]) gave a description for the C^*-algebra of the group $\mathrm{Aff}\mathbb{C}$ and some other groups. In 1977, D. N. Diep ([3]) further gave a complete system of invariants of C^*-algebras of type I by using the BDF K-homology functors. Hence, it is natural to propose the following two general problems:

- Generalize the K-homology functors so that these functors can be applied to a larger class of C^*-algebras.

*AMS Mathematics Subject Classification (2000): Primary 22E45, Secondary 46E25, 20C20.

- Find the C*-algebras which can be described by using the generalized K-functors.

Concerning the first problem, we note that G. G. Kasparov ([5]) in 1980 introduced the concept of KK-functors which is a generalized concept of BDF K-homology functors. Then by using KK-functors, G. G. Kasparov described theC^*-algebra of the Heisenberg groups H_{2n+1}.

For the second problem, it was noticed that this problem is closely related with the Orbit Method proposed by A. A. Kirillov ([6]) in 1962. After studying the Kirillov's Orbit Method, Diep in 1980 suggested to consider the class of Lie groups and Lie algebras MD and $\overline{MD}$ ([4]) so that the C^*-algebras of them can be described by using KK-functors. If G is an n-dimensional real Lie group, then G is called a MDn-group or a MD-group with dimension n iff the orbits of G in the K-representation (K-orbits) are orbits of dimension zero or orbits of maximal dimension (i.e. dimension k, where k is some even constant, $k \leq n$). When $k = n$, we call G an $\overline{MDn}$-group or $\overline{MD}$-group of dimension n. The corresponding Lie algebra $Lie(G)$ of G is said to be an MDn-algebra or $\overline{MDn}$-algebra, respectively. It is clear that the class $\overline{MD}$ is a subclass of the class MD. Thus, the problem of classifying MD-algebras, describing the K-representation of MD-groups and characterizing the C^*-algebras of MD-groups is significant. Note that all the Lie algebras and the Lie groups of dimension n with $n < 4$ are MD-algebras and MD-groups, and moreover they can be listed easily. So we only take interest in MDn-groups and MDn-algebras for $n \geq 4$.

We remark here that all $\overline{MD}$-algebras (of arbitrary dimension) was classified, up to isomorphism, by H. H. Viet in [9]. This class includes only the following algebras:

- $\mathbb{R}^n$ - The commutative Lie Algebra of dimension n;
- Lie(Aff$\mathbb{R}$) - The Lie algebra of the group of affine traformations of the real straight line;
- Lie(Aff$\mathbb{C}$) - The Lie algebra of the group of affine transformations of the complex straight line.

It is noteworthy that Viet [9] also described the C^*-algebras of the universal covering of group Aff$\mathbb{C}$ by using KK-functors. Thus, the C^*-algebras of all groups of the class $\overline{MD}$ were described by Diep, Rosenberg and Viet.

The problem for the class of MD-algebras is much more complicated than $\overline{MD}$-algebras. In 1984, Dao Van Tra [11] listed all $MD4$-algebras. In 1990, all $MD4$-algebras were classified, up to isomorphism, by Vu (see [12–14]). Until quite recently, Vu together with Nguyen Cong Tri, Duong

Minh Thanh and Duong Quang Hoa introduced some $MD5$ - algebras and $MD5$ - groups (see [15–20]). Until the present moment, there is no complete classification for MDn-algebras with $n \geq 5$.

On the other hand, by studying the foliated manifold, Connes ([1]) in 1982 proposed the notion of C^*-algebras associated with a measured foliation. The following question naturally arises: Can we describe the Connes C^*-algebras by using KK-functors? In fact, Torpe has shown in [10] that the KK-functors are very useful and effective to describe the structure of Connes C^*-algebras associated with the Reeb foliations.

The other reason for studying the class MD is based on the following fact: if G is a certain MD-group, then the family of its K-orbits with maximal dimension forms a measured foliation. This foliation is called MD-foliation associated with G. Furthermore, the C^*-algebra of G can be easily described when the Connes C^*-algebra of MD-foliation associated with G is known. Hence, the problem of classifying the topology and describing the Connes C^*-algebras of the class of MD-foliations is worth to study.

On this aspect, Vu in 1992 gave a topological classification of all $MD4$-foliations and described all Connes C^*-algebras of them by using the KK-functors (see [12–14]). We noticed that the Connes C^*-algebras of MDn-foliations with $n > 4$ has not yet been described. Following [9], if $\mathcal{G}$ is an MD-algebra then the second derived ideal $\mathcal{G}^2 = [\mathcal{G}^1, \mathcal{G}^1] = [[\mathcal{G}, \mathcal{G}], [\mathcal{G}, \mathcal{G}]]$ is commutative, however, the converse is not true. Therefore, we need to consider only $\mathcal{G}$ for which $\mathcal{G}^2$ is commutative. In particular, if $\mathcal{G}^2 = 0$ (i.e. $\mathcal{G}^1$ is commutative) then $\mathcal{G}$ could be an MD-algebra. Hence, we will restrict ourself only to this case. Our main result is to classify, up to an isomorphism, all $MD5$-algebras $\mathcal{G}$ having commutative derived ideal $\mathcal{G}^1 = [\mathcal{G}, \mathcal{G}]$. The topology of $MD5$-foliations associated with the $MD5$-groups and the description of Connes C^*-algebras of these foliations will be considered and studied later on.

2. Preliminaries

We first recall in this Section some preliminary results and notations which will be used in the sequel. For more detailed information, the reader is referred to [4] and [6].

2.1. *The co-adjoint Representation and K-orbits*

Let G be a Lie group. Let $\mathcal{G} = Lie(G)$ be the Lie algebra of G and we use $\mathcal{G}^*$ to denote the dual space of $\mathcal{G}$. For every $g \in G$, we denote the internal

automorphism associated with g by $A_{(g)}$, and whence, $A_{(g)} : G \longrightarrow G$ can be defined as follows

$$A_{(g)}(x) := g.x.g^{-1}, \forall x \in G.$$

The above automorphism induces the following mapping:

$$A_{(g)_*} : \mathcal{G} \longrightarrow \mathcal{G}$$

$$X \longmapsto A_{(g)_*}(X) := \frac{d}{dt}[g.exp(tX)g^{-1}] \mid_{t=0}$$

which is called *the tangent mapping* of $A_{(g)}$.

We now formulate the following definitions.

Definition 2.1. The action

$$Ad : G \longrightarrow Aut(\mathcal{G})$$

$$g \longmapsto Ad(g) := A_{(g)_*}$$

is called the adjoint representation of G in $\mathcal{G}$.

Definition 2.2. The action

$$K : G \longrightarrow Aut(\mathcal{G}^*)$$

$$g \longmapsto K_{(g)}$$

such that

$$\langle K_{(g)}F, X\rangle := \langle F, Ad(g^{-1})X\rangle; \quad (F \in \mathcal{G}^*, X \in \mathcal{G})$$

is called the co-adjoint representation or K-representation of G in $\mathcal{G}^*$.

Definition 2.3. Each orbit of the co-adjoint representation of G is called a K-orbit of G.

Thus, for every $F \in \mathcal{G}^*$, the K-orbit containing F defined above can be written by

$$\Omega_F := \{K_{(g)}F/g \in G\}.$$

The dimension of every K-orbit of an arbitrary Lie group G is always even. In order to define the dimension of the K-orbits Ω_F for each F from the dual space $\mathcal{G}^*$ of the Lie algebra $\mathcal{G} = Lie(G)$ of G, it is useful to consider the following skew-symmetric bilinear form B_F on $\mathcal{G}$

$$B_F(X, Y) := \langle F, [X, Y]\rangle; \forall X, Y \in \mathcal{G}.$$

Denote the stabilizer of F under the co-adjoint representation of G in $\mathcal{G}^*$ by G_F and $\mathcal{G}_F := \mathrm{Lie}(G_F)$.

We shall need in the sequel the following result.

Proposition 2.1. *(see [6], Section 15.1)* $KerB_F = \mathcal{G}_F$ *and* $dim\Omega_F = dim\mathcal{G} - dim\mathcal{G}_F$.

2.2. *MDn-Groups and MDn-Algebras*

Definition 2.4. (see [4] Chapter 4, definition 1.1) An MDn-group is an n-dimensional real solvable Lie group such that its K-orbits are orbits of dimension zero or maximal dimension. The Lie algebra of an MDn-group is called an MDn-algebra.

The following proposition gives a necessary condition for a Lie algebra belonging to the class of all MD-algebras.

Proposition 2.2. *(see [9], Theorem 4) Let $\mathcal{G}$ be an MD-algebra. Then its second derived ideal $\mathcal{G}^2 := [[\mathcal{G},\mathcal{G}],[\mathcal{G},\mathcal{G}]]$ is commutative.*

We point out here that the converse of the above result is in general not true. In other words, the above necessary condition is not a sufficient condition. We now only consider the 5-dimensional Lie algebras $\mathcal{G}$ having a second derived ideal $\mathcal{G}^2 = \{0\}$, i.e. the derived ideal $\mathcal{G}^1$ is commutative. Thus, the $\mathcal{G}$ could be an $MD5$-algebra.

3. The Main Result

From now on, we use $\mathcal{G}$ to denote an Lie algebra of dimension 5. We always choose a suitable basis $(X_1, X_2, X_3, X_4, X_5)$ in $\mathcal{G}$ so that $\mathcal{G}$ is isomorphic to $\mathbb{R}^5$ as a real vector space. The notation $\mathcal{G}^*$ will be used to denote the dual space of $\mathcal{G}$. Clearly, $\mathcal{G}^*$ can be identified with $\mathbb{R}^5$ by fixing in it the basis $(X_1^*, X_2^*, X_3^*, X_4^*, X_5^*)$ which is the dual of the basis $(X_1, X_2, X_3, X_4, X_5)$.

Theorem 3.1. *Let $\mathcal{G}$ be an MD5-algebra whose $\mathcal{G}^1 := [\mathcal{G},\mathcal{G}]$ is commutative. Then the following assertions hold.*

(i) If $\mathcal{G}$ is decomposable, then $\mathcal{G} \cong \mathcal{H} \oplus \mathbb{R}$, where $\mathcal{H}$ is an MD4-algebra.

(ii) If $\mathcal{G}$ is indecomposable, then we can choose a suitable basis $(X_1, X_2, X_3, X_4, X_5)$ of $\mathcal{G}$ such that $\mathcal{G}$ is isomorphic to one and only one of the following Lie algebra.

1. $g^1 = R.X_5 \cong R.$

i&,,l : [Xl,X2] = [X3,X4] = X_5*; the others Lie Brackets are trivial.*

2. $G^1 = \mathbf{R.X_4} \oplus R.X_5 \cong R^2$

2.1. &,2,1 : *[Xl,Xz] = X4, [*X_2*,*X_3*] =* X_5*; the others Lie brackets are trivial.*

2.2. a_5,2,2(A) : *[Xi,*X_2*] = [*X_3*,*X_4*] =* X_5*, [*X_2*,*X_3*] =* X_4*,* A G R\{0}; *the others Lie Brackets are trivial.*

3. $g^1 = R.X_3 \oplus R.X_4 \oplus R.X_5 \cong R^3, ad_{XI} = 0, ad_{XI}$ G *End{*g^l*)* ≅ *Mah(R);* [X_1,X_2*)=*X_3*.*

3-1. &5,3,l(Ai,A2) :

$$ad_{X2} = \begin{pmatrix} A\,l & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad Ai, A_2 \text{ G } R \setminus \{1\}, \; Ai \wedge A_2 \neq 0.$$

3.2. $G_{5,3,2(\lambda)}$:

$$ad_{X2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & A \end{pmatrix}; \quad A \text{ G } R \setminus \{0, 1\}.$$

3.3. G5,3,3(X) :

$$ad_{X2} = \begin{pmatrix} A & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad A \in R \setminus \{1\}.$$

3.4. G5,3,4 :

$$ad_{X_2} = 0 \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & \\ 0 & 0 & 1 \end{pmatrix}.$$

3.5. $\mathcal{G}_{5,3,5(\lambda)}$:

$$ad_{X_2} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}; \quad A \in R\setminus\{1\}.$$

3.6. $\mathcal{G}_{5,3,6(\lambda)}$:

$$ad_{X_2} = \begin{pmatrix} a & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & A \end{pmatrix}; \quad A \in R\setminus\{0,1\}.$$

3.7. $S_{5,3,?}$:

$$ad_{X_2} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

3.8. $\mathcal{G}_{5,3,8(\lambda,\varphi)}$:

$$adx_2 = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & \lambda \end{pmatrix}; \quad A \in R\setminus\{0\}, \varphi \in (0,\pi).$$

4. $G^1 = \mathbb{R}.X_3 \oplus \mathbb{R}.X_3$

$$adx, \in Bnd(G^1) \cong Mat_4(R).$$

4.1. $\mathcal{G}_{5,4,1(\lambda_1,\lambda_2,\lambda_3)}$:

$$ad_{X_4} = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ 0 & 0 & X_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$A_1, A_2, A_3 \in R\setminus\{0,1\}$, $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_1$.

4.2. $\mathcal{G}_{5,4,2(\lambda_1,\lambda_2)}$:

$$ad_{Xl} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad \lambda_1, \lambda_2 \in \mathbb{R} \setminus \{0,1\}, \lambda_1 \neq \lambda_2.$$

4.3. $\mathcal{G}_{5,4,3(\lambda)}$:

$$ad_{X_1} = \begin{pmatrix} /A & 0 & 0 & 0 \\ 0 & A & O & O \\ 0 & 0 & 1 & 0 \\ \backslash 0 & 0 & 0 & I) \end{pmatrix}; \quad A G E \backslash \{0,1\}.$$

4.4. $\mathcal{G}_{5,4,4(\lambda)}$:

$$ad_{Xl} = \begin{pmatrix} /A & 0 & 0 & ON \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \backslash 0 & 0 & 0 & 1 \end{pmatrix}; \quad A e R \backslash \{0,l\}.$$

4.5. 85,4,5 :

$$ad_{Xl} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

4.6. $\mathcal{G}_{5,4,6(\lambda_1,\lambda_2)}$:

$$ad_{X_1} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad Ai, A_2 G R \backslash \{0,l\}, Ai \wedge A_2.$$

4.7. $\mathcal{G}_{5,4,7(\lambda)}$:

$$ad_{x,} = \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad A E R \backslash \{O,l\}.$$

4-8. (75,4,8(A) :

$$ad_{X_1} = \begin{pmatrix} (\backslash 1\ 0\ Q\backslash \\ 0\ \mathrm{AOO} \\ 0\ 0\ 1\ 1 \\ \backslash O\ 0\ 0\ 1/ \end{pmatrix}; \quad \mathrm{AeE}\backslash\{0,1\}.$$

4.9. $\mathcal{G}_{5,4,9(\lambda)}$:

$$ad_{XI} = \begin{pmatrix} \lambda\ 0\ 0\ 0 \backslash \\ 0\ 1\ 1\ 0 \\ 0\ 0\ 1\ 1 \\ 0\ 0\ 0\ 1/ \end{pmatrix}; \quad \mathrm{AsR}\backslash\{0,1\}.$$

4-10- 05,4,10 :

$$ad_{XI} = \begin{pmatrix} (I\ 1\ 0\ O\backslash \\ 0110 \\ 0\ 0\ 1\ 1 \\ \backslash 0\ 0\ 0\ 1/ \end{pmatrix}.$$

4.11. $\mathcal{G}_{5,4,11(\lambda_1,\lambda_2,\varphi)}$:

$$ad_{XI} = \begin{pmatrix} coscp—sinp & 0 & 0 \\ sirup\ \ cosip & 0 & 0 \\ 0 \quad 0 & \mathrm{Ai} & 0 \\ 0 \quad 0 & 0 & \mathrm{A}_2 \end{pmatrix};$$

$$\lambda_{1;\,\mathrm{A}_2} \in \mathbb{R} \setminus \{0\}, \lambda_1 \neq \lambda_2, \varphi \in (0,7\mathrm{r}).$$

4.12. $\mathcal{G}_{5,4,12(\lambda,\varphi)}$:

$$ad_{X_1} = \begin{pmatrix} cos\varphi & —sincp & 0 & 0 \\ sin\varphi & \mathrm{costp} & 0 & 0 \\ 0 & 0 & \mathrm{A}\,0 \\ 0 & 0 & \mathrm{o\ x} \end{pmatrix}; \quad \lambda \in \mathbb{R} \setminus \{0\}, \varphi \in (0, \pi).$$

4-13. $\mathcal{G}_{5,4,13(\lambda,\varphi)}$:

$$ad_{XI} = \begin{pmatrix} cos<p & —sirup & 0 & 0 \\ simp & cos<p & 0 & 0 \\ 0 & 0 & A & 1 \\ . 0 & 0 & 0 & A/ \end{pmatrix}; \quad A \in R N \{ 0 > > v \in (0 | i r) \cdot$$

4.14. $\mathcal{G}_{5,4,14(\lambda,\mu,\varphi)}$:

$$ad_{X_1} = \begin{pmatrix} cosip & —sirup & 0 & 0 \\ sin\varphi & costp & 0 & 0 \\ & & & \mu \\ & & & \end{pmatrix};$$

$$\begin{matrix} 0 & 0 & A - \\ \lambda, \mu \in \mathbb{R} \ 0 & 0 & fi \ A \end{matrix}$$

$$,/j>0,((J\in (0,TT).$$

In proving Theorem 3.1, we need some lemmas.

Lemma 3.1. *For X, Y E Q\B[1], X #Y, by considering adx, ady as operators on Q^1 we have adx ° [a]dy = [a]dy ° adx.*

Proof. By using the Jacobi identity for X, Y and consider an arbitrary element *Z* E G^1, we have

$$\begin{aligned} & [[X,YI,Z] + \text{“}Y,ZI,Xl + \text{“}\sim,XI,Yl=0 \\ * \quad & [X,[Y,Zll-[Y,[X,211=0 \\ <=> \quad & adx°ad_y(2)=ad_yoad_x(2),VZ \ E \ G^1 \\ <\$\bullet \ & adx \circ {}^a dy = ady \ o \ adx \ . \end{aligned}$$ 0

Lemma 3.2. *(see [2], Chapter 2, Proposition 2.1) Let G be an MD-algebra with F* E *Q* is not vanishing perfectly in Q^1, i.e. there exists U E Q^1 such that (F, U) #* 0. *Then the K-orbit RF as one of the K-orbits having maximal dimension.*

Proof. Assume that RF is not a K-orbit with maximal dimension, that is, $dim\Omega_F = 0$. Then we have

$$dim\mathcal{G}_F = dim6 — dimRF = dimQ.$$

Consequently, $KerB_F = \mathcal{G}_F = \mathcal{G} \supset \mathcal{G}^1$ and F is perfectly vanishing in $\mathcal{G}^1$. This contradicts the hypotheses of the Lemma. Therefore, Ω_F must be a K-orbit with maximal dimension. □

Lemma 3.3. *Let F be an arbitrary element of $\mathcal{G}^*$. Then $dim\Omega_F = rank(B)$, where $B = (b_{ij})_5 := (\langle F, [X_j, X_i]\rangle), 1 \leq i, j \leq 5$, is the matrix of the skew-symmetric bilinear form B_F in the basis $(X_1, X_2, X_3, X_4, X_5)$ of $\mathcal{G}$.*

Proof. Let $U = aX_1 + bX_2 + cX_3 + dX_4 + eX_5 \in \mathcal{G}$. Then we have

$$\begin{aligned} \mathcal{G}_F &= KerB_F \\ &= \{U \in \mathcal{G} / \langle F, [U, X_i]\rangle = 0;\ i = 1, 2, , 3, 4, 5\}. \end{aligned}$$

By simple computation, we obtain

$$U \in \mathcal{G}_F \Leftrightarrow B \begin{pmatrix} a \\ b \\ c \\ d \\ f \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hence, $dim\Omega_F = dim\mathcal{G} - dim\mathcal{G}_F = rank(B)$. □

Lemma 3.4. *If $\mathcal{G}$ is a real solvable Lie algebra of dimension 5 with the first derived ideal $\mathcal{G}^1 \cong \mathbb{R}^4$ then $\mathcal{G}$ is a MD5-algebra.*

Proof. Let $\mathcal{G}$ be a real solvable Lie algebra with dimension 5 such that $\mathcal{G}^1$ is the commutative Lie algebra with dimension 4. Without loss of generality, we may assume that $\mathcal{G}^1 = \mathbb{R}.X_2 \oplus \mathbb{R}.X_3 \oplus \mathbb{R}.X_4 \oplus \mathbb{R}.X_5 \equiv \mathbb{R}^4$, $ad_{X_1} = (a_{ij})_4 \in End(\mathcal{G}^1) \equiv Mat_4(\mathbb{R})$; $a_{ij} \in \mathbb{R}, 1 \leq i, j \leq 4$.

Let $F = \alpha {X_1}^* + \beta {X_2}^* + \gamma {X_3}^* + \delta {X_4}^* + \sigma {X_5}^* \equiv (\alpha, \beta, \gamma, \delta, \sigma)$ be an arbitrary element from $\mathcal{G}^* \equiv \mathbb{R}^5; \alpha, \beta, \gamma, \delta, \sigma \in \mathbb{R}$. Then, by simple computation, we can see that the matrix B of the bilinear form B_F in the basis $(X_1, X_2, X_3, X_4, X_5)$ of $\mathcal{G}$ is a matrix of the following

$$\begin{pmatrix} 0 & -\sum_{i=2}^5 a_{i2}\alpha_i & -\sum_{i=2}^5 a_{i3}\alpha_i & -\sum_{i=2}^5 a_{i4}\alpha_i & -\sum_{i=2}^5 a_{i5}\alpha_i \\ \sum_{i=2}^5 a_{i2}\alpha_i & 0 & 0 & 0 & 0 \\ \sum_{i=2}^5 a_{i3}\alpha_i & 0 & 0 & 0 & 0 \\ \sum_{i=2}^5 a_{i4}\alpha_i & 0 & 0 & 0 & 0 \\ \sum_{i=2}^5 a_{i2}\alpha_i & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is now clear that $rank(B) \in \{0, 2\}$. Hence, according to Lemma 3.3, Ω_F is the orbit with dimension 0 or 2, i.e. $\mathcal{G}$ is an $MD5$-algebra. □

We now prove the main theorem.

Proof. It is clear that assertion (i) of Theorem 3.1 holds obviously. We only need to prove assertion (ii). Assume that $\mathcal{G}$ is an indecomposable MD5-algebra with basis $(X_1, X_2, X_3, X_4, X_5)$ and its first derived ideal $\mathcal{G}^1$ is commutative. Then $dim\,\mathcal{G}^1 \in \{1, 2, 3, 4\}$. In [16], Theorem 2.1 and [19], Theorem 3.2, the cases had been considered when $dim\,\mathcal{G}^1 \in \{3, 4\}$. Therefore, we only need to consider the remaining cases when $dim\,\mathcal{G}^1 \in \{1, 2\}$. However, for the sake of completeness, we now consider here all cases.

1. $dim\mathcal{G}^1 = 1$. Without loss of generality , we may assume that $\mathcal{G}^1 = \mathbb{R}.X_5 \equiv \mathbb{R}$.

1.1. Assume that there exists $i \in \{1, 2, 3, 4\}$ with $[X_i, X_5] \neq 0$. Renumber the given basis, if necessary, and we suppose that $[X_4, X_5] = aX_5$, for some $a \in \mathbb{R}\backslash\{0\}$. Then, by changing X_4 with $X_4^{'} = \frac{1}{a}X_4$, we obtain $[X_4^{'}, X_5] = X_5$. Now, without any restriction of generality, we can assume that $[X_4, X_5] = X_5$.
Let $[X_i, X_5] = a_iX_5, [X_i, X_4] = b_iX_5; a_i, b_i \in \mathbb{R}; i = 1, 2, 3$. Then, by changing $X_i^{'} = X_i - a_iX_4 + b_iX_5 (i = 1, 2, 3)$, we get $[X_i^{'}, X_5] = [X_i^{'}, X_4] = 0; i = 1, 2, 3..$ Hence, we can always suppose right from the start that $[X_i, X_5] = [X_i, X_4] = 0; i = 1, 2, 3$.
Now, let $[X_i, X_j] = c_{ij}X_5$, $c_{ij} \in \mathbb{R}; 1 \leq i < j \leq 3$. Then, by using the Jacobi identity, we get $c_{ij} = 0$ for all i, j, $1 \leq i < j \leq 3$. But this shows that $\mathcal{G}$ is decomposable, which is a contradiction. Thus, this case cannot happen.

1.2. Assume that $[X_i, X_5] = 0$ for all i = 1, 2, 3, 4. Then, there exists $[X_i, X_j] = c_{ij}X_5$, $c_{ij} \neq 0$ for some $i, j \in \{1, 2, 3, 4\}, i \neq j$. By applying the same argument as in Case 1.1, we can suppose that $[X_1, X_2] = [X_3, X_4] = X_5$ and $[X_i, X_3] = [X_i, X_4] = 0; i = 1, 2$. Therefore, $\mathcal{G} \cong \mathcal{G}_{5,1}$.

2. $dim\mathcal{G}^1 = 2$. Without loss of generality, we now assume that $\mathcal{G}^1 = \mathbb{R}.X_4 \oplus \mathbb{R}.X_5 \equiv \mathbb{R}^2$; $ad_{X_1}, ad_{X_2}, ad_{X_3} \in End(\mathcal{G}^1) \equiv Mat_2(\mathbb{R})$.

2.1. $[X_i, X_j] = 0, 1 \leq i < j \leq 3$.
If there exists $ad_{X_i} = 0$ then $\mathcal{G}$ is decomposable, which is a contradiction. Hence, $ad_{X_i} \neq 0, i = 1, 2, 3$. We now show that we can always obtain $ad_{X_2} = 0$ by changing the basis. Indeed, we can let ad_{X_i} be $\begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \neq$

$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$; $i = 1, 2, 3$. We first assume that $a_3 \neq 0$. Then, by writing $X_i{}' = X_i - \frac{a_i}{a_3} X_3$, we get $ad_{X_i'} = \begin{pmatrix} 0 & b_i{}' \\ c_i{}' & d_i{}' \end{pmatrix}$, $i = 1, 2$. Hence, we can suppose that $ad_{X_i} = \begin{pmatrix} 0 & b_i \\ c_i & d_i \end{pmatrix}$, $i = 1, 2$. According to Lemma 3.2, $ad_{X_1} \circ ad_{X_2} = ad_{X_2} \circ ad_{X_1}$. It follows that $ad_{X_1} = k.ad_{X_2}$, for some $k \in \mathbb{R}\backslash\{0\}$. By changing $X_2{}' = X_2 - k.X_1$, we get $ad_{X_2'} = 0$, a contradiction. When $d_3 \neq 0$, then by using the same argument, we can also obtain a contradiction. Finally, assume that $a_3 = d_3 = 0$, $b_3^2 + c_3^2 \neq 0$. In view of Lemma 3.2, we get $ad_{X_i} \circ ad_{X_3} = ad_{X_3} \circ ad_{X_i} (i = 1, 2)$

Hence, it follows that $ad_{X_i} = k_i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $0 \neq k_i \in \mathbb{R}$, $i = 1, 2$. In particular, $ad_{X_2} = k.ad_{X_1}, k = \frac{k_2}{k_1}$. Now, by changing $X_2{}' = X_2 - k.X_1$, we get $ad_{X_2'} = 0$, again a contradiction. Hence, Case 2.1 can not happen.

2.2. Assume that there exists $[X_i, X_j] \neq 0, 1 \leq i < j \leq 3$ and $ad_{X_i} = 0$ $i = 1, 2, 3$.

It is clear that $\mathcal{G}^1 = \langle [X_1, X_2], [X_1, X_3], [X_2, X_3] \rangle$, and whence, the rank of $\{[X_1, X_2], [X_1, X_3], [X_2, X_3]\}$ is 2 and without restriction of generality, we can assume that $\{[X_1, X_2], [X_2, X_3]\}$ is a basis of $\mathcal{G}^1$. Let $[X_1, X_2] = aX_4 + bX_5$, $[X_2, X_3] = cX_4 + dX_5$ with $D := det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$. By changing basis as follows

$$X_4 = \frac{1}{D}(dX_4{}' - bX_5{}'),\ X_5 = \frac{1}{D}(-cX_4{}' + aX_5{}')$$

we get $[X_1, X_2] = X_4{}'$, $[X_2, X_3] = X_5{}'$. Hence, we can assume that $[X_1, X_2] = X_4$, $[X_2, X_3] = X_5$.

Let $[X_1, X_3] = \alpha X_4 + \beta X_5$. Then, by changing the basis as follows:

$$X_1{}' = X_1 - \beta X_2, X_2{}' = X_2, X_3{}' = -\alpha X_2 + X_3$$

we get

$$[X_1{}', X_2{}'] = X_4,\ [X_2{}', X_3{}'] = X_5,\ [X_1{}', X_3{}'] = 0.$$

Thus, we can always assume that

$$[X_1, X_2] = X_4,\ [X_2, X_3] = X_5,\ [X_1, X_3] = 0.$$

Therefore $\mathcal{G} \cong \mathcal{G}_{5,2,1}$.

2.3. Assume that there exists $[X_i, X_j] \neq 0$ and $ad_{X_k} \neq 0$, $1 \leq i \neq j \leq 3$, $1 \leq k \leq 3$. Then, without loss of generality, we may assume that $ad_{X_3} \neq 0$. We can always change basis of $\mathcal{G}^1$ such that ad_{X_3} becomes one of the following matrices

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} cos\varphi & -sin\varphi \\ sin\varphi & cos\varphi \end{pmatrix}; \lambda \in \mathbb{R}, \varphi \in (0, \pi).$$

2.3a. Assume that $ad_{X_3} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then by using an argument analogous to that in Subsection 2.2, we get $ad_{X_1} = ad_{X_2} = 0$. Again, by Jacobi identity, we obtain $[X_1, X_2] = aX_5$, $a \in \mathbb{R}$.
Let $[X_i, X_3] = a_iX_4 + b_iX_5$; $a_i, b_i \in \mathbb{R}$, $i = 1, 2$. If $a = 0$, then by changing $X_i' = X_i + b_iX_4$, we get $[X_i', X_3] = a_iX_4$, $i = 1, 2$. Hence, we can always assume from the outset that $[X_i, X_3] = a_iX_4$; $i = 1, 2$; ${a_1}^2 + {a_2}^2 \neq 0$. Without loss of generality, we may assume that $a_2 \neq 0$. Now, we change again the basis as follows

$$X_1' = X_1 - \frac{a_1}{a_2}X_2, \; X_2' = \frac{1}{a_2}X_2.$$

Then we get $[X_1', X_3] = 0, [X_2', X_3] = X_4$, i.e. $\mathcal{G}$ is decomposable, a contradiction. Hence, $a \neq 0$.

In the same way, we obtain

$$[X_1, X_2] = [X_3, X_4] = X_5, \; [X_2, X_3] = \lambda X_4, \; 0 \neq \lambda \in \mathbb{R}.$$

Therefore $\mathcal{G} \cong \mathcal{G}_{5,2,2(\lambda)}$.

2.3b. Assume that $ad_{X_3} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$, $\lambda \in \mathbb{R}$. Then, by using a similar argument as above, we get $\mathcal{G}_{5,2,3}$: $[X_1, X_2] = X_5, [X_3, X_4] = X_4$. By using Lemmas 3.1 and Lemma 3.2, and by direct computation, we can show that $\mathcal{G}_{5,2,3}$ is not an MD5-algebra. Hence, this case has to be rejected.

2.3c. Assume that $ad_{X_3} \in \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} cos\varphi & -sin\varphi \\ sin\varphi & cos\varphi \end{pmatrix}; \varphi \in (0, \pi) \right\}$. By using a similar argument as above, these cases have to be also rejected.

3. $dim\mathcal{G}^1 = 3$. We can always change basis to obtain $\mathcal{G}^1 = \mathbb{R}.X_3 \oplus \mathbb{R}.X_4 \oplus \mathbb{R}.X_5 \equiv \mathbb{R}^3$; $ad_{X_1}, ad_{X_2} \in End(\mathcal{G}^1) \equiv Mat_3(\mathbb{R})$.

It is obvious that ad_{X_1} and ad_{X_2} cannot be the trivial operators concurrently because $\mathcal{G}^1 \cong \mathbb{R}^3$. Without loss of generality, we may assume that $ad_{X_2} \neq 0$. Then, by changing basis, if necessary, we obtain a similar classification of ad_{X_2} as follows

- $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $(\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{1\}, \lambda_1 \neq \lambda_2 \neq 0)$;
- $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}$, $(\lambda \in \mathbb{R} \setminus \{0, 1\})$;
- $\begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $(\lambda \in \mathbb{R} \setminus \{1\})$;
- $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$;
- $\begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, $(\lambda \in \mathbb{R} \setminus \{1\})$;
- $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}$, $(\lambda \in \mathbb{R} \setminus \{0, 1\})$;
- $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$;
- $\begin{pmatrix} cos\varphi & -sin\varphi & 0 \\ sin\varphi & cos\varphi & 0 \\ 0 & 0 & \lambda \end{pmatrix}$, $(\lambda \in \mathbb{R} \setminus \{0\}, \varphi \in (0, \pi))$.

Assume that $[X_1, X_2] = mX_3 + nX_4 + pX_5; m, n, p \in \mathbb{R}$. We can always change basis to have $[X_1, X_2] = mX_3$. Indeed, if

$$ad_{X_2} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, (\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{1\}, \lambda_1 \neq \lambda_2 \neq 0),$$

then by changing X_1 for $X_1{}' = X_1 + \frac{n}{\lambda_2} X_4 + pX_5$ we get $[X_1{}', X_2] = mX_3$, $m \in \mathbb{R}$. For the other values of ad_{X_2}, we can also change basis in the same way. Hence, without restriction of generality, we can assume that

$[X_1, X_2] = mX_3$, $m \in \mathbb{R}$.

There are three cases which contradict each other as follows.

3.1. $[X_1, X_2] = 0$ (i.e. $m = 0$) and $ad_{X_1} = 0$. Then $\mathcal{G} = \mathcal{H} \oplus \mathbb{R}.X_1$, where $\mathcal{H}$ is the subalgebra of $\mathcal{G}$ generated by $\{X_2, X_3, X_4, X_5\}$, i.e. $\mathcal{G}$ is decomposable. Hence, this case is rejected.

3.2. $[X_1, X_2] = 0$ and $ad_{X_1} \neq 0$.

3.2a. Assume that $ad_{X_2} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; $\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{1\}$, $\lambda_1 \neq \lambda_2 \neq 0$. In view of Lemma 3.1, it follows by a direct computation that

$$ad_{X_1} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \nu & 0 \\ 0 & 0 & \xi \end{pmatrix}; \mu, \nu, \xi \in \mathbb{R}; \mu^2 + \nu^2 + \xi^2 \neq 0.$$

If $\xi \neq 0$, by changing $X_1{}' = X_1 - \xi X_2$, we get

$$ad_{X_1{}'} = \begin{pmatrix} \mu' & 0 & 0 \\ 0 & \nu' & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

where $\mu' = \mu - \xi\lambda_1, \nu' = \nu - \xi\lambda_2$. Thus, we can assume that

$$ad_{X_1} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \nu & 0 \\ 0 & 0 & 0 \end{pmatrix}; \mu, \nu \in \mathbb{R}; \mu^2 + \nu^2 \neq 0.$$

Using Lemmas 3.2, 3.3, and by direct computation, we can show that $\mathcal{G}$ will not be an $MD5$-algebra in Case 3.2a . So this case must be rejected.

3.2b. In exactly the same way, but replacing the considered value of ad_{X_2} with the others, we can easily see that Case 3.2 cannot occur.

3.3. $[X_1, X_2] \neq 0$ (i.e. $m \neq 0$). By changing X_1 by $X_1{}' = \frac{1}{m}X_1$, we have $[X_1{}', X_2] = X_3$. Hence, without loss of generality, we may assume that $[X_1, X_2] = X_3$. By using a similar argument as the one in Case 3.2a, we obtain again a contradiction if $ad_{X_1} \neq 0$. In other words, $ad_{X_1} = 0$. Therefore, in the dependence on the value of ad_{X_2}, $\mathcal{G}$ must be isomorphic to one of the following algebras:

- $\mathcal{G}_{5,3,1(\lambda_1,\lambda_2)}$, $(\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{1\}, \lambda_1 \neq \lambda_2 \neq 0)$;
- $\mathcal{G}_{5,3,2(\lambda)}$, $(\lambda \in \mathbb{R} \setminus \{0, 1\})$;

- $\mathcal{G}_{5,3,3(\lambda)}$, $(\lambda \in \mathbb{R} \setminus \{1\})$;
- $\mathcal{G}_{5,3,4}$;
- $\mathcal{G}_{5,3,5(\lambda)}$, $(\lambda \in \mathbb{R} \setminus \{1\})$;
- $\mathcal{G}_{5,3,6(\lambda)}$, $(\lambda \in \mathbb{R} \setminus \{0,1\})$;
- $\mathcal{G}_{5,3,7}$;
- $\mathcal{G}_{5,3,8(\lambda,\varphi)}$, $(\lambda \in \mathbb{R} \setminus \{0\}), \varphi \in (0,\pi))$.

Obviously, these algebras are not mutually isomorphic to each other.

4. $dim\mathcal{G}^1 = 4$. Without loss of generality, we may assume that $\mathcal{G}^1 = \mathbb{R}.X_2 \oplus \mathbb{R}.X_3 \oplus \mathbb{R}.X_4 \oplus \mathbb{R}.X_5 \equiv \mathbb{R}^4$, $ad_{X_1} \in End(\mathcal{G}^1) \equiv Mat_4(\mathbb{R})$.

According to Lemma 3.4, the final assertions of Theorem 3.1 can be obtained by using similar classification of ad_{X_1}.

In view of Lemma 3.3, it follows by direct computation that all algebras listed in Theorem 3.1 are $MD5$-algebras. This completes the proof. □

3.1. *Concluding Remark*

Recall that every real Lie algebra $\mathcal{G}$ defines only one connected and simply connected Lie group G such that $Lie(G) = \mathcal{G}$. Therefore, we obtain a collection of twenty - five families of connected and simply connected $MD5$-groups corresponding to given indecomposable $MD5$-algebras in Theorem 3.1. For the sake of convenience, we denote every $MD5$-group from this collection by using the same indices as its corresponding $MD5$-algebra. For example, $G_{5,3,1(\lambda_1,\lambda_2)}$ is the connected and simply connected MD5-group which corresponds to $\mathcal{G}_{5,3,1(\lambda_1,\lambda_2)}$. All of these groups are indecomposable $MD5$-groups. In the next papers, we shall compute the invariants of given $MD5$-algebras, describe the geometry of K-orbits of its corresponding $MD5$-groups and also we shall classify topologically the $MD5$-foliations associated with these $MD5$-groups. In addition, characterization theorems of Connes C^*-algebras corresponding to these $MD5$-foliations will also be established.

Acknowledgement

The first author would like to thank Professor Do Ngoc Diep for giving him excellent advice and support. He also want to thank the Organizing Committee of The Second International Congress In Algebras and Combinatorics - July 2007, Beijing, China for inviting him to give a talk on this topic at the congress.

References

1. A. Connes, *A Survey of Foliations and Operator Algebras*, Proc. Symp. Pure Math., 38 (1982), 521 - 628, Part I.
2. Do Ngoc Diep, *The Structure of the Group C^*-algebras of the Group of Affine Transformations of the Straight line*, Funkt. Anal. i Priloz. 9 (1975), No. 1, 63 - 64.
3. Do Ngoc Diep, *The Structure of C^*-algebras of Type I*, Vestnik Moskow Uni. (1978), No. 2, 81 - 87.
4. Do Ngoc Diep, *Method of Noncommutative Geometry for Group C^*-algebras*, Chapman and Hall / CRC Press Reseach Notes in Mathematics Series, 416, (1999).
5. G. G. Kasparov, *The Operator K-functors and Extensions of C^*-algebras*, Math. USSR Izvestija, 16 (1981), No. 3, 513 - 572.
6. A. A. Kirillov, *Elements of the Theory of Representations*, (Springer - Verlag, Berlin - Heidenberg - New York, 1976).
7. J. Rosenberg, *The C^*-algebras of Some Real and p-adic Solvable Groups*, Pacific J. Math., 65 (1976), 175 - 192.
8. J. Rosenberg, *Homological Invariants of Extensions of C^*-algebras*, Proc. Symp. Pure Math., Vol. 38, 35 - 75, A.M.S., R. I.
9. Vuong Manh Son et Ho Huu Viet, *Sur La Structure Des C^* – algèbres D'une Classe De Groupes De Lie*, J. Operator Theory, 11 (1984), 77-90.
10. A. M. Torpe, *K-theory for the Leaf Space of Foliations by Reeb Components*, J. Funct. Anal., 61 (1985), 15 - 71.
11. D. V. Tra, *On the Lie Algebras of low dimension*, Sci. Papes of the 12th College of Institute of Math. Vietnam, Hanoi (1984) (in Vietnamese).
12. Le Anh Vu, *On the Structure of the C^*-algebra of the Foliation Formed by the K-orbits of Maximal Dimension of the Real Diamond Group*, J. Operator Theory, 24 (1990), 227 - 238.
13. Le Anh Vu, *On the Foliations Formed by the Generic K-orbits of the MD4-Groups*, Acta Math. Vietnam, No., 2 (1990), 39 - 55.
14. Le Anh Vu, *Foliations Formed by Orbits of Maximal Dimension in the Co-adjoint Representation of a Class of Solvable Lie Groups*, Vest. Moscow Uni., Math. Bulletin, Vol. 48 (1993), No. 3, 24 - 27.
15. Le Anh Vu, *Foliations Formed by Orbits of Maximal Dimension of Some MD5-Groups*, East-West J. of Mathematics, Vol. 5, No. 2 (2003), 159 - 168.
16. Le Anh Vu, *On a Subclass of 5-dimensional Lie Algebras Which Have 3-dimensional commutative derived ideals*, East-West J. of Mathematics, Vol. 7, No. 1 (2005), 13 - 22.
17. Le Anh Vu and Nguyen Cong Tri, *Some Examples on MD5-algebras and MD5-mesured Foliations Associated to Corresponding MD5-groups*, Scientific Journal of University of Pedagogy of Ho Chi Minh City, No. 8 (42) (2006), 14 - 32 (In Vietnamese).
18. Le Anh Vu and Duong Minh Thanh, *The Geometry of K-orbits of a Subclass of MD5-groups and Foliations Formed by Their Generic K-orbits*, Contributions in Mathematics and Applications - ICMA, December 2005, Mahidol University, Bangkok, Thailand (A special volume published by East-West

Journal of Mathematics), ISSN 1513-489X, 169-184.
19. Le Anh Vu,*Classification of 5-dimensional MD-algebras Which have 4-dimensional Commutative Derived Ideals*, Scientific Journal of University of Pedagogy of Ho Chi Minh City, No. 12 (46) (2007), 3 - 15 (In Vietnamese).
20. Le Anh Vu and Duong Quang Hoa, *The Geometricaly Picture of K-orbits of Connected and Simply Connected MD5-groups such that thier MD5-algebras have 4-dimensional Commutative Derived Ideals*, Scientific Journal of University of Pedagogy of Ho Chi Minh City, No. 12 (46) (2007), 16 - 28 (In Vietnamese).
21. Vyacheslav Boyko, Jiri Patera and Roman Popovych, *Computation of Invariants of Lie Algebras by Means of Moving Frames*, ARXiv: math-ph/0602046 v2, 11 Apr 2006.

This page intentionally left blank

Advances in Algebra and Combinatorics

This volume is a compilation of lectures on algebras and combinatorics presented at the Second International Congress in Algebra and Combinatorics. It reports on not only new results, but also on open problems in the field. The proceedings volume is useful for graduate students and researchers in algebras and combinatorics. Contributors include eminent figures such as V Artamonov, L Bokut, J Fountain, P Hilton, M Jambu, P Kolesnikov, Li Wei and K Ueno.

World Scientific
www.worldscientific.com
6694 hc

ISBN-13 978-981-279-000-2
ISBN-10 981-279-000-4

