

ISSN: 1337-6365

© Slovak University of Technology in Bratislava

All rights reserved

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

VOLUME 5 (2012), NUMBER 3

APLIMAT – JOURNAL OF APPLIED MATHEMATICS

VOLUME 5 (2012), NUMBER 3

Edited by: Slovak University of Technology in Bratislava

Editor - in - Chief: KOVÁČOVÁ Monika (Slovak Republic)

Editorial Board: CARKOVŠ Jevgenijs (Latvia)
CZANNER Gabriela (Great Britain)
CZANNER Silvester (Great Britain)
DOLEŽALOVÁ Jarmila (Czech Republic)
FEČKAN Michal (Slovak Republic)
FERREIRA M. A. Martins (Portugal)
FRANCAVIGLIA Mauro (Italy)
KARPÍŠEK Zdeněk (Czech Republic)
KOROTOV Sergey (Finland)
LORENZI Marcella Giulia (Italy)
MESIAR Radko (Slovak Republic)
VELICHOVÁ Daniela (Slovak Republic)

Editorial Office: Institute of natural sciences, humanities and social sciences
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava
Námestie slobody 17
812 31 Bratislava

Correspondence concerning subscriptions, claims and distribution:

F.X. spol s.r.o
Dúbravská cesta 9
845 03 Bratislava 45
journal@aplimat.com

Frequency: One volume per year consisting of three issues at price of 120 EUR, per volume, including surface mail shipment abroad.
Registration number EV 2540/08

Information and instructions for authors are available on the address:

<http://www.journal.aplimat.com/>

Printed by: FX spol s.r.o, Azalková 21, 821 00 Bratislava

Copyright © STU 2007-2012, Bratislava

All rights reserved. No part may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission from the Editorial Board. All contributions published in the Journal were reviewed with open and blind review forms with respect to their scientific contents.

APLIMAT – JOURNAL OF APPLIED MATHEMATICS

VOLUME 5 (2012), NUMBER 3

ALGEBRA AND ITS APPLICATIONS

JASEM Milan: A CAUCHY COMPLETION OF DUALY RESIDUATED LATTICE ORDERED SEMIGROUPS	15
SEIBERT Jaroslav: ON THE HYPER-WIENER INDEX AND POLYNOMIAL OF A GRAPH.	25
VANŽUROVÁ Alena : HEXAGONAL AND GOLDEN QUASIGROUPS	31
VANŽUROVÁ Alena, BARTOŠKOVÁ Zuzana: TOYODA'S THEOREM AND QUASIGROUP MODES	41

APLIMAT – JOURNAL OF APPLIED MATHEMATICS

VOLUME 5 (2012), NUMBER 3

GEOMETRY AND ITS APPLICATIONS

ARSAN Güler Gürpınar, ÇİVİ Gülçin: GEODESIC MAPPINGS PRESERVING THE EINSTEIN TENSOR OF WEYL SPACES	51
BASTL Bohumír, LÁVIČKA Miroslav: SMOOTH CURVES APPROXIMATION BY CHORD-LENGTH CURVES	57
BIZZARRI Michal, LÁVIČKA Miroslav: ON COMPUTING APPROXIMATE PARAMETERIZATIONS OF ALGEBRAIC SURFACES	67
ÇIVI Gülçin , ARSAN Güler Gürpınar: ON HOLOMORPHICALLY PROJECTIVE MAPPINGS OF KÄHLER- WEYL SPACES	79
CHEPURNA Olena, MIKEŠ Josef: HOLOMORPHICALLY PROJECTIVE MAPPINGS OF HYPERBOLICALLY KÄHLER SPACES PRESERVING THE EINSTEIN TENSOR	85
CHODOROVÁ Marie, CHUDÁ Hana, SHIHA Mohsen: ON COMPOSITION OF CONFORMAL AND HOLOMORPHICALLY PROJECTIVE MAPPINGS BETWEEN CONFORMALLY KÄHLERIAN SPACES	91
JUKL Marek, JUKLOVÁ Lenka, MIKEŠ Josef: MULTIPLE COVARIANT DERIVATIVE AND DECOMPOSITION PROBLEMS	97
KAŇKA Miloš, ELIÁŠOVÁ Lada: THE CURVATURES OF SPECIAL FUNCTIONS IN ECONOMY	105
PETRASOVA Alena, CZANNER Silvester, CHALMERS Alan, WOLKE Dieter: THE UTILITY OF THE VIRTUAL REALITY IN DEEPER UNDERSTANDING OF PEOPLE'S EXPERIENCES OF INFANT FEEDING	113
SLABÁ Kristýna, BASTL Bohumír: CIRCLE-PRESERVING SUBDIVISION SCHEME BASED ON APOLLONIUS' CIRCLE	123
VANŽUROVÁ Alena, JUKL Marek: PARALLELOGRAM SPACES AND MEDIAL QUASIGROUPS	133
VANŽUROVÁ Alena, PIRKLOVÁ Petra: METRIZABLE CONNECTIONS AND RESTRICTIVELY VARIATIONAL CONNECTIONS IN AFFINE MANIFOLDS	141

VELICHOVÁ Daniela: MULTIDIMENSIONAL MANIFOLDS AS
MINKOWSKI OPERATION PRODUCTS **151**

VOICU Nicoleta: FINSLERIAN CONNECTIONS AND THE EQUATIONS
OF SPINNING CHARGED PARTICLES IN GENERAL RELATIVITY **159**

APLIMAT – JOURNAL OF APPLIED MATHEMATICS

VOLUME 5 (2012), NUMBER 3

STATISTICAL METHODS IN TECHNICAL AND ECONOMIC SCIENCES AND PRACTICE

ANDRADE Marina, FERREIRA Manuel Alberto M.: CRIME SCENE INVESTIGATION THROUGH DNA TRACES USING BAYESIAN NETWORKS	167
ANDRADE, Marina, FERREIRA, Manuel Alberto M.: CIVIL AND CRIMINAL IDENTIFICATION WITH BAYESIAN NETWORKS	173
ARSHINOVA Tatyana: RISK MANAGEMENT OF EQUITY PORTFOLIO CONSTRUCTION ON THE BASIS OF DATA ENVELOPMENT ANALYSIS APPROACH	185
BARTOŠOVÁ Jitka, FORBELSKÁ Marie: GMM MODEL OF AT-RISK-OF-POVERTY CZECH HOUSEHOLDS DEPENDING ON THE AGE AND SEX OF THE HOUSEHOLDER (EU-SILC 2005-2009)	195
BEZRUCKO Aleksandrs: LATVIAN GDP: TIME SERIES FORECASTING USING VECTOR AUTO REGRESSION	205
FERREIRA Manuel Alberto M., ANDRADE Marina: A METHOD TO APPROXIMATE FIRST PASSAGE TIMES DISTRIBUTIONS IN DIRECT TIME MARKOV PROCESSES	217
FERREIRA, Manuel Alberto M., ANDRADE, Marina: SOJOURN TIMES IN JACKSON NETWORKS	225
FJODOROVŠ Jegors, MATVEJEVS Andrejs: COPULA BASED SEMIPARAMETRIC REGRESSIVE MODELS	241
HECKENBERGEROVÁ Jana, MAREK Jaroslav, SOUČKOVÁ Jitka, TUČEK Pavel: NONSMOOTH FUNCTION APPROXIMATION IN PRACTICAL CHANGE POINT PROBLEM	249
JAROŠOVÁ Eva: COMPARISON OF TWO BAYESIAN APPROACHES TO SPC	259
MALÁ Ivana: QUANTILE CHARACTERISTICS OF CONDITIONAL DISTRIBUTIONS OF FINITE MIXTURES	269

MISKOLCZI Martina, LANGHAMROVÁ Jitka: MULTISTATE LIFE TABLES: APPLICATION OF THE METHOD ON THE MARRIAGE CAREER	279
MOŠNA František: TWO APPLICATIONS OF PROBABILITY IN THE THEORY OF RELIABILITY AND MAINTENANCE	287
NEUBAUER Jiří: INFLATION MODELING AND COINTEGRATION	293
ŽIŽKA David: APPLICATION OF RELEVANCE VECTOR MACHINE TO FORECASTING VOLATILITY IN CZECH FINANCIAL TIME SERIES	301

LIST OF REVIEWERS

Andrade Marina , Professor Auxiliar	University Institute of Lisbon, Lisboa, Portugal
Bartošová Jitka , RNDr., PhD	University of Economics, Jindřichův Hradec, Czech Republic
Baštinec Jaromír , doc. RNDr., CSc.	FEEC, Brno University of Technology, Brno, Czech Republic
Beránek Jaroslav , doc. RNDr., CSc.	Masaryk University, Brno, Czech Republic
Biswas Md. Haider Ali , Associate Professor	Engineering and Technology School, Khulna University, Belize
Bittnerová Daniela , RNDr., CSc.	Technical Univerzity of Liberec, Liberec, Czech Republic
Brabec Marek , Ing., PhD	Academy of Sciences of the Czech Republic, Praha, Czech Republic
Buikis Maris , Prof. Dr.	Riga Technical University, Riga, Latvia
Cyhelský Lubomír , Prof. Ing., DrSc.	Vysoká škola finanční a správní, Praha, Czech Republic
Dorociaková Božena , RNDr., PhD	University of Žilina, Žilina, Slovak Republic
Emanovský Petr , Doc. RNDr., PhD	Palacky University, Olomouc, Czech Republic
Ferreira Manuel Alberto M. , Professor Catedrático	University Institute of Lisbon, Lisboa, Portugal
Filipe José António , Professor Auxiliar	IBS - IUL, ISCTE - IUL, Lisboa , Portugal
Habiballa Hashim , RNDr. PaedDr., PhD	University of Ostrava, Ostrava, Czech Republic
Habiballa Hashim , RNDr. PaedDr., PhD	University of Ostrava, Ostrava, Czech Republic
Hošková-Mayerová Šárka , doc. RNDr., PhD	University of Defence, Brno, Czech Republic

Hošpesová Alena , doc. PhDr., PhD	Jihočeská univerzita, České Budějovice, Czech Republic
Iorfida Vincenzo	Lamezia Terme, Italy
Iveta Stankovičová , PhD	UK Bratislava, Bratislava, Slovak Republic
Jancarík Antonín , PhD	Charles University, Prague, Czech Republic
Jukl Marek , RNDr., PhD	Palacky University, Olomouc, Czech Republic
Kráľ Pavol , RNDr., PhD	Matej Bel University, Banská Bystrica, Slovak Republic
Kunderová Pavla , doc. RNDr., CSc.	Palacky University, Olomouc, Czech Republic
Kvasz Ladislav , Prof.	Charles University, Prague, Czech Republic
Langhamrová Jitka , doc. Ing., CSc	University of Economics in Prague, Prague Czech Republic
Linda Bohdan , doc. RNDr., CSc.	University of Pardubice, Pardubice, Czech Republic
Maroš Bohumil , doc. RNDr., CSc.	University of Technology, Brno, Czech Republic
Matvejevs Andrejs , DrSc., Ing.	Riga Technical university, Riga, Latvia
Mikeš Josef , Prof. RNDr., DrSc.	Palacky University, Olomouc, Czech Republic
Milerová Helena , Bc.	Charles University, Prague, Czech Republic
Miroslav Husek	Charles University, Prague, Czech Republic
Miskolczi Martina , Mgr., Ing.	University of Economics in Prague, Prague, Czech Republic
Morkisz Paweł ,	AGH University of Science and Technology, Krakow, Poland
Mošna František , RNDr., PhD	Czech Univ. of Life Sciences, Praha, Czech Republic
Paláček Radomír , RNDr., PhD	VŠB - Technical University of Ostrava, Ostrava, Czech Republic
Pospíšil Jiří , Prof. Ing., CSc.	Czech Technical University of Prague, Prague, Czech Republic

Potůček Radovan , RNDr., PhD	University of Defence, Brno, Czech Republic
Radova Jarmila , doc. RNDr., PhD	University of Economics, Prague, Czech Republic
Rus Ioan A. , Professor	Babes-Bolyai University of Cluj-Napoca, Cluj-Napoca, Romania
Růžicková Miroslava , doc. RNDr., CSc.	University of Žilina, Žilina, Slovak Republic
Segeth Karel , Prof. RNDr., CSc.	Academy of Sciences of the Czech Republic, Prague , Czech Republic
Slaby Antonin , Prof. RNDr., PhDr., CSc.	University of Hradec Kralove, Hradec Kralove, Czech Republic
Sousa Cristina Alexandra , Master	Universidade Portucalense Infante D. Henrique, Porto, Portugal
Svoboda Zdeněk , RNDr., CSc.	FEEC, Brno University of Technology, Brno, Czech Republic
Šamšula Pavel , doc. PaedDr., CSc	Charles University, Prague, Czech Republic
Torre Matteo , Laurea in Matematica	Scula Secondaria Superiore, Alessandria, Italy
Trojovsky Pavel , RNDr., PhD	University of Hradec Kralove, Hradec Kralove, Czech Republic
Trokanová Katarína , Doc.	Slovak Technical University, Bratislava, Slovak Republic
Ulrychová Eva , RNDr.	University of Finance and Administration, Prague, Czech Republic
Vanžurová Alena , doc. RNDr., CSc.	Palacký University, Olomouc, Czech Republic
Velichová Daniela , doc. RNDr., CSc. mim.prof.	Slovak University of Technology, Bratislava, Slovak Republic
Vítovec Jiří , Mgr., PhD	Brno University of Technology, Brno, Czech Republic
Voicu Nicoleta , Dr.	Transilvania University of Brasov, Romania, Brasov, Romania
Volna Eva , doc. RNDr. PaedDr. PhD	University of Ostrava, Ostrava, Czech Republic

Wimmer Gejza, Professor

Slovak Academy of Sciences, Bratislava, Slovak
Republic

Zeithamer Tomáš R., Ing., PhD

University of Economics, Prague, Czech Republic

A CAUCHY COMPLETION OF DUALY RESIDUATED LATTICE ORDERED SEMIGROUPS

JASEM Milan, (SK)

Abstract. In this paper convergence with a fixed regulator in dually residuated lattice ordered semigroups is investigated and a u-Cauchy completion of a strong dually residuated lattice ordered semigroup is constructed. It is also shown that this completion is uniquely determined up to isomorphism.

Key words and phrases. u-uniform convergence, Cauchy sequence, dually residuated lattice ordered semigroups.

Mathematics Subject Classification. Primary 06F05.

1 Introduction

Dually residuated lattice ordered semigroups (DRL-semigroups) were introduced and studied by Swamy in [14], [15], [16]. DRL-semigroups were also investigated by Kovář [10], [11], Kühr [12] and by the author [8]. Galatos and Tsinakis [7] proved that DRL-semigroups are equivalent to commutative GBL-algebras.

Birkhoff [1] and Luxemburg and Zaanen [13] studied relatively uniform convergence of sequences in vector lattices. Relatively uniform convergence in lattice ordered groups was dealt with by Černák and Lihová [5] and Černák and Jakubík [6]. Convergence with a fixed regulator in lattice ordered groups was studied by Černák [2], [3] and Černák and Lihová [4].

This paper is a continuation of the paper [9] where convergence with a fixed regulator in DRL-semigroups was introduced and studied. In the present paper Cauchy sequences are investigated and a u-Cauchy completion of a strong DRL-semigroup has been constructed.

2 Preliminaries

We review some notions and notations used in the paper.

A system $A = (A, +, \leq, -)$ is called a dually residuated lattice ordered semigroup if and only if

- (1) $(A, +, \leq)$ is a commutative lattice ordered semigroup with zero element 0, i. e. $(A, +)$ is a commutative semigroup with zero 0 and (A, \leq) is a lattice with lattice operations \wedge and \vee such that $a + (b \vee c) = (a + b) \vee (a + c)$ and $a + (b \wedge c) = (a + b) \wedge (a + c)$,
- (2) given a, b in A there exists a least x in A such that $b + x \geq a$, and this x is denoted by $a - b$,
- (3) $(a - b) \vee 0 + b \leq a \vee b$ for all $a, b \in A$,
- (4) $(a - a) \leq 0$ for each $a \in A$.

Any DRL-semigroup can be equationally defined as an algebra with the binary operations $+, \vee, \wedge, -$, by replacing (2) by the equations:

$$x + (y - x) \geq y, \quad x - y \leq (x \vee z) - y, \quad (x + y) - y \leq x \quad [14, \text{Theorem 1}].$$

For any a and b in a DRL-semigroup A we shall write $|a - b| = (a - b) \vee (b - a)$ ($|a - b|$ is called the symetric difference of a and b .) This notation arising from lattice ordered groups is different from one used by Swamy in [14], however it is most suitable for our case.

The symetric difference satisfies the following conditions:

- (i) $|a - b| \geq 0$, $|a - b| = 0$ if and only if $a = b$,
- (ii) $|a - b| = |b - a|$,
- (iii) $|a - c| \leq |a - b| + |b - c|$.

Any DRL-semigroup is an autometrized algebra with the symetric difference [14, Theorem 9].

We use \mathbb{N} for the set of all positive integers.

Let A be a DRL-semigroup. We denote $A^+ = \{x \in A; x \geq 0\}$.

An element $x \in A^+$ is said to be Archimedean if whenever $y \in A^+$ and $ny \leq x$ for each $n \in \mathbb{N}$, then $y = 0$.

A DRL-semigroup A is called strong if $x, y \in A$ and $2x \leq 2y$ implies $x \leq y$.

Any abelian lattice ordered group is a strong DRL-semigroup and hence any Archimedean l-group is also a strong DRL-semigroup.

We shall need the following propositions from [14].

Let A be a DRL-semigroup, $a, b, c \in A$. Then

- (P₁) $a \leq b$ if and only if $a - b \leq 0$ (Lemma 7),
- (P₂) $a \leq b$ implies $(b - a) + a = b$ (Lemma 8),
- (P₃) $a \leq b$ implies $a - c \leq b - c$ and $c - b \leq c - a$ (Lemma 3),
- (P₄) $(a \vee b) - c = (a - c) \vee (b - c)$ (Lemma 4).

Luxemburg and Zaanen [13] introduced notions of a u-uniform convergence and of a relatively uniform convergence of sequences for vector lattices and Černák and Lihová [5] for lattice ordered groups.

Analogous definition of a u-uniform convergence we shall use for DRL-semigroups.

Definition 2.1 Let A be a DRL-semigroup, (x_n) a sequence in A , $u \in A^+$. It is said that a sequence (x_n) in A converges u-uniformly to an element $x \in A$, written $x_n \xrightarrow{u} x$, if the following condition is satisfied:

(C₃) for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$, such that $k|x_n - x| \leq u$ for each $n \in \mathbb{N}$, $n \geq n_k$.

The element u in the Definition 2.1 is called a convergence regulator.

If $x_n \xrightarrow{u} x$, we say that x is a u -limit of (x_n) .

If we take the same regulator for all sequences we get convergence which is called convergence with a fixed regulator.

Černák and Lihová have shown that if convergence regulator u in lattice ordered group is not Archimedean then a sequence can have more u -limits. So, it is convenient to have an Archimedean element in the role of convergence regulator.

Basic properties of the convergence with a fixed regulator in DRL-semigroups were established in [9].

It was shown that if convergence regulator u in a strong DRL-semigroup B is an Archimedean element, then u -limits are uniquely determined (Theorem 1) and if $(x_n), (y_n)$ are sequences in B and $x_n \xrightarrow{u} x$ and $y_n \xrightarrow{u} y$, then $x_n + y_n \xrightarrow{u} x + y$, $x_n - y_n \xrightarrow{u} x - y$, $x_n \vee y_n \xrightarrow{u} x \vee y$, $x_n \wedge y_n \xrightarrow{u} x \wedge y$ (Theorem 2).

3 Cauchy sequences and a u -Cauchy completion

Definition 3.1 Let B be a DRL-semigroup, $u \in B^+$. A sequence (x_n) in B is called a u -Cauchy sequence, if for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $u \geq k|x_m - x_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$.

Throughout the rest of the paper A will be a strong DRL-semigroup and an Archimedean element u of A will be a fixed regulator for all sequences in A .

In [9] it was showed that each u -convergent sequence in A is a u -Cauchy sequence (Theorem 6) and that if (x_n) and (y_n) are u -Cauchy sequences, then $(x_n + y_n)$, $(x_n - y_n)$, $(x_n \vee y_n)$, $(x_n \wedge y_n)$ are u -Cauchy sequences, too (Theorem 7).

Let C be the set of all u -Cauchy sequences in A .

Let $(x_n), (y_n) \in C$. We put $(x_n) + (y_n) = (x_n + y_n)$. Further we set $(x_n) \leq (y_n)$ if and only if $x_n \leq y_n$ for each $n \in \mathbb{N}$.

If $(x_n), (y_n) \in C$ and $2(x_n) \leq 2(y_n)$, then $(x_n) \leq (y_n)$ [9, Theorem 2(v)].

We denote by (\bar{x}) the sequence (x, x, x, \dots) in A . Clearly, the u -limit of this sequence is x . In [9] it was also proved that $(C, +, \leq)$ is a strong DRL-semigroup with zero $(\bar{0})$ and lattice operations \vee and \wedge such that $(x_n) \vee (y_n) = (x_n \vee y_n)$, $(x_n) \wedge (y_n) = (x_n \wedge y_n)$ for all $(x_n), (y_n) \in C$. Further, $(x_n) - (y_n) = (x_n - y_n)$ for all $(x_n), (y_n) \in C$.

Swamy [16, p. 71] defined an ideal and a convex sub-DRL-semigroup of a DRL-semigroup as follows.

Definition 3.2 A non-empty subset I of A is called an ideal of A if and only if:

- (i) $a, b \in I$ implies $a + b \in I$,
- (ii) $a \in I, b \in A$ and $|b - 0| \leq |a - 0|$ imply $b \in I$.

Definition 3.3 Non-empty subset S of A is called a convex sub-DRL-semigroup if and only if the following conditions are satisfied:

- (i) if $a, b \in S$, then $a + b, a - b, a \wedge b, a \vee b \in S$,
- (ii) if $a, b \in S, x \in A$ and $a \wedge b \leq x \leq a \vee b$, then $x \in S$.

The set E of all sequences (x_n) in A such that $x_n \xrightarrow{u} 0$ is an ideal of C and a convex sub-DRL-semigroup of C . [9, Theorems 9 and 10]

By a congruence relation on a DRL-semigroup B we mean an equivalence relation, having the substitution property with respect to all the operations: $+$, \vee , \wedge and $-$. Swamy [16, Theorem 1.2] showed that ideals of any DRL-semigroup correspond one to one to its congruence relations. Further, he showed that if I is an ideal of B , then the binary relation $\vartheta(I)$ defined by $(a, b) \in \vartheta(I)$ if and only if $|a - b| \in I$ is a congruence relation on B [16, p. 72].

Hence, the factor semigroup $A^* = C/\vartheta(E)$ is a DRL-semigroup.

We denote by $(x_n)^*$ the congruence class modulo $\vartheta(E)$ containing the sequence (x_n) . Recall that $(x_n)^* + (y_n)^* = (x_n + y_n)^*$, $(x_n)^* - (y_n)^* = (x_n - y_n)^*$, $(x_n)^* \wedge (y_n)^* = (x_n \wedge y_n)^*$, $(x_n)^* \vee (y_n)^* = (x_n \vee y_n)^*$, for all $(x_n)^*, (y_n)^* \in A^*$.

Lemma 3.4 (i) $E = (\bar{0})^*$.

(ii) If $(x_n) \in C$, $(x'_n) \in (x_n)^*$, then $(x'_n) - (x_n), (x_n) - (x'_n) \in E$.

Proof. (i) Let $(x_n) \in E$. Since E is a DRL-semigroup, we have $|(x_n) - (\bar{0})| \in E$. Hence $(x_n) \in (\bar{0})^*$ and thus $E \subseteq (\bar{0})^*$.

Let $(y_n) \in (\bar{0})^*$. Then $|(y_n) - (\bar{0})| \in E$. Since $|(y_n) - (\bar{0})| = ||(y_n) - (\bar{0})| - (\bar{0})| \in E$, we have $(y_n) \in E$. Hence $(\bar{0})^* \subseteq E$.

(ii) If $(x'_n) \in (x_n)^*$, then $(x'_n)^* = (x_n)^*$. Thus $(x'_n - x_n)^* = (x'_n)^* - (x_n)^* = (x_n)^* - (x_n)^* = (x_n - x_n)^* = (\bar{0})^* = E$. Hence $(x'_n) - (x_n) \in E$. Analogously, $(x_n) - (x'_n) \in E$.

Lemma 3.5 Let $(x_n), (y_n) \in C$. Then the following conditions are equivalent:

(i) $(x_n)^* \leq (y_n)^*$,

(ii) $(x_n) \leq (y_n) + (t_n)$ for some $(t_n) \in E$, $(t_n) \geq (\bar{0})$,

(iii) for each $(x'_n) \in (x_n)^*$ there exists $(y'_n) \in (y_n)^*$ such that $(x'_n) \leq (y'_n)$.

Proof. (i) \Leftrightarrow (ii) Clearly, $(x_n)^* \leq (y_n)^*$ iff $(x_n)^* \vee (y_n)^* = (y_n)^*$ iff $((x_n) \vee (y_n)) - (y_n) = |((x_n) \vee (y_n)) - (y_n)| \in E$. Let $(z_n) = ((x_n) \vee (y_n)) - (y_n)$. Thus $(z_n) \geq (\bar{0})$. If $(x_n)^* \leq (y_n)^*$, then $(z_n) \in E$. In view of (P₂) we obtain $(x_n) \leq (x_n) \vee (y_n) = (((x_n) \vee (y_n)) - (y_n)) + (y_n) = (z_n) + (y_n) = (y_n) + (z_n)$.

Conversely, if $(x_n) \leq (y_n) + (z_n)$, where $(\bar{0}) \leq (z_n) \in E$, then from (P₃) it follows that $(x_n) - (y_n) \leq ((z_n) + (y_n)) - (y_n) \leq (z_n)$. According to (P₄), $((x_n) \vee (y_n)) - (y_n) = ((x_n) - (y_n)) \vee (\bar{0}) \leq (z_n) \vee (\bar{0}) = (z_n) \in E$.

The equivalence of (i) and (iii) is obvious because $\vartheta(E)$ is a lattice congruence.

Lemma 3.6 A^* is a strong DRL-semigroup.

Proof. Let $(x_n)^*, (y_n)^* \in A^*$, $2(x_n)^* \leq 2(y_n)^*$. Then $(2x_n)^* \leq (2y_n)^*$. By Lemma 3.5, $(2x_n) \leq (2y_n) + (t_n) \leq (2y_n) + (2t_n)$, where $(t_n) \in E$, $(\bar{0}) \leq (t_n)$. This implies $2x_n \leq 2y_n + 2t_n = 2(y_n + t_n)$ for each $n \in \mathbb{N}$. Then $x_n \leq y_n + t_n$ for each $n \in \mathbb{N}$. This implies $(x_n) \leq (y_n) + (t_n)$. By Lemma 3.5, $(x_n)^* \leq (y_n)^*$. Hence DRL-semigroup A^* is strong.

Lemma 3.7 $(\bar{u})^*$ is an Archimedean element in A^* .

Proof. Let $(x_n)^* \in A^*$, $(\bar{0})^* \leq (x_n)^*$, $m(x_n)^* \leq (\bar{u})^*$ for each $m \in \mathbb{N}$. From $(\bar{0})^* \leq (x_n)^*$ and Lemma 3.5 it follows that $(\bar{0}) \leq (x'_n)$ for some $(x'_n) \in (x_n)^*$. Thus $0 \leq x'_n$ for each $n \in \mathbb{N}$, $(x'_n)^* = (x_n)^*$. Then $(mx'_n)^* = m(x'_n)^* = m(x_n)^* \leq (\bar{u})^*$ for each $m \in \mathbb{N}$. By Lemma 3.5, $(mx'_n) \leq (\bar{u}) + (t_n)$ for some $(t_n) \in E$, $(t_n) \geq (\bar{0})$. Since $t_n \xrightarrow{u} 0$ for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $t_n = |t_n - 0| \leq k|t_n - 0| \leq u$ for each $n \in \mathbb{N}$, $n \geq n_k$. Thus for each $m \in \mathbb{N}$ we have $mx'_n \leq u + t_n \leq 2u$, where $n \in \mathbb{N}$, $n \geq n_k$. Hence $2px'_n \leq 2u$ for each $p \in \mathbb{N}$. Then $px'_n \leq u$ for each $p \in \mathbb{N}$. This yields $x'_n = 0$ for each $n \in \mathbb{N}$, $n \geq n_k$. Thus $x'_n \xrightarrow{u} 0$ and hence $(x'_n) \in E$. Then $(x_n)^* = (x'_n)^* = E$. Therefore $(\bar{u})^*$ is an Archimedean element in A^* .

Lemma 3.8 Let (x_n) be a sequence in A . If $x_n \xrightarrow{u} 0$, then $(\overline{x_n})^* \xrightarrow{(\bar{u})^*} (\bar{0})^*$.

Proof. Let (x_n) be a sequence in A , $x_n \xrightarrow{u} 0$. Hence for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $k|x_n - 0| \leq u$ for each $n \in \mathbb{N}$, $n \geq n_k$. Then $(\bar{u}) \geq (k|x_n - 0|) = k(|x_n - 0|) = k|(\overline{x_n}) - (\bar{0})|$ for each $n \in \mathbb{N}$, $n \geq n_k$. Thus $(\bar{u})^* \geq k|(\overline{x_n})^* - (\bar{0})^*|$ for each $n \in \mathbb{N}$, $n \geq n_k$. Hence $(\overline{x_n})^* \rightarrow (\bar{0})^*$.

Lemma 3.9 Let $(x_n) \in C$, $l \in \mathbb{N}$, $a_1, \dots, a_l \in A$. Let $y_n = x_{l+n-1}$ for each $n \in \mathbb{N}$ i. e. $(y_n) = (x_l, x_{l+1}, x_{l+2}, \dots)$. Let $z_1 = a_1, \dots, z_l = a_l$, $z_n = x_n$ for each $n \in \mathbb{N}$, $n \geq l+1$ i. e. $(z_n) = (a_1, \dots, a_l, x_{l+1}, x_{l+2}, \dots)$. Then
(i) $(y_n), (z_n) \in C$,
(ii) $(x_n)^* = (y_n)^* = (z_n)^*$.

Proof. Since $(x_n) \in C$, for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $k|x_m - x_n| \leq u$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$.

(i) Since $l + m - 1 \geq n_k$, $l + n - 1 \geq n_k$, we have $u \geq k|x_{l+m-1} - x_{l+n-1}| = k|y_m - y_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$. Therefore $(y_n) \in C$.

Further, if we take $n_k \in \mathbb{N}$, $n_k \geq l$, then we have $k|z_m - z_n| = k|x_m - x_n| \leq u$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$. Hence $(z_n) \in C$.

(ii) If we take $m = l + n - 1$, we have $u \geq k|x_{l+n-1} - x_n| = k|y_n - x_n| = k||y_n - x_n| - 0|$ for each $n \in \mathbb{N}$, $n \geq n_k$. Thus $|y_n - x_n| \xrightarrow{u} 0$. Hence $|(y_n) - (x_n)| \in E$. This implies $(x_n)^* = (y_n)^*$.

Clearly $|z_n - x_n| \xrightarrow{u} 0$ and hence $|(z_n) - (x_n)| \in E$. Therefore $(x_n)^* = (z_n)^*$.

Denote by C^* the set of all $(\bar{u})^*$ -Cauchy sequences in A^* .

Theorem 3.10 Let (x_n) be a sequence in A . Then $(x_n) \in C$ if and only if $((\overline{x_n})^*) \in C^*$.

Proof. Let $(x_n) \in C$. Then for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $k|x_m - x_n| \leq u$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$. Thus $(\bar{u}) \geq (k|x_m - x_n|)$ and hence $(\bar{u})^* \geq (k|x_m - x_n|)^* = k|(\overline{x_m})^* - (\overline{x_n})^*|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$. Therefore $((\overline{x_n})^*) \in C^*$.

Let $((\overline{x_n})^*) \in C^*$. Then for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $k|(\overline{x_m})^* - (\overline{x_n})^*| = |(k(\overline{x_m})^* - (\overline{x_n})^*)| \leq (\bar{u})^*$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$.

From Lemma 3.5 it follows that $(k|x_m - x_n|) \leq (\bar{u}) + (t_l)$, where $(t_l) \in (\bar{0})^*$, $(\bar{0}) \leq (t_l)$. Hence $t_l \xrightarrow{u} 0$. Thus for each $p \in \mathbb{N}$ there exists $n_p \in \mathbb{N}$, such that $p|t_l| \leq u$ for each $l \in \mathbb{N}$, $l \geq n_p$. For $k = 1$ we have $t_l \leq u$ for each $l \in \mathbb{N}$, $l \geq n_1$.

Hence $k|x_m - x_n| \leq 2u$, for each $m, n \in \mathbb{N}$, $m, n \geq n_{k_0} = \max\{n_k, n_1\}$. Then for any $k \in \mathbb{N}$ there exists $n_{k_1} \in \mathbb{N}$ such that $2k|x_m - x_n| \leq 2u$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{k_1}$. This implies that for each $k \in \mathbb{N}$ there exists $n_{k_1} \in \mathbb{N}$ such that $k|x_m - x_n| \leq u$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{k_1}$. Therefore $(x_n) \in C$.

Theorem 3.11 *Let $(x_n) \in C$. Then $(\overline{x_n})^* \xrightarrow{(\overline{u})^*} (x_n)^*$.*

Proof. If $(x_n) \in C$, then for each $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ such that $u \geq k|x_m - x_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_k$. If we take $n = n_k$, we get $u \geq k|x_m - x_{n_k}|$ for each $m \in \mathbb{N}$, $m \geq n_k$. Similarly, if we take $n = n_k + 1, n = n_k + 2, \dots$, we can get $u \geq k|x_m - x_{n_k+1}|, u \geq k|x_m - x_{n_k+2}|, \dots$, for each $m \in \mathbb{N}$, $m \geq n_k$. Hence $(\overline{u})^* \geq (k|x_m - x_{n_k}|, k|x_m - x_{n_k+1}|, k|x_m - x_{n_k+2}|, \dots)^*$. In view of Lemma 3.9 we obtain $(\overline{u})^* \geq (k|x_m - x_1|, k|x_m - x_2|, \dots, k|x_m - x_{n_k}|, k|x_m - x_{n_k+1}|, \dots)^* = (k|(\overline{x_m}) - (x_n)|)^* = k|(\overline{x_m})^* - (x_n)^*|$ for each $m \in \mathbb{N}$, $m \geq n_k$. Therefore $(\overline{x_n})^* \xrightarrow{(\overline{u})^*} (x_n)^*$.

Theorem 3.12 *Let $\varphi : A \rightarrow A^*$ be the mapping such that $\varphi(x) = (\overline{x})^*$ for each $x \in A$. Then*
(i) φ is a monomorphism,
(ii) every element of A^ is the $(\overline{u})^*$ -limit of some sequence in $\varphi(A)$.*

Proof. (i) Let $x, y \in A$. Clearly $\varphi(x + y) = \varphi(x) + \varphi(y)$, $\varphi(x - y) = \varphi(x) - \varphi(y)$, $\varphi(x \wedge y) = \varphi(x) \wedge \varphi(y)$, $\varphi(x \vee y) = \varphi(x) \vee \varphi(y)$.

If $\varphi(x) = \varphi(y)$, then $(\overline{x})^* = (\overline{y})^*$. This implies $(\overline{x}) \in (\overline{y})^*$. Then $|(\overline{x}) - (\overline{y})| = (\overline{|x - y|}) \in E$. Since 0 is the u -limit of the sequence $(\overline{|x - y|})$, we have $|x - y| = 0$. Hence $x = y$.

(ii) Let $(x_n)^* \in A^*$. Then $((\overline{x_n})^*)$ is a sequence in $\varphi(A)$. Since $(x_n) \in C$, from Theorem 3.11 it follows that $(\overline{x_n})^* \xrightarrow{(\overline{u})^*} (x_n)^*$.

Definition 3.13 *Let B be a DRL-semigroup. If every u -Cauchy sequence (x_n) in B is u -convergent, then B is called u -Cauchy complete.*

Definition 3.14 *Let B be a DRL-semigroup.*

A DRL-semigroup D is said to be a u -Cauchy completion of B , if the following conditions are satisfied:

- (i) B is a sub-DRL-semigroup of D ,*
- (ii) D is u -Cauchy complete,*
- (iii) Every element of D is a u -limit of some sequence in B .*

Theorem 3.15 *A^* is a $(\overline{u})^*$ -Cauchy complete DRL-semigroup.*

Proof. Let $X^1 = (x_m^1)^*, X^2 = (x_m^2)^*, \dots$ be a sequence in C^* . Let $n \in \mathbb{N}$. Let $X_m^n = (\overline{x_m^n})^*$ for each $m \in \mathbb{N}$. Hence $X_m^n \in \varphi(A)$ for each $m \in \mathbb{N}$. By Theorem 3.11, the sequence $(X_1^n, X_2^n, X_3^n, \dots)$ $(\overline{u})^*$ -converges to X^n . Hence for choosen $n \in \mathbb{N}$ there exists $m_n \in \mathbb{N}$, such that $n|X_m^n - X^n| \leq (\overline{u})^*$ for each $m \in \mathbb{N}$, $m \geq m_n$. If we let run n over \mathbb{N} , we can take $m_1 \leq m_2 \leq \dots$. Further, if we take $m = m_n$, we get $n|X_{m_n}^n - X^n| \leq (\overline{u})^*$ for each $n \in \mathbb{N}$. Let $Z_n = X_{m_n}^n$ for each $n \in \mathbb{N}$. Then $n|Z_n - X^n| \leq (\overline{u})^*$ for any $n \in \mathbb{N}$.

Now we show that $(Z_n) \in C^*$. Since (X^n) is an $(\bar{u})^*$ -Cauchy sequence, for each $l \in \mathbb{N}$ there exists $n_l \in \mathbb{N}$, $n_l > l$ such that $l|X^m - X^n| \leq (\bar{u})^*$ for each $m, n \in \mathbb{N}$, $m, n \geq n_l$. Since $l \leq m, n$, we get $l|Z_m - X^m| \leq m|Z_m - X^m| \leq (\bar{u})^*$, $l|X^n - Z_n| \leq n|X^n - Z_n| = n|Z_n - X^n| \leq (\bar{u})^*$ for each $m, n \in \mathbb{N}$, $m, n \geq n_l$. Then we have $2(\bar{u})^* \geq l|Z_m - X^m| + l|X^m - X^n| = l(|Z_m - X^m| + |X^m - X^n|) \geq l(|Z_m - X^n|)$ for each $m, n \in \mathbb{N}$, $m, n \geq n_l$. Hence for any $l \in \mathbb{N}$ there exists $n_{l_1} \in \mathbb{N}$, $n_{l_1} \geq n_l$ such that $2(\bar{u})^* \geq 2l|Z_m - X^n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{l_1}$. Therefore $(\bar{u})^* \geq l|Z_m - X^n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{l_1}$. Further, we obtain $2(\bar{u})^* \geq l|Z_m - X^n| + l|X^n - Z_n| = l(|Z_m - X^n| + |X^n - Z_n|) \geq l|Z_m - Z_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{l_1}$. Thus for any $l \in \mathbb{N}$ there exists $n_{l_2} \in \mathbb{N}$, $n_{l_2} \geq n_{l_1}$ such that $2(\bar{u})^* \geq 2l|Z_m - Z_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{l_2}$. Therefore for each $l \in \mathbb{N}$ there exists $n_{l_2} \in \mathbb{N}$ such that $(\bar{u})^* \geq l|Z_m - Z_n|$ for each $m, n \in \mathbb{N}$, $m, n \geq n_{l_2}$. Therefore $(Z_n) \in C^*$.

If we put $z_n = x_{m_n}^n$, then $Z_n = (\bar{z}_n)^*$. Since $(Z_n) \in C^*$, from Theorem 3.10 it follows that $(z_n) \in C$. By Theorem 3.11, $Z_n \rightarrow (z_n)^*$.

Let $t \in \mathbb{N}$. Since $Z_n \rightarrow (z_n)^*$, there exist $n_t \in \mathbb{N}$, $n_t \geq t$ such that $t|Z_k - (z_n)^*| \leq (\bar{u})^*$ for each $k \in \mathbb{N}$, $k \geq n_t$.

Since $t \leq k$, we have $t|X^k - Z_k| = t|Z_k - X^k| \leq k|Z_k - X^k| \leq (\bar{u})^*$ for each $k \in \mathbb{N}$, $k \geq n_t$. Then we have $2(\bar{u})^* \geq t|X^k - Z_k| + t|Z_k - (z_n)^*| = t(|X^k - Z_k| + |Z_k - (z_n)^*|) \geq t|X^k - (z_n)^*|$ for each $k \in \mathbb{N}$, $k \geq n_t$. Thus for any $t \in \mathbb{N}$ there exists $n_{t_1} \in \mathbb{N}$, $n_{t_1} \geq n_t$ such that $2(\bar{u})^* \geq 2t|X^k - (z_n)^*|$ for each $k \in \mathbb{N}$, $k \geq n_{t_1}$. Hence for each $t \in \mathbb{N}$ there exists $n_{t_1} \in \mathbb{N}$ such that $(\bar{u})^* \geq t|X^k - (z_n)^*|$ for each $k \in \mathbb{N}$, $k \geq n_{t_1}$. Therefore $X^n \xrightarrow{(\bar{u})^*} (z_n)^*$.

If x and $\varphi(x) = (\bar{x})^*$ will be identified for each $x \in A$, then A is a subgroup of A^* and $(\bar{u})^* = u$. Then we get as a consequence of Theorems 3.12 and 3.15 the following proposition.

Theorem 3.16 A^* is an u -Cauchy completion of A .

Theorem 3.17 If A_1 and A_2 are u -Cauchy completions of A , then there exists a semigroup l -isomorphism of A_1 onto A_2 leaving all elements of A fixed.

Proof. From the assumptions it follows that A is a sub-DRI-semigroup of A_1 and A_2 . Let $x' \in A_1$. Then there exists a sequence (x_n) in A such that $x_n \xrightarrow{u} x'$ in A_1 . Since (x_n) is a convergent, (x_n) is an u -Cauchy sequence in A_1 and then also in A and A_2 . Since A_2 is u -Cauchy complete, there exists $x'' \in A_2$ such that $x_n \xrightarrow{u} x''$ in A_2 . We put $\psi(x') = x''$.

Now we show that the mapping ψ is correctly defined. Let (y_n) be another sequence in A such that $y_n \xrightarrow{u} x'$ in A_1 . Analogously as above we can get that there exists $x''' \in A_2$ such that $y_n \xrightarrow{u} x'''$ in A_2 . Then we have $x_n - y_n \xrightarrow{u} 0$, $y_n - x_n \xrightarrow{u} 0$ in A_1 and $x_n - y_n \xrightarrow{u} x'' - x'''$, $y_n - x_n \xrightarrow{u} x''' - x''$ in A_2 . Since $0 \in A \subseteq A_2$, we have $x_n - y_n \xrightarrow{u} 0$ in A_2 , $y_n - x_n \xrightarrow{u} 0$ in A_2 . Because u -limits are uniquely determined we obtain $x'' - x''' = 0$, $x''' - x'' = 0$. Then (P_1) yields $x'' \leq x'''$, $x''' \leq x''$. Therefore $x'' = x'''$.

Let $z' \in A_2$. Then there exists a sequence $(z_n) \in A$ such that $z_n \xrightarrow{u} z'$ in A_2 and $z' \in A_1$ such that $z_n \xrightarrow{u} z'$ in A_1 . Then $\psi(z') = z'$. Therefore ψ is a surjective mapping. Let $a', b' \in A_1$, $\psi(a') = a''$, $\psi(b') = b''$. Hence there are sequences (a_n) and (b_n) in A such that $a_n \xrightarrow{u} a'$, $b_n \xrightarrow{u} b'$ in A_1 and $a_n \xrightarrow{u} a''$, $b_n \xrightarrow{u} b''$ in A_2 . Then $a_n + b_n \xrightarrow{u} a' + b'$, $a_n - b_n \xrightarrow{u} a' - b'$, $a_n \vee b_n \xrightarrow{u} a' \vee b'$, $a_n \wedge b_n \xrightarrow{u} a' \wedge b'$ in A_1 , $a_n + b_n \xrightarrow{u} a'' + b''$, $a_n - b_n \xrightarrow{u} a'' - b''$, $a_n \vee b_n \xrightarrow{u} a'' \vee b''$, $a_n \wedge b_n \xrightarrow{u} a'' \wedge b''$ in A_2 . Therefore $\psi(a' + b') = a'' + b'' = \psi(a') + \psi(b')$, $\psi(a' - b') = a'' - b'' = \psi(a') - \psi(b')$, $\psi(a' \vee b') = a'' \vee b'' = \psi(a') \vee \psi(b')$, $\psi(a' \wedge b') = a'' \wedge b'' = \psi(a') \wedge \psi(b')$.

If $a'' = b''$, then $a_n - b_n \xrightarrow{u} 0$, $b_n - a_n \xrightarrow{u} 0$ in A_2 and also in A and A_1 . Since $a_n - b_n \xrightarrow{u} a' - b'$, $b_n - a_n \xrightarrow{u} b' - a'$ in A_1 , we get $a' - b' = 0$, $b' - a' = 0$. By (P_1) , $a' \leq b'$, $b' \leq a'$. Therefore $a' = b'$. Clearly, $\psi(x) = x$ for each $x \in A$.

Acknowledgement

Support of Slovak Vega Grants 1/0198/09 and 1/0071/09 is acknowledged.

References

- [1] BIRKHOFF, G.: *Lattice Theory*. Third edition, Amer. Math. Soc., Providence 1967.
- [2] ČERNÁK, Š.: *Convergence with a fixed regulator in Archimedean lattice ordered groups*. Math. Slovaca 56, 2006, 167-180.
- [3] ČERNÁK, Š.: *Convergence with a fixed regulator in lattice ordered groups and applications to MV-algebras*. Soft Computing 12, 2008, 453-462.
- [4] ČERNÁK, Š., LIHOVÁ, J.: *Convergence with a fixed regulator in lattice ordered groups*. Tatra Mt. Math. Publ. 30, 2005, 35-45.
- [5] ČERNÁK, Š., LIHOVÁ, J.: *Relatively uniform convergence in lattice ordered groups*. Selected question of algebra. Collection of papers dedicated to the memory of N. Ja. Medvedev, Altai State Univ. Barnaul, Barnaul, 2007, 218-241.
- [6] ČERNÁK, Š., JAKUBÍK, J.: *Relatively uniform convergences in archimedean lattice ordered groups*. Math. Slovaca 60, 2010, 447-460.
- [7] GALATOS, N., TSINAKIS, C.: *Generalized MV-algebras*. J. Algebra 283, 2005, 245-291.
- [8] JASEM, M.: *Weak isometries and direct decompositions of dually residuated lattice ordered semigroups*. Math. Slovaca 43, 1993, 119-136.
- [9] JASEM, M.: *On convergence with a fixed regulator in dually residuated lattice ordered semigroups*. Proceedings of International Conference Presentation of Mathematics'11, Technical university of Liberec, 2011, 69-77.
- [10] KOVÁŘ, T.: *Any DRL-semigroup is the direct product of a commutative l-group and a DRL-semigroup with the least element*. Discuss. Math. Algebra & Stochastic Methods 16, 1996, 99-105.
- [11] KOVÁŘ, T.: *A general theory of dually residuated lattice ordered monoids*, Ph. D. Thesis, Palacký Univ., Olomouc 1996.
- [12] KÜHR, J.: *Dually residuated lattice ordered monoids*, Ph. D. Thesis, Palacký Univ., Olomouc 2003.
- [13] LUXEMBURG, W. A. J., ZAAANEN, A. C.: *Riesz spaces. Vol. I*. North Holland Publ., Amsterdam-London 1971
- [14] SWAMY K. L. N.: *Dually Residuated Lattice ordered Semigroups*, Math. Annalen 159, 1965, 105-114.
- [15] SWAMY K. L. N.: *Dually Residuated Lattice ordered Semigroups, II* Math. Annalen 160, 1965, 64-71.
- [16] SWAMY K. L. N.: *Dually Residuated Lattice ordered Semigroups, III* Math. Annalen 167, 1966, 71-74.

Current address

Milan Jasem, PhD.

Institute of Information Engineering, Automation and Mathematics
Faculty of Chemical and Food Technology, Slovak Technical University
Radlinského 9, 812 37 Bratislava, Slovak Republic
E-mail: milan.jasem@stuba.sk

ON THE HYPER-WIENER INDEX AND POLYNOMIAL OF A GRAPH

SEIBERT Jaroslav, (CZ)

Abstract. There exist more than thousand topological indices and various polynomials which characterize the structure of graphs. Most of them have some applications in chemistry, physics, biochemistry, computer and communication sciences. The Wiener and hyper – Wiener indices are the most studied topological indices, both for algebraic aspects and rich applications. In this contribution the hyper – Wiener index and polynomial of some specific classes of graphs are found. Further, one of the most interesting extension of the Wiener index introduced by Eliasi and Taeri is shown.

Key words and phrases. Simple graph, distance matrix, hyper-Wiener index, hyper-Wiener polynomial, extension of Wiener index

Mathematics Subject Classification: Primary 05C12, 05C85; Secondary 05C50

1 Introduction

The advantage of using the graph theory in chemical studies lies in the possibility to apply directly its mathematical apparatus and proof techniques. A given problem may be considered on a higher level of abstraction which enables a relatively simple insight into the structural features of the molecule. The obtained graph-theoretical results have a general validity and may be formulated as theorems and rules which can then be applied to any similar group of molecules without any further numerical or conceptual work.

Various topological indices are real numbers related to a molecular graph. They must be structural invariants which do not depend on the labeling of the concrete representation of a graph. Some of these topological indices have found applications as means to model chemical, pharmaceutical and other properties of molecules.

roughout this contribution we will use connected graphs without loops and multiple edges. Several invariants were derived from the distance matrix of a graph.

Definition 1. Let $G = (V, E)$ be a graph with the vertex set $V = \{v_1, \dots, v_n\}$. Then the distance matrix of G is defined as the $n \times n$ matrix $D(G) = (d_{ij})$, where $d_{ij} = 0$ for $i = j$ and d_{ij} is the distance between vertices v_i and v_j for $i \neq j$.

Definition 2. The Wiener index $W(G)$ of a graph G is the sum of all entries above the main diagonal of the distance matrix, so that $W(G) = \sum_{i < j} d_{ij}$. Let $d(G, k)$ be the number of pairs of vertices in G that are distance k apart, and $d(G, 0) = n$. The graph polynomial defined as $H(G, x) = \sum_{k=0}^N d(G, k)x^k$, where N is the diameter of the graph G , is called the Wiener polynomial.

The Wiener polynomial was introduced by Hosoya in 1988. It is easy to see that the first derivative of $H(G, x)$ at $x = 1$ equals the Wiener index.

We let P_n, C_n, K_n and L_n denote the path, circuit, complete graph and wheel on n vertices, $K_{r,s}$ the complete bipartite graph on parts of size r, s . The Wiener polynomials for these types of graphs are given e. g. in [6, Theorem 1.2.].

Theorem 1. (Theorem 1.3 in [6], Theorem 3 in [7]).

The following relations hold for the Wiener index of the above-mentioned types of graphs:

$$\begin{aligned} W(P_n) &= \binom{n+1}{3}, \\ W(C_n) &= \frac{1}{8}n^3 \text{ for } n \text{ even, } W(C_n) = \frac{3}{4}\binom{n+1}{3} \text{ for } n \text{ odd,} \\ W(K_n) &= \binom{n}{2}, \\ W(K_{r,s}) &= 2\binom{r}{2} + rs + 2\binom{s}{2}, \\ W(L_n) &= 2\binom{n-1}{2}. \end{aligned}$$

In the case of large graphs it is practical to use some of „decomposition“ rules to compute the Wiener index of such graphs. Two statements were proved in [7].

Theorem 2. ([7], Theorem 1]

Let G be a graph having a bridge incident with the vertices v_p, v_{p+1} . The deletion of this bridge disconnects G into two connected subgraphs G_p and G_q , $p \geq 1, q \geq 1$, with the sets of vertices $\{v_1, \dots, v_p\}$ and $\{v_{p+1}, \dots, v_{p+q}\}$, respectively. Then

$$W(G) = W(G_p) + W(G_q) + pq + q \sum_{i=1}^p d_{ip} + p \sum_{j=p+1}^{p+q} d_{p+1j}.$$

A similar statement was also proved for using an articulation vertex instead of a bridge.

2 The hyper – Wiener index

Many authors have tried to generalize the Wiener index and the Wiener polynomial of a graph. Randić in 1993 introduced an extension of the Wiener index for trees, and this has come to be known as the hyper–Wiener index. In 1995 Klein, Lucovits and Gutman generalized this extension to cyclic structures (see more in [3]). Then the hyper–Wiener index $W_h(G)$ of a graph G

is expressed in the form $W_h(G) = \frac{1}{2}W(G) + \frac{1}{2}\sum_{i<j} d_{ij}^2$ or equivalently

$$W_h(G) = \frac{1}{2}\sum_{i<j} (d_{ij} + d_{ij}^2) = \sum_{i<j} \frac{d_{ij}(d_{ij} + 1)}{2} = \sum_{i<j} \binom{d_{ij} + 1}{2}.$$

It is easy to see that the following relations between $W_h(G)$ and polynomial $H(G, x)$ hold.

If we denote $\frac{d^2}{dx^2}[xH(G, x)] = f(x)$ and $\frac{d}{dx}H(G, x) + \frac{1}{2}\frac{d^2}{dx^2}H(G, x) = g(x)$ then $W_h(G) = f(1) = g(1)$. All attempts to extend the Wiener polynomial were done with respect to the previous relations.

In [2] Cash has defined the hyper-Wiener polynomial by $H_h(G, x) = \sum_{k=0}^N \frac{k+1}{2} d(G, k) x^k$.

The value of its derivative at $x = 1$ is $W_h(G)$ as

$$\frac{d}{dx}H_h(G, x) = H'_h(G, x) = \sum_{k=1}^N \frac{k+1}{2} d(G, k) k x^{k-1}.$$

Then $H'_h(G, 1) = \sum_{k=1}^N \binom{k+1}{2} d(G, k)$ which is equivalent to the sum $\sum_{i<j} \binom{d_{ij} + 1}{2} = W_h(G)$.

Now we express the hyper–Wiener polynomials and indices for the path P_n , complete graph K_n , circuit C_n and wheel L_n on n vertices and the complete bipartite graph $K_{r,s}$.

Theorem 3. The following statements hold for the hyper–Wiener indices.

1. $W_h(P_n) = \binom{n+2}{4},$
2. $W_h(K_n) = \binom{n}{2},$
3. $W_h(C_n) = n \binom{\frac{n}{2} + 2}{3} - \binom{\frac{n}{2} + 1}{2}$ if $n \geq 4$ is even,

$$W_h(C_n) = n \binom{\frac{n+3}{2}}{2} \quad \text{if } n \geq 3 \text{ is odd,}$$

$$4. W_h(L_n) = 2(n-1) + \frac{3}{2}(n-1)(n-4) \quad \text{for } n \geq 4,$$

$$5. W_h(K_{r,s}) = rs + 3 \left[\binom{r}{2} + \binom{s}{2} \right].$$

Proof. The formulas can be proved by direct computation of the hyper-Wiener index. We will do it only for a circuit C_n using the basic relation $W_h(G) = \sum_{i < j} \binom{d_{ij} + 1}{2}$.

First, let $n \geq 4$ be an even integer. Then $d_{ij} = 1$ for n pairs of vertices (v_i, v_j) , $d_{ij} = 2$ also for n pairs of vertices and so on, but $d_{ij} = \frac{n}{2}$ for $\frac{n}{2}$ pairs of vertices.

It means that

$$W_h(C_n) = n \binom{2}{2} + n \binom{3}{2} + \dots + n \binom{\frac{n}{2}}{2} + \frac{n}{2} \binom{\frac{n}{2} + 1}{2} = n \sum_{j=2}^{\frac{n}{2}+1} \binom{j}{2} - \frac{n}{2} \binom{\frac{n}{2} + 1}{2} = n \binom{\frac{n}{2} + 2}{3} - \frac{n}{2} \binom{\frac{n}{2} + 1}{2}$$

using the well-known combinatorial identity $\sum_{i=a}^b \binom{i}{a} = \binom{b+1}{a+1}$, where $b \geq a$.

Now let $n \geq 3$ be an odd integer. Then we have by the same reasons

$$W_h(C_n) = n \binom{2}{2} + n \binom{3}{2} + \dots + n \binom{\frac{n+1}{2}}{2} = n \sum_{j=2}^{\frac{n+1}{2}} \binom{j}{2} = n \binom{\frac{n+3}{2}}{2}.$$

Theorem 4. The following relations hold for the hyper-Wiener polynomials of the mentioned types of graphs.

$$1. H_h(P_n, x) = \sum_{k=0}^{n-1} \frac{k+1}{2} (n-k) x^k,$$

$$2. H_h(K_n, x) = \frac{1}{2}n + \binom{n}{2}x,$$

$$3. H_h(C_n, x) = \frac{1}{2}n + n \sum_{k=1}^{\frac{n-1}{2}} \frac{k+1}{2} x^k + \frac{1}{4}n \left(\frac{n}{2} + 1\right) x^{\frac{n}{2}} \quad \text{if } n \geq 4 \text{ is even,}$$

$$H_h(C_n, x) = \frac{1}{2}n + n \sum_{k=1}^{\frac{n-1}{2}} \frac{k+1}{2} x^k \quad \text{if } n \geq 3 \text{ is odd,}$$

$$4. H_h(L_n, x) = \frac{1}{2}n + 2(n-1)x + \frac{3}{4}(n-1)(n-4)x^2 \quad \text{for } n \geq 4,$$

$$5. H_h(K_{r,s}, x) = \frac{1}{2}(r+s) + rsx + \frac{3}{2} \left[\binom{r}{2} + \binom{s}{2} \right] x^2.$$

Proof. These polynomials are derived from the definition given by Cash.

A direct calculation of the index $W_h(G)$ and the polynomial $H_h(G, x)$ is possible only for several classes of graphs or for graphs on a small number of vertices. In the case of large graphs it is necessary to use a computer program. Such program can be found in [8].

3 Other extension of the Wiener index

In this section we will comment on one of the most interesting way for a generalization of the Wiener and hyper-Wiener indices of graphs. Eliasi and Taeri [3] introduced the notion of the y -Wiener index of graphs. They used the Gamma function as a generalization of the well-known factorial function. It is defined as $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ for any nonnegative real number x . Remember

two basic properties of this function:

(1) $\Gamma(x+1) = x\Gamma(x)$ and (2) $\Gamma(k+1) = k!$ for a nonnegative integer k .

Definition 3. ([3], Definition 1).

Let y be a positive real number. Then the y -Wiener index $W(G, y)$ of a graph G is defined as follows:

$$W(G, y) = \sum_{i < j} \frac{\Gamma(d_{ij} + y)}{y\Gamma(d_{ij})} \quad \text{or equivalently} \quad W(G, y) = \sum_{k=1}^N \frac{\Gamma(k + y)}{y\Gamma(k)} d(G, k).$$

Putting $y = 1$ in the relation for $W(G, y)$ we have

$$W(G, 1) = \sum_{i < j} \frac{\Gamma(d_{ij} + 1)}{\Gamma(d_{ij})} = \sum_{i < j} \frac{d_{ij} \Gamma(d_{ij})}{\Gamma(d_{ij})} = \sum_{i < j} d_{ij} = W(G),$$

which means that the 1-Wiener index is the Wiener index.

Putting $y = 2$ in the definition of $W(G, y)$ we have

$$W(G, 2) = \sum_{i < j} \frac{\Gamma(d_{ij} + 2)}{2\Gamma(d_{ij})} = \sum_{i < j} \frac{(d_{ij} + 1)\Gamma(d_{ij} + 1)}{2\Gamma(d_{ij})} = \sum_{i < j} \frac{(d_{ij} + 1)d_{ij}}{2} = \sum_{i < j} \binom{d_{ij} + 1}{2} = W_h(G),$$

which means that the 2-Wiener index is the hyper-Wiener index.

By a similar way Eliasi and Taeri introduced the y -Wiener polynomial $H(G, x, y)$ of a graph G .

Concretely $H(G, x, y) = \sum_{k=1}^N \frac{\Gamma(k + y)}{y\Gamma(k + 1)} d(G, k) x^k$.

We can easy see that $H(G, x, 1) = \sum_{k=1}^N \frac{\Gamma(k + 1)}{\Gamma(k + 1)} d(G, k) x^k = H(G, x) - n$

and $H(G, x, 2) = \sum_{k=1}^N \frac{\Gamma(k + 2)}{2\Gamma(k + 1)} d(G, k) x^k = \sum_{k=1}^N \frac{k + 1}{2} d(G, k) x^k = H_h(G, x) - \frac{n}{2}$.

The value of the first derivative of $H(G, x, y)$ at $x = 1$ is $W(G, y)$. The authors also obtained some mathematical properties of this new topological index.

4 Concluding remarks

Some authors, e.g. [6], have found as more practical and natural for applications in chemistry to express the Wiener polynomial in the ordered form. Then $\overline{H}(G, x) = \sum_{i,j} d_{ij} x^k$ where the sum is now over all ordered pairs (v_i, v_j) of vertices in G , including those where $v_i = v_j$. It means that $\overline{H}(G, x) = 2H(G, x) + n$. Further invariants can be obtained of this definition.

It is also possible to introduce the hyper-Wiener index for trees by another way, e.g. [5]. Let P be the unique path connecting the vertices v_i, v_j and let n_i, n_j be the number of vertices on the two sides of the path P . Then $W_h(G) = \sum_{i,j} n_i n_j$ with summation going over all pairs of vertices of a tree G .

References

- [1] ABU GHNEIM, O. A., AL-EZEH, H., AL-EZEH, M.: The Wiener Polynomial of the k-th Power Graph. *International Journal of Mathematics and Mathematical Sciences*, Hindawi Publishing Corporation, Vol. 2007, Article ID 24873, 6 pp.
- [2] CASH, G. G.: Relationship Between the Hosoya Polynomial and the Hyper-Wiener Index. *Applied Mathematical Letters*, 15 (2002), 893 – 895.
- [3] ELIASI, M., TAERI, B.: Extension of the Wiener Index and Wiener Polynomial. *Applied Mathematical Letters*, 21 (2008), 916 – 921.
- [4] GUTMAN, I.: Some Properties of the Wiener Polynomial. *Graph Theory Notes of New York* XXV, 1993, 13 – 18.
- [5] GUTMAN, I.: A New Hyper-Wiener Index. *Croatica Chemica Acta* 77 (2004), 61– 64.
- [6] SAGAN, B. E., YEH, Y. N., ZHANG, P.: The Wiener Polynomial of a Graph. *International Journal of Quantum Chemistry*, 60 (1996), 959 – 969.
- [6] SEIBERT, J., TROJOVSKY, P.: Double Invariants and Wiener Index of Graphs. *Proceedings of the 5th International Conference APLIMAT 2006*, 151 – 156.
- [7] SEIBERT, J., TROJOVSKY, P.: Wiener Polynomial and Some Topological Indices of a Graph. *Proceedings of the 6th International Conference APLIMAT 2007*, 115 – 121.

Current address

Jaroslav Seibert

Institute of Mathematics
Faculty of Economics and Administration
University of Pardubice
Czech Republic
e-mail: jaroslav.seibert@upce.cz

HEXAGONAL AND GOLDEN QUASIGROUPS

VANŽUROVÁ Alena, (CZ)

Abstract. Our aim is to investigate two subvarieties in the variety of idempotent medial quasigroups, namely hexagonal quasigroups and golden section quasigroups. For both classes, we present here a construction of special finite examples of low order, particularly those arising from an additive group of a finite field and a suitable left translation (with respect to multiplication). As useful tools, we use the concept of isotopism and a modified version of the Toyoda theorem.

Key words and phrases. Magma, quasigroup, medial law, hexagonal quasigroup, golden section quasigroup, finite field.

Mathematics Subject Classification. Primary 20N05, 05B15; Secondary 12E20.

1 Basic concepts

We consider mostly binary systems $\mathcal{Q} = (Q, \cdot)$ here, algebras of type $(\cdot, 2)$, with one binary operation on the non-empty carrier set Q , that are called *groupoids* or *magmas*. If not otherwise stated, we write the operation multiplicatively, and additive notation is reserved for groups only. Juxtaposition is preferred to composition where the product is explicitly written.

The *left* and *right translation* in (Q, \cdot) by a fixed element u are the maps $L_u, R_u: Q \rightarrow Q$, $L_u: x \mapsto ux$, respectively $R_u: x \mapsto xu$. A *dual* magma $(Q, *)$ to (Q, \cdot) has the operation related by $a * b = b \cdot a$.

Under a *pointed* magma (or quasigroup) we mean the algebraic system together with a distinguished element from the underlying set. We use the notation $(Q, \cdot; q)$ etc. In the case of a pointed group or loop, we agree to distinguish just the identity element.

1.1 Isotopism of quasigroups

Quasigroups can be characterized in two ways. Either equationally, as algebras with three binary operation $\cdot, /$ and \backslash (called multiplication, right and left division) which are related by

the identities [9], [7]

$$xy/y = y \backslash yx = (x/y)y = y(y \backslash x) = x,$$

or as magmas in which the equations $ay = b$, $xa = b$ admit unique solution, denoted $a \backslash b$ or b/a , respectively, for which the above identities can be verified; left and right translations are permutations of Q . In the finite case, both approaches are interchangeable. For the sake of simplicity, we follow the second view-point here.

A *loop* is a quasigroup with two-sided identity element. Recall that a *group* can be introduced as a quasigroup which is at the same time a semigroup, i.e. is associative, [3]. Denote by $\text{Aut}(\mathcal{G})$ the automorphism group of the given group \mathcal{G} .

Under an *isotopy* of a magma (Q, \cdot) onto (Q', \cdot') we mean a triplet of bijections

$$(\alpha, \beta, \gamma): Q \rightarrow Q'$$

such that $\alpha(x) \cdot' \beta(y) = \gamma(x \cdot y)$ for all $x, y \in Q$, or equivalently, $x' \cdot' y' = \gamma(\alpha^{-1}(x') \cdot \beta^{-1}(y'))$ for all $x', y' \in Q'$, [9], [8]. In the case $\alpha = \beta = \gamma$, we speak about *isomorphism*.

The class of isotopisms coincides with the class of isomorphisms for groups, hence for groups, isotopism plays no role whatever. But for quasigroups and loops, isotopism plays the central role; isomorphism is too restrictive since too strong. We speak about the *principal isotopy* of a quasigroup if $\gamma = \text{id}$. The *LP-isotopy* of a quasigroup is a principal isotopy onto a loop.

Under a *pointed* magma or quasigroup we mean the algebraic system together with a distinguished element from the underlying set. We use the notation $(Q, \cdot; q)$ etc. In the case of a pointed group, we agree to distinguish just the identity element.

1.2 Idempotent medial quasigroups

In the term algebra of type $(\cdot, 2)$, consider the unary term $j(x) = x \cdot x = x^2$, the ternary term $t(x, y, z) = xy \cdot z$ and the quaternary terms $m(x, y, z, u) = xy \cdot zu$, $p(x, y, z, u) = x(yz \cdot u)$.

We are interested in magmas $\mathcal{Q} = (Q, \cdot)$ which are reducts of quasigroups and in which the identity $j^{\mathcal{Q}}(x) = x$ holds together with some of the identities $m^{\mathcal{Q}}(x, y, z, u) = m^{\mathcal{Q}}(x, z, y, u)$, $p^{\mathcal{Q}}(x, y, z, u) = p^{\mathcal{Q}}(y, x, z, u)$, or $t^{\mathcal{Q}}(x, t(x, y, z), z) = y$.

An element $q \in Q$ is *idempotent* in the magma (Q, \cdot) if $q^2 = q$ holds. If the identity $j^{\mathcal{Q}}(x) = x$ holds, i.e. all elements are idempotent, the magma itself is called *idempotent*. If the identity

$$xy \cdot zu = xz \cdot yu \tag{1}$$

is satisfied, i.e. if $m^{\mathcal{Q}}(x, y, z, u) = m^{\mathcal{Q}}(x, z, y, u)$ holds for the term function, the magma \mathcal{Q} is called *medial* [7], *entropic* [11], [12], also *abelian*, [13], [14]. The identity $p^{\mathcal{Q}}(x, y, z, u) = p^{\mathcal{Q}}(y, x, z, u)$ will be called the *left hexagonal identity*; explicitly,

$$x(yz \cdot u) = y(xz \cdot u). \tag{2}$$

The identity $t^{\mathcal{Q}}(x, t(x, y, z), z) = y$ is the *left GS-identity*; explicitly,

$$(x \cdot (xy \cdot z)) \cdot z = y. \tag{3}$$

Mediality together with idempotency imply elasticity since $xy \cdot xx = xx \cdot yx$, and also left and right distributivity:

Lemma 1.1 *Any idempotent and medial quasigroup is distributive.*

Proof. Indeed, $xy \cdot xz = xx \cdot yz = x(yz)$ and $xz \cdot yz = xy \cdot zz = (xy)z$.

2 Toyoda-like Theorems

It is known for some time that idempotent medial quasigroups are in fact isotopes of abelian groups. In what follows we make use of the following results, [13], [16], [8]:

Lemma 2.1 *Let $\mathcal{G} = (Q, +; e)$ be a commutative group with the identity element e and let $\alpha, \beta \in \text{Aut}(\mathcal{G})$ be commuting automorphisms of the group. Introduce the operation*

$$x \cdot y = \alpha(x) + \beta(y) \quad \text{for all } x, y \in Q. \quad (4)$$

Then $\mathcal{Q} = (Q, \cdot)$ is a medial quasigroup with the idempotent element e , and the given mappings can be interpreted as translations of the quasigroup, namely, $\alpha = R_e$ and $\beta = L_e$.

Lemma 2.2 *If $\mathcal{Q} = (Q, \cdot; e)$ is a pointed medial quasigroup then the binary operation $+_e$ defined on Q by*

$$x +_e y = R_e^{-1}(x) \cdot L_e^{-1}(y) = (x/e) \cdot (e \backslash y) \quad \text{for } x, y \in Q \quad (5)$$

is associative and commutative. Moreover, if $e \in Q$ is an idempotent element of (Q, \cdot) then $(Q, +_e; e)$ is a commutative group with the identity element e , and the translations R_e and L_e are commuting automorphisms of the group \mathcal{G} , [13], [16].

Proof. In [13], [16], it is checked that $(Q, +_e)$ is a commutative semigroup. Verify that $(Q, +_e)$ is a quasigroup: (R_e, L_e, id_Q) is a principal isotopy of (Q, \cdot) onto $(Q, +_e)$ since we can write (5) equivalently as

$$xy = R_e(x) +_e L_e(y) = xe +_e ey \quad \text{for all } x, y \in Q.$$

Let $e \in Q$ be idempotent. Then for all $x \in Q$, $xe = xe +_e ee = xe +_e e$ holds, and $ex = ee +_e ex = e +_e ex$. Hence e is the identity element, and $(Q, +_e)$ is an LP -isotope of \mathcal{Q} .

Together, we have

Theorem 2.3 *A pointed magma $(Q, \cdot; e)$ is a medial quasigroup with idempotent element e if and only if there is a commutative group $\mathcal{G} = (Q, +; e)$ and a pair of commuting automorphisms $\alpha, \beta \in \text{Aut}(\mathcal{G})$ such that*

$$x \cdot y = \alpha(x) + \beta(y).$$

If this is the case then $\alpha = R_e$ and $\beta = L_e$.

We say that a magma (Q, \cdot) is *linear over the commutative group $\mathcal{G} = (Q, +; e)$ under the automorphism φ* when $\varphi \in \text{Aut}(\mathcal{G})$ and the binary dot operation on Q is related to the group operation by

$$x \cdot y = x + \varphi(y - x) = (\text{id}_Q - \varphi)(x) + \varphi(y) \quad \text{for } x, y \in Q. \quad (6)$$

Theorem 2.4 *Let $\mathcal{G} = (Q, +; e)$ be a commutative group, $\varphi \in \text{Aut}(\mathcal{G})$ a non-trivial automorphism of \mathcal{G} . Let $\mathcal{Q} = (Q, \cdot)$ be a magma linear over \mathcal{G} with the automorphism φ . Then \mathcal{Q} is an idempotent medial quasigroup and $\varphi = L_e$.*

Proof. We can check that $\psi = \text{id}_Q - \varphi$ is also the group automorphism: $\psi(x + y) = x + y - \varphi(x + y) = (x - \varphi(x)) + (y - \varphi(y)) = \psi(x) + \psi(y)$. Further, $\psi\varphi(x) = \varphi(x) - \varphi^2(x) = \varphi\psi(x)$, hence the automorphisms commute. We use the Theorem 2.3 and set $\alpha = R_e = \text{id}_Q - \varphi$, $\beta = L_e = \varphi$ to finish the proof.

Theorem 2.5 *If (Q, \cdot) is a non-trivial ($\text{card}(Q) \geq 1$) idempotent medial quasigroup, then for any fixed element $e \in Q$ there is a commutative group $\mathcal{G} = (Q, +; e)$ satisfying $xy = xe + ey$ such that the translations R_e, L_e are commuting automorphisms of the group \mathcal{G} and for all $x \in Q$, $x + e \cdot x = x$ holds. Moreover, (Q, \cdot) is linear over \mathcal{G} with the automorphism $\varphi = L_e$. The groups $(Q, +_e; e)$ are pairwise isomorphic for all possible choices of element $e \in Q$.*

Proof. Let (Q, \cdot) satisfy the assumptions and $e \in Q$ be a fixed element. First check that $R_e, L_e \in \text{Aut}(\mathcal{G})$. We have $R_e(x) +_e R_e(y) = xe +_e ye = (x +_e y) \cdot (e +_e e) = (x +_e y) \cdot e = R_e(x +_e y)$, analogously for L_e . We get $x = xx = xe +_e ex$ and $xy = xe +_e ey = xe +_e ex -_e ex +_e ey = x +_e L_e(y -_e x)$ for all $x, y \in Q$, where $L_e \neq \text{id}_Q$. Hence (Q, \cdot) is linear over \mathcal{G} with the automorphism $\varphi = L_e$. For any $x \in Q$, $R_e L_e(x) = ex \cdot ee = ee \cdot xe = L_e R_e(x)$, hence R_e and L_e commute. Since each of the groups $(Q, +_e)$, $e \in Q$, is isotopic to (Q, \cdot) they are necessarily pairwise isotopic, and therefore isomorphic.

Note that the assumption $L_e = \text{id}_Q$ would imply $ex = x = xx$, $x = e$, hence $Q = \{e\}$ would be trivial.

3 Hexagonal quasigroups

The class of hexagonal quasigroups, introduced by V. Volenec in [22], is an interesting subclass in the variety of idempotent medial quasigroups.

According to [22], a quasigroup (Q, \cdot) is called *hexagonal* if it satisfies the “hexagonal” identity

$$x(yz \cdot ww) = y(xz \cdot w). \quad (7)$$

Idempotency can be checked by taking $x = y$ in (7) and using left cancellation. In what follows we show that hexagonal quasigroups are, among others, elastic, medial, (left and right) distributive, and can be equivalently characterized also in another way, by idempotency, mediality, elasticity and “semisymmetry”. A quasigroup is called *semisymmetric* in [22] if it satisfies the identities $(xy)x = y$ and $x(yx) = y$. If this is the case then the pair of quasi-identities $xy = z \Leftrightarrow x = yz$ hold, [22, p. 113].

Theorem 3.1 *In a quasigroup $\mathcal{Q} = (Q, \cdot)$, the following conditions are equivalent:*

- (i) \mathcal{Q} is hexagonal;
- (ii) \mathcal{Q} is idempotent and satisfies left hexagonality (2), i.e. $x(yz \cdot u) = y(xz \cdot u)$;
- (iii) \mathcal{Q} is idempotent and satisfies

$$(x \cdot yz)w = (x \cdot yw)z \quad (\text{“right hexagonality”}). \quad (8)$$

Proof. We can easily see that (7) and (2) are equivalent provided idempotency holds; mirror arguments prove (7) \Leftrightarrow (8).

Lemma 3.2 *In an idempotent quasigroup \mathcal{Q} , the right hexagonality implies the identity*

$$(xy)x = y, \quad (9)$$

and left hexagonality implies the dual identity

$$x(yx) = y. \quad (10)$$

Proof. If we put $x = y = w$ in (8) we get $(x \cdot xz)x = (x \cdot xx)z$, which is equivalent with

$$(x \cdot xz)x = xz \quad (11)$$

provided idempotency holds. If $a, b \in Q$ are arbitrary fixed elements and g is the unique solution of the equation $ay = b$ in Q (i.e. $g = a \backslash b$) then $(ab)a = (a \cdot ag)a = ag = b$. Hence (9) holds. The proof of the second part is mirror.

As a consequence, a hexagonal quasigroup is elastic, i.e. it satisfies:

$$(xy)x = x(yx). \quad (12)$$

Lemma 3.3 *If the identity $(xy)x = y$ is valid in a quasigroup $\mathcal{Q} = (Q, \cdot)$, then the following two identities are equivalent in \mathcal{Q} :*

(i) *left hexagonality,*

(ii) *mediality.*

Proof. Let $x, y, z, w \in Q$ and let the mediality is satisfied. Let us check left hexagonality. Denote $u = x(yz \cdot w)$, $v = y(xz \cdot w)$. By $(xy)x = y$,

$$ux = (x(yz \cdot w))x = yz \cdot w, \quad ux \cdot yz = (yz \cdot w) \cdot yz = w,$$

and similarly

$$vy = (y(xz \cdot w))y = xz \cdot w, \quad vy \cdot xz = (xz \cdot w) \cdot xz = w.$$

It follows $ux \cdot yz = vy \cdot xz$. By mediality, $ux \cdot yz = uy \cdot xz$. Comparing right hand sides and using cancellation we get $u = v$, hence (2) is satisfied. Vice versa, if (2) holds then $u = v$, and the mediality follows.

The mirror proof checks the following.

Lemma 3.4 *If $x(yx) = y$ holds in \mathcal{Q} then the right hexagonality is equivalent with mediality.*

4 Golden quasigroups

A magma (Q, \cdot) is said to be a *golden magma* if it is idempotent, satisfies (3) and the following condition (“right golden ratio identity”):

$$x \cdot ((x \cdot yz) \cdot z) = y. \quad (13)$$

Obviously, (Q, \cdot) satisfies (3) (or (13), respectively), if and only if its dual magma satisfies the “dual” identity. In a cancellative magma $\mathcal{Q} = (Q, \cdot)$, (3) and (13) are equivalent, [21]. In fact, plugging yz for y , by (3) we get $(x \cdot ((x \cdot yz) \cdot z)) \cdot z = yz$. Accordingly, (13) follows by right cancellation. Similarly for the converse implication.

Note that at least theoretically, the identity (3) might have some application in cryptography when both parties use the same key.

Lemma 4.1 *Any magma satisfying both (3) and (13) is a medial quasigroup.*

Proof. Indeed, (3) guarantees solvability in \mathcal{Q} of the equations of the form $xc = b$ for $b, c \in Q$. To verify uniqueness, let $ax_1 = ax_2$. Then, according to (3), $x_1 = (a(ax_1 \cdot a)) \cdot a = (a(ax_2 \cdot a)) \cdot a = x_2$. Hence $x_1 = x_2$, and \mathcal{Q} is left cancellative. Similarly, solvability of the equations of the form $cy = b$ and right cancellation are consequences of (13). Let us check the mediality: $ac \cdot (ab \cdot cd)d = a(ab \cdot (ab \cdot cd)d) \cdot (ab \cdot cd)d = b = ac \cdot (ac \cdot bd)d$ holds for arbitrary elements, and it is sufficient to use cancellation again.

It follows from the discussion above that we can adopt the following definition of a *golden quasigroup*, or a *GS-quasigroup*, as an idempotent quasigroup that satisfies (3), or equivalently, (13), [8], [19]. Originally, the term golden section, instead of more correct *golden ratio*, was used, which explains the abbreviation.

5 Specialization of Toyoda theorem

Now let us show the consequences of the hexagonal identity, or the golden ratio identity, respectively.

Theorem 5.1 *For any GS-quasigroup $\mathcal{Q} = (Q, \cdot)$ there is a commutative group $\mathcal{G} = (Q, +; e)$ and there is an automorphism $\varphi \in \text{Aut}(\mathcal{G})$ such that (Q, \cdot) is linear over the group \mathcal{G} with the non-trivial automorphism φ , i.e. (6) holds, and the following is satisfied:*

$$\varphi^2 - \varphi - \text{id}_Q = \epsilon \quad (14)$$

where $\epsilon : Q \rightarrow Q$ is the constant mapping $\epsilon(a) = e$ for any $a \in Q$.

Proof. Consider a GS-quasigroup (Q, \cdot) ; it is idempotent and medial, hence by Theorem 2.4, there is a commutative group $\mathcal{G} = (Q, +; e)$ and $\varphi \in \text{Aut}(\mathcal{G})$ such that $ab = a + \varphi(b - a)$ for all $a, b \in Q$. Step by step, we get $ab \cdot c = a + \varphi(b - a) + \varphi(c - a) - \varphi^2(b - a)$, similarly $a(ab \cdot c) = a + \varphi^2(b - a) + \varphi^2(c - a) - \varphi^3(b - a)$, and $a(ab \cdot c) \cdot c = a + \varphi(c - a) + \varphi^2(b - a) + \varphi^2(c - a) - 2\varphi^3(b - a) - \varphi^3(c - a) + \varphi^4(b - a)$. According to (13) we calculate $e = -(b - a) + \varphi(c - a) + \varphi^2((b - a) + (c - a)) - \varphi^3(2(b - a) + (c - a)) + \varphi^4(b - a)$. Let us set

$x = b - a$, $y = c - a$. We obtain $\varphi^4(x) - \varphi^3(2x + y) + \varphi^2(x + y) + \varphi(y) - x = e$ for any $x, y \in Q$. Particularly for $x = e$, it follows $-\varphi^3(y) + \varphi^2(y) + \varphi(y) = e$. Setting $z := \varphi(y)$ we get $z + \varphi(z) - \varphi^2(z) = e$, and this equality holds for all $z \in Q$ as φ is bijective. Hence (14), i.e. $\varphi^2 - \varphi - \text{id}_Q = \varepsilon$, holds.

Theorem 5.2 *For any hexagonal quasigroup $\mathcal{Q} = (Q, \cdot)$ there is a commutative group $\mathcal{G} = (Q, +; e)$ and an automorphism $\varphi \in \text{Aut}(\mathcal{G})$ such that (Q, \cdot) is linear over \mathcal{G} with the (non-trivial) automorphism φ , and the following is satisfied:*

$$\varphi^2 - \varphi + \text{id}_Q = \epsilon. \quad (15)$$

Proof. We already know by the Theorem 2.4 that the operation takes the form $ab = a + \varphi(b - a)$ for some automorphism φ of a commutative group $\mathcal{G} = (Q, +; e)$. By $y = xy \cdot x$, we get $b = ab \cdot a = a + \varphi(b - a) - \varphi^2(b - a)$, hence $\varphi^2(b - a) - \varphi(b - a) + (b - a) = e$. Any element c from Q can be written as $c = b - a$ for convenient $a, b \in Q$. Hence (15), i.e. $\varphi^2 - \varphi + \text{id}_Q = e$, is satisfied.

To prove the converse is easier:

Theorem 5.3 *Let $\mathcal{Q} = (Q, \cdot)$ be a magma linear over a commutative group $\mathcal{G} = (Q, +; e)$ with an automorphism $\varphi \in \text{Aut}(\mathcal{G})$.*

(i) *If φ satisfies $\varphi^2 - \varphi + \text{id}_Q = \epsilon$ then \mathcal{Q} is a hexagonal quasigroup.*

(ii) *If φ satisfies $\varphi^2 - \varphi - \text{id}_Q = \epsilon$ then \mathcal{Q} is a GS-quasigroup.*

Proof. According to Theorem 2.3, \mathcal{Q} is an idempotent and medial quasigroup. If (14) holds then $ab \cdot c = 2a - b + \varphi(c - a)$, $a(ab \cdot c) = c + \varphi(c - b)$, and $a(ab \cdot c) \cdot c = c + \varphi(c - b) - \varphi^2(c - b) = b$. If (15) holds then $ab \cdot a = a + \varphi(b - a) - \varphi^2(a - b) = b$ for all $a, b \in Q$, $a \cdot ba = a + \varphi(b - a) - \varphi^2(a - b) = b$ holds, and we use Lemma 3.3.

6 Examples

The following examples explain a motivation and justify the introduced geometrical terminology, [21]:

Example 6.1 Let $(F, +, \cdot)$ be a field in which the equation

$$x^2 - x - 1 = 0, \quad \text{or} \quad x^2 - x + 1 = 0 \quad (16)$$

has a root $q \in F$, and define on F a binary operation

$$a \circ b = (1 - q)a + qb = a + q(b - a) \quad a, b \in F. \quad (17)$$

Then the left translation $\varphi(x) = qx$, $\varphi = L_q: F \rightarrow F$, which will be called here the *standard dilatation by q* , is an additive automorphism of the group $(F, +; 0)$ of the field, $\varphi \in \text{Aut}(F, +)$, and we can write the equality (17) as $a \circ b = a + \varphi(b - a)$. Hence (F, \circ) is linear over $(F, +; 0)$ with the automorphism φ ,

$$\varphi^2 - \varphi - \text{id}_Q = \epsilon \quad (18)$$

holds in the first case, and

$$\varphi^2 - \varphi + \text{id}_Q = \epsilon \quad (19)$$

is satisfied in the second case (since q is a root of the corresponding equation). The Theorem 5.3 guarantees that (F, \circ) is a hexagonal quasigroup, or a GS -quasigroup, respectively.

Example 6.2 Consider the field of complex numbers $(\mathbb{C}, +, \cdot)$; the elements are represented by points in the Gaussian plane. Define an operation (17) with $q = \exp(i\pi/3)$. Then (\mathbb{C}, \circ) is a hexagonal quasigroup. The points $0, 1, q$ are vertices of a positively oriented equilateral triangle, and $a, b, a \circ b$ are vertices of a similar triangle since we have

$$(a \circ b - a) : (b - a) = (q - 0) : (q - 0).$$

Hence $a \circ b$ can be considered as the centre of a (positively oriented) regular hexagon with vertices a, b .

Example 6.3 In \mathbb{C} , take q as one of the roots of the equation $x^2 - x - 1 = 0$, $q = \frac{1}{2}(1 \pm \sqrt{5})$. The corresponding magma (\mathbb{C}, \circ) is a GS -quasigroup. In the Gaussian plane, the equality (17) with $a \neq b$ may be written as $\frac{a \circ b - a}{b - a} = q$, which means that the point $a \circ b$ divides the segment between a and b in the ratio q . If $q = \frac{1}{2}(1 - \sqrt{5})$ then the golden ratio of the pair $(b, a \circ b)$ is a . If $q = \frac{1}{2}(1 + \sqrt{5})$ then the golden ratio of the pair $(a, a \circ b)$ is b .

Example 6.4 Over $GF(3)$, we have a 3-element hexagonal quasigroup with the operation \circ_2 which arises from left multiplication by the element 2:

\circ_2	0	1	2
0	0	2	1
1	2	1	0
2	1	0	2

Example 6.5 Over $GF(4)$, we get two hexagonal quasigroups which are also GS -quasigroups, are dual to each other, and isomorphic; one of them is given by the operation \circ given in the Table below.

Example 6.6 Over $GF(5)$, there is a GS -quasigroup with star operation $*$ which is self-dual, i.e. commutative.

\circ	0	1	c	d
0	0	d	1	c
1	c	1	d	0
c	d	0	c	1
d	1	c	0	d

$*$	0	1	2	3	4
0	0	3	1	4	2
1	3	1	4	2	0
2	1	4	2	0	3
3	4	2	0	3	1
4	2	0	3	1	4

By a computer search, using Maple, we succeeded to discuss existence and in the affirmative case, constructed examples whith orders up to 1000.

Acknowledgement

The paper was supported by the grant from Grant Agency of Czech Republic GAČR no. P201/11/0356 with the title: "Riemannian, pseudo-Riemannian and affine differential geometry" and by the project of specific university research of the Brno University of Technology, No. FAST-S-11-47.

References

- [1] BARTOŠKOVÁ, Z.: *Medial Quasigroups and Geometry*. Mgr. Thesis, Palacký University Olomouc, Fac. Sci., 2011 (in Czech).
- [2] BARTOŠKOVÁ, Z.: *Commutative Groups and Medial Quasigroups*. Bc. Thesis, Palacký University Olomouc, Fac. Sci., 2009 (in Czech).
- [3] BELOUSOV, V.D.: *Foundations of the theory of quasigroups and loops*. Moscow, Nauka, 1967.
- [4] BRUCK, R.H.: *Some results in the theory of quasigroups*. Trans. Amer. Math. Soc., Vol. 55, pp. 19-52, 1944.
- [5] RUCK, R.H.: *A Survey of Binary Systems*. Berlin, Springer, 1958.
- [6] CHEIN, O., PFLUGFELDER, H.O, SMITH, J.D.H. (eds.): *Quasigroups and Loops: Theory and Applications*. Heldermann Verlag, Berlin, 1990.
- [7] JEŽEK, J., KEPKA, T.: *Medial quasigroups*. Academia Praha, 1983.
- [8] HAVEL, V.J., VANŽUROVÁ, A.: *Foundations of the theory of quasigroups and loops*. Palacký University in Olomouc, Olomouc, 2006.
- [9] PFLUGFELDER, H.O.: *Quasigroups and Loops, Introduction*. Heldermann Verlag, Berlin, 1990.
- [10] PICKERT, G.: *Projektive Ebenen*. Springer, Berlin, Heidelberg, New York, 1975.
- [11] ROMANOVSKA, A., SMITH, J.D.H.: *Modal Theory, An Algebraic Approach to Order, Geometry, and Convexity*. Berlin, Heldermann Verlag, 1985.
- [12] ROMANOVSKA, A., SMITH, J.D.H.: *Modes*. World Scientific, New Jersey, London, Singapore, Hong Kong, 2002.
- [13] TOYODA, K.: *On axioms of mean transformations and automorphic transformations of abelian groups*. Tôhoku Math. J. Vol. 46, pp. 239-251, 1940.
- [14] TOYODA, K.: *On affine geometry of abelian groups*. Proc. Imp. Acad. Tokyo Vol. 16, No. 5, pp. 161-164, 1940.
- [15] VANŽUROVÁ, A.: *Medial Quasigroups of Type (n, k)* . Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica 49, pp. 107-122, 2010.
- [16] VANŽUROVÁ, A.: *Golden Section Quasigroups, Finite Examples*. Proc. 10th International Conf. APLIMAT 2011, Part IV, pp. 183-190, 2011.
- [17] VANŽUROVÁ, A.: *Algebraic systems in cryptology and the security of information*. In International Conference in Military Technology Proceeding, Section 12 Security technology (ed. Z. Vránová and V. Plátěnka), ICMT'11, Brno : University of Defence, pp. 1251-1256, 2011. ISBN 978-80-7231-787-5 (ISBN 978-80-7231-788-2 for CD).

- [18] VANŽUROVÁ, A., DOLEŽALOVÁ, J.: *Hexagonal quasigroups over finite fields*. Proc. of 7th Conf.on Mathematics and Physics at Technical Universities with internat. participation, Brno, September 22, 2011, part 1 - mathematics, pp. 411-419, 2011. ISBN 978-80-7231-815-5 (ISBN 978-80-7231-818-6 for CD).
- [19] VOLENEC, V.: *Geometry of medial quasigroups*. Rad. Jugosl. Akad. Znam. Umjet 421, pp. 79-91, 1986.
- [20] VOLENEC, V., *Admissible identities in complex IM-quasigroups*. Rad. Jugosl. Akad. Znam. Umjet, Knjiga 421 Sv. 5, pp. 45-78, 1986.
- [21] VOLENEC, V.: *GS-quasigroups*. Časop. pěst. mat. 115, pp. 307-318, 1990.
- [22] VOLENEC, V.: *Hexagonal quasigroups*. Archivum Mathematicum (Brno) 27a, pp. 123-132, 1991.

Current address

VANŽUROVÁ Alena, Doc. RNDr. CSc.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University, Tř. 17. listopadu 12, 771 46 Olomouc, alena.vanzurova@upol.cz;

Department of Mathematics, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic, vanzurova.a@fce.vutbr.cz.

TOYODA'S THEOREM AND QUASIGROUP MODES

VANŽUROVÁ Alena, (CZ) BARTOŠKOVÁ Zuzana, (CZ)

Abstract. The paper is concerned with the representative principle describing medial quasigroups by means of corresponding commutative groups. We present Toyoda-like theorems: how to pass from a medial quasigroup to the corresponding commutative group, and conversely. Then we give a modification for the variety of quasigroup modes (idempotent medial quasigroups).

Key words and phrases. Quasigroup; mediality; isotopy; quasigroup mode; Toyoda theorem.

Mathematics Subject Classification. Primary 20N05, 05B15; Secondary 12E20.

1 Introduction

First we give a short survey on quasigroups, recall the concept of quasigroup isotopy which plays an important role in the quasigroup theory, and particularly we mention loop principal isotopy that is used as a useful tool in what follows. Especially, we show the form of all loop principal isotopes of a given quasigroup (Lemma 1.4). For groups, the concept of isotopism coincides with the concept of isomorphism.

1.1 Groupoids, quasigroups

A *groupoid* is an ordered pair $\mathcal{G} = (G, \cdot)$ where G is a non-empty set, the *carrier set* of \mathcal{G} , and “ \cdot ” a binary operation on G . Often, the dot for a binary operation is omitted when there is no danger of confusion.

Let (G, \cdot) be a groupoid. An element $e \in G$ is said to be *left-neutral* (*right-neutral*) if it satisfies the equality $ex = x$ ($xe = x$) for all $x \in G$. Being both left-neutral and right-neutral the element is said to be *neutral*; a groupoid cannot have two different neutral elements.

A groupoid (G, \cdot) is *left cancellative* if $(ca = cb \Rightarrow a = b)$ for $a, b, c \in G$, and *right cancellative* if $(ac = bc \Rightarrow a = b)$ for $a, b, c \in G$; *cancellative* if it is left and right cancellative. A groupoid

(G, \cdot) is a *quasigroup* if for any $a, b \in G$ there is just one $x \in G$ solving the equation $xa = b$ and denoted $x = b/a$; and just one $y \in G$ solving the equation $ay = b$ and denoted $y = a \backslash b$. Any quasigroup is cancellative.

A quasigroup with a neutral element is called a *loop*. A loop with the carrier set L and the neutral element e will be denoted $(L, \cdot; e)$.

For a groupoid (G, \cdot) we define a *left translation* by $a \in G$, $L_a: x \mapsto ax$, and a *right translation*, $R_a: x \mapsto xa$. A groupoid is a quasigroup if and only if all left translations as well as all right translations of the groupoid are permutations of the carrier set.

If (G, \cdot) is a groupoid then $a \in G$ is called an *idempotent element* if $a^2 = a$. A groupoid (G, \cdot) is *idempotent* if all its elements are idempotent. A groupoid (G, \cdot) is *commutative* if for any $a, b \in G$, the equality $ab = ba$ is satisfied. A groupoid (G, \cdot) is *medial* if the following identity is satisfied

$$(xy)(uv) = (xu)(yv); \quad (1)$$

sometimes, the law is also called surcommutative (Soublin), entropic (Etherington), alternative (Sholander), abelian (Murdoch), symmetric (Frink), or bisymmetric (Aczél). A groupoid (G, \cdot) is *associative* if it satisfies the identity $(xy)z = x(yz)$; an associative groupoid is called a *semigroup*. A groupoid (G, \cdot) is a *group* if it is an associative quasigroup. Each group has a uniquely determined neutral element 1, and to every element $x \in G$ there exists one and only one element x^{-1} , the *inverse* of x , satisfying $xx^{-1} = 1 = x^{-1}x$; the group will be denoted by $(G, \cdot; 1)$.

Given a groupoid (G, \cdot) , under a *dual groupoid* we understand the groupoid $(G, *)$ (with the same underlying set) where “ $*$ ” is the so-called *dual operation* to “ \cdot ” in G , $a * b := ba$ for $a, b \in G$.

1.2 Substructures

A groupoid (G_1, \cdot_1) is a *subgroupoid* of the groupoid (G_2, \cdot_2) if $G_1 \subseteq G_2$ and $x \cdot_1 y = x \cdot_2 y$ for all $x, y \in G_1$. A subgroupoid of an associative groupoid is associative. A nonempty set $G' \subseteq G$ is a carrier set of a subgroupoid of a groupoid (G, \cdot) if and only if G' is *closed* under the dot operation. A nonempty set $G' \subseteq G$ is a carrier set of an associative subgroupoid of an associative groupoid (G, \cdot) if and only if $(x, y \in G' \Rightarrow x \cdot y \in G')$.

Given quasigroups (G_1, \cdot_1) , (G_2, \cdot_2) , the first one is said to be a *subquasigroup* of the second one if $G_1 \subseteq G_2$ and for all $x, y \in G_1$, $x \cdot_1 y = x \cdot_2 y$, $x \backslash^1 y = x \backslash^2 y$, $x / ^1 y = x / ^2 y$. A nonempty set $G' \subseteq G$ is a carrier set of a subquasigroup of a quasigroup (G, \cdot) if and only if $(x, y \in G' \Rightarrow x \cdot y, x \backslash y, x / y \in G')$, i.e. G' is closed under all three operations $\cdot, \backslash, /$. In the preceding, if G is a finite set then the implication $(x, y \in G' \Rightarrow x \backslash y, x / y \in G')$ is a consequence of $(x, y \in G' \Rightarrow x \cdot y \in G')$.

Given two loops $(L_1, \cdot_1; e_1)$, $(L_2, \cdot_2; e_2)$ the first one is said to be a *subloop* of the second one if $L_1 \subseteq L_2$ and $x \cdot_1 y = x \cdot_2 y$ for all $x, y \in L_1$.

A nonempty set $L' \subseteq L$ is a carrier set of a subloop of a loop $(L, \cdot; e)$ if and only if $(x, y \in L' \Rightarrow x \cdot y, x \backslash y, x / y \in L')$. If L is a finite set then $(x, y \in L' \Rightarrow x \backslash y, x / y \in L')$ is a consequence of $(x, y \in L' \Rightarrow x \cdot y \in L')$.

Given groups $(G_1, \cdot_1; e_1)$, $(G_2, \cdot_2; e_2)$ the first one is a *subgroup* of the second one if $G_1 \subseteq G_2$ and $x \cdot_1 y = x \cdot_2 y$ for all $x, y \in G_1$. A nonempty set $G' \subseteq G$ is a carrier set of a subgroup

of a group $(G, \cdot; e)$ if and only if $x, y \in G' \Rightarrow e/x, x \cdot y \in G'$. If G is a finite set the condition $(x, y \in G' \Rightarrow e/x \in G')$ is a consequence of the implication $(x, y \in G' \Rightarrow x \cdot y \in G')$.

1.3 Isotopy, isomorphism

If (G, \cdot) , (G', \cdot') are quasigroups, then an *isotopy* of (G, \cdot) onto (G', \cdot') is an ordered triple (α, β, γ) of bijections of G onto G' such that $\alpha(x) \cdot' \beta(y) = \gamma(x \cdot y)$ holds for all $x, y \in G$. Thus $x \cdot y = \gamma^{-1}(\alpha(x) \cdot' \beta(y))$, and $x' \cdot' y' = \gamma(\alpha^{-1}(x') \cdot \beta^{-1}(y'))$ for all $x', y' \in G'$. If there is an isotopism of (G, \cdot) onto (G', \cdot') then we call (G', \cdot') an *isotope* of (G, \cdot) ; (G, \cdot) and (G', \cdot') are then *isotopic quasigroups*. Note that isotopism is an equivalence relation on the class of all quasigroups which defines the decomposition into *isotopy subclasses* as maximal subclasses of isotopic quasigroups.

In the case $\alpha = \beta = \gamma$, we speak on *isomorphism* of quasigroups, and we write α only. The isomorphic image of a loop (group) with neutral element 1 under an isomorphism α is a loop (group) with the neutral element $\alpha(1)$.

An isotopy (α, β, id_Q) between two quasigroups over the same carrier set Q is called *principal*. An isotopy onto a loop is called a *loop isotopy*. Especially, if (α, β, id_Q) gives a loop then we speak of a *loop principal isotopy* (briefly *LP-isotopy*, and of *LP-isotopes*.

For quasigroups, isotopisms are much more important than isomorphisms. Isotopy plays the central role in loop theory.

1.4 Isotopes of quasigroups

First let us extend informations on quasigroup isotopes by the following:

Lemma 1.1 *Let (Q, \cdot) be a quasigroup and (α, β, γ) be bijective mappings of Q onto Q' . Let $x' \cdot' y' = \gamma(\alpha^{-1}(x') \cdot \beta^{-1}(y'))$ for all $x', y' \in Q'$. Then (Q', \cdot') is a quasigroup.*

Proof. Let us examine the equation $a \cdot' y = b$ for fixed $a, b \in Q'$. Obviously, $y = \beta(\alpha^{-1}(a) \setminus \gamma^{-1}(b))$ is the unique solution of the equation under consideration. Indeed, $a \cdot' y = b \Leftrightarrow \alpha^{-1}(a) \cdot \beta^{-1}(y) = \gamma^{-1}(b) \Leftrightarrow \beta^{-1}(y) = \alpha^{-1}(a) \setminus \gamma^{-1}(b) \Leftrightarrow y = \beta(\alpha^{-1}(a) \setminus \gamma^{-1}(b))$. Similarly for the equation $x \cdot' a = b$ with $a, b \in Q'$.

Composition of isotopisms is an isotopism, too:

Lemma 1.2 *Let (α, β, γ) be an isotopy of the quasigroup (Q, \cdot) onto (Q', \cdot') , and $(\alpha', \beta', \gamma')$ an isotopy of (Q', \cdot') onto (Q'', \cdot'') . Then $(\alpha\alpha', \beta\beta', \gamma\gamma')$ is an isotopy of (Q, \cdot) onto (Q'', \cdot'') .*

Lemma 1.3 *Every isotope of a quasigroup is isomorphic to its principal isotope.*

Proof. Let (α, β, γ) be an isotopy of a quasigroup (Q, \cdot) onto a quasigroup (Q_1, \cdot_1) . Thus we have the equality $\alpha(a) \cdot_1 \beta(b) = \gamma(a \cdot b)$ for all $a, b \in Q$. Let us take the isotopy $(\gamma^{-1}\alpha, \gamma^{-1}\beta, id_Q)$ which maps (Q, \cdot) onto a quasigroup (Q, \cdot_2) such that $\gamma^{-1}\alpha(a) \cdot_2 \gamma^{-1}\beta(b) = a \cdot b$ holds for all $a, b \in Q$. Hence $\gamma^{-1}(\alpha(a) \cdot_1 \beta(b)) = a \cdot b = \gamma^{-1}\alpha(a) \cdot_2 \gamma^{-1}\beta(b)$ holds for all $a, b \in Q$. Since α, β are bijective mappings the first of the above equalities can be rewritten as $\gamma^{-1}(x \cdot_1 y) = \gamma^{-1}(x) \cdot_2 \gamma^{-1}(y)$ for all $x, y \in Q$, and we obtain an isomorphism γ^{-1} of (Q, \cdot) onto (Q_1, \cdot_1) .

Loop principal isotopes of a quasigroup can be characterized as follows:

Lemma 1.4 Any LP -isotope of a quasigroup (Q, \cdot) is just an isotope of the form (R_b, L_a, id_Q) for some $a, b \in Q$; the corresponding neutral element is ab .

Proof. Let (Q, \cdot) be a given quasigroup. All principal isotopisms of (Q, \cdot) are of the form (α, β, id_Q) where α, β are arbitrary permutations (i.e. bijections onto itself) of the carrier set Q . The corresponding isotopes will be denoted by $(Q, \cdot_{\alpha, \beta})$, thus $\alpha(a) \cdot_{\alpha, \beta} \beta(b) = a \cdot b$ for all $a, b \in Q$. Now we are interested in the case when $(Q, \cdot_{\alpha, \beta})$ is a loop (i.e. has a neutral element). The element $\alpha(a)$ is left-neutral in $(Q, \cdot_{\alpha, \beta})$ if and only if $a \cdot b = \alpha(a) \cdot_{\alpha, \beta} \beta(b) = \beta(b)$ holds for all $b \in Q$. This means that $\beta = L_a$. Similarly, $\beta(b)$ is right-neutral in $(Q, \cdot_{\alpha, \beta})$ if and only if $\alpha = R_b$. So there is a neutral element in $(Q, \cdot_{\alpha, \beta})$ if and only if $\alpha = R_b$ and $\beta = L_a$ for some $a, b \in Q$; this neutral element is just $R_b(a) = L_a(b) = ab$.

Lemma 1.5 Any loop \mathcal{L} isotopic to a quasigroup $\mathcal{Q} = (Q, \cdot)$ is isomorphic to an LP -isotope of \mathcal{Q} .

Proof. According to lemma 1.3, \mathcal{L} is isomorphic to a loop $\mathcal{Q}' = (Q, \cdot'; e')$ derived from \mathcal{Q} by means of the principal isotopy (α, β, id_Q) where $\alpha(x) \cdot' \beta(y) = xy$ for all $x, y \in Q$. Equivalently, $x \cdot' y = \alpha^{-1}(x) \cdot \beta^{-1}(y)$. Plugging $y = e'$ we have

$$x = x \cdot' e' = \alpha^{-1}(x) \cdot \beta^{-1}(e') = R_{\beta^{-1}(e')}(\alpha^{-1}(x)) \quad \text{for all } x \in Q,$$

that is $\alpha = R_{\beta^{-1}(e')}$. Now setting $x = e'$ we get

$$y = e' \cdot' y = \alpha^{-1}(e') \cdot \beta^{-1}(y) = L_{\alpha^{-1}(e')}(\beta^{-1}(y)) \quad \text{for all } y \in Q.$$

That is, $\beta = L_{\alpha^{-1}(e')}$. Hence \mathcal{Q}' is an LP -isotope of \mathcal{Q} under the isotopism $(R_{\beta^{-1}(e')}, L_{\alpha^{-1}(e')}, id_Q)$.

1.5 Isotopes of commutative groups

The following results explain why isotopy plays no role in group theory. The first of them is sometimes called the “Albert’s theorem”, the second one the “Isotopic groups rule”.

Lemma 1.6 Every loop isotopic to a group is isomorphic to it.

Proof. Suppose that $\mathcal{Q} = (Q, +; 0)$ is a group and consider its arbitrary LP -isotope $\mathcal{L} = (Q, \cdot_{ab}; ab)$ where $xb \cdot_{ab} ay = xy$ for all $x, y \in Q$, or equivalently, $x \cdot_{ab} y = x \cdot b^{-1} \cdot a^{-1} \cdot y$ for all $x, y \in Q$. Now consider the right translation $\theta = R_{ab}$. We obtain

$$\theta(x) \cdot_{ab} \theta(y) = xabb^{-1}a^{-1}yab = xyab = \theta(xy)$$

for all $x, y \in Q$ which shows that θ is an isomorphism of \mathcal{Q} onto \mathcal{L} . We obtain the result by means of Lemma 1.3.

Lemma 1.7 Isotopic groups are isomorphic.

2 Medial quasigroups

Now let us pay attention to quasigroups that satisfy mediality (1).

Corollary 2.1 *Any two LP-isotopes of a medial quasigroup are isomorphic.*

Lemma 2.2 *Any medial loop is a commutative group.*

Proof. If (Q, \cdot) is a medial loop with the neutral element e we put $a = e$, $d = e$ into the mediality law and get $bc = cb$, the commutativity of the dot operation. To check the associativity it is sufficient to put $c = e$: $(ab)d = a(bd)$.

Lemma 2.3 *Let (Q, \cdot) be a medial quasigroup, $e \in Q$ some element, and define $x \cdot_e y = R_e^{-1}(x) \cdot y$. Then (Q, \cdot_e) is a medial quasigroup for which e is a right-neutral element.*

Proof. As R_e is a permutation, (Q, \cdot_e) is a quasigroup. The element e is right-neutral as for any $x \in Q$, $x \cdot_e e = R_e^{-1}(x) \cdot e = R_e R_e^{-1}(x) = x$ holds. Now check mediality. For any $\tilde{a}, \tilde{b} \in Q$ the mediality of “ \cdot ” implies

$$(\tilde{a}\tilde{b})e = (\tilde{a}\tilde{b})((e/e) \cdot e) = (\tilde{a} \cdot (e/e)) \cdot (\tilde{b}e)$$

which can be written as $R_e(\tilde{a}\tilde{b}) = R_{e/e}(\tilde{a}) \cdot R_e(\tilde{b})$. Let us take $a, b \in Q$ such that $R_{e/e}^{-1}(R_e^{-1}(a)) = \tilde{a}$, $R_e^{-1}(b) = \tilde{b}$. We can write $R_e(\tilde{a}\tilde{b}) = R_{e/e}(\tilde{a}) \cdot R_e(\tilde{b})$ equivalently as $R_e(R_{e/e}^{-1}(R_e^{-1}(a)) \cdot R_e^{-1}(b)) = R_e^{-1}(a) \cdot b$. Finally we get

$$R_{e/e}^{-1}(R_e^{-1}(a)) \cdot R_e^{-1}(b) = R_e^{-1}(R_e^{-1}(a) \cdot b). \quad (2)$$

For any $a, b, c, d \in Q$ we check step by step (by means of (2), (1) and the definition of \cdot_e)

$$\begin{aligned} (a \cdot_e b) \cdot_e (c \cdot_e d) &= (R_e^{-1}(a) \cdot b) \cdot_e (R_e^{-1}(c) \cdot d) \\ &= R_e^{-1}(R_e^{-1}(a) \cdot b) \cdot (R_e^{-1}(c) \cdot d) \\ &= ((R_{e/e}^{-1}(R_e^{-1}(a)) \cdot R_e^{-1}(b)) \cdot (R_e^{-1}(c) \cdot d) \\ &= ((R_{e/e}^{-1}(R_e^{-1}(a)) \cdot R_e^{-1}(c)) \cdot (R_e^{-1}(b) \cdot d) \\ &= (R_e^{-1}(a) \cdot c) \cdot_e (R_e^{-1}(b) \cdot d) = (a \cdot_e c) \cdot_e (b \cdot_e d). \end{aligned}$$

Note that an analogous statement holds if we define $x \cdot_e y = x \cdot L_e^{-1}(y)$.

Theorem 2.4 *Every LP-isotope of a medial quasigroup is a commutative group.*

Proof. Let (Q, \cdot) be a medial quasigroup. Fix an arbitrary element $q \in Q$. Then the groupoid (Q, \circ_q) , where \circ_q is a binary operation on Q defined by $x \circ_q y = x \cdot L_q^{-1}(y)$, is a medial quasigroup in which q is the left-neutral element. The proof is similar to those in Lemma 2.3. Further, we know that every LP-isotope (Q, \diamond) of a given medial quasigroup (Q, \cdot) has its binary operation \diamond of the form $x \diamond y = R_v^{-1}(x) \cdot L_u^{-1}(y)$ with convenient $u, v \in Q$, and has the neutral element uv by Lemma 1.4. By Lemmas 2.2, 2.3, (Q, \diamond) is a medial loop; note that $x \diamond y = x \cdot_v L_u^{-1}(y) = R_v^{-1}(x) \cdot L_u^{-1}(y)$. Moreover, by Lemma 2.1, it is a commutative group.

Example 2.5 On the other hand, there exist LP -isotopes of commutative groups that are *not medial*. Indeed, consider the multiplication group $\mathcal{G} = (\mathbb{R}^+, \cdot; 1)$ of positive reals and a bijection $\sigma: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ that fixes all reals with the only exception of the pair of numbers 1, 2 that are interchanged. The image (\mathbb{R}^+, \circ) of the group \mathcal{G} under the isotopism (id, σ, id) has the operation of the form $x \circ \sigma(y) = x \cdot y$, or equivalently, $x \circ y = x \cdot \sigma^{-1}(y)$. It can be checked that (\mathbb{R}^+, \circ) is the LP -isotope with the identity element 2 which is not medial (since $(1 \circ 3) \circ (1 \circ 4) = (1 \cdot 3) \cdot (1 \cdot 4) = 12$ while $(1 \circ 1) \circ (3 \circ 4) = (1 \cdot 2) \cdot (3 \cdot 4) = 24$).

Theorem 2.6 *If $(G, +; 0)$ is a commutative group, $q \in G$ a fixed element, and f, g are commuting automorphisms of the group, then (G, \cdot) with $x \cdot y = f(x) + g(y) + q$, $x, y \in G$, is a medial quasigroup isotopic to the reduct $(G, +)$, $q = 0 \cdot 0$ holds, and the automorphisms take the form $f = R_{q \setminus q}$, $g = L_{q/q}$.*

Proof. Let $a, b, c, d \in G$. We deduce (by means of the definition of “ \cdot ”, commutability of automorphisms, and distributivity of the composition of maps)

$$\begin{aligned} (ab)(cd) &= \varphi(\varphi(a) + \psi(b) + h) + \psi(\varphi(c) + \psi(d) + h) + h \\ &= \varphi^2(a) + \varphi\psi(b) + \varphi(h) + \psi\varphi(c) + \psi^2(d) + \psi(h) + h \\ &= \varphi^2(a) + \varphi\psi(c) + \varphi(h) + \psi\varphi(b) + \psi^2(d) + \psi(h) + h \\ &= \varphi(\varphi(a) + \psi(c) + h) + \psi(\varphi(b) + \psi(d) + h) + h = (ac)(bd). \end{aligned}$$

Due to the definition of the dot operation, we get the unique solutions of the equations $xa = b$ and $ay = b$ in the form

$$x = \varphi^{-1}(b - \psi(a) - h), \quad y = \psi^{-1}(b - \varphi(a) - h)$$

where minus is the inverse of the group addition. Hence (G, \cdot) is a quasigroup. Further, we have $qq = \varphi(q) + \psi(q) + h = q + q + h = h$ and $q = (q/q)q = \varphi(q/q) + \psi(q) + h = \varphi(q/q) + q + h = \varphi(q/q) + h$. Since $L_{q/q}(x) = (q/q)x = \varphi(q/q) + \psi(x) + h$ we obtain $\psi = L_{q/q}$. Similarly, $\varphi = R_{q \setminus q}$. Finally, we can write $xy = \varphi(x) + \psi(y) + h$ as $\varphi^{-1}(\hat{x}) \cdot \psi^{-1}(\hat{y}) = R_h^+(\hat{x} + \hat{y})$, which shows that $(\varphi^{-1}, \psi^{-1}, R_h^+)$ is an isotopism of $(G, +)$ onto (G, \cdot) .

Corollary 2.7 *If $h = q$ is an idempotent element in (G, \cdot) then $xy = R_q(x) + L_q(y)$.*

Theorem 2.8 *Every medial isotope (G, \cdot) of a commutative group $(G, +; 0)$ has its binary operation “ \cdot ” of the form $x \cdot y = \varphi(x) + \psi(y) + h$ for a suitable element $h \in G$ and suitable commuting automorphisms φ, ψ of $(G, +; 0)$.*

Proof. Due to Lemma 1.3, we can restrict ourselves onto a principal isotope of a commutative group $(G, +; 0)$, without loss of generality. Let (α, β, id_G) be the corresponding isotopy. Hence the operations satisfy $xy = \alpha(x) + \beta(y)$, $x, y \in G$. From mediality, it follows

$$\begin{aligned} &\alpha(\alpha(a) + \beta(b)) + \beta(\alpha(c) + \beta(d)) = \\ &= \alpha(\alpha(a) + \beta(c)) + \beta(\alpha(b) + \beta(d)) \quad \text{for all } a, b, c, d \in G. \end{aligned} \tag{3}$$

Let us put $\hat{a} = \alpha(a)$, $\hat{d} = \beta(d)$, $c = 0$. We obtain

$$\alpha(\hat{a} + \beta(b)) + \beta(\alpha(0) + \hat{d}) = \alpha(\hat{a} + \beta(0)) + \beta(\alpha(b) + \hat{d}) \quad \text{for all } \hat{a}, b, \hat{d} \in G.$$

Adding $-\beta(\alpha(0) + \hat{d}) - \alpha(\hat{a} + \beta(0))$ to both sides and using commutativity we get

$$\alpha(\hat{a} + \beta(b)) - \alpha(\hat{a} + \beta(0)) = \beta(\hat{d} + \alpha(b)) - \beta(\hat{d} + \alpha(0)) \quad \text{for all } \hat{a}, b, \hat{d} \in G. \quad (4)$$

The right-hand side is independent of \hat{a} . Once we use the formula (4) as it is, and next time we substitute 0 instead of \hat{a} into the left-hand side; after subtraction we have

$$\alpha(\hat{a} + \beta(b)) - \alpha(\hat{a} + \beta(0)) = \alpha\beta(b) - \alpha(\beta(0)) \quad \text{for all } \hat{a}, b \in G.$$

Hence

$$\alpha(\hat{a} + \beta(b)) = -\alpha(\beta(0)) + \alpha(\hat{a} + \beta(0)) + \alpha(\beta(b)) \quad \text{for all } \hat{a}, b \in G. \quad (5)$$

Plugging $\beta(b) = \hat{b}$ we get

$$\alpha(\hat{a} + \hat{b}) = -\alpha(\beta(0)) + \alpha(\beta(0) + \hat{a}) + \alpha(\hat{b}) \quad \text{for all } \hat{a}, \hat{b} \in G. \quad (6)$$

Similarly,

$$\beta(\hat{a} + \hat{b}) = -\beta(\alpha(0)) + \beta(\alpha(0) + \hat{a}) + \beta(\hat{b}) \quad \text{for all } \hat{a}, \hat{b} \in G. \quad (7)$$

According to commutativity, interchanging \hat{a} , \hat{b} in (6) and (7) we have

$$\alpha(\beta(0) + \hat{a}) + \alpha(\hat{b}) = \alpha(\beta(0) + \hat{b}) + \alpha(\hat{a}).$$

Particularly for $\hat{b} = 0$,

$$\alpha(\beta(0) + \hat{a}) - \alpha(\hat{a}) = \alpha(\beta(0) + \hat{b}) - \alpha(\hat{b}) = \alpha(\beta(0) + 0) - \alpha(0) = \alpha(\beta(0)) - \alpha(0),$$

and consequently

$$\alpha(\beta(0) + \hat{a}) = -\alpha(0) + \alpha(\beta(0)) + \alpha(\hat{a}) \quad \text{for all } \hat{a} \in G. \quad (8)$$

Interganging the role of α and β we get

$$\beta(\alpha(0) + \hat{a}) = -\beta(0) + \beta(\alpha(0)) + \beta(\hat{a}) \quad \text{for all } \hat{a} \in G. \quad (9)$$

After substituting (8), (9) into the right hand sides of (6), (7) we get

$$\alpha(\hat{a} + \hat{b}) = -\alpha(\beta(0)) - \alpha(0) + \alpha(\beta(0)) + \alpha(\hat{a}) + \alpha(\hat{b}) = -\alpha(0) + \alpha(\hat{a}) + \alpha(\hat{b}),$$

and further, adding $-\alpha(0)$,

$$-\alpha(0) + \alpha(\hat{a} + \hat{b}) = (-\alpha(0) + \alpha(\hat{a})) + (-\alpha(0) + \alpha(\hat{b})) \quad \text{for all } \hat{a}, \hat{b} \in G. \quad (10)$$

Similar considerations give

$$-\beta(0) + \beta(\hat{a} + \hat{b}) = (-\beta(0) + \beta(\hat{a})) + (-\beta(0) + \beta(\hat{b})) \quad \text{for all } \hat{a}, \hat{b} \in G. \quad (11)$$

Consider permutations of G introduced by $\varphi: x \mapsto \alpha(x) - \alpha(0)$, $\psi: x \mapsto \beta(x) - \beta(0)$. We easily check that φ and ψ are automorphisms of $(G, +; 0)$. It remains to show that φ and ψ commute. Since $xy = \alpha(x) + \beta(y) = \varphi(x) + \psi(y) + \hat{e}$ where $\hat{e} = \alpha(0) + \beta(0)$ we can write

$$(ac)(bd) = \hat{e} + \varphi(\hat{e}) + \psi(\hat{e}) + \varphi^2(a) + \psi^2(d) + \varphi\psi(b) + \psi\varphi(c),$$

$$(ab)(cd) = \hat{e} + \varphi(\hat{e}) + \psi(\hat{e}) + \varphi^2(a) + \psi^2(d) + \varphi\psi(c) + \psi\varphi(b).$$

Mediality of the dot operation, i.e. the equality of left sides, implies the equality of right sides. Consequently, $\varphi\psi(b) + \psi\varphi(c) = \varphi\psi(c) + \psi\varphi(b)$. Hence $\varphi\psi(b) - \psi\varphi(b) = 0$ for all $b \in G$, that is, $\varphi\psi = \psi\varphi$.

3 Medial quasigroups and Toyoda's theorem

Theorem 3.1 (Toyoda's theorem for medial quasigroups) *For every medial quasigroup (Q, \cdot) there exists a commutative group $(Q, +)$, an element $q \in Q$ and a pair of commuting automorphisms φ, ψ of $(Q, +)$ such that*

$$xy = \varphi(x) + \psi(y) + q \quad \text{for all } x, y \in Q.$$

Proof. Every LP -isotope of a given medial quasigroup (Q, \cdot) is a commutative group (Theorem 2.4). Given such a commutative group $(Q, +)$, there must exist its commuting automorphisms φ, ψ and an element $q \in Q$ such that $xy = \varphi(x) + \psi(y) + q$ (Theorem 2.8).

According to Theorems 2.6, 2.8 and 3.1 we can formulate a deeper version of the Toyoda's theorem as follows:

Theorem 3.2 *Let (Q, \cdot) be a medial quasigroup. Then for every $q \in Q$ there is a commutative group $(Q, +_q; q)$ such that $xy = R_{q \setminus q}(x) +_q L_{q/q}(y) +_q (q \cdot q)$ holds for all $x, y \in Q$ where $R_{q \setminus q}, L_{q/q}$ are commuting automorphisms of $(Q, +_q; q)$. All the groups $(Q, +_q; q), q \in Q$, are isomorphic.*

Corollary 3.3 *If a medial quasigroup (Q, \cdot) contains an idempotent element e let us introduce a binary operation $+_e$ on Q by $x +_e y = R_e^{-1}(x) \cdot L_e^{-1}(y)$ for all $x, y \in Q$. Then $(Q, +_e; e)$ is a commutative group, and R_e, L_e are commuting automorphisms of the group. Moreover, if (Q, \cdot) is idempotent then the groups $(Q, +_e; e), e \in Q$ are pairwise isomorphic.*

4 Quasigroup modes

An idempotent medial quasigroup is called a *quasigroup mode*. We easily verify:

Lemma 4.1 *Any quasigroup mode is elastic, i.e. it satisfies $(xy)x = x(yx)$.*

The last part of Corollary 3.3 is closely related also to the following theorem:

Theorem 4.2 *Let $\mathcal{Q} = (Q, \cdot)$ be a quasigroup mode. For any element $q \in Q$ let us introduce a binary operation $+_q$ on Q by $x +_q y = (R_q^{-1}(x)) \cdot (L_q^{-1}(y))$ for $x, y \in Q$. Then $(Q, +_q; q)$ is a commutative group, principally isotopic to \mathcal{Q} , and R_q, L_q are commuting automorphisms of the group such that $R_q(x) +_q L_q(x) = x$ holds identically in $x \in Q$. All the groups $(Q, +_q; q), q \in Q$, are isomorphic.*

Proof. By Theorem 2.4, $(Q, +_q; q)$ is a commutative group. Let us check that R_q and L_q are commuting automorphisms of $(Q, +_q; q)$. Indeed, by elasticity, mediality and idempotency, we have $R_q(R_q(x) +_q L_q(y)) = R_q(xy) = (xy)q$, $R_q(R_q(x)) +_q R_q(L_q(y)) = R_q(R_q(x)) +_q L_q(R_q(y)) = R_q(x) \cdot R_q(y) = (xq)(yq) = (xy)(qq) = (xy)q$ for all $x, y \in Q$. Hence R_q is an automorphism of $(Q, +_q; q)$. Similarly, also L_q is an automorphism of $(Q, +_q; q)$. Moreover, $L_q R_q(x) = q(xq)$, $R_q L_q(x) = (qx)q$ for all $x \in Q$. Hence by elasticity, $L_q R_q = R_q L_q$. Further,

using the definition of $+_q$ and elasticity, $R_q(x) +_q L_q(x) = x \cdot x = x$ for all $x \in Q$. Now fix elements $a, b \in Q$ and write, for any $x, y \in Q$,

$$xy = R_a(x) +_a L_a(y) = R_b(x) +_b L_b(y).$$

Now putting $\hat{x} = R_a(x)$, $\hat{y} = L_a(y)$ we get $\hat{x} +_a \hat{y} = R_b R_a^{-1}(\hat{x}) +_b L_b L_a^{-1}(\hat{y})$, which defines an isotopy of $(Q, +_a)$ onto $(Q, +_b)$, and hence an isomorphism of the group $(Q, +_a; a)$ onto the group $(Q, +_b; b)$.

As a corollary of Theorem 2.6, with $h = 0$, we get the converse:

Theorem 4.3 *If $\mathcal{G} = (G, +; 0)$ is a commutative group and φ, ψ is a pair of commuting automorphisms of \mathcal{G} such that $\varphi(x) + \psi(x) = x$ holds for all $x \in G$ then on the carrier set G , the operation \cdot introduced by the formula $x \cdot y = \varphi(x) + \psi(y)$ determines a quasigroup mode $\mathcal{Q} = (G, \cdot)$ which is a principal isotope of \mathcal{G} under the isotopy $(\varphi^{-1}, \psi^{-1}, id_Q): \mathcal{G} \rightarrow \mathcal{Q}$.*

Acknowledgement

This paper has been supported by the project of specific university research of the Brno University of Technology, No. FAST-S-11-47, grant no. PrF 2011 022 and by the grant from Grant Agency of Czech Republic GAČR no. P201/11/0356 with the title: "Riemannian, pseudo-Riemannian and affine differential geometry".

References

- [1] BARTOŠKOVÁ, Z.: *Medial Quasigroups and Geometry*. Mgr. Thesis, Palacký University Olomouc, Fac. Sci., 2011 (in Czech).
- [2] BARTOŠKOVÁ, Z.: *Commutative Groups and Medial Quasigroups*. Bc. Thesis, Palacký University Olomouc, Fac. Sci., 2009 (in Czech).
- [3] BELOUSOV, V.D.: *Foundations of the theory of quasigroups and loops*. Moscow, Nauka, 1967.
- [4] BRUCK, R.H.: *Some results in the theory of quasigroups*. Trans. Amer. Math. Soc., Vol. 55, pp. 19-52, 1944.
- [5] RUCK, R.H.: *A Survey of Binary Systems*. Berlin, Springer, 1958.
- [6] CHEIN, O., PFLUGFELDER, H.O, SMITH, J.D.H. (eds.): *Quasigroups and Loops: Theory and Applications*. Heldermann Verlag, Berlin, 1990.
- [7] HAVEL, V.J., VANŽUROVÁ, A.: *Medial Quasigroups and Geometry*. Palacký University in Olomouc, Olomouc, 2006.
- [8] PFLUGFELDER, H.O.: *Quasigroups and Loops, Introduction*. Heldermann Verlag, Berlin, 1990.
- [9] PICKERT, G.: *Projektive Ebenen*. Springer, Berlin, Heidelberg, New York, 1975.
- [10] TOYODA, K.: *On axioms of mean transformations and automorphic transformations of abelian groups*. Tôhoku Math. J. Vol. 46, pp. 239-251, 1940.
- [11] TOYODA, K.: *On affine geometry of abelian groups*. Proc. Imp. Acad. Tokyo Vol. 16, No. 5, pp. 161-164, 1940.

- [12] VANŽUROVÁ, A.: *Golden Section Quasigroups, Finite Examples*. Proc. 10th International Conf. APLIMAT 2011, Part IV, pp. 183-190, 2011.
- [13] VANŽUROVÁ, A.: *Algebraic systems in cryptology and the security of information*. In International Conference in Military Technology Proceeding, Section 12 Security technology (ed. Z. Vránová and V. Plátěnka), ICMT'11, Brno : University of Defence, pp. 1251-1256, 2011.
- [14] VANŽUROVÁ, A., DOLEŽALOVÁ, J.: *Hexagonal quasigroups over finite fields*. Proc. of 7th Conf. on Mathematics and Physics at Technical Universities with internat. participation, Brno, September 22, 2011, part 1 - mathematics, pp. 411-419, 2011.

Current address

VANŽUROVÁ Alena, Doc. RNDr. CSc.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University, Tř. 17. listopadu 12, 771 46 Olomouc, alena.vanzurova@upol.cz;

Department of Mathematics, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic, vanzurova.a@fce.vutbr.cz.

BARTOŠKOVÁ Zuzana, Mgr.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University, Tř. 17. listopadu 12, 771 46 Olomouc, ZuzanaBartoskova@seznam.cz

GEODESIC MAPPINGS PRESERVING THE EINSTEIN TENSOR OF WEYL SPACES

ARSAN Gürpınar Güler, (TR), ÇİVİ Gülçin, (TR)

Abstract. In this paper, we consider geodesic mappings preserving the Einstein tensor between Weyl spaces. We prove that the generalized concircular curvature tensor is invariant under the geodesic mapping preserving the Einstein tensor if and only if the covector field of the mapping is locally a gradient.

Key words and phrases. Weyl space, geodesic mapping, the Einstein tensor, generalized concircular curvature tensor.

Mathematics Subject Classification. Primary 53A40.

1 Introduction

Geodesic mappings of manifolds with affine connection and Riemannian spaces were studied by many authors (e.g., [1,3,6,9,10,11]). In [12], we studied geodesic mappings between Kahler-Weyl spaces. In [4], I. Hinterleitner and J. Mikes investigated geodesic mappings of special manifolds with affine connection and with general condition of recurrency onto Weyl spaces. In [2], authors considered geodesic mappings preserving the Einstein tensor between Pseudo-Riemannian spaces and they proved that the concircular curvature tensor is invariant under the geodesic mappings preserving the Einstein tensor.

In this work, we consider geodesic mappings preserving the Einstein tensor between Weyl spaces and we obtain a necessary and sufficient condition for generalized concircular curvature tensor to be invariant under the geodesic mapping preserving the Einstein tensor between Weyl spaces.

2 Preliminaries

An n -dimensional differentiable manifold W_n is said to be a Weyl space if it has a conformal metric tensor g and a symmetric connection ∇ satisfying the compatibility condition given by the equation

$$\nabla_k g_{ij} - 2 T_k g_{ij} = 0, \quad (1)$$

where T_k denotes a covariant vector field. Under the renormalization

$$\tilde{g} = \lambda^2 g \quad (2)$$

of the metric tensor g , T is transformed by the law

$$\tilde{T}_k = T_k + \partial_k \ln \lambda$$

where λ is a function defined on W_n .

An object A defined on $W_n(g, T)$ is called a satellite of weight $\{p\}$ of the tensor g_{ij} , if it admits a transformation of the form

$$\tilde{A} = \lambda^p A$$

under the renormalization of the metric tensor g_{ij} [5,7].

The prolonged covariant derivative of a satellite A is defined by

$$\dot{\nabla}_k A = \nabla_k A - p T_k A. \quad (3)$$

We note that the prolonged covariant derivative preserves the weight.

Writing (1) in local coordinates and expanding it we find that

$$\partial_k g_{ij} - g_{hj} \Gamma_{ik}^h - g_{ih} \Gamma_{jk}^h - 2 T_k g_{ij} = 0$$

where Γ_{jk}^i are the coefficients of the Weyl connection ∇ obtained by

$$\Gamma_{jk}^i = \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} - (\delta_j^i T_k + \delta_k^i T_j - g^{im} g_{jk} T_m), \quad (4)$$

in which $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$ are the coefficients of the Levi-Civita connection.

Consider two Weyl spaces W_n and \overline{W}_n with connection ∇ and $\overline{\nabla}$, respectively.

A diffeomorphism between two Weyl space W_n and \overline{W}_n is called geodesic mapping if it is geodesic preserving, that is, when it maps any geodesic of W_n into an arbitrarily parametrized geodesic of \overline{W}_n again.

It is known that, the curve

$$L : x^i = x^i(t)$$

is a geodesic of a afin space if and only if

$$\frac{d^2x^i}{dt^2} + \Gamma_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = \rho(t) \frac{dx^i}{dt}$$

is hold, where $\rho(t)$ is a determinate function of t [3].

Let $f : W_n \rightarrow \overline{W}_n$ is a geodesic mapping. Then, by f , the geodesic L of W_n passes to the geodesic \bar{L} of \overline{W}_n and

$$\frac{d^2x^i}{dt^2} + \bar{\Gamma}_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = \bar{\rho}(t) \frac{dx^i}{dt}$$

is hold where $\bar{x}^i = x^i$.

In [12], we obtained that a necessary and sufficient condition for existence of a geodesic mapping between W_n and \overline{W}_n to satisfy the following condition

$$\bar{\Gamma}_{jk}^i = \Gamma_{jk}^i + \delta_k^i \psi_j + \delta_j^i \psi_k \quad (5)$$

holds, where Γ_{jk}^i and $\bar{\Gamma}_{jk}^i$ are components of the Christoffel symbols in W_n and \overline{W}_n , respectively, and

$$\psi_j = \frac{1}{n+1} \left[\partial_j \ln \sqrt{\frac{\bar{g}}{g}} - n P_j \right], \quad (6)$$

where $\bar{g} = \det \bar{g}_{ij}$, $g = \det g_{ij}$ and $P_j = \bar{T}_j - T_j$

If $\psi_j \neq 0$, then a geodesic mapping is called nontrivial; otherwise it is said to be trivial or affine.

Applying the formula (5) we find the relationships for the curvature and Ricci tensors of W_n and \overline{W}_n as

$$\bar{R}_{ijk}^h = R_{ijk}^h + \delta_i^h (\psi_{kj} - \psi_{jk}) + \delta_k^h \psi_{ij} - \delta_j^h \psi_{ik}, \quad (7)$$

$$\bar{R}_{ij} = R_{ij} + n\psi_{ij} - \psi_{ji}, \quad (8)$$

where $\psi_{ij} = \dot{\nabla}_j \psi_i - \psi_i \psi_j$ and $R_{ij} = R_{ij}^h$.

The tensor Z of type (1,3) whose components are given by [8]

$$Z_{ijk}^h = R_{ijk}^h - \frac{R}{n(n-1)} (g_{ij} \delta_k^h - g_{ik} \delta_j^h) \quad (9)$$

is called the generalized concircular curvature tensor of W_n .

Contraction on the indices h and k in (9) gives the generalized concircularly tensor

$$Z_{ij} = R_{ij} - \frac{R}{n} g_{ij} \quad (10)$$

where $R = R_{ab} g^{ab}$ is the scalar curvature of W_n .

3 Geodesic mappings preserving the Einstein tensor of Weyl spaces

If a geodesic mapping $f : W_n \rightarrow \overline{W}_n$ satisfies

$$\bar{E}_{ij} = E_{ij} \quad (11)$$

it is said to be the geodesic mapping preserving the Einstein tensor.

Since the Ricci tensor of W_n is not symmetric, the Einstein tensor E_{ij} is

$$E_{ij} = R_{(ij)} - \frac{R}{n} g_{ij}, \quad (12)$$

where $R_{(ij)}$ denotes the symmetric part of the Ricci tensor.

Let $W_n(g, T)$ and $\overline{W}_n(\bar{g}, \bar{T})$ be two manifolds which are in geodesic mapping preserving the Einstein tensor. Then, from (8) we obtain

$$\bar{R}_{(ij)} = R_{(ij)} + (n-1)\psi_{(ij)}. \quad (13)$$

In view of (11) and (12) we get

$$\psi_{(ij)} = \frac{1}{n(n-1)}(\bar{R}\bar{g}_{ij} - Rg_{ij}). \quad (14)$$

Using (9), the generalized concircular curvature tensor of $\overline{W}_n(\bar{g}, \bar{T})$ can be written as

$$\bar{Z}_{ijk}^h = \bar{R}_{ijk}^h - \frac{\bar{R}}{n(n-1)}(\bar{g}_{ij}\delta_k^h - \bar{g}_{ik}\delta_j^h). \quad (15)$$

By virtue of (7), (15) becomes

$$\bar{Z}_{ijk}^h = R_{ijk}^h + \delta_i^h(\psi_{kj} - \psi_{jk}) + \delta_k^h\psi_{ij} - \delta_j^h\psi_{ik} - \frac{\bar{R}}{n(n-1)}\bar{g}_{ij}\delta_k^h + \frac{\bar{R}}{n(n-1)}\bar{g}_{ik}\delta_j^h. \quad (16)$$

Inserting (14) into (16) and making the necessary arrangements we obtain

$$\bar{Z}_{ijk}^h = Z_{ijk}^h + \delta_i^h(\psi_{kj} - \psi_{jk}) + \frac{1}{2}\delta_k^h(\psi_{ij} - \psi_{ji}) + \frac{1}{2}\delta_j^h(\psi_{ki} - \psi_{ik}). \quad (17)$$

Contraction on the indices h and k in (17) gives

$$\bar{Z}_{ij} = Z_{ij} + \left(\frac{n+1}{2}\right)(\psi_{ij} - \psi_{ji}). \quad (18)$$

Also, from (15) we obtain

$$\bar{Z}_{ij} = \bar{R}_{ij} - \frac{\bar{R}}{n}\bar{g}_{ij} = \bar{E}_{ij} + \bar{R}_{[ij]}. \quad (19)$$

Using (10),(11),(18) and (19) we have

$$\bar{R}_{[ij]} - R_{[ij]} = \left(\frac{n+1}{2}\right)(\psi_{ij} - \psi_{ji}). \quad (20)$$

Remembering that the anti-symmetric part of the Ricci tensor of $W_n(g, T)$ satisfies

$$R_{[ij]} = n \nabla_{[i} T_{j]},$$

we get

$$\bar{\nabla}_{[i} \bar{T}_{j]} - \nabla_{[i} T_{j]} = \left(\frac{n+1}{n} \right) \psi_{[ij]}. \quad (21)$$

Considering $P_j = \bar{T}_j - T_j$ and using (21), we may conclude that

$$\nabla_{[i} P_{j]} = \frac{1}{2} (\partial_i P_j - \partial_j P_i) = \left(\frac{n+1}{n} \right) \psi_{[ij]}. \quad (22)$$

Suppose now that $\bar{Z}_{ijk}^h = Z_{ijk}^h$. Then, we have from (17) that

$$\delta_i^h (\psi_{kj} - \psi_{jk}) + \frac{1}{2} \delta_k^h (\psi_{ij} - \psi_{ji}) + \frac{1}{2} \delta_j^h (\psi_{ki} - \psi_{ik}) = 0$$

which follows that

$$\psi_{ij} = \psi_{ji}. \quad (23)$$

In view of (22) and (23) we obtain

$$\partial_i P_j - \partial_j P_i = 0 \quad (24)$$

which means that P is locally a gradient so, by (6), the covector field ψ is also locally a gradient.

Conversely, if the equation (24) holds, then from (22) we have $\psi_{ij} = \psi_{ji}$. In this case, using (17) we find that $\bar{Z}_{ijk}^h = Z_{ijk}^h$.

Consequently, we have proved

Theorem 3.1 *The generalized concircular curvature tensor is invariant under geodesic mapping preserving the Einstein tensor if and only if the covector field of the mapping is locally a gradient.*

References

- [1] AMINOVA, A.V.: *Projective transformations of pseudo-Riemannian manifolds*. In J. Math. Sci., New York, Vol.113, No.3, 367-470, 2003.
- [2] CHEPURNA, E.O., KIOSAK A.V., MIKES, J.: *On geodesic mappings preserving the Einstein tensor*. In Acta Univ. Palacki. olomuc., Fac. rer. nat., Mathematica 49, No. 2, pp. 49-52, 2010.
- [3] EISENHART, L.P.: *Riemannian geometry*. Princenton Univ. Press, 1926.
- [4] HINTERLEITNER, I., MIKES, J.: *Geodesic mappings onto Weyl manifolds*. In Aplimat-Journal of Applied Mathematics, Vol. 2, No. 1, pp. 125-134, 2009.
- [5] HLAVATY, V.: *Theorie d'immersion d'une W_m dans W_n* . In Ann Soc Polon, Math, 21, pp. 196-206, 1949.

- [6] MIKES, J.: *Geodesic mappings of affine-connected and Riemannian spaces*. In J.Math.Sci. New York, Vol.78, No. 3, pp.311-333, 1996.
- [7] NORDEN, A.: *Affinely connected spaces (in Russian)*. Moscow, GRMFML,1976.
- [8] ÖZDEĞER, A., SENTÜRK, Z.: *Generalized circles in Weyl spaces and their conformal mapping*. In Publ Math., Deprecen, 60(1/2), pp. 75-87, 2002.
- [9] PETROV, A. Z.: *New methods in the theory of general relativity*. Nauka, Moscow, 1966.
- [10] SCHOUTEN, J. A, STRUIK, D.,J.:*Introduction into new methods in Differential Geometry*.(Germ. *Einführung in die neueren Methoden der Differentialgeometrie*). 1935.
- [11] SINYUKOV, N.S.: *Geodesic mappings of Riemannian spaces*. Nauka, Moscow, 1979.
- [12] YILDIRIM, G.Ç., ARSAN, G.G.: *Geodesic mappings between Kahler-Weyl spaces*. Diff. Geometry-Dynamical Systems, Vol. 9, pp. 156-159, 2007.

Current address

Güler Gürpınar Arsan, Assoc.Prof.Dr.

Istanbul Technical University, Faculty of Science and Letters,
Department of Mathematics,
34469, Maslak-Istanbul, Turkey,
tel.+(90)2122853269 and e-mail: ggarsan@itu.edu.tr

Gülçin Çivi, Assoc.Prof.Dr.

Istanbul Technical University, Faculty of Science and Letters,
Department of Mathematics,
34469, Maslak-Istanbul, Turkey,
tel.+(90)2122856959 and e-mail: civi@itu.edu.tr

SMOOTH CURVES APPROXIMATION BY CHORD-LENGTH CURVES

BASTL Bohumír, (CZ), LÁVIČKA Miroslav, (CZ)

Abstract. This paper is devoted to one practical application of planar rational curves with chord length parameterization (shortly RCL curves). Rational curves with chord length parameterizations are a chord-length analogy to the so-called Pythagorean-hodograph curves characterized by closed form formulas for their arc-lengths. They represent a new representation of objects in CAGD which can be used for formulating alternative modeling techniques. Using the universal formula for planar RCL curves, we design a simple G^1 Hermite interpolation algorithm based on solving a small system of linear equations. In particular, we show how to approximate a general planar curve using arcs of RCL curves. The efficiency of the designed method is presented on two particular examples.

Key words and phrases. Rational curves, chord length parameterizations, Hermite interpolation.

Mathematics Subject Classification. Primary 51N35; Secondary 14H45.

1 Introduction and related work

The investigation of rational varieties with chord length parameterization (shortly RCL varieties) has become a popular research field in recent years. It started in [5] for curves in plane by proving that the chord length parameter condition is fulfilled for circular arcs in standard rational quadratic form. An independent geometric proof of this fact is also presented in [11], where a circle-preserving variant of the four-point subdivision scheme is thoroughly analyzed.

Studying chord length parameterizations in geometric modelling was mainly influenced by the use of chord length or chordal method for interpolation and approximation of discrete data points, cf. [9]. RCL parameterizations can be considered as an alternative to arc-length parameterizations because (as an arc-length parameter) the chord-length parameter is also uniquely determined by the loci of the curve. Thus, RCL parameterizations can be seen, in some sense, as a chord-length analogy to the so called Pythagorean hodograph (PH) curves

characterized by closed form formulas for their arc-lengths, see [6, 7] and references therein for more details. We would like to recall that RCL curves are worth studying especially because of the following advantages: they provide a simple inversion formula applicable e.g. for computing their implicit form, they do not have self-intersections, and they are suitable for point-curve evaluation.

Later, a thorough analysis of curves with RCL property followed. The close connection between bipolar coordinates and curves with RCL parameterization was studied in [13]. It was shown that these curves are exactly those whose parameter coincides with one of the bipolar coordinates. Independently, computations in the complex plane for studying rational curves which can be parameterized by chord length were used in [10]. Finally, some curves with chord length parameterization were found among curves possessing a complex rational form, see [12]. Besides straight lines and circles in standard form, the family of RCL curves contains e.g. Bernoulli's lemniscate, Pascal's Limaçon and equilateral hyperbolas.

Interesting observations about RCL curves led to the idea to extend this approach also to rational surfaces. First, it was proved in [1] that the equal chord property holds for certain quadratic rational Bézier patches describing a segment of a sphere. This result is a direct surface analogy to the planar result of [5]. In addition, it was shown how to characterize the RCL property of surfaces using tripolar coordinates in space, which extend the results of [13] concerning the bipolar coordinates (see also [4, 8]). A thorough analysis of surfaces with RCL property was provided in [2]. Rational triangular Bézier surfaces of an arbitrary degree were considered and conditions under which they are rationally parameterized by chord lengths with respect to the reference circle were investigated.

Finally, the RCL property was extended to arbitrary k -dimensional rational varieties and a general approach to rational chord length parameterizations in any dimension was formulated in [3]. It was shown that the observations from [1, 2, 5, 10, 13] can be identified as special cases of the results provided by the general approach.

In this paper, we show that the RCL curves are very suitable for constructing G^1 Hermite interpolating splines. The extra feature of these splines consist of all benefits of the RCL curves, i.e., they do not possess self-intersections, and they are suitable for point-curve testing (of course, piecewise now). The main advantage of the designed algorithm is its simplicity as the designed method is based only on solving a simple system of linear equations.

2 Construction and properties of RCL curves

Given a parametric curve $\mathbf{p}(u)$ over a certain domain $u \in [a, b]$, its *chord-length* at a given point $\mathbf{p}(u)$ is defined as

$$\text{chord}(u) := \frac{\|\mathbf{p}(u) - \mathbf{A}\|}{\|\mathbf{p}(u) - \mathbf{A}\| + \|\mathbf{p}(u) - \mathbf{B}\|}, \quad \mathbf{A} = \mathbf{p}(a), \mathbf{B} = \mathbf{p}(b). \quad (1)$$

The curve is said to be *chord length parameterized* if

$$\text{chord}(u) = u \quad (2)$$

Rational curves with chord length parameterization have recently been studied as special instances of RCL k -varieties in d -dimensional Euclidean space. We recall the definition from [3].

Definition 2.1 Let $S \subset \mathbb{E}_k$ be a $(k-1)$ -dimensional sphere and $\gamma(\mathbf{u})$, where $\mathbf{u} = (u_1, \dots, u_k)$, a rational parameterization of the space \mathbb{E}_k . We say that a k -dimensional surface $\mathbf{P}(\mathbf{u})$ is RCL with respect to S and γ if

$$|\mathbf{P}(\mathbf{u}) - \tilde{A}| : |\mathbf{P}(\mathbf{u}) - \bar{A}| = |\gamma(\mathbf{u}) - \tilde{A}| : |\gamma(\mathbf{u}) - \bar{A}| \quad (3)$$

holds for any $\tilde{A}, \bar{A} \in S$ and any \mathbf{u} .

As the notion of being RCL is covariant with respect to similarities, we can choose the reference sphere S as the unit sphere in the (x_1, \dots, x_k) -plane. It was proved in [3]:

Theorem 2.2 A variety $\mathbf{P}(\mathbf{u})$ is a rational chord length parameterization with respect to the $(k-1)$ -dimensional reference sphere S and the rational parameterization $\gamma(\mathbf{u})$ if and only if there exists a rational mapping $q(\mathbf{u}) : \mathbb{R}^k \rightarrow \mathbb{R}^{n-k}$ such that

$$\mathbf{P}(\mathbf{u}) = M(\tilde{x}_1(\mathbf{u}), \dots, \tilde{x}_{k-1}(\mathbf{u}), \tilde{x}_k(\mathbf{u})N(\mathbf{q}(\mathbf{u}))), \quad (4)$$

where M is the inversion with respect to the $(n-1)$ -dimensional sphere centered at $(0, \dots, 0, \underbrace{-1}_{k\text{-th}}, 0, \dots, 0)^\top$ with radius $\sqrt{2}$, \tilde{x}_i is defined by

$$M(\mathbf{P}(\mathbf{u})) = (\tilde{x}_1(\mathbf{u}), \dots, \tilde{x}_{k-1}(\mathbf{u}), \bar{x}_k(\mathbf{u}), \dots, \bar{x}_n(\mathbf{u})), \quad (5)$$

where

$$\bar{x}_k(\mathbf{u})^2 + \dots + \bar{x}_n(\mathbf{u})^2 = \tilde{x}_k(\mathbf{u})^2,$$

and

$$N(q_1, \dots, q_{n-k}) = \frac{(1 - q_1^2 - \dots - q_{n-k}^2, 2q_1, \dots, 2q_{n-k})}{1 + q_1^2 + \dots + q_{n-k}^2}. \quad (6)$$

For $n=2, k=1$, we obtain RCL curves in the plane. The reference sphere consists of two points $(\pm 1, 0)^\top$ and the inversion has the equation

$$M(x, y) = \frac{1}{(x+1)^2 + y^2} \begin{pmatrix} 1 - x^2 - y^2 \\ 2y \end{pmatrix}. \quad (7)$$

Let us consider the trivial parameterization of the x -axis $\gamma(u) = (u, 0)^\top$. Then

$$M(\gamma(u)) = \left(\frac{1 - u^2}{(u+1)^2}, 0 \right)^\top. \quad (8)$$

Moreover, N depends only on one rational function q and takes the form

$$N = \left(\frac{1 - q^2}{1 + q^2}, \frac{2q}{1 + q^2} \right)^\top. \quad (9)$$

Substituting (8) and (9) to (4) we obtain the following explicit formula with respect to the trivial parameterization of the line, which is equivalent to the construction obtained in [13].

Theorem 2.3 A planar curve \mathbf{p} is a rational chord length parameterization with respect to the reference points $(\pm 1, 0)^\top$ if and only if there exists a rational function $q(u)$ such that

$$\mathbf{p}(u) = \left(\frac{(1 + q^2)u}{1 + q^2 u^2}, \frac{q(1 - u^2)}{1 + q^2 u^2} \right)^\top. \quad (10)$$

3 G^1 Hermite interpolation with arcs of RCL curves

In this section, we describe an algorithm for construction of G^1 Hermite interpolant represented by an arc of a suitably chosen RCL curve.

Let two points \mathbf{A} , \mathbf{B} with associated unit tangent vectors $\mathbf{t}_\mathbf{A}$, $\mathbf{t}_\mathbf{B}$ be given G^1 Hermite data which we want to interpolate. In the first step, we need to transform given data using a similarity such that \mathbf{A} is mapped to the point $(-1, 0)^\top$ and \mathbf{B} is mapped to $(1, 0)^\top$ (see Theorem 2.3). Concurrently, we obtain also the transformed unit tangent vectors $\tilde{\mathbf{t}}_\mathbf{A}$, $\tilde{\mathbf{t}}_\mathbf{B}$. Then, we can use the general formula (10) describing all chord length parameterizations of planar curves. It is obvious that it is enough to find a suitable function $q(u)$ such that given data are interpolated. Substituting the boundary values of the parameter interval of the parameterization (10), we arrive at

$$\begin{aligned} \mathbf{p}(-1) &= (-1, 0)^\top, & \mathbf{p}'(-1) &= \left(\frac{1 - q(-1)^2}{1 + q(-1)^2}, \frac{2q(-1)}{1 + q(-1)^2} \right)^\top, \\ \mathbf{p}(1) &= (1, 0)^\top, & \mathbf{p}'(1) &= \left(\frac{1 - q(1)^2}{1 + q(1)^2}, -\frac{2q(1)}{1 + q(1)^2} \right)^\top. \end{aligned} \quad (11)$$

Thus, we need to focus only on the given unit tangent vectors and determine the function $q(u)$ such that

$$\mathbf{p}'(-1) = \tilde{\mathbf{t}}_\mathbf{A}, \quad \mathbf{p}'(1) = \tilde{\mathbf{t}}_\mathbf{B}.$$

It can be seen that expressions for $\mathbf{p}'(-1)$ and $\mathbf{p}'(1)$ in fact represent a rational parameterization of the unit circle with the center at the origin obtained via the stereographic projection and that $q(-1)$ and $q(1)$, respectively, correspond to the choice of a parameter determining one particular point on this unit circle. Therefore, using the inverse stereographic projection we can obtain the following system of equations relating $q(u)$ and $\tilde{\mathbf{t}}_\mathbf{A} = (\tilde{t}_\mathbf{A}^x, \tilde{t}_\mathbf{A}^y)$, $\tilde{\mathbf{t}}_\mathbf{B} = (\tilde{t}_\mathbf{B}^x, \tilde{t}_\mathbf{B}^y)$

$$q(-1) = \frac{\tilde{t}_\mathbf{A}^y}{1 + \tilde{t}_\mathbf{A}^x}, \quad q(1) = -\frac{\tilde{t}_\mathbf{B}^y}{1 + \tilde{t}_\mathbf{B}^x}. \quad (12)$$

The simplest choice for $q(u)$ providing a unique solution of G^1 Hermite interpolation problem is linear, i.e., we choose

$$q(u) = q_0 + q_1 u. \quad (13)$$

By substituting (13) to (12) we obtain a simple system of linear equations which has an explicit solution in the form

$$q_0 = \frac{1}{2} \left(\frac{\tilde{t}_\mathbf{A}^y}{1 + \tilde{t}_\mathbf{A}^x} - \frac{\tilde{t}_\mathbf{B}^y}{1 + \tilde{t}_\mathbf{B}^x} \right), \quad q_1 = -\frac{1}{2} \left(\frac{\tilde{t}_\mathbf{A}^y}{1 + \tilde{t}_\mathbf{A}^x} + \frac{\tilde{t}_\mathbf{B}^y}{1 + \tilde{t}_\mathbf{B}^x} \right). \quad (14)$$

This leads to the explicit form of $q(u)$, i.e.,

$$q(u) = -\frac{\tilde{t}_\mathbf{A}^y}{2(1 + \tilde{t}_\mathbf{A}^x)}(u - 1) - \frac{\tilde{t}_\mathbf{B}^y}{2(1 + \tilde{t}_\mathbf{B}^x)}(u + 1). \quad (15)$$

By computing $q(u)$ for given data, the interpolant parameterization is known. The final step is the transformation back to initial position. The whole process is summarized in Algorithm 1.

Algorithm 1 G^1 Hermite interpolation by RCL curves

Require: Points \mathbf{A} , \mathbf{B} with associated unit tangent vectors $\mathbf{t}_\mathbf{A}$, $\mathbf{t}_\mathbf{B}$

Ensure: RCL curve $\mathbf{p}(u)$, $u \in [-1, 1]$, interpolating the given data

- 1: Find a similarity ϕ which maps \mathbf{A} to $(-1, 0)^\top$ and \mathbf{B} to $(1, 0)^\top$
- 2: Transform the given unit tangent vectors

$$\tilde{\mathbf{t}}_\mathbf{A} = \text{Normalize}(\phi(\mathbf{t}_\mathbf{A})), \quad \tilde{\mathbf{t}}_\mathbf{B} = \text{Normalize}(\phi(\mathbf{t}_\mathbf{B}))$$

- 3: Compute the function $q(u)$

$$q(u) = -\frac{\tilde{t}_\mathbf{A}^y}{2(1 + \tilde{t}_\mathbf{A}^x)}(u - 1) - \frac{\tilde{t}_\mathbf{B}^y}{2(1 + \tilde{t}_\mathbf{B}^x)}(u + 1)$$

- 4: Compute the parameterization of the interpolating RCL curve

$$\tilde{\mathbf{p}}(u) = \left(\frac{(1 + q^2)u}{1 + q^2u^2}, \frac{q(1 - u^2)}{1 + q^2u^2} \right)^\top, u \in [-1, 1]$$

- 5: Transform $\tilde{\mathbf{p}}(u)$ back to the initial position

$$\mathbf{p}(u) = \phi^{-1}(\tilde{\mathbf{p}}(u))$$

Remark 3.1 Because of the division in (15) $\tilde{\mathbf{t}}_\mathbf{A} \neq (-1, 0)^\top$ and $\tilde{\mathbf{t}}_\mathbf{B} \neq (1, 0)^\top$ must hold. This fact is caused by the derivation of a rational parameterization of the unit circle via the stereographic projection involved in (11).

Example 3.2 Let us consider G^1 Hermite data

$$\mathbf{A} = (-1, 0)^\top, \mathbf{B} = (1, 0)^\top, \mathbf{t}_\mathbf{A} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)^\top, \mathbf{t}_\mathbf{B} = (0, 1)^\top.$$

Using Algorithm 1 we can find an interpolating RCL arc matching given data, see Fig. 1 (left).

Example 3.3 Let us consider G^1 Hermite data

$$\mathbf{A} = (-1, -1)^\top, \mathbf{B} = (1, 0)^\top, \mathbf{t}_\mathbf{A} = \left(-\frac{3}{\sqrt{13}}, \frac{2}{\sqrt{13}} \right)^\top, \mathbf{t}_\mathbf{B} = \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}} \right)^\top.$$

Using Algorithm 1 we can find an interpolating RCL arc matching given data, see Fig. 1 (right).

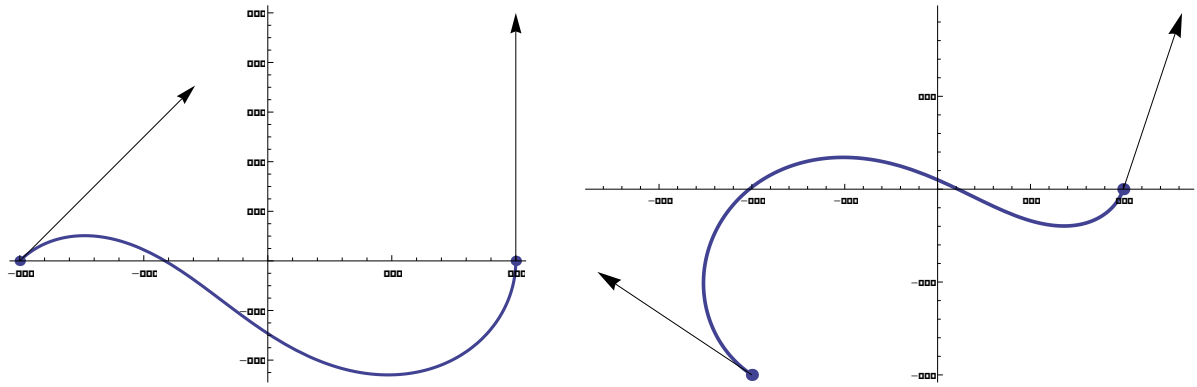


Figure 1: Left: RCL interpolant for data from Example 3.2; Right: RCL interpolant for data from Example 3.3.

4 Approximation order and examples

Using Algorithm 1 we can formulate a simple non-adaptive method for conversion of an arbitrary planar curve $\mathbf{c}(t)$, $t \in [0, 1]$, with well-defined tangent vector everywhere into G^1 RCL spline.

We choose the number of segments n on $\mathbf{c}(t)$ and sample G^1 Hermite data

$${}^i\mathbf{A} = \mathbf{c}\left(\frac{i-1}{n}\right), {}^i\mathbf{B} = \mathbf{c}\left(\frac{i}{n}\right), {}^i\mathbf{t}_\mathbf{A} = \frac{\mathbf{c}'\left(\frac{i-1}{n}\right)}{\|\mathbf{c}'\left(\frac{i-1}{n}\right)\|}, {}^i\mathbf{t}_\mathbf{B} = \frac{\mathbf{c}'\left(\frac{i}{n}\right)}{\|\mathbf{c}'\left(\frac{i}{n}\right)\|}, \quad i = 1, \dots, n.$$

Then we compute the corresponding RCL curve ${}^i\mathbf{p}(u)$ matching these data.

Further, we evaluate the approximation error by measuring the Hausdorff distance

$$\varepsilon = \max \left(\max_{u \in [0,1]} \left(\min_{t \in [0,1]} \|\mathbf{c}(t) - \mathbf{p}^n(u)\| \right), \max_{t \in [0,1]} \left(\min_{u \in [0,1]} \|\mathbf{c}(t) - \mathbf{p}^n(u)\| \right) \right), \quad (16)$$

where $\mathbf{p}^n(u)$, $u \in [0, 1]$, is RCL spline curve obtained by the linear reparameterization. If ε is greater than the prescribed error ϵ , we set $n = 2n$ and repeat the whole process.

Example 4.1 Let us consider two Bézier quartic curves on the interval $t \in [0, 1]$. The first curve has the control points $(0, 0)^\top$, $(0, 1)^\top$, $(1, 2)^\top$, $(2, 1)^\top$ and $(1, 0)^\top$ and no inflections (see Fig. 2). The second curve has the control points $(0, 0)^\top$, $(2, 1)^\top$, $(2, -1)^\top$, $(3, 2)^\top$ and $(4, 0)^\top$ and two inflection points (see Fig. 3). Table 1 summarizes the approximation error and its improvement (ratio of two consecutive errors) for the first Bézier curve. The error was obtained by measuring the Hausdorff distance. The improvement ratio tends to $8 = 2^3$ which indicates that the approximation order of the curve approximation by the above mentioned procedure is 3.

5 Conclusion

In this paper, a simple G^1 Hermite interpolation algorithm with arcs of RCL curves was presented. The method is based only on solving a small system of linear equations. We also showed

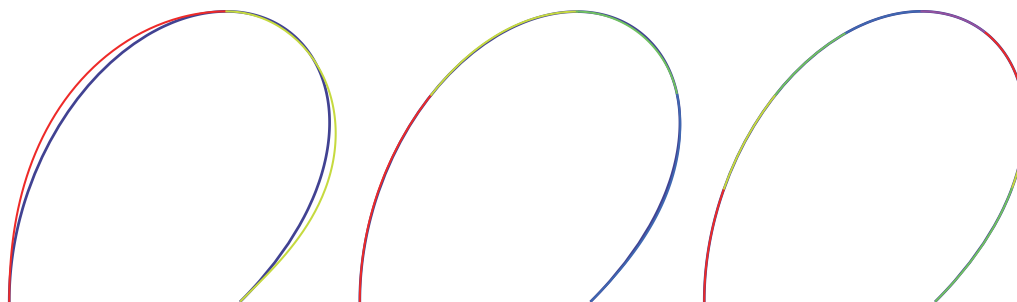


Figure 2: Bézier curve and its approximation by G^1 RCL spline curve for $n = 1$ (left), $n = 2$ (middle) and $n = 3$ (right).

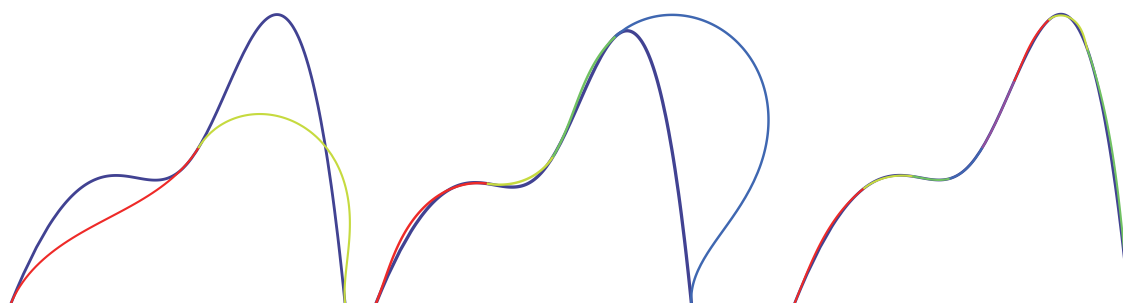


Figure 3: Bézier curve with inflection points and its approximation by G^1 RCL spline curve for $n = 1$ (left), $n = 2$ (middle) and $n = 3$ (right).

how to approximate an arbitrary planar curve with well-defined tangent vectors everywhere by RCL spline curve and numerically derived the approximation order of this procedure which turned out to be 3.

Based on the results from [3], the method could be also generalized for an interpolation/approximation of space curves and also even of surfaces because general formulas for all space RCL curves and also for all RCL surfaces are presented there. This will be our future work on this topic.

Acknowledgement

The work on this topic was supported by the Research Plan MSM 4977751301.

References

- [1] BASTL, B., JÜTTLER, B., LÁVIČKA, M., SCHICHO, J., and ŠÍR, Z. Spherical quadratic Bézier triangles with chord lengths parameterization and tripolar coordinates in space. *Comput. Aided Geom. Des.* 28, 2 (2011), 127–134.
- [2] BASTL, B., JÜTTLER, B., LÁVIČKA, M., and ŠÍR, Z. Surfaces with Rational Chord Length Parameterization. *Lecture Notes in Computer Science 6130* (2010), 19–28. Mourrain, B, Schaefer, S, and Xu, G (eds.): GMP 2010.

Table 1: Errors and their improvements (ratios of two consecutive errors) for Example 4.1.

Parts	Error	Ratio
2	4.02435×10^{-2}	
4	1.93839×10^{-3}	20.76123
8	1.47523×10^{-4}	13.13961
16	4.24730×10^{-5}	3.47333
32	2.08335×10^{-5}	2.03868
64	2.98628×10^{-6}	6.97641
128	3.66785×10^{-7}	8.14177
256	4.54321×10^{-8}	8.07325
512	5.65271×10^{-9}	8.03722
1024	7.04936×10^{-10}	8.01876
2048	8.80134×10^{-11}	8.00941
4096	1.09951×10^{-11}	8.00471

- [3] BASTL, B., JÜTTLER, B., LÁVIČKA, M., and ŠÍR, Z. Curves and surfaces with rational chord lengths parameterization. *Computer Aided Geometric Design* (20xx). To appear.
- [4] BATEMAN, H. Spheroidal and bipolar coordinates. *Duke Mathematical Journal* 4, 1 (1938), 39–505.
- [5] FARIN, G. Rational quadratic circles are parametrized by chord length. *Computer Aided Geometric Design* 23, 9 (2006), 722–724.
- [6] FAROUKI, R. *Pythagorean-Hodograph Curves: Algebra and Geometry Inseparable*. Springer, 2008.
- [7] FAROUKI, R., and SAKKALIS, T. Pythagorean hodographs. *IBM Journal of Research and Development* 34, 5 (1990), 736–752.
- [8] FAROUKI, R. T., and MOON, H. P. Bipolar and multipolar coordinates. In *Proceedings of the 9th IMA Conference on the Mathematics of Surfaces* (London, UK, 2000), Springer-Verlag, pp. 348–371.
- [9] FLOATER, M. S., and SURAZHISKY, T. Parameterization for curve interpolation. In *Topics in Multivariate Approximation and Interpolation*, W. H. R. S. Kurt Jetter, Martin D. Buhmann and J. Stekler, Eds., vol. 12 of *Studies in Computational Mathematics*. Elsevier, 2006, pp. 39–54.
- [10] LÜ, W. Curves with chord length parameterization. *Computer Aided Geometric Design* 26, 3 (2009), 342–350.
- [11] SABIN, M. A., and DODGSON, N. A. A circle-preserving variant of the four-point subdivision scheme. In *Mathematical Methods for Curves and Surfaces: Tromsø 2004, Modern Methods in Mathematics* (2005), Nashboro Press, pp. 275–286.
- [12] SÁNCHEZ-REYES, J. Complex rational Bézier curves. *Computer Aided Geometric Design* 26, 8 (2009), 865–876.
- [13] SÁNCHEZ-REYES, J., and FERNÁNDEZ-JAMBRINA, L. Curves with rational chord-length parametrization. *Computer Aided Geometric Design* 25, 4-5 (2008), 205–213.

Current address

Bastl Bohumír, Ing., Ph.D.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic,
tel. ++420 377 632655, e-mail: bastl@kma.zcu.cz

Lávička Miroslav, doc., RNDr., Ph.D.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic,
tel. ++420 377 632619, e-mail: lavicka@kma.zcu.cz

ON COMPUTING APPROXIMATE PARAMETERIZATIONS
OF ALGEBRAIC SURFACES

BIZZARRI Michal, (CZ), LÁVIČKA Miroslav, (CZ)

Abstract. In this paper we present a novel approach which can be used for computing approximate parameterizations of selected algebraic surfaces. The algorithm is based on determining the *topological graph* of a given surface \mathcal{S} and replacing its triangular faces by suitable triangular Bézier patches. The computation of the topological graph of \mathcal{S} uses the *critical curves* of \mathcal{S} and their topological graphs, see [1]. The union of the specially constructed topological graphs of the critical curves of \mathcal{S} leads to the topological graph of \mathcal{S} .

Key words and phrases. Algebraic curve, algebraic surface, critical point, topological graph, triangular Bézier patch.

Mathematics Subject Classification. Primary 60A05, 08A72; Secondary 28E10.

1 Introduction

When the computers allowed machining of 3D shapes, a necessity to define a computer-compatible description of those objects appeared. The most promising representation was soon identified to be in forms of parametric curves and surfaces. Parameterizations allow us to generate points on curves and surfaces, they are also suitable for surface plotting, computing transformations, determining offsets, computing curvatures e.g. for shading and colouring, for surface-surface intersection problems etc. – see e.g. [2, 3]. Among all parameterizations, the most important ones are those that can be described with the help of polynomials or rational functions since these representations can be easily included into standard CAD systems. However, not every geometric object (curve/surface/volume) can be described using rational parameterizations and therefore approximate techniques must be applied.

For algebraic curves and surfaces, various symbolic techniques can be found – see e.g. [4, 5, 6] for parameterization of curves, and [7, 8, 9] for surfaces. However, these techniques are algorithmically quite difficult and they cannot be used for all algebraic curves and surfaces, since

an exact rational parameterization does not exist in the generic case. Approximate algorithms, which generate a parameterization within a certain region of interest, are used to overcome these problems. Hence, it is worth exploring also approximate techniques for parameterizing planar curves, space curves, and surfaces. Various related results for planar curves exist, cf. [10, 11]. Numerical methods for space curves have been discussed e.g. in [1, 12].

In this paper we present a novel approach which can be used for computing approximate parameterizations of selected algebraic surfaces. The algorithm is based on determining the *topological graph* of a given surface \mathcal{S} without curves of singular points and replacing its triangular faces by suitable triangular Bézier patches. The first step for finding the topological graph of \mathcal{S} is an identification of the *critical curves* of \mathcal{S} followed by computing their topological graphs, see [1]. Then, the union of the specially constructed topological graphs of the critical curves of \mathcal{S} leads to the topological graph of \mathcal{S} . The functionality of the designed method is presented on two examples.

The remainder of this paper is organized as follows. The next section summarizes several fundamental facts concerning the algebraic varieties, especially algebraic curves and surfaces. The topology of space algebraic curves is studied in Section 3. Section 4 is devoted to the computation of the topological graphs of algebraic surfaces. Finally, after presenting the method for constructing the approximate parameterizations of algebraic surfaces based on their topological graphs in Section 5, we conclude this paper and mention some open problems.

2 Preliminaries

In this section, we recall some fundamental properties of algebraic and rational curves and surfaces which are then used in the following sections. More details can be found e.g. in [5, 6, 13]. Throughout this paper, let \mathbb{K} be an algebraically closed field of characteristic zero and the affine space of dimension n over the field \mathbb{K} will be denoted by \mathbb{A}^n .

An *affine algebraic variety* \mathcal{V} in \mathbb{A}^n is defined as the set of all points satisfying $f_1(x_1, \dots, x_n) = \dots = f_k(x_1, \dots, x_n) = 0$, i.e.,

$$\mathcal{V} = \{(a_1, \dots, a_n)^\top \in \mathbb{A}^n \mid f_i(a_1, \dots, a_n) = 0 \text{ for all } i = 1, \dots, k\}. \quad (1)$$

The polynomials $f_1, \dots, f_k \in \mathbb{K}[x_1, \dots, x_n]$ are called *defining polynomials* of the variety \mathcal{V} . The degree of the variety \mathcal{V} is $d_1 \cdots d_k$, where d_1, \dots, d_k are the algebraic degrees of f_1, \dots, f_k , respectively. The *dimension* of \mathcal{V} is the transcendence degree over \mathbb{K} of the function field $\mathbb{K}(\mathcal{V})$ of all rational functions on \mathcal{V} , with values in \mathbb{K} , see [13].

In this paper we will focus on algebraic plane and space curves (the algebraic varieties of dimension 1) in \mathbb{A}^2 and \mathbb{A}^3 , respectively) and on algebraic surfaces (the algebraic varieties of dimension 2) in \mathbb{A}^3 . An *affine plane algebraic curve* \mathcal{D} is the set of zeros of a polynomial, i.e.,

$$\mathcal{D} = \{(a_1, a_2)^\top \in \mathbb{A}^2 \mid f(a_1, a_2) = 0\}. \quad (2)$$

An *affine space algebraic curve* \mathcal{C} is defined as the set all solutions of a system of two polynomial equations, i.e.,

$$\mathcal{C} = \{(a_1, a_2, a_3)^\top \in \mathbb{A}^3 \mid f(a_1, a_2, a_3) = g(a_1, a_2, a_3) = 0\}. \quad (3)$$

Finally, an *affine algebraic surface* \mathcal{S} is the set of zeros of a polynomial, i.e.,

$$\mathcal{S} = \{(a_1, a_2, a_3)^\top \in \mathbb{A}^3 \mid f(a_1, a_2, a_3) = 0\}. \quad (4)$$

Let \mathcal{V} be a variety of dimension d over a field \mathbb{K} . Then \mathcal{V} is said to be *unirational*, or *parametric*, if there exists a rational map $\mathcal{P} : \mathbb{K}^d \rightarrow \mathcal{V}$ such that $\mathcal{P}(\mathbb{K}^d)$ is dense in \mathcal{V} . We speak about a (*rational*) *parameterization* $\mathcal{P}(t_1, \dots, t_d)$ of \mathcal{V} . Furthermore, if \mathcal{P} defines a birational map then \mathcal{V} is called *rational*, and we say that $\mathcal{P}(t_1, \dots, t_d)$ is a *proper parametrization*. By a theorem of Riemann, a planar curve has a parametrization iff it has a proper parametrization iff its genus (see [13] and Section 2.1 for a definition of this notion) is zero. Hence, for planar curves the notions of rationality and unirationality are equivalent for any field. In the surface case, the theory differs as Castelnuovo's theorem holds only for algebraically closed fields of characteristic zero. By this theorem, a surface has a parametrization iff it has a proper parametrization iff the arithmetical genus p_a and the second plurigenus P_2 are both zero (see [13] for a definition of these notions).

Finally, we recall basic property of triangular Bézier patches, see [3]. Bézier triangular patch, constructed from a triangular domain is defined as

$$\mathbf{x}(\mathbf{u}) = \sum_{i+j+k=n} \mathbf{b}_{ijk} B_i^n(\mathbf{u}), \quad (5)$$

where $\mathbf{u} = (u, v, w)^\top$ are barycentric coordinates fulfilling $u+v+w = 1$, $B_i^n(\mathbf{u})$ are the Bernstein polynomials in the form

$$B_i^n(\mathbf{u}) = \frac{n!}{i!j!k!} u^i v^j w^k, \quad (6)$$

and \mathbf{b}_{ijk} are a given control points determining the shape of the surface. A Bézier triangle consists of all points $\mathbf{x}(\mathbf{u})$ with the barycentric coordinates \mathbf{u} within the domain triangle, $0 < u, v, w < 1$.

3 Topology of space algebraic curves

In this section we give a brief sketch of computing the topology of implicitly given space algebraic curves in some region of interest. By determining the topology of a space curve \mathcal{C} we understand a construction of a certain arrangement of polylines which is topologically equivalent to the given curve. Such arrangement of polylines is called a *topological graph* of \mathcal{C} and is denoted by $G(\mathcal{C})$. It is beyond the scope of this paper to go into details, so we recall only basic steps – the reader more interested in this topic is kindly referred to [1, 14].

Let \mathcal{C} be a space curve defined by $f = g = 0$ and $\mathbf{t}_{\mathbf{p}} = \nabla f(\mathbf{p}) \times \nabla g(\mathbf{p})$ denote the tangent vector of \mathcal{C} at \mathbf{p} . Then the point $\mathbf{p} \in \mathcal{C}$ is called:

- (i) an *x-critical point* if $\mathbf{t}_{\mathbf{p}} \cdot (1, 0, 0)^\top = 0$;
- (ii) a *y-critical point* if $\mathbf{t}_{\mathbf{p}} \cdot (0, 1, 0)^\top = 0$;
- (iii) a *z-critical point* if $\mathbf{t}_{\mathbf{p}} \cdot (0, 0, 1)^\top = 0$;
- (iv) a *singular point* if $\mathbf{t}_{\mathbf{p}} = (0, 0, 0)^\top$.

All these points are called by a unified name *critical points*. A *non-critical point* is a point which is not critical.

Now, we shortly outline the steps of the construction of the topological graph $G(\mathcal{C})$ in the region of interest $\mathfrak{R} = [X_1, X_2] \times [Y_1, Y_2] \times [Z_1, Z_2]$ of a space algebraic curve \mathcal{C} . Firstly, we need \mathcal{C} to be in the so called *general space position*, see [1] for an explanation of this notion. Next we compute all critical points of \mathcal{C} – in the algorithm for computing the topology of algebraic surfaces, we need the topological graphs of some specially chosen space curves containing all their critical points, see Section 4. Next we add all the intersections of \mathcal{C} with the boundary of \mathfrak{R} to the set of the vertices of $G(\mathcal{C})$. Thus, we obtain new points $\mathbf{b}_1, \dots, \mathbf{b}_s$ by solving the following six systems of three non-linear equations:

$$\begin{aligned} f(x, y, z) = 0, \quad \frac{\partial f(x, y, z)}{\partial z} = 0, \quad x = X_i, \quad i = 1, 2; \\ f(x, y, z) = 0, \quad \frac{\partial f(x, y, z)}{\partial z} = 0, \quad y = Y_i, \quad i = 1, 2; \\ f(x, y, z) = 0, \quad \frac{\partial f(x, y, z)}{\partial z} = 0, \quad z = Z_i, \quad i = 1, 2. \end{aligned} \tag{7}$$

Obviously, only one branch of \mathcal{C} (assuming \mathbf{b}_i is not an x -critical point of \mathcal{C}) goes from each point \mathbf{b}_i . Next, we find the corresponding projection $\pi_z(\mathcal{C})$ of \mathcal{C} and construct the planar topological graph $G(\pi_z(\mathcal{C}))$ containing the projections of all vertices of $G(\mathcal{C})$. Finally, we delete some redundant vertices of $G(\pi_z(\mathcal{C}))$ and lift its edges back to space, see Algorithm 1.

Algorithm 1 TOP_GRAPH_CURVE [f, g, \mathfrak{R}]

INPUT: A space curve \mathcal{C} defined by the polynomials f and g and some region of interest $\mathfrak{R} = [X_1, X_2] \times [Y_1, Y_2] \times [Z_1, Z_2]$.

- 1: Compute the x -coordinates $X_1 \leq x_1 < \dots < x_k \leq X_2$ of the x - y - and z -critical points of \mathcal{C} , of the singular points of $\pi_z(\mathcal{C})$ obtained only by the projection π_z and of the boundary points (by solving (7));
- 2: For every x_i , compute the real roots of $h(x_i, y)$: $Y_1 \leq y_{i,1} < y_{i,2} < \dots < y_{i,s} \leq Y_2$, where $h = \text{Res}_z(f, g)$;
- 3: Delete such points $(x_i, y_{i,j})^\top$ which have not the corresponding z -coordinates in the interval $[Z_1, Z_2]$;
- 4: For all points $(x_i, y_{i,j})^\top$ compute the number of left and right branches going from these points;
- 5: Connect the points $(x_i, y_{i,j})^\top$ with the points $(x_{i+1}, y_{i+1,j})^\top$ appropriately;
- 6: Delete such singular points of $\pi_z(\mathcal{C})$ which are not the projections of the singular points of \mathcal{C} , see [1];
- 7: Delete all vertices of $G(\pi_z(\mathcal{C}))$ which are not the projections of x - y - z -critical and boundary points;
- 8: Lift the edges of $G(\pi_z(\mathcal{C}))$ to the space and hence obtain $G(\mathcal{C})$;

OUTPUT: The topological graph $G(\mathcal{C})$ of \mathcal{C} having its x - y - z -critical and the boundary points as its vertices.

Thus, we need to compute the number of the left branches and right branches going from a particular point \mathbf{p} (Step 4 in Algorithm 1) on the plane curve $\pi_z(\mathcal{C})$. This can be done by the following method: We enclose \mathbf{p} by a small box \mathfrak{B} such that the curve $\pi_z(\mathcal{C})$ does not intersect the box \mathfrak{B} in the bottom and in the top and, moreover, there exists exactly one intersection

point (the point \mathbf{p}) of the vertical line going through the point \mathbf{p} with the curve $\pi_z(\mathcal{C})$ in the box. Then the number of the right and left intersection points yields the number of half-branches to the right and to the left at \mathbf{p} , respectively, cf. [15] for more details.

Finally, we show how the vertices of $G(\pi_z(\mathcal{C}))$ can be connected. We write the sequence R_i of vertices $(x_i, y_{i,j})^\top$ where each vertex occurs as many times as it has half-branches to the right, and the sequence L_{i+1} of vertices $(x_{i+1}, y_{i+1,j})^\top$ where each vertex again occurs as many times as it has half-branches to the left, for particular i . Then, the m -th vertex from R_i is connected with the m -th vertex from L_{i+1} . Note that the way how to connect the vertices by edges is uniquely determined since any incorrect connecting vertices leads to at least one intersection of two edges at a non-critical point.

Remark 3.1 Let us emphasize that if a given curve is in the general space position, the lifting process is determined uniquely, see [1] for the explanation.

4 Topology of algebraic surfaces

This section is devoted to the construction of the topological graphs of algebraic surfaces without the curves of singular points. We compute the critical points of a given algebraic surface \mathcal{S} defined by the polynomial $f(x, y, z)$ and connect them appropriately (with the help of the topologies of some specially chosen curves on \mathcal{S}).

Let $\mathbf{n}_\mathbf{p} = \nabla f(\mathbf{p})$ denote the normal vector of the surface \mathcal{S} at \mathbf{p} . Then the point $\mathbf{p} \in \mathcal{S}$ is called:

- (i) an *x-critical point* if $\mathbf{n}_\mathbf{p} \cdot (0, 1, 0)^\top = \mathbf{n}_\mathbf{p} \cdot (0, 0, 1)^\top = 0$;
- (ii) a *y-critical point* if $\mathbf{n}_\mathbf{p} \cdot (1, 0, 0)^\top = \mathbf{n}_\mathbf{p} \cdot (0, 0, 1)^\top = 0$;
- (iii) a *z-critical point* if $\mathbf{n}_\mathbf{p} \cdot (1, 0, 0)^\top = \mathbf{n}_\mathbf{p} \cdot (0, 1, 0)^\top = 0$;
- (iv) a *singular point* if $\mathbf{n}_\mathbf{p} = (0, 0, 0)^\top$.

All these points are called by a unified name *critical points*. A *non-critical point* is a point which is not critical.

Then the algorithm for computing the topology of a given algebraic surface proceeds as follows: First, we compute the topological graphs $G(\mathcal{C}_x)$, $G(\mathcal{C}_y)$ and $G(\mathcal{C}_z)$ of the so called *critical curves* \mathcal{C}_x , \mathcal{C}_y and \mathcal{C}_z of a given surface \mathcal{S} (defined by the polynomial f), i.e.,

$$\begin{aligned} \mathcal{C}_x : & \quad f(x, y, z) = f_x(x, y, z) = 0; \\ \mathcal{C}_y : & \quad f(x, y, z) = f_y(x, y, z) = 0; \\ \mathcal{C}_z : & \quad f(x, y, z) = f_z(x, y, z) = 0, \end{aligned} \tag{8}$$

where f_α denotes the derivative of f with respect to α . Next, we compute the topological graphs $G(\mathcal{C}_1), \dots, G(\mathcal{C}_6)$ of the boundary curves $\mathcal{C}_1, \dots, \mathcal{C}_6$ – the intersection curves of \mathcal{S} with the boundary of $\mathfrak{R} = [X_1, X_2] \times [Y_1, Y_2] \times [Z_1, Z_2]$, i.e.

$$\begin{aligned} \mathcal{C}_{1,2} : & \quad f(x, y, z) = 0, & x = X_i, & \quad i = 1, 2; \\ \mathcal{C}_{3,4} : & \quad f(x, y, z) = 0, & y = Y_i, & \quad i = 1, 2; \\ \mathcal{C}_{5,6} : & \quad f(x, y, z) = 0, & z = Z_i, & \quad i = 1, 2. \end{aligned} \tag{9}$$

Then all 9 topological graphs together generate the topological graph of a given surface \mathcal{S} . Since the x -critical points of \mathcal{C}_y , \mathcal{C}_z and \mathcal{S} coincide, see Theorem 4.1 (analogously for y - and z -critical points) the topological graphs $G(\mathcal{C}_x)$, $G(\mathcal{C}_y)$ and $G(\mathcal{C}_z)$ are connected together at these points (the critical points of \mathcal{S}). Next, since all topological graphs $G(\mathcal{C}_x)$, $G(\mathcal{C}_y)$ and $G(\mathcal{C}_z)$ contain the boundary points as its vertices, the boundary curves will connect them at these points.

Theorem 4.1 *Let an algebraic surface \mathcal{S} is defined by the polynomial f and $\mathcal{C}_x, \mathcal{C}_y$ and \mathcal{C}_z be its corresponding critical curves. Then it holds:*

- (1) x -critical points of \mathcal{C}_y , \mathcal{C}_z and \mathcal{S} coincide;
- (2) y -critical points of \mathcal{C}_x , \mathcal{C}_z and \mathcal{S} coincide;
- (3) z -critical points of \mathcal{C}_x , \mathcal{C}_y and \mathcal{S} coincide.

Proof. We prove only (1) since (2) and (3) can be proved by analogy. Thus, x -critical points of \mathcal{C}_y are defined by the equations

$$f = f_y = f_y f_{yz} - f_z f_{yy} = 0. \quad (10)$$

Since $f_y = 0$, the equation $f_y f_{yz} - f_z f_{yy} = 0$ degenerates to the equation $f_z = 0$. Next, x -critical points of \mathcal{C}_z are defined by the equations

$$f = f_z = f_y f_{zz} - f_z f_{yz} = 0. \quad (11)$$

Again, since $f_z = 0$, the equation $f_y f_{zz} - f_z f_{yz} = 0$ degenerates to the equation $f_y = 0$. Thus in both case we arrived at the equations

$$f = f_y = f_z = 0, \quad (12)$$

which are exactly the equations of the x -critical points of \mathcal{S} . □

Remark 4.2 Let us note, that some of the curves $\mathcal{C}_x, \mathcal{C}_y, \mathcal{C}_z, \mathcal{C}_1, \dots, \mathcal{C}_6$ can have a common component. We can delete some of these common components by considering new curves $\mathcal{K}_1, \dots, \mathcal{K}_s$ defined by the tuples of polynomials $(f, g_1), \dots, (f, g_s)$, where g_1, \dots, g_s are irreducible factors of $f_x, f_y, f_z, x - X_1, x - X_2, y - Y_1, y - Y_2, z - Z_1, z - Z_2$.

The method of constructing the topological graph is summarized in Algorithm 2.

Example 4.3 We construct the topological graph of the ellipsoid given by

$$f = \frac{x^2}{3} + \frac{y^2}{2} + z^2 - 1. \quad (13)$$

The critical curves are shown in Fig. 1 (left) and the topological graph $G(\mathcal{S})$ of \mathcal{S} in Fig. 1 (right).

Algorithm 2 TOP_GRAPH_SURFACE $[f, \mathfrak{R}]$

INPUT: An algebraic surface \mathcal{S} defined by the polynomial f and some region of interest $\mathfrak{R} = [X_1, X_2] \times [Y_1, Y_2] \times [Z_1, Z_2]$.

- 1: Construct curves $\mathcal{K}_1, \dots, \mathcal{K}_s$ defined by the tuples of polynomials $(f, g_1), \dots, (f, g_s)$, where g_1, \dots, g_s are irreducible factors of $f_x, f_y, f_z, x - X_1, x - X_2, y - Y_1, y - Y_2, z - Z_1, z - Z_2$.
- 2: Compute the topological graphs for all \mathcal{K}_i , i.e., $G(\mathcal{K}_i) = \text{TOP_GRAPH_CURVE}[f, g_i, \mathfrak{R}]$;
- 3: The topological graph $G(\mathcal{S}) = \bigcup_{i=1, \dots, s} G(\mathcal{K}_i)$;

OUTPUT: The topological graph $G(\mathcal{S})$ of \mathcal{S} having its x - y - z -critical and the boundary points as its vertices.

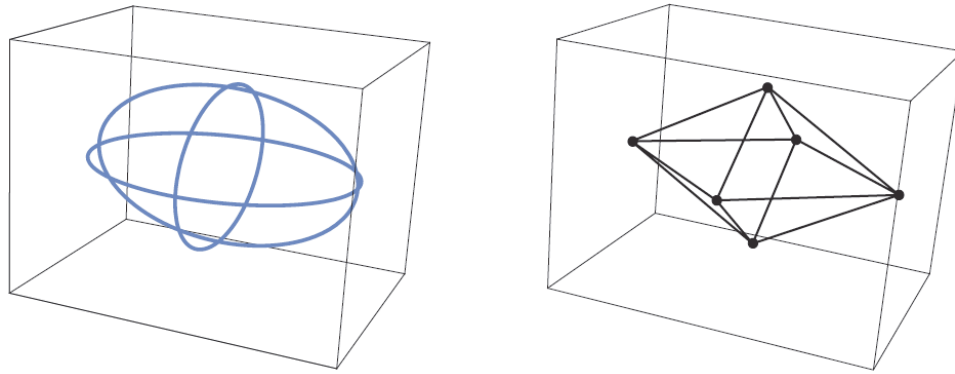


Figure 1: The critical and boundary curves from Example 4.3 (left), and their corresponding topological graphs which give us the topological graph $G(\mathcal{S})$ of \mathcal{S} (right).

Example 4.4 We construct the topological graph of the algebraic surface (see Fig. 2) defined by the polynomial

$$f = x^3 - 9y^2 - 9z^2 - 3x + 2. \quad (14)$$

The critical and boundary curves are defined by equations, see Fig. 3 (left):

$$\begin{aligned} \mathcal{C}_x : & -x^3 + 3x + 9y^2 + 9z^2 - 2 = 1 - x^2 = 0; \\ \mathcal{C}_y : & -x^3 + 3x + 9y^2 + 9z^2 - 2 = y = 0; \\ \mathcal{C}_z : & -x^3 + 3x + 9y^2 + 9z^2 - 2 = z = 0; \\ \mathcal{C}_1 : & -x^3 + 3x + 9y^2 + 9z^2 - 2 = x - 3 = 0. \end{aligned} \quad (15)$$

Thus, we construct the topological graphs $G(\mathcal{C}_x), G(\mathcal{C}_y), G(\mathcal{C}_z), G(\mathcal{C}_1)$ of $\mathcal{C}_x, \mathcal{C}_y, \mathcal{C}_z, \mathcal{C}_1$, which give us the topological graph $G(\mathcal{S})$ of \mathcal{S} , see Fig. 3 (right).

5 An approximate parameterization algorithm

In this section we describe a simple method how the triangular faces of the constructed topological graph of a given algebraic surface can be replaced by cubic Bézier triangular patches. The method is formulated as an optimization problem when the objective function approximates the integral of the squared Euclidean distance of the constructed approximate surface to the implicit one.

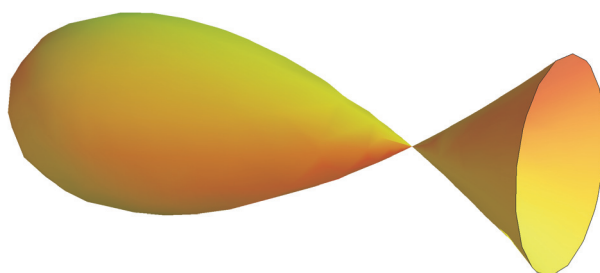


Figure 2: Tschirnhausen cubic surface \mathcal{S} from Example 4.4.

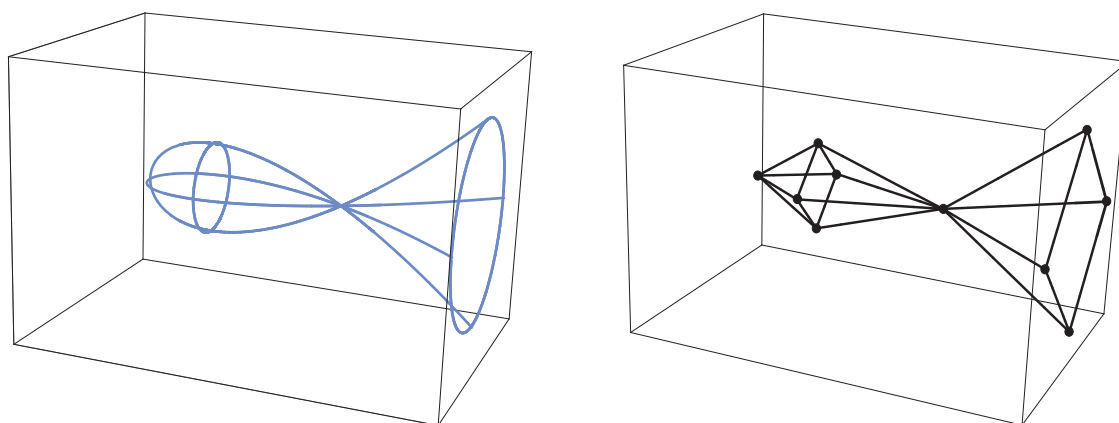


Figure 3: The critical and boundary curves from Example 4.4 (left), and their corresponding topological graphs which give us the topological graph $G(\mathcal{S})$ of \mathcal{S} (right).

Suppose, that the three points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ and the normal vectors $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ at these points are given. Then the cubic Bézier triangle patch is given by the equation

$$\begin{aligned}
 \mathbf{x}(u, v, w) &= \sum_{i+j+k=3} \mathbf{b}_{ijk} \frac{3!}{i!j!k!} u^i v^j w^k = \\
 &= \mathbf{b}_{300} u^3 + 3\mathbf{b}_{210} u^2 v + 3\mathbf{b}_{201} u^2 w + 3\mathbf{b}_{120} u v^2 + 6\mathbf{b}_{111} u v w + \\
 &\quad + 3\mathbf{b}_{102} u w^2 + \mathbf{b}_{030} v^3 + 3\mathbf{b}_{021} v^2 w + 3\mathbf{b}_{012} v w^2 + \mathbf{b}_{003} w^3,
 \end{aligned} \tag{16}$$

where $u + v + w = 1$. According to [16] the control points \mathbf{b}_{ijk} can be computed as follows:

$$\begin{aligned} \mathbf{b}_{300} &= \mathbf{p}_1; \\ \mathbf{b}_{030} &= \mathbf{p}_2; \\ \mathbf{b}_{003} &= \mathbf{p}_3; \\ \mathbf{b}_{210} &= (2\mathbf{p}_1 + \mathbf{p}_2 - w_{12}\mathbf{n}_1)/3; \\ \mathbf{b}_{120} &= (2\mathbf{p}_2 + \mathbf{p}_1 - w_{21}\mathbf{n}_2)/3; \\ \mathbf{b}_{021} &= (2\mathbf{p}_2 + \mathbf{p}_3 - w_{23}\mathbf{n}_2)/3; \\ \mathbf{b}_{012} &= (2\mathbf{p}_3 + \mathbf{p}_2 - w_{32}\mathbf{n}_3)/3; \\ \mathbf{b}_{102} &= (2\mathbf{p}_3 + \mathbf{p}_1 - w_{31}\mathbf{n}_3)/3; \\ \mathbf{b}_{201} &= (2\mathbf{p}_1 + \mathbf{p}_3 - w_{13}\mathbf{n}_1)/3; \\ \mathbf{b}_{111} &= E + (E - V)/2, \end{aligned} \quad (17)$$

where $w_{ij} = (\mathbf{p}_j - \mathbf{p}_i) \cdot \mathbf{n}_i$, $E = (\mathbf{b}_{12} + \mathbf{b}_{21} + \mathbf{b}_{102} + \mathbf{b}_{120} + \mathbf{b}_{201} + \mathbf{b}_{210})/6$ and $V = (\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3)/3$.

Then, the whole process of approximate parameterization of algebraic surfaces can be formulated as follows: For each triangle face T_i of the topological graph $G(\mathcal{S})$ construct its corresponding Bézier patch $\mathbf{x}(u, v, \alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3})$ such that we substitute to (17) the normal vectors with the lengths as free parameters i.e., $\alpha_{i_1}\mathbf{n}_{i_1}$, $\alpha_{i_2}\mathbf{n}_{i_2}$ and $\alpha_{i_3}\mathbf{n}_{i_3}$. Then, the following objective function measures the Euclidean distance of the constructed approximate surface to the implicit one

$$F(\alpha_1, \dots, \alpha_k) = \int_0^1 \left(\int_0^{1-u} \left(\sum_{i=1}^{k/3} \frac{f^2(\mathbf{x}(u, v, \alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}))}{\|\nabla f(\mathbf{x}(u, v, \alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}))\|^2} \right) dv \right) du. \quad (18)$$

Finally, minimizing the objective function (18) can be done by the Newton's method which is characterized by the fast convergence, see for instance [17].

Example 5.1 We construct one Bézier triangular patch on the ellipsoid from Example 4.4. The Bézier patch is given by the following data:

$$\begin{aligned} \mathbf{p}_1 &= (1.732, 0, 0)^\top, & \mathbf{p}_2 &= (0, 1.414, 0)^\top, & \mathbf{p}_3 &= (0, 0, 1)^\top \\ \mathbf{n}_1 &= (1, 0, 0)^\top, & \mathbf{n}_2 &= (0, 1, 0)^\top, & \mathbf{n}_3 &= (0, 0, 1)^\top. \end{aligned} \quad (19)$$

The optimization process gives us the following lengths of the normal vectors

$$\alpha_1 = \alpha_2 = \alpha_3 \doteq 1.33108. \quad (20)$$

The approximate Bézier triangular patch is shown in Fig. 4.

6 Conclusion

In this paper, we presented a novel approach for computing approximate parameterizations of algebraic surfaces based on computing their topological graphs and consequently replacing their triangular faces by triangular Bézier patches. The optimization of the lengths of the normal vectors at the join points is realized using the classical Newton's method. The main advantage of the presented method, combining symbolical and numerical steps to the approximation problem, lies in its straightforwardness and simplicity.

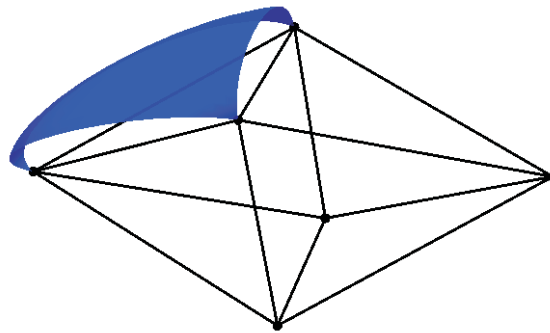


Figure 4: B ezier triangular patch from Example 5.1.

On the other hand, there is not guaranteed that the constructed topological graph will contain only triangular faces. In such cases some additional connections of its vertices must be constructed – this is the subject for our further research. Furthermore, it is an open problem how to deal with the surfaces having the curves of singular points.

Acknowledgement

The work on this paper was supported by the Research Plan MSM 4977751301.

References

- [1] Michal BIZZARRI and Miroslav L AVI KA. A symbolic-numerical approach to approximate parameterizations of space curves using reduced topological graphs. Submitted to Journal of Computational and Applied Mathematics, 2011.
- [2] Gerald FARIN. Curves and Surfaces for Computer-Aided Geometric Design. Academic Press, 1988.
- [3] Gerald FARIN, Josef HOSCHEK and Myung-Soo KIM, editors. Handbook of Computer Aided Geometric Design. Elsevier, 2002.
- [4] Michal BIZZARRI and Miroslav L AVI KA. Algorithm for parameterization of rational curves revisited. Journal for Geometry and Graphics, 15(1):118, 2011.
- [5] Juan Rafael SENDRA and Franz WINKLER. Symbolic parametrization of curves. Journal of Symbolic Computation, 12(6):607631, 1991.
- [6] Robert J. WALKER. Algebraic Curves. Princeton University Press, 1950.
- [7] Martin PETERNELL and Helmut POTTMANN. Computing rational parametrizations of canal surfaces. Journal of Symbolic Computation, 23:255266, February 1997.
- [8] Josef SCHICHO. Rational parametrization of surfaces. Journal of Symbolic Computation, 26(1):129, 1998.
- [9] Ibolya SZIL GYI, Bert J UTTNER, and Josef Schicho. Local parametrization of cubic surfaces. Journal of Symbolic Computation, 41:3048, January 2006.
- [10] Chandrajit BAJAJ and Guoliang XU. Piecewise rational approximations of real algebraic curves. Journal of Computational Mathematics, 15:5571, 1997.

- [11] Bert J"UTTLE and Pavel CHALMOVIANSKÝ. Approximate parameterization by planar rational curves. In Proceedings of the 20th spring conference on Computer graphics, SCCG04, pages3441, New York, NY,USA,2004.ACM.
- [12] Bert J"UTTLE and Pavel CHALMOVIANSKÝ. A predictor-corrector-type technique for the approximate parameterization of intersection curves. Applicable Algebra in Engineering, Communication and Computing, 18:151168, January2007.
- [13] Igor Rostislavovich SHAFAREVICH. Basic algebraic geometry. Springer Verlag, 1974.
- [14] Juan Gerardo ALCÁZAR and Juan Rafael SENDRA. Computation of the topology of real algebraic space curves. Journal of Symbolic Computation, 39:719744, June2005.
- [15] Jinsan CHENG, Sylvain LAZARD, Luis PENARANDA, Marc POUGET, Fabrice ROUILIER, and Elias TSIGARIDAS. On the topology of planar algebraic curves. In Proceedings of the 25th annual symposium on Computational geometry, SCG09, pages361370, NewYork, NY, USA, 2009.ACM.
- [16] Alex VLACHOS, Jörg PETERS, Chas BOYD, and Jason L. MITCHELL. Curved pn triangles. pages 159166, 2001.
- [17] Richard W. DANIELS. Introduction to Numerical Methods and Optimization Techniques. Elsevier Science Ltd,1978.

Current address

BIZZARRI Michal, Mgr.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic, tel. ++420 377 632662, e-mail: bizzarri@kma.zcu.cz

LÁVIČKA Miroslav, doc., RNDr., Ph.D.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic, tel. ++420 377 632619, e-mail: lavicka@kma.zcu.cz

ON HOLOMORPHICALLY PROJECTIVE MAPPING OF KÄHLER-WEYL SPACES

ÇİVİ Gülçin, (TR), ARSAN Gürpınar Güler, (TR)

Abstract.

In this paper, we obtained the necessary and sufficient conditions for a Kähler-Weyl space to admit a nontrivial holomorphically projective mapping onto another Kähler-Weyl space and we found the relations between the curvature and Ricci tensors of two Kähler Weyl spaces admitting a holomorphically projective mapping.

Key words and phrases. Holomorphically projective, Kähler-Weyl space.

Mathematics Subject Classification. Primary 53B35, 53B15; Secondary 53B20.

1 Introduction

Many authors studied holomorphically projective mapping between special Kähler spaces and generalization of the theory of holomorphically projective mapping [1,2,4,5,8,11,12]. In [6], it is investigated objects which are invariant under the holomorphically projective mappings of two parabolically Kähler spaces. In [10], the authors determined that conformally Kähler spaces being a generalization of Kähler spaces do not admit a nontrivial holomorphically projective mapping preserving the complex structure onto almost Hermitian spaces.

In this work, our main aim is to obtain the necessary and sufficient conditions for a Kähler-Weyl space to admit a holomorphically projective mapping onto another Kähler-Weyl space and to find the relationships between curvature tensors and Ricci tensors of two these spaces.

An n -dimensional Weyl manifold having a conformal metric g and a symmetric connection ∇ satisfies the compatibility condition

$$\dot{\nabla}_k g_{ij} = \nabla_k g_{ij} - 2T_k g_{ij} = 0, \quad (1)$$

where T_k denotes a covariant vector field and $\dot{\nabla}$ denotes the prolonged derivative of g .

Under the renormalization

$$\widetilde{g} = \lambda^2 g \quad (2)$$

of the metric tensor g , T is transformed by the law

$$\widetilde{T}_k = T_k + \partial_k \ln \lambda ,$$

where λ is a scalar function defined on W_n [3,7].

If an object A in the Weyl transforms of the form $\widetilde{A} = \lambda^p A$ under the renormalization of g , A is said to be a satellite of weight $\{p\}$ of g .

Let W_n be a Weyl space of dimension $n = 2m$ and let W_n be endowed with an almost complex structure F_i^j of weight $\{0\}$, i.e.,

$$F_i^j F_j^k = -\delta_i^k, \quad (3)$$

suppose further that

$$g_{ij} F_h^i F_k^j = g_{hk} , \quad (4)$$

$$\dot{\nabla}_k F_i^j = 0 , \quad \text{for all } i, j, k. \quad (5)$$

and

$$F_{ij} = g_{jk} F_i^k = -F_{ji} , \quad F^{ij} = g^{ih} F_h^j = -F^{ji} \quad (6)$$

where the tensors F_{ij} and F^{ij} of weight $\{2\}$ and $\{-2\}$, respectively.

Such a Weyl space is called a Kahler-Weyl space and we will denote it by KW_n .

2 Holomorphically Projective Mappings of Kahler-Weyl Spaces

A curve C given by the equations $x^i = x^i(t)$ ($i = 1, 2, \dots, n$) with tangent vector $\frac{dx^i}{dt}$ in KW_n is a analytically planar curve if the conditions [4]

$$\frac{d^2 x^i}{dt^2} + \Gamma_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = a(t) \frac{dx^i}{dt} + b(t) F_j^i \frac{dx^j}{dt} \quad (7)$$

hold, where $a = a(t)$ and $b = b(t)$ are functions of t and Γ_{jk}^i are the Christoffel symbols of the form

$$\Gamma_{jk}^i = \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} - (\delta_j^i T_k + \delta_k^i T_j - g^{li} g_{jk} T_l). \quad (8)$$

Consider a diffeomorphism ρ from a Kahler-Weyl space KW_n onto another Kahler space $K\bar{W}_n$ preserving the complex structure and having a common coordinate system $x^i = \bar{x}^i$. If $\rho : KW_n \rightarrow K\bar{W}_n$ maps all analytically planar curves of KW_n into analytically planar curve in $K\bar{W}_n$ then the mapping ρ is a holomorphically projective mapping [4,6,12].

Under the holomorphically projective mapping $\rho : KW_n \rightarrow K\bar{W}_n$ we have

$$\frac{d^2 x^i}{dt^2} + \bar{\Gamma}_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = a(\bar{t}) \frac{dx^i}{dt} + b(\bar{t}) \bar{F}_j^i \frac{dx^j}{dt} \quad (9)$$

where $\bar{\Gamma}_{jk}^i$ is the coefficients of connection $\bar{\nabla}$.

Substracting (7) and (9) and by considering the fact $F_k^i = \bar{F}_k^i$ we obtain

$$\left(\bar{\Gamma}_{jk}^i - \Gamma_{jk}^i\right) \frac{dx^j}{dt} \frac{dx^k}{dt} = 2\psi_j \frac{dx^j}{dt} \frac{dx^i}{dt} + 2\varphi_j F_k^i \psi_j \frac{dx^j}{dt} \frac{dx^k}{dt} \quad (10)$$

which implies that

$$\bar{\Gamma}_{jk}^i = \Gamma_{jk}^i + \psi_j \delta_k^i + \psi_k \delta_j^i + \varphi_j F_k^i + \varphi_k F_j^i, \quad (11)$$

where $2\psi_j \frac{dx^j}{dt} = \bar{a} - a$ and $2\varphi_j \frac{dx^j}{dt} = \bar{b} - b$.

It can be easily seen that under the condition (11), the equations (7) and (9) are satisfied which denotes $\rho : KW_n \rightarrow K\bar{W}_n$ is a holomorphically projective planar mapping.

By (11), in the local coordinates, the compatibility condition of the metric tensor \bar{g} of $K\bar{W}_n$ with respect to $\bar{\nabla}$ can be written as the equation

$$\begin{aligned} \dot{\nabla}_k \bar{g}_{ij} &= \nabla_k \bar{g}_{ij} - 2T_k \bar{g}_{ij} \\ &= 2(\psi_k + P_k) \bar{g}_{ij} + \bar{g}_{kj} \psi_i + \bar{g}_{ik} \psi_j + \varphi_i F_k^m \bar{g}_{mj} + \varphi_j F_k^m \bar{g}_{im}, \end{aligned} \quad (12)$$

where $P_k = \bar{T}_k - T_k$.

Similarly, the prolonged derivative of complex structure with respect to $\bar{\nabla}$ gives

$$\begin{aligned} \dot{\nabla}_k \bar{F}_i^h &= \nabla_k \bar{F}_i^h \\ &= -[\delta_k^h (\psi_m \bar{F}_i^m + \varphi_i) + \bar{F}_k^h (\varphi_m \bar{F}_i^m + \psi_i)]. \end{aligned} \quad (13)$$

By remembering that

$$\dot{\nabla}_k \bar{F}_i^j = \dot{\nabla}_k F_i^j = 0,$$

we obtain

$$\psi_i = \varphi_m F_i^m. \quad (14)$$

Thus, we proved that, if the relation given by (11) holds, the equations (12) and (14) are satisfied.

Conversely, if we have the equations given by (12) and (14), one can see that

$$\begin{aligned} \left(\bar{\Gamma}_{ik}^m - \Gamma_{ik}^m\right) \bar{g}_{mj} + \left(\bar{\Gamma}_{jk}^m - \Gamma_{jk}^m\right) \bar{g}_{im} &= (\psi_k \delta_i^m + \psi_i \delta_k^m + \varphi_i F_k^m) g_{mj} + (\psi_k \delta_j^m + \psi_j \delta_k^m + \varphi_j F_k^m) g_{im} \\ &= (\psi_k \delta_i^m + \psi_i \delta_k^m + \varphi_i F_k^m + \varphi_k F_i^m) g_{mj} \\ &\quad + (\psi_k \delta_j^m + \psi_j \delta_k^m + \varphi_j F_k^m + \varphi_k F_j^m) g_{im} \end{aligned} \quad (15)$$

or

$$\bar{\Gamma}_{jk}^i = \Gamma_{jk}^i + \psi_j \delta_k^i + \psi_k \delta_j^i + \varphi_j F_k^i + \varphi_k F_j^i,$$

which denotes the mapping between two Kahler Weyl spaces is a holomorphically projective mapping.

In (11), if $\varphi_i \neq 0$ then the mapping is nontrivial. Otherwise, it is said to be trivial or affine. Thus, we can state the following theorem.

Theorem 2.1 *A Kahler-Weyl space admits a nontrivial holomorphically projective mapping onto another Kahler-Weyl space if and only if one of the following equivalent conditions holds.*

$$a) \quad \bar{\Gamma}_{jk}^i = \Gamma_{jk}^i + \psi_j \delta_k^i + \psi_k \delta_j^i + \varphi_j F_k^i + \varphi_k F_j^i.$$

$$b) \quad \dot{\nabla}_k \bar{g}_{ij} = 2(\psi_k + P_k) \bar{g}_{ij} + \bar{g}_{kj} \psi_i + \bar{g}_{ik} \psi_j + \varphi_i F_k^m \bar{g}_{mj} + \varphi_j F_k^m \bar{g}_{im} \quad , \quad \psi_i = \varphi_m F_i^m \quad .$$

Theorem 2.2 *The covector field ψ_i of the holomorphically projective mapping between two Kahler-Weyl spaces is in the form*

$$\psi_i = \frac{1}{n+2} \left[\partial_i \ln \sqrt{\frac{\bar{g}}{g}} - n P_i \right]$$

where $g = \det(g_{ij})$ and $\bar{g} = \det(\bar{g}_{ij})$ and $P_j = \bar{T}_j - T_j$.

Proof. Contraction (11) on the indices i and k gives

$$\bar{\Gamma}_{ji}^i - \Gamma_{ji}^i = (n+2) \psi_j$$

from which follows

$$\psi_j = \frac{1}{n+2} \left[\partial_j \ln \sqrt{\frac{\bar{g}}{g}} - n P_j \right] \quad , \quad (16)$$

where $P_j = \bar{T}_j - T_j$.

It can be stated that, unlike the covector field of a holomorphically projective mapping between two Kahler spaces, instead of being gradient, the covector field ψ is different from a gradient for Kahler-Weyl spaces in general case.

Let R_{ijk}^h and \bar{R}_{ijk}^h be the mixed curvature tensors of KW_n and $K\bar{W}_n$, respectively.

After some calculations, the relation between the curvature tensors of two Kahler-Weyl spaces under the holomorphically projective mapping are obtained as

$$\bar{R}_{ijk}^h = R_{ijk}^h + \delta_j^h \psi_{ik} - \delta_k^h \psi_{ij} + \delta_i^h (\psi_{jk} - \psi_{kj}) + F_j^h \varphi_{ik} - F_k^h \varphi_{ij} + F_i^h (\varphi_{jk} - \varphi_{kj}) \quad (17)$$

where

$$\psi_{ij} = \dot{\nabla}_j \psi_i + \varphi_i \varphi_j - \psi_i \psi_j \quad , \quad \varphi_{ij} = -\psi_{mj} F_i^m \quad .$$

The Ricci tensors \bar{R}_{ij} and R_{ij} are related by the equation

$$\begin{aligned} \bar{R}_{ij}^h &= \bar{R}_{ij} \\ &= R_{ij} + \psi_{ji} - (n+1) \psi_{ij} - (\psi_{hk} + \psi_{kh}) F_i^h F_j^k \end{aligned} \quad (18)$$

The antisymmetric part of the Ricci tensor, from (18) and (16) obtained as

$$\bar{R}_{[ij]} = R_{[ij]} + (n+2) \psi_{[ji]} \quad (19)$$

$$= \frac{(n+2)}{2} \left(\frac{\partial_i \psi_j}{\partial x^i} - \frac{\partial_j \psi_i}{\partial x^j} \right) \quad (20)$$

$$= n P_{[i,j]} \quad ,$$

where brackets indicate antisymmetrization.

Hence we have

Theorem 2.3 *For a nontrivial holomorphically projective mapping between two Kahler-Weyl spaces, the antisymmetric part of the Ricci tensors satisfies the relation*

$$\bar{R}_{[ij]} = R_{[ij]} + \frac{n}{n+2} \psi_{[ij]} .$$

References

- [1] BACSO, S., ILOSVAY, F.: *On holomorphically projective mappings of special Kahler spaces.* Acta Mathematica Academia Paedagogicae Nyiregyhaziensis, 15 , 41-44, 1999.
- [2] DOMASHEV , V.V., MIKES, J.: *To theory of holomorphically projective mappings of Kahlerian spaces.* Mat. Zametki 28,297-303, 1978.
- [3] HLAVATY, V.: *Theorie d'immersion d'une W_m dans W_n* , Ann. Soc. Polon. Math., 21, 196-206, 1949.
- [4] ISHIHARA, S., TACHIBANA , S.: *On infinitesimal holomorphically projective transformations in Kahlerian manifolds.* Tohoku Math. J. 12, 77-101, 1960.
- [5] MIKES, J.: *On holomorphically projective mappings of Kahler spaces.* Ukr, Geom. Sb., Kharkov, 23, 90-98 , 1980.
- [6] MIKES, J. , SHIHA M., VANZUROVA A.: *Invariant Objects by Holomorphically Projective Mappings of Parabolically Kahler Spaces* Journal of Applied Mathematics, Aplimat , Volume II, number I, 2009.
- [7] NORDEN, A.: *Affinely connected spaces.* GRMFML, Moscow, (in Russian), 1976.
- [8] OTSUKI, T., TASHIRO, Y.: *On curves in Kahlerian spaces.* Math. J. Okayama Univ. 4, No. 1, 57-78, 1954.
- [9] PRVANOVIC, M.: *Holomorphically projective transformations of the Kahler spaces.* Tensor, New ser.35, 99-104, 1981.
- [10] RADULOVIC, Z. , MIKES, J. : *Geodesic and holomorphically projective mappings of conformally-Kahlerian spaces.* Differential Geometry and its Applications, Proc. Conf. Opava (Czechoslovakia) , silesian univ. 151-156, 1993.
- [11] TASHIRO, Y.: *On Holomorphically projective correspondences in an almost complex space.* Math. J. Okayama Univ. ,6, No 2, 147-152, 1957.
- [12] YANO, K.: *Differential Geometry on complex and Almost Complex Spaces.* Pergomon press, 1965.

Current address

Gülçin Çivi, Assoc. Prof. Dr.

Istanbul Technical University, Faculty of Science and Letters,
Department of Mathematics 34469, Maslak-İstanbul ,TURKEY
tel.: +(90) 2122856959 , e-mail: civi@itu.edu.tr

Güler Gürpınar Arsan, Assoc. Prof. Dr.

Istanbul Technical University, Faculty of Science and Letters,
Department of Mathematics, 34469, Maslak-Istanbul, TURKEY
tel.+(90) 2122853269 , e-mail: ggarsan@itu.edu.tr

HOLOMORPHICALLY PROJECTIVE MAPPINGS OF HYPERBOLICALLY KÄHLER SPACES PRESERVING THE EINSTEIN TENSOR

CHEPURNA Olena, (UKR), MIKEŠ Josef, (CZ)

Abstract. In this paper there are discussed the holomorphically projective mappings of hyperbolically Kähler spaces which preserved the Einstein tensor. We proved that the tensor of h -conircular curvature is invariant under Einstein tensor-preserving holomorphically projective mappings.

Key words and phrases. holomorphically projective mapping, hyperbolically Kähler space, Einstein tensor.

Mathematics Subject Classification. Primary 53B30, 53B35.

1 Introduction

From the very beginning, the theory of geodesic and holomorphically projective mappings of classical and hyperbolically Kähler spaces attracted attention by a wide scale of possibilities for applications, not only in geometry itself, but also as a useful tool of modeling various processes in mechanics and physics [1]–[24].

If we distinguish some class of mappings between spaces from a fixed class, a natural questions arises, what objects and properties of spaces are preserved, invariant, under all mappings under consideration.

As far as invariant objects under holomorphically projective mappings are concerned, let us mention generalized Thomas' parameters and the tensor of holomorphically projective curvature. To mention some invariant properties, note that the class of spaces of constant curvature and the class of Einstein spaces are closed under holomorphically projective mappings.

In this paper, we examine nontrivial holomorphically projective mappings of hyperbolically Kähler spaces preserving the Einstein tensor. We prove that the tensor of h -conircular

curvature is an invariant of holomorphically projective mappings. Further, we examine some geometric properties of such spaces. These results are analogous, as in theory of holomorphically projective mappings of Kähler spaces, see [4].

2 Basic concepts

A (pseudo-) Riemannian space K_n is called a *hyperbolically Kähler space* if it is endowed, besides a metric tensor g , with an affinor structure F ($\neq \text{Id}$) satisfying the following relations [17, 19, 20, 21]

$$F^2 = \text{Id}, \quad g(X, FX) = 0, \quad \nabla F = 0.$$

Here X are all tangent vectors of TK_n and ∇ is a connection of K_n . The structure F is a product structure.

It is known, that the diffeomorphism f between hyperbolic Kähler spaces K_n and \bar{K}_n is called a *holomorphically projective mapping*, if f maps any analytical planar curve of K_n onto a analytical planar curve of \bar{K}_n . Due to the diffeomorphism f , we can suppose that $\bar{M} = M$, where M is “common” manifolds on which the metrics g and \bar{g} and the complex structure F on K_n and \bar{K}_n are defined.

A holomorphically projective mapping f from K_n onto \bar{K}_n preserves the structures and is characterized by the following condition

$$(\bar{\nabla} - \nabla)(X, Y) = \psi(X)Y + \psi(Y)X + \bar{\psi}(X)FY + \bar{\psi}(Y)FX \quad (1)$$

for any vector fields X, Y , where $\bar{\nabla}$ and ∇ are affine connections of K_n and \bar{K}_n , ψ is a linear form and $\bar{\psi}(X) = \psi(FX)$.

The mapping from K_n onto \bar{K}_n is *holomorphically projective* if the equations hold

$$\begin{aligned} \nabla_Z \bar{g}(X, Y) &= 2\psi(Z)\bar{g}(X, Y) + \psi(X)\bar{g}(Y, Z) + \psi(Y)\bar{g}(X, Z) \\ &+ \bar{\psi}(X)\bar{F}(Y, Z) + \bar{\psi}(Y)\bar{F}(X, Z), \end{aligned} \quad (2)$$

where ∇ is Levi-Civita connection of K_n , ψ is a linear form and X, Y, Z are tangent vectors, $\bar{F}(X, Z) = -\bar{g}(X, FZ)$. If $\psi = 0$, then a holomorphically projective mapping is called *trivial* or *affine*. The equations (2) we rewrite in local coordinates:

$$\bar{g}_{ij,k} = 2\psi_k \bar{g}_{ij} + \psi_i \bar{g}_{jk} + \psi_j \bar{g}_{ik} + \bar{\psi}_i \bar{F}_{jk} + \bar{\psi}_j \bar{F}_{ik}, \quad (3)$$

where $\bar{g}_{ij}(x)$, $\psi_k(x)$, $\bar{\psi}_k(x)$ and \bar{F}_{ij} are components of \bar{g} , ψ , $\bar{\psi}$, \bar{F} and “,” is a covariant derivative on K_n , $x = (x^1, x^2, \dots, x^n)$ is a point of coordinate neighbourhood $U \subset M$. Equations (2) and (3) hold when K_n and $\bar{K}_n \in C^1$, i.e. $g_{ij}(x)$ and $\bar{g}_{ij}(x) \in C^1$ in any coordinate neighbourhood U .

The following conditions are necessary for a holomorphically projective mapping:

$$\bar{R}_{ijk}^h = R_{ijk}^h + \psi_{ij}\delta_k^h - \psi_{ik}\delta_j^h - \psi_{i\alpha}F_j^\alpha F_k^h + \psi_{i\alpha}F_k^\alpha F_j^h - 2\psi_{j\alpha}F_k^\alpha F_i^h, \quad (4)$$

$$\bar{R}_{ij} = R_{ij} + (n-1)\psi_{ij}. \quad (5)$$

Here R_{ijk}^h is the Riemannian curvature tensor, R_{ij} is the Ricci tensor, and

$$\psi_{ij} = \psi_{i,j} - \psi_i \psi_j + \bar{\psi}_i \bar{\psi}_j.$$

On the other hand, necessary and sufficient condition for existence of holomorphically projective mappings of the given hyperbolically Kähler space onto hyperbolically Kähler spaces is existence of a solution for the system of equations [8, 14, 17, 19, 20]

$$a_{ij,k} = \lambda_i g_{jk} + \lambda_j g_{ik} - \bar{\lambda}_i F_{jk} - \bar{\lambda}_j F_{ik}, \quad (6)$$

$$n\lambda_{i,j} = \mu g_{ij} + a_{\alpha i} R_j^\alpha - a_{\alpha \beta} R_{ij}^\alpha \cdot^\beta, \quad (7)$$

$$\mu_{,k} = 2\lambda_\alpha R_k^\alpha \quad (8)$$

with respect to a regular symmetric tensor a_{ij} , a co-vector λ_i and a function μ . Here $R_j^i = R_{\alpha j} g^{\alpha i}$; $R_{ij}^k \cdot^h = R_{\alpha i j \beta} g^{\alpha k} g^{\beta h}$; $F_{ij} = F_i^\alpha g_{\alpha j}$; and g^{ij} are elements of the matrix inverse to g_{ij} .

According to the known solutions of the above system of differential equations the metrics of the resulting image spaces under holomorphically projective mappings can be determined from the equations [14, 20]:

$$a_{ij} = e^{2\psi} \bar{g}^{\alpha\beta} g_{\alpha i} g_{\beta j}; \quad (9)$$

$$\lambda_i = -e^{2\psi} \psi_{,\alpha} \bar{g}^{\alpha\beta} g_{\beta i}. \quad (10)$$

The important invariants under holomorphically projective mappings of hyperbolically Kähler spaces are the generalized Thomas' parameters

$$\bar{T}_{ij}^h = T_{ij}^h; \quad T_{ij}^h = \Gamma_{ij}^h - \frac{1}{n+2} (\delta_i^h \Gamma_{j\alpha}^\alpha + \delta_j^h \Gamma_{i\alpha}^\alpha + F_i^h F_j^\beta \Gamma_{\beta\alpha}^\alpha + F_j^h F_i^\beta \Gamma_{\beta\alpha}^\alpha) \quad (11)$$

and the tensor of holomorphically projective curvature

$$\bar{P}_{ijk}^h = P_{ijk}^h; \quad P_{ijk}^h = R_{ijk}^h - \frac{1}{n+2} (\delta_k^h R_{ji} - \delta_j^h R_{ki} - F_k^h F_j^\alpha R_{\alpha i} - F_j^h F_k^\alpha R_{\alpha i} - 2F_i^h F_j^\alpha R_{\alpha k}). \quad (12)$$

3 Basic equations for Einstein tensor-preserving holomorphically projective mappings

We call a holomorphically projective mapping *Einstein tensor-preserving* if it satisfies:

$$\bar{E}_{ij} = E_{ij}, \quad (13)$$

where

$$E_{ij} = R_{ij} - \frac{R}{n} g_{ij} \quad (14)$$

is the Einstein tensor and $R = R_{\alpha\beta} g^{\alpha\beta}$ is the scalar curvature.

If this is the case, the deformation tensor for the Ricci tensor takes the form:

$$T_{ij} = \bar{R}_{ij} - R_{ij} = \frac{\bar{R}}{n} \bar{g}_{ij} - \frac{R}{n} g_{ij}. \quad (15)$$

On the other hand, accounting (5) we obtain

$$T_{ij} = \bar{R}_{ij} - R_{ij} = (n-1)\psi_{ij}. \quad (16)$$

Comparing we get:

$$\psi_{ij} = \frac{\bar{R}}{n(n-1)} \bar{g}_{ij} - \frac{R}{n(n-1)} g_{ij}. \quad (17)$$

Substituting the last expression into (4) and using the notation

$$Y_{ijk}^h = R_{ijk}^h - \frac{R}{n(n-1)} (\delta_k^h g_{ij} - \delta_j^h g_{ik} - F_k^h F_i^\alpha g_{\alpha j} + F_j^h F_i^\alpha g_{\alpha k} - 2 F_i^h F_j^\alpha g_{\alpha k}) \quad (18)$$

(and similarly with bar) we find

$$\bar{Y}_{ijk}^h = Y_{ijk}^h. \quad (19)$$

Here Y_{ijk}^h are components of the tensor of *h-concircular curvature* on hyperbolically Kähler space, where Y is an analog of the tensor of concircular curvature [14, 18, 19, 20, 25].

Hence we have proved:

Theorem 3.1 *The tensor of h-concircular curvature is invariant under Einstein tensor-preserving holomorphically projective mappings of hyperbolically Kähler space.*

Let us apply covariant differentiation to the formula (10):

$$\lambda_{i,j} = -e^{2\psi} \psi_{\alpha,j} \bar{g}^{\alpha\beta} g_{\beta i} + e^{2\psi} \psi_\alpha \psi_\beta \bar{g}^{\alpha\beta} g_{ji} + e^{2\psi} \psi_j \psi_\alpha \bar{g}^{\alpha\beta} g_{\beta i}. \quad (20)$$

By (9) and (17), we get

$$\lambda_{i,j} = \mu g_{ij} + \frac{R}{n(n-1)} a_{ij}, \quad (21)$$

where

$$\mu = e^{2\psi} \left(\psi_\alpha \psi_\beta \bar{g}^{\alpha\beta} - \frac{\bar{R}}{n(n-1)} \right). \quad (22)$$

Obviously using (9), (10), from (20) and (21) we get (17), and consequently also (13), hence we have proved:

Theorem 3.2 *A hyperbolically Kähler space admits an Einstein tensor-preserving holomorphically projective mapping on hyperbolically Kähler spaces if and only if the conditions (6), (21) and (22) are satisfied.*

We say that a hyperbolically Kähler space K_n belongs to the class $K_n[B]$ if it admits a holomorphically projective mapping and the corresponding vector satisfies [13, 15, 19]

$$\lambda_{i,j} = \mu g_{ij} + B a_{ij} \quad (23)$$

for some function B . Further analysis, make sure that B is the constant.

So we have actually proved that a hyperbolically Kähler space K_n admitting Einstein tensor-preserving holomorphically projective mappings belongs to the class $K_n[B]$ where $B = -\frac{R}{n(n-1)} = \text{const.}$

Acknowledgement

The paper was supported by grant P201/11/0356 of The Czech Science Foundation.

References

- [1] MIKEŠ, J., BÁCSÓ, S., ZEDNÍK, J.: *On f -planar mappings of affine-connection spaces with infinite dimension*. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 36, pp. 157-162, 1997.
- [2] CHEPURNA, O., KIOSAK, V., MIKEŠ, J.: *Conformal mappings of Riemannian spaces which preserve the Einstein tensor*. Proceedings of the 8th Int. Conf. APLIMAT, Part II, pp. 461-466, 2010.
- [3] CHEPURNA, O., KIOSAK, V., MIKEŠ, J.: *On geodesic mappings preserving the Einstein tensor*. Acta UPOL, Mathematica, pp. 49-52, 2010.
- [4] CHEPURNA, O., MIKEŠ, J.: *Holomorphically projective mappings preserving the Einstein tensor*. Proceedings of the 10th Int. Conf. APLIMAT, pp. 661-666, 2011.
- [5] CHODOROVÁ, M., MIKEŠ, J.: *A note to K -torse-forming vector fields on compact manifolds with complex structure*. Acta Physica Debrecina, 42, pp. 11-18, 2008.
- [6] CHUDÁ, H., MIKEŠ, J.: *On F -planar mappings with a certain initial conditions*. In APLIMAT 2006 - 5th Int. Conf., Part II, 2006.
- [7] CHUDÁ, H., MIKEŠ, J.: *On first quadratic integral of geodesics with a certain initial conditions*. In APLIMAT 2007 - 6th Int. Conf., Part II, pp. 85-88, 2007.
- [8] DOMASHEV, V.V., MIKEŠ, J.: *Theory of holomorphically projective mappings of Kählerian spaces*. (English. Russian original) Math. Notes 23, pp. 160-163, 1978; translation from Mat. Zametki 23, pp. 297-303, 1978.
- [9] HINTERLEITNER, I., MIKEŠ, J.: *Fundamental equations of geodesic mappings and their generalisations*. (English. Russian original) J. Math. Sci. New York, 2010; translation from Itogi Nauki Tekh., Ser. Sovrem. Mat. Prilozh., Temat. Obz., 124, pp. 7-34, 2010.
- [10] JUKL, M., JUKLOVÁ, L., MIKEŠ, J.: *Some results on the decomposition of the traceless tensors*. (English. Russian original) J. Math. Sci. New York, 2010; translation from Itogi Nauki Tekh., Ser. Sovrem. Mat. Prilozh., Temat. Obz., 124, pp. 139-158, 2010.
- [11] JUKL, M., LAKOMÁ, L.: *The Decomposition of Tensors Spaces with Almost Complex Structure*. Suppl. ai Rend. del Circ. Matematico di Palermo, Serie II, No 72, pp. 145-150, 2004.
- [12] KURBATOVA, I.N.: *HP-mappings of H -spaces*. (Russian) Ukr. Geom. Sb. 27, pp. 75-83, 1984.
- [13] MIKEŠ, J.: *Geodesic mappings of special Riemannian spaces*. (Russian) PhD Thesis, Odessa State University, 1979.
- [14] MIKEŠ, J.: *On holomorphically projective mappings of Kählerian spaces*. (Russian) Ukr. Geom. Sb. 23, pp. 90-98, 1980.
- [15] MIKEŠ, J.: *Geodesic, F -planar and holomorphically projective mappings of Riemannian spaces and spaces with affine connection*. (Russian) DrSc. Thesis, Palacky University Olomouc, (1995), Charles University Prague, 1996.
- [16] MIKEŠ, J.: *Geodesic mappings of affine-connected and Riemannian spaces*. J. Math. Sci. New York, Vol. 78, No. 3, pp. 311-333, 1996.
- [17] MIKEŠ, J.: *Holomorphically projective mapping and their generalizations*. J. Math. Sci. New York, Vol. 89, No. 3, pp. 1334-1353, 1998.
- [18] MIKEŠ, J., KIOSAK, V., VANŽUROVÁ, A.: *Geodesic mappings of manifolds with affine connection*. Publ. Palacky University, Olomouc, 2008.

- [19] MIKEŠ, J., VANŽUROVÁ, HINTERLEITNER I.: *Geodesic mappings and some generalizations*. Publ. Palacky University, Olomouc, 2009.
- [20] SINYUKOV, N.S.: *Geodesic mappings of Riemannian spaces*. Nauka, Moscow, 1979.
- [21] SINYUKOV, N.S., KURBATOVA, I.N.; MIKEŠ, J.: *Holomorphically projective mappings of Kähler spaces*. Odessk. Univ., Odessa, 1985.
- [22] AL LAMY, RAAD J., MIKEŠ, J., ŠKODOVÁ, M.: *On holomorphically projective mappings from equiaffine generally recurrent spaces onto Kählerian spaces*. Arch. Math. (Brno) 42, suppl., pp. 291-299, 2006.
- [23] STANKOVIĆ, M.S., MINČIČ, S.M., VELIMIROVIĆ, L.S.: *On equitorsion holomorphically projective mappings of generalized Kählerian spaces*. Czech. Math. J. 54, No. 3, pp. 701-715, 2004.
- [24] STEPANOV, S.E.: *New methods of the Bochner technique and their applications*. J. Math. Sci., New York 113, No. 3, pp. 514-536, 2003.
- [25] YANO, K.: *Differential Geometry on Complex and Almost Complex Spaces*. Pergamon Press, Oxford, 1965.

Current address

Olena Chepurna

Department of Mathematics
Odessa State Economical University
Odessa, Ukraine
E-mail: culeshowa@ukr.net

Prof. RNDr. Josef Mikeš, DrSc.

Dept. Algebra and Geometry, Faculty of Science, Palacký University
17. listopadu 12, 77900 Olomouc, Czech Republic,
tel. +420-585 634 656 and e-mail: mikes@inf.upol.cz.

ON COMPOSITION OF CONFORMAL AND HOLOMORPHICALLY PROJECTIVE MAPPINGS BETWEEN CONFORMALLY KÄHLERIAN SPACES

CHODOROVÁ Marie, (CZ), CHUDÁ Hana, (CZ), SHIHA Mohsen, (SY)

Abstract. In this paper we study general properties of composition conformal and holomorphically projective mappings f (i.e. $f = f_1 \circ f_2 \circ f_3$, where f_1, f_3 are conformal mappings and f_2 is a holomorphically projective mapping) between conformally Kählerian spaces K_n and \bar{K}_n .

Key words and phrases. conformal mapping, holomorphically projective mapping, conformally holomorphically projective mapping, Kählerian space.

Mathematics Subject Classification. Primary 60A05, 08A72; Secondary 28E10.

1 Introduction

Many monographs and papers are devoted to the theory of conformal, geodesic and holomorphically projective mappings, see [1] - [28]. Compositions of conformal and geodesic mappings were studied in the papers by Zudina, Stepanov [29] and Hinterleitner, Mikeš [7], [8].

We suppose, that metrics of the considered Riemannian spaces V_n have a general signature, i.e. we talk about Riemannian or (pseudo-) Riemannian spaces.

2 Main Properties of Kählerian and Conformally Kählerian Spaces

It is known [17, 21, 22, 27, 28] that an n -dimensional (pseudo-) Riemannian space H_n with a metric tensor g will be called an *almost Hermitian space* if there an almost hermitian structure is defined, i.e. there is an affinor field F such that

$$F^2 = -Id, \quad g(X, FX) = 0 \quad (1)$$

for all tangent vector X .

An almost hermitian space K_n will be called *Kählerian*, if the structure F is covariantly constant, i.e. $\nabla F = 0$. Where ∇ is the Levi-Civita connection on K_n .

Any Riemannian space which may be conformally mapped onto Kähler space, will be called *a conformally Kählerian space*.

Clearly, any conformally Kählerian space K_n may be considered as an almost Hermitian space and it may be characterised by an almost Hermitian structure (1). This structure has the following properties, see [25]:

$$\nabla_Y(F(X)) = \varphi(X) \cdot Y - g(X, Y) \cdot \Phi + \varphi(FX) \cdot F(Y) - g(FX, Y) \cdot F\Phi,$$

where $\varphi(X) = g(X, \Phi) = \nabla_X \mathcal{F}$, \mathcal{F} is a function on K_n and X, Y are tangent vector fields.

3 Main Properties of the Holomorphically Projective Mappings

It is known [17, 22, 27, 28] that a diffeomorphism f between Kählerian spaces K_n and \bar{K}_n is called a *holomorphically projective mapping*, if f maps any analytic planar curve in K_n onto an analytic planar in \bar{K}_n .

Using properties of a diffeomorphism f , we can suppose that M is a “common” manifold on which metrics g and \bar{g} of K_n and \bar{K}_n are defined.

A mapping of K_n onto \bar{K}_n is holomorphically projective if and only if the following equations holds

$$(\bar{\nabla} - \nabla)_X X = 2\psi(X) \cdot X - 2\psi(FX) \cdot FX, \quad (2)$$

for all tangent vector field X . Here ∇ and $\bar{\nabla}$ are the Levi-Civita connections on K_n and \bar{K}_n , ψ is a linear form. If $\psi = 0$ then the holomorphically projective mapping is called *trivial* or *affine*.

The equations (2) we re-write in local coordinates:

$$\bar{\Gamma}_{ij}^h = \Gamma_{ij}^h + \psi_i \delta_j^h + \psi_j \delta_i^h - \bar{\psi}_i F_j^h - \bar{\psi}_j F_i^h, \quad (3)$$

where $\Gamma_{ij}^h(x)$ and $\bar{\Gamma}_{ij}^h(x)$ are the Christoffel symbols on K_n and \bar{K}_n , $\bar{g}_{ij}(x)$ and $\psi_k(x)$ are components of \bar{g} and ψ , δ_i^h is the Kronecker symbol. $x = (x^1, x^2, \dots, x^n)$ is a point of a coordinate neighbourhood $U \subset M$. Equations (2) and (3) hold when $K_n \in C^1$ and $\bar{K}_n \in C^1$, i.e. $g_{ij}(x) \in C^1$, $\bar{g}_{ij}(x) \in C^1$ in any coordinate neighbourhood U .

It is known that

$$\psi_i = \nabla \Psi, \quad \Psi = \frac{1}{2(n+2)} \ln \left| \frac{\bar{g}}{g} \right|.$$

It follows that $\psi_i(x) \in C^0$.

The tensor of the holomorphically projective curvature of Kählerian space has the form:

$$P_{ijk}^h = R_{ijk}^h - \frac{1}{n+2} (\delta_k^h R_{ij} - \delta_j^h R_{ik} F_k^h R_{i\alpha} F_j^\alpha - F_j^h R_{i\alpha} F_k^\alpha - 2F_i^h R_{j\alpha} F_k^\alpha), \quad (4)$$

where R_{ijk}^h are components of the Riemannian tensor and $R_{ij} = R_{ij\alpha}^\alpha$ are components of the Ricci tensor.

The tensor of the holomorphically projective curvature is an invariant under holomorphically projective mappings $K_n \rightarrow \bar{K}_n$, i.e. $P = \bar{P}$.

It is known, that a Kähler manifold K_n is a space of the constant curvature if and only if the tensor of the holomorphically projective curvature is vanishing ($P = 0$).

4 Main Properties of the Conformal Mappings

A diffeomorphism f between Riemannian spaces V_n and \bar{V}_n is called a *conformal mapping*, if f preserves angles among all curves on V_n .

The mapping of $V_n = (M, g)$ onto $\bar{V}_n = (M, \bar{g})$ is conformal if and only if, in each coordinate system x , metrics are proportional, i.e. the following condition holds [6, 22, 24, 27]:

$$\bar{g} = \rho \cdot g, \quad (5)$$

where ρ is a nonzero function, which depends on a point in M .

From the equation (5) it follows, that

$$(\bar{\nabla} - \nabla)_X X = 2\sigma(X) \cdot X - g(X, X) \cdot \Sigma, \quad (6)$$

where $\sigma(X) = \frac{1}{2} \nabla_X \ln |\rho|$, $\sigma(X) = g(X, \Sigma)$ and X is an arbitrary tangent vector. In local coordinates:

$$\bar{\Gamma}_{ij}^h = \Gamma_{ij}^h + \sigma_i \delta_j^h + \sigma_j \delta_i^h - \sigma^h g_{ij}, \quad (7)$$

where $g_{ij}(x)$, $\sigma_i(x)$ and $\sigma^h(x)$ are components of g , σ and Σ on a coordinate neighborhood U .

When we have a conformal mapping $V_n \rightarrow \bar{V}_n$ the Weyl tensor of the conformal curvature is an invariant (i.e. $\bar{C} = C$) and it has the form:

$$C_{ijk}^h = R_{ijk}^h + \delta_j^h L_{ik} - \delta_k^h L_{ij} + L_j^h g_{ik} + L_k^h g_{ij} \quad (8)$$

where $L_{ij} = \frac{1}{n-2} (R_{ij} - \frac{R}{2(n-1)} g_{ij})$, $L_i^h = g^{h\alpha} L_{\alpha i}$, $R = R_{\alpha\beta} g^{\alpha\beta}$ is the scalar curvature and g^{ij} are components of the inverse matrix of g_{ij} .

It is known, that for $n = 2$ and $n = 3$ the Weyl tensor of the conformal curvature C always vanishes identically.

For $n > 3$ a Riemannian space is (locally) *conformally flat* if and only if, the Weyl tensor of the conformal curvature is vanishing ($C = 0$).

5 On Conformally Holomorphically Projective Mappings

In papers by I. Hinterleitner [7, 8] there were studied mappings, which are the composition of conformal and holomorphically projective mappings – *conformally projective mappings* and further H. Chudá, J. Mikeš conformally geodesic mappings [3].

In our case we study more general situation, when the conformally holomorphically projective mapping

$$f: K_n = (M, g, F) \rightarrow \bar{K}_n = (M, \bar{g}, \bar{F})$$

is a composition of following mappings

$$f = f_1 \circ f_2 \circ f_3$$

$$\begin{aligned} f_1 : K_n = (M, g) &\rightarrow K_n^1 = (M, \overset{1}{g}) && \text{conformal mapping} \\ f_2 : K_n^1 = (M, \overset{1}{g}) &\rightarrow K_n^2 = (M, \overset{2}{g}) && \text{holomorphically projective mapping} \\ f_3 : K_n^2 = (M, \overset{2}{g}) &\rightarrow \bar{K}_n = (M, \bar{g}) && \text{conformal mapping.} \end{aligned}$$

We suppose, that a holomorphically projective mapping $f \in C^r$, if K_n and \bar{K}_n are conformally Kählerian manifolds and K_n^1 and K_n^2 are Kählerian manifolds.

Evidently, for a conformal mappings f_1 and f_2 the following conditions hold

$$\overset{1}{g} = \overset{1}{\sigma} \cdot g \quad \text{and} \quad \bar{g} = \overset{2}{\sigma} \cdot \overset{2}{g}, \quad (9)$$

where $\overset{1}{\sigma}$ and $\overset{2}{\sigma}$ are the functions on M belonging to the class C^r .

Remark 5.1 *Conformally holomorphically projective mappings do not generate equivalent classes of Kählerian manifolds.*

Evidently, $\text{id}: K_n \rightarrow K_n$ is a trivial conformally holomorphically projective mapping. Conformally holomorphically projective mappings are also symmetric relations, i.e. if $f = f_1 \circ f_2 \circ f_3: K_n \rightarrow \bar{K}_n$ is a conformally holomorphically projective mapping, then $f^{-1} = f_3^{-1} \circ f_2^{-1} \circ f_1^{-1}: \bar{K}_n \rightarrow K_n$ is also a conformally holomorphically projective mapping.

Finally, unfortunately the composition of conformally holomorphically projective mappings might not be conformally holomorphically projective, i.e. this relation is not transitive.

Theorem 5.2 *A diffeomorphism $f \in C^1$: $K_n = (M, g) \rightarrow \bar{K}_n = (M, \bar{g})$ is a conformally holomorphically projective mapping if and only if the following condition holds*

$$(\bar{\nabla} - \nabla)_X X = 2\psi(X) \cdot X - 2\psi(FX) \cdot FX + g(X, X) \cdot \Sigma + \bar{g}(X, X) \cdot \Omega, \quad (10)$$

where ψ is a linear form, Σ and Ω are vector fields and there exist functions ϱ_1^* , ϱ_2^* and ϱ_3^* on M for which

$$\nabla_X \varrho_1^* = g(X, \Sigma), \quad \nabla_X \varrho_2^* = \bar{g}(X, \Omega), \quad \nabla_X \varrho_3^* = \psi(X).$$

Proof. The necessary relation (10) follows from equations (2) and (6). The sufficient condition follows from additional analysis of these relations. Further we prove with suitable ordering, that a conformally holomorphically projective mapping f is expressed as the composition $f_1 \circ f_2 \circ f_3$, where f_1 and f_3 are conformal mappings and f_2 is a holomorphically projective one, see the scheme

$$K_n \xrightarrow{f_1} K_n^1 \xrightarrow{f_2} K_n^2 \xrightarrow{f_3} \bar{K}_n.$$

We construct metrics

$$\overset{1}{g} = \exp(-2\varrho_1^*) \cdot g, \quad \text{and} \quad \overset{2}{g} = \exp(2\varrho_2^*) \cdot \bar{g}.$$

After computing the difference between connections of spaces $\overset{1}{K}_n = (M, \overset{1}{g})$ and $\overset{2}{K}_n = (M, \overset{2}{g})$, we see that

$$(\overset{2}{\nabla} - \overset{1}{\nabla})_X X = (2\psi(X) - \nabla_X \varrho_1^* - \nabla_X \varrho_2^*) \cdot X - 2\psi(FX) \cdot FX. \quad (11)$$

It follows that spaces $\overset{1}{K}_n$ and $\overset{2}{K}_n$ are in a holomorphically projective correspondence.

References

- [1] AMINOVA, A.V.: *Projective transformations of Riemannian manifolds*, J. Math. Sci., New York 113, No. 3, 367-470, 2003.
- [2] CHUDÁ, H.; MIKEŠ, J.: *On F-planar mappings with a certain initial conditions*, Proceedings of the 5th Int. Conf. APLIMAT, Part II, 2006.
- [3] CHUDÁ, H.; MIKEŠ, J.: *On composition of conformal and geodesic mappings*, 7th Conf. on Math. and Phys. on Technical Univesities, Brno, Sept. 22, 2011, Proc. of Contrib. 2011.
- [4] CHUDÁ, H.; MIKEŠ, J.: *Conformally geodesic mappings satisfying a certain initial condition*, Archivum math. (Brno), T. 47, 389-394, 2011.
- [5] DOMASHEV, V.V.; MIKEŠ, J.: *Theory of holomorphically projective mappings of Kählerian spaces*, Math. Notes **23** (2), 160-163, 1978; translation from Mat. Zametki, 23:2, 297-303, 1978.
- [6] EISENHART, L.P.: *Non-Riemannian Geometry*, Princeton Univ. Press. 1926. Amer. Math. Soc. Colloquium Publications **8**, 2000.
- [7] HINTERLEITNER, I.: *Selected Special Vector Fields and Mappings in Riemannian Geometry*. Vědecké spisy Vysokého učení technického v Brně, Edice PhD Thesis, 525, 1-20, 2009.
- [8] HINTERLEITNER, I.; MIKEŠ, J.: *On the equations of conformally-projective harmonic mappings*, AIP Conf. Proc. 956, 141-148, 2007.
- [9] HINTERLEITNER, I.; MIKEŠ, J.: *Geodesic mappings onto Weyl manifolds*, J. of Appl. Math. Aplimat, 2, 1, 125-133, 2009.
- [10] HINTERLEITNER, I.; MIKEŠ, J.: *On fundamental equations of geodesic mappings and their generalizations*, J. Math. Sci. 174, 5, 537-554 (2011); translation from Itogi Nauki i Tekhniki. Ser. Sovrem. Mat. Pril. Temat. Obz. 124, 7-34, 2010.
- [11] HINTERLEITNER, I.; MIKEŠ, J.: *Projective equivalence and manifolds with equiaffine connection*, J. Math. Sci. (2011); translation from Fundam. Prikl. Mat., 16:1, 47-54, 2010.
- [12] JUKL, M.; JUKLOVÁ, L.; MIKEŠ, J.: *The decomposition of tensor spaces with quaternionic structure*, 6th Int. Conf. on Aplimat, Feb. 06-09, 2007 Bratislava, SLOVAKIA. APLIMAT 2007, 6th Int. Conf. I, 217-222, 2007.
- [13] LAKOMÁ, L.; JUKL, M.: *The decomposition of tensor spaces with almost complex structure*, Suppl. Rend. Circ. Mat. Palermo, II. Ser. 72, 145-150, 2004.
- [14] AL LAMI, R.J.K.; ŠKODOVÁ, M.; MIKEŠ, J.: *On holomorphically projective mappings from equiaffine generally recurrent spaces onto Kählerian spaces*, Arch. Math., Brno 42, No. 5, 291-299, 2006.
- [15] MIKESH, I.: *Equidistant Kähler spaces*, Math. Notes **38** (4), 855-858 (1985); translation from Mat. Zametki, 38:4, 627-633, 1985.

- [16] MIKESH, I.; STARKO, G.A.: *Hyperbolically Sasakian and equidistant hyperbolically Kahlerian spaces*, J. of Sov. Math. **59** (2), 756-760, 1992; translation from Ukr. Geom. Sb. 32, 92-98, 1989.
- [17] MIKEŠ, J.: *Holomorphically projective mappings and their generalizations*, Journal of Mathematical Sciences (New York), 89:3, 1334-1353 (1998); translation from Geometry 3, Itogi Nauki i Tekhniki. Ser. Sovrem. Mat. Pril. Temat. Obz., 30, VINITI, Moscow, 258-289, 2002.
- [18] MIKEŠ, J.; HINTERLEITNER, I.: *On geodesic mappings of manifolds with affine connection*, Acta Math. Acad. Paedagog. Nyházi. (N.S.) 26, no. 2, 343-347, 2010.
- [19] MIKEŠ, J.; CHODOROVÁ, M.: *On concircular and torse-forming vector fields on compact manifolds*, Acta Math. Acad. Paedagog. Nyházi. (N.S.) 26, no. 2, 329-335, 2010.
- [20] MIKEŠ, J.; JUKL, M.; JUKLOVÁ, L.: *Some results on traceless decomposition of tensors*, Journal of Mathematical Sciences, 1-14, (2011); translation from Itogi Nauki i Tekhniki. Ser. Sovrem. Mat. Pril. Temat. Obz., 124, 139-158, 2010.
- [21] MIKEŠ, J.; KIOSAK, V.; VANŽUROVÁ, A.: *Geodesic mappings of manifolds with affine connection*, Palacky University Press, Olomouc, 220p, 2008.
- [22] MIKEŠ, J.; VANŽUROVÁ, A.; HINTERLEITNER, I.: *Geodesic mappings and some generalizations*, Palacky University Press, Olomouc, 304p, 2009.
- [23] MIKEŠ, J.; STRAMBACH, K.: *Differentiable structures on elementary geometries*, Result. Math. 53, No. 1-2, 153-172, 2009.
- [24] PETROV, A.Z.: *New methods in the general theory of relativity*, Moscow, Nauka, 1966.
- [25] RADULOVIĆ, Ž. and MIKEŠ, J.: *Geodesic mappings of conformal Kähler spaces*, Russian Math. (Iz. VUZ) 38, 4850, 1994 Izv. Vyssh. Uchebn. Zaved. Mat., no. 3, 50-52, 1994.
- [26] SHIHA, M.; MIKEŠ, J.: *On equidistant parabolically Kählerian spaces*, Tr. Geom. Semin., 22, 97-107, 1994.
- [27] SINYUKOV, N.S.: *Geodesic mappings of Riemannian spaces*, Nauka, Moscow, 1979.
- [28] YANO, K.: *Differential geometry of complex and almost complex spaces*, Pergamon Press, 1965.
- [29] ZUDINA, T.V.; STEPANOV, S.E.: *On a class of equiaffine mappings*, (Russian. English summary) Differ. Geom. Mnogoobr. Figur 36, 43-49, 2005.

Current address

Chodorová Marie, Mgr. Ph.D.

Dept. Algebra and Geometry, Faculty of Science,
Palacky University in Olomouc, 17. listopadu 12, 77900 Olomouc, Czech Republic,
+420 58 563 4647, marie.chodorova@upol.cz

Chudá Hana, Mgr. Ph.D.

Dept. of Mathematics, Faculty of Applied Informatics,
Tomas Bata University in Zlín, T.G. Masaryka 5555, 760 01 Zlín, Czech Republic,
+420 57 603 5007, chuda@fai.utb.cz

Shiha Mohsen, Ph.D.

Department of Mathematics, University of Homs
Homs, Syria
e-mail: mohsen_sheha@yahoo.com

MULTIPLE COVARIANT DERIVATIVE AND DECOMPOSITION PROBLEMS

JUKL Marek, (CZ), JUKLOVÁ Lenka, (CZ), MIKEŠ Josef, (CZ)

Abstract. This paper deals the properties of m -covariant derivative with respect to the general decomposition of tensor fields on manifolds with affine connection. It is shown that properties of multiple covariant derivative $\underbrace{\nabla \nabla \cdots \nabla}_{m \times}$ are transferred to the components of given decomposition.

Key words and phrases. multiple covariant derivative, decomposition of tensor.

Mathematics Subject Classification. Primary 53A55, 15A69; Secondary 53B05.

1 Introduction

The trace decomposition of tensors on an n -dimensional Riemannian manifold M_n was described by Weyl [19]. This decomposition problem may be naturally generalized for a number of certain other cases (see for example [1, 3, 4, 7, 8, 11, 12, 18, 20]).

The theory of decompositions of tensors has been used in the studying geodesic and other mappings of special Riemannian spaces (see [10, 13, 14, 15, 16, 17]).

In this paper, we bring certain more general results for decompositions of tensor (and also tensor fields) than they are contained in [5, 19]. Crasmareanu [2] has studied decompositions of tensor fields for the case covariant derivative of such tensors satisfies certain conditions.

These results are generalized for multiple derivative $\underbrace{\nabla \nabla \cdots \nabla}_{m \times}$. We show that multiple derivative $\underbrace{\nabla \nabla \cdots \nabla}_{m \times}$ of given tensor is transferred to components of studied decompositions.

2 General decomposition of tensors

It is well known that on every manifold M_n a positive definite metric g determining the structure of Riemannian manifold on M_n may be introduced. For practical reasons (e.g. in theoretical

physics) the pseudo-Riemannian manifolds with indefinite pseudometric g are considered, too. Through the following we will in both cases use terms *Riemannian manifold* and *metric*, only.

Let a manifold M_n with a metric g be given. We introduce the following denotation: For every ordered system of indices i_1, i_2, \dots, i_q , $1 \leq i_\rho \leq n$, and every couple (i_ρ, i_σ) , $\rho < \sigma$, of them we will by

$$M_{i_1, \dots, i_{\rho-1} \overset{\rho}{\vee} i_{\rho+1} \dots i_{\sigma-1} \overset{\sigma}{\vee} i_{\sigma+1} \dots i_q}^{(\rho, \sigma)}$$

denote the tensor of the type $(0, q-2)$, where the ρ -th and σ -th indices are omitted.

Then we have the following theorem:

Theorem 2.1 *Let a tensor $T_{i_1 \dots i_q}$ of the type $(0, q)$ and s ($\leq \frac{1}{2} q(q-1)$) couples of indices*

$$(i_{k_1}, i_{l_1}), (i_{k_2}, i_{l_2}), \dots, (i_{k_s}, i_{l_s}), \quad \text{with } k_\sigma < l_\sigma, \quad (1)$$

be given. Then there exists the following decomposition of the tensor $T_{i_1 \dots i_q}$:

$$T_{i_1 \dots i_q} = \tilde{T}_{i_1 \dots i_q} + \sum_{\sigma=1}^s g_{i_{k_\sigma} i_{l_\sigma}} \cdot M_{i_1, \dots, i_{k_\sigma-1} \overset{k_\sigma}{\vee} i_{k_\sigma+1} \dots i_{l_\sigma-1} \overset{l_\sigma}{\vee} i_{l_\sigma+1} \dots i_q}^{(k_\sigma, l_\sigma)}, \quad (2)$$

where the tensor \tilde{T} fulfilling

$$\tilde{T}_{i_1 \dots i_{k_\sigma} \dots i_{l_\sigma} \dots i_q} \cdot g^{i_{k_\sigma} i_{l_\sigma}} = 0, \quad (3)$$

for every indices (1), is determined uniquely and M are certain tensors of the type $(0, q-2)$.

Proof. Let M_n be a Riemannian manifold with metric g . Now, let us consider an a priori choosen point x_0 at M_n . Values of metric g as well as of all given tensors will be considered in this choosen point.

Considering such coordination system that the matrix of g fulfils

$$g_{ij} = g_0 \cdot D, \quad \text{where } g_0 \in R, \quad \text{and } D = \text{diag}(\pm 1, \pm 1, \dots, \pm 1), \quad (4)$$

we obtain that the inverse matrix g^{-1} has the form:

$$g^{ij} = g_0^{-1} \cdot D. \quad (5)$$

The existence of the coordinated system above is guaranteed only for a concrete point x_0 . A metric (4) exists globally in Euclidean and pseudo-Euclidean spaces.

Therefore the relation (3) may be written by

$$\sum_{k_\sigma, l_\sigma=1}^n \tilde{T}_{i_1 \dots i_{k_\sigma} \dots i_{l_\sigma} \dots i_q} \cdot g_{i_{k_\sigma} i_{l_\sigma}} = 0. \quad (6)$$

Let us define the inner product \circ on T_q^0 by

$$\overset{1}{T} \circ \overset{2}{T} \stackrel{\text{def}}{=} \sum_{i_1, \dots, i_q=1}^n \overset{1}{T}_{i_1 \dots i_q} \cdot \overset{2}{T}_{i_1 \dots i_q}. \quad (7)$$

Now, let us construct a linear subspace $\overset{*}{T} \subseteq T_p^0$ which is generated by all tensors in the form

$$\sum_{\sigma=1}^s g_{i_{k_\sigma} i_{l_\sigma}} \cdot M_{i_1, \dots, i_{k_\sigma-1} \underset{\vee}{i_{k_\sigma+1} \dots i_{l_\sigma-1}} \underset{\vee}{i_{l_\sigma+1} \dots i_q}}^{(k_\sigma, l_\sigma)}. \quad (8)$$

Considering a linear subspace in T_p^0 which contains all tensors fulfilling (3) we clearly for any tensors \tilde{T} of this subspace get $\tilde{T} \circ \overset{*}{T} = 0$. Therefore this subspace is contained in the orthogonal complement of the subspace $\overset{*}{T}$, which implies the unicity of the decomposition (2) of the tensor $T_{i_1 \dots i_p}$.

3 Special decompositions of tensors

Comparing the decomposition in the following theorem with decompositions in Theorem 2.1 (and Theorem 3.3) as well as the fundamental theorem proved by Weyl (see [19]) we may remark that in the following one there is uniquely determined not only tensor \tilde{T} but also all tensors $\overset{(*)}{M}$.

Theorem 3.1 *Let a tensor $T_{i_1 i_2 \dots i_q}$ of the type $(0, q)$ and following couples of indices*

$$(i_1, i_2), (i_1, i_3), \dots, (i_1, i_q) \quad (9)$$

be given. Then for $n > q - 1$ there exists exactly one decomposition of the tensor $T_{i_1 \dots i_q}$ in the form

$$T_{i_1 \dots i_q} = \tilde{T}_{i_1 \dots i_q} + \sum_{\sigma=2}^n g_{i_1 i_\sigma} \cdot M_{i_2, \dots, i_{s-1} \underset{\vee}{i_{s+1} \dots i_q}}^{(1, \sigma)}, \quad (10)$$

where the tensor $\tilde{T}_{i_1 \dots i_q}$ fulfilling

$$\tilde{T}_{i_1 \dots i_\sigma \dots i_q} \cdot g^{i_1 i_\sigma} = 0, \quad (11)$$

for any couples of indices (9), and all tensors $\overset{(i_1 i_\sigma)}{M}$ of the type $(0, q-2)$ are determined uniquely.

Proof. With respect to the Theorem 2.1 we only need to prove the unicity of the tensors $\overset{(i_1 i_\sigma)}{M}$. Contracting (10) with $g^{i_1 i_s}$ for $s = 2, \dots, n$ and using (11) for the tensor \tilde{T} we obtain the following system of linear equations

$$T_{i_1 \dots i_q} g^{i_1 i_s} = 0 + n \cdot M_{i_2 \dots i_{s-1} i_{s+1} \dots i_q}^{(1, \sigma)} + \sum_{\sigma=2, \sigma \neq s}^q M_{i_2 \dots i_{\sigma-1} i_{\sigma+1} \dots i_{s-1} i_{s+1} \dots i_q}^{(1, \sigma)} \quad (12)$$

It follows from the fixed point theorem that this system of equations with variables $M^{(i_1 i_2)}, \dots, M^{(i_1 i_n)}$ has for $n > q - 1$ exactly one solution.

Let us recall the raising indices of tensor. If a tensor T of the type $(0, p + q)$ is given then we may construct a tensor of the type (p, q) by the following:

$$T_{j_1 \dots j_q}^{i_1 \dots i_p} \stackrel{\text{def}}{=} g^{i_1 \alpha_1} g^{i_2 \alpha_2} \dots g^{i_p \alpha_p} T_{\alpha_1 \dots \alpha_p j_1 \dots j_q} \quad (13)$$

Raising indices we get from Theorem 3.1 known unique decomposition of the tensor $T_{i_2 \dots i_p}^{i_1}$:

$$T_{i_2 \dots i_p}^{i_1} = \tilde{T}_{i_2 \dots i_p}^{i_1} + \delta_{i_2}^{i_1} M_{i_3 \dots i_p}^{12} + \delta_{i_3}^{i_1} M_{i_2 i_4 \dots i_p}^{13} + \dots + \delta_{i_p}^{i_1} M_{i_2 \dots i_{p-1}}^{1p}, \quad (14)$$

where the tensor \tilde{T} is traceless, i.e.

$$\tilde{T}_{\dots \alpha \dots}^{\alpha} = 0 \quad (15)$$

and tensor \tilde{T} as well as tensors M^{1*} of the type $(0, p - 1)$ are uniquely determined.

As we have mentioned above decomposition (14) is presented by Weyl in [19], but the unicity of tensors M^{1*} is not contained there.

Using (13) we get from Theorem 2.1, immediately

Theorem 3.2 *Let a tensor $T_{i_1 \dots i_p}^{j_1 \dots j_q}$ of the type (p, q) and $s (\leq p, q)$ couples of indices*

$$(i_{k_1}, j_{l_1}), (i_{k_2}, j_{l_2}), \dots, (i_{k_s}, j_{l_s}) \quad \text{with} \quad 1 \leq k_\sigma \leq p, \quad 1 \leq l_\sigma \leq q, \quad (16)$$

be given. Then there exists the following decomposition of the tensor $T_{i_1 \dots i_p}$:

$$T_{j_1 \dots j_q}^{i_1 \dots i_p} = \tilde{T}_{j_1 \dots j_q}^{i_1 \dots i_p} + \sum_{\sigma=1}^s \delta_{j_{l_\sigma}}^{i_{k_\sigma}} \cdot M_{(l_\sigma) j_1 \dots j_{l_\sigma-1} j_{l_\sigma+1} \dots j_q}^{(k_\sigma) i_1 \dots i_{k_\sigma-1} i_{k_\sigma+1} \dots i_p}, \quad (17)$$

where the tensor \tilde{T} fulfilling

$$\tilde{T}_{j_1 \dots j_{l_\sigma} \dots j_q}^{i_1 \dots i_{k_\sigma} \dots i_p} \cdot \delta_{i_{k_\sigma}}^{j_{l_\sigma}} = 0, \quad (18)$$

for every indices (16), is determined uniquely and $M_{(l_\sigma)}^{(k_\sigma)}$ are certain tensors of the type $(p-1, q-1)$.

The condition (18) means that the tensor \tilde{T} is traceless over any pair of indices (16). This may be expressed by

$$\tilde{T}_{j_1 \dots j_{k\sigma-1} \alpha j_{k\sigma+1} \dots j_q}^{i_1 \dots i_{k\sigma-1} \alpha i_{k\sigma+1} \dots i_p} = 0. \quad (19)$$

Let us remark that Theorem 3.2 holds on every manifold since the metric tensor (which was used in the proof of Theorem 2.1) may be constructed in any case.

Immediately, we obtain from this the following Theorem which is presented in [11, 5].

Theorem 3.3 *Let $T_{i_1 \dots i_p}^{j_1 \dots j_p}$ be a tensor of the type (p, q) . If $n+1 \geq p+q$ then there exists unique decomposition of the $T_{i_1 \dots i_p}^{j_1 \dots j_p}$ in the following form.*

$$T_{i_1 \dots i_p} = \tilde{T}_{i_1 \dots i_p} + \sum_{t=1}^{\min\{p,q\}} \sum_{\oplus} \delta_{j_{\sigma_1}}^{i_{\rho_1}} \delta_{j_{\sigma_2}}^{i_{\rho_2}} \dots \delta_{j_{\sigma_t}}^{i_{\rho_t}} M^{\star} \dots, \quad (20)$$

where

$$\oplus = \left\{ \begin{array}{ll} \rho_1, \rho_2, \dots, \rho_t = 1, 2, \dots, p & (\rho_1 < \rho_2 < \dots < \rho_t) \\ \sigma_1, \sigma_2, \dots, \sigma_t = 1, 2, \dots, q & (\sigma_i \text{ are mutually different}), \end{array} \right. \quad (21)$$

$$\star = \left\{ \begin{array}{cccc} \rho_1 & \rho_2 & \dots & \rho_t \\ \sigma_1 & \sigma_2 & \dots & \sigma_t \end{array} \right\}$$

and tensors $\tilde{T}_{i_1 \dots i_p}$ a M^{\star} are tracers.

4 Multiple derivative and decomposition

Let us consider m -tuple covariant derivative $\underbrace{\nabla \nabla \dots \nabla}_{m \times}$.

Clearly, we may write

$$\begin{aligned} 1) \quad & \underbrace{\nabla \nabla \dots \nabla}_{m \times} (\overset{1}{T} \pm \overset{2}{T}) = \underbrace{\nabla \nabla \dots \nabla}_{m \times} \overset{1}{T} \pm \underbrace{\nabla \nabla \dots \nabla}_{m \times} \overset{2}{T}; \\ 2) \quad & \nabla \delta = 0; \\ 3) \quad & \nabla g = 0, \end{aligned} \quad (22)$$

where $\overset{1}{T}, \overset{2}{T}$ are arbitrary tensors of the same type and δ is a Kronecker tensor.

Respecting these properties and using Theorem 2.1 we may prove following lemmas.

Lemma 4.1 *Let T be a tensor of the type $(0, p)$ with the decomposition (2), for choosen couples of indices (1).*

If $\underbrace{\nabla \nabla \dots \nabla}_{m \times} \tilde{T} = 0$ and $\underbrace{\nabla \nabla \dots \nabla}_{m \times} M^{\star} = 0$, where $\star = (k_{\sigma}, l_{\sigma})$, then $\underbrace{\nabla \nabla \dots \nabla}_{m \times} T = 0$.

Lemma 4.2 Let T be a tensor of the type $(0, p)$ with the decomposition (2), for choosen couples of indices (1).

$$\text{If } \underbrace{\nabla \nabla \cdots \nabla T}_{m \times} = 0 \text{ then } \underbrace{\nabla \nabla \cdots \nabla \tilde{T}}_{m \times} = 0.$$

Let a $C^\infty(M)$ -module of m -differential forms $\Omega^m(M)$ be given on the manifold M_n . Considering a m -differential form $\omega \in \Omega^m(M)$ (see [2]) with

$$\nabla_{X_m} \cdots \nabla_{X_1} T = \omega(X_1, \dots, X_m) \cdot T, \quad (23)$$

and choosing indices according to the Theorem 3.1 we may prove the following theorem.

Theorem 4.3 Let a tensor T of the type $(0, p)$ and the decomposition of the T in the form (10) for couples of indices (9) be given. Then

$$\underbrace{\nabla \nabla \cdots \nabla T}_{m \times} = \omega T, \quad (24)$$

if and only if

$$\underbrace{\nabla \nabla \cdots \nabla \tilde{T}}_{m \times} = \omega \tilde{T} \quad \text{and} \quad \underbrace{\nabla \nabla \cdots \nabla M}_{m \times} = \omega M^{1\sigma} \quad (25)$$

where ω is a m -differential form in $\Omega^m(M)$.

Proof. Let a decomposition of the tensor T in the form (10) for couples of indices (9) be given. For purposes of this proof, let us for an arbitrary tensor T denote $\underbrace{\nabla \nabla \cdots \nabla T}_{m \times}$ by LT , only.

Firstly, let us suppose that

$$L\tilde{T} = \omega \tilde{T} \text{ and } LM^{1\sigma} = \omega M^{1\sigma}. \quad (26)$$

Substituting (26) into (10) we get $LT = \omega T$, immediately.

Now, let us suppose $LT = \omega T$. Applying L on the expression of T in the form (10) we have

$$LT_{i_1 \cdots i_p} = L\tilde{T}_{i_1 \cdots i_p} + g_{i_1 i_2} L\tilde{M}_{i_3 \cdots i_p}^{12} + g_{i_1 i_3} L\tilde{M}_{i_2 i_4 \cdots i_p}^{13} + \cdots + g_{i_1 i_p} L\tilde{M}_{i_2 \cdots i_{p-1}}^{1p} \quad (27)$$

i.e.

$$\omega T = L\tilde{T}_{i_1 \cdots i_p} + g_{i_1 i_2} L\tilde{M}_{i_3 \cdots i_p}^{12} + g_{i_1 i_3} L\tilde{M}_{i_2 i_4 \cdots i_p}^{13} + \cdots + g_{i_1 i_p} L\tilde{M}_{i_2 \cdots i_{p-1}}^{1p} \quad (28)$$

Applying a form ω on the expression of T in the form (10) we obtain

$$\omega T = L\tilde{T}_{i_1 \cdots i_p} + g_{i_1 i_2} \omega \tilde{M}_{i_3 \cdots i_p}^{12} + g_{i_1 i_3} \omega \tilde{M}_{i_2 i_4 \cdots i_p}^{13} + \cdots + g_{i_1 i_p} \omega \tilde{M}_{i_2 \cdots i_{p-1}}^{1p}. \quad (29)$$

Comparing 28 and 29 we have

$$\begin{aligned} (L\tilde{T} - \omega \tilde{T}) + (g_{i_1 i_2} L\tilde{M}_{i_3 \cdots i_p}^{12} - g_{i_1 i_2} \omega \tilde{M}_{i_3 \cdots i_p}^{12}) + (g_{i_1 i_3} L\tilde{M}_{i_2 i_4 \cdots i_p}^{13} - g_{i_1 i_3} \omega \tilde{M}_{i_2 i_4 \cdots i_p}^{13}) + \\ + \cdots + (g_{i_1 i_p} L\tilde{M}_{i_2 \cdots i_{p-1}}^{1p} - g_{i_1 i_p} \omega \tilde{M}_{i_2 \cdots i_{p-1}}^{1p}) = 0 \end{aligned} \quad (30)$$

Contracting (30) with $g^{i_1 i_\sigma}$ for $\sigma = 2, \dots, n$ we may derive a system of linear equations

$$n \cdot (L\overset{1\sigma}{M} - \omega \overset{1\sigma}{M}) + \sum_{\tau=2, \tau \neq \sigma}^p (L\overset{1\tau}{M} - \omega \overset{1\tau}{M}) = 0 \quad (31)$$

For $n > p - 1$ this system has the following unique solution $L\overset{1\sigma}{M} - \omega \overset{1\sigma}{M} = 0$ pro $\sigma = 2, \dots, n$, which implies that $L\tilde{T} - \omega T = 0$.

Further, it follows from Theorem 4.3

Theorem 4.4 *Let a tensor T of the type $(0, p)$ and the decomposition of the T in the form (10) for couples of indices (9) be given. Then*

$$\underbrace{\nabla \nabla \dots \nabla}_{m \times} T = 0 \quad (32)$$

if and only if

$$\underbrace{\nabla \nabla \dots \nabla}_{m \times} \tilde{T} = 0 \text{ and } \underbrace{\nabla \nabla \dots \nabla}_{m \times} \overset{1\sigma}{M} = 0. \quad (33)$$

Acknowledgement

The authors are supported by the grant from Grant Agency of Czech Republic GA ĆR P201/11/0356.

References

- [1] CRASMAREANU, M.: *Particular trace decompositions and applications of trace decomposition to almost projective invariants*. Mathematica Bohemica, Vol. 126, pp. 631–637, 2001.
- [2] CRASMAREANU, M.: *The decomposition and recurrency*. Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica, Vol. 40, pp. 43–46, 2001.
- [3] JUKL, M., JUKLOVÁ, L., MIKEŠ, J.: *The decomposition of tensor spaces with quaternionic structure*. APLIMAT 2007, Bratislava, pp. 217–222, 2007.
- [4] JUKL, M., JUKLOVÁ, L., MIKEŠ, J.: *On generalized trace decompositions problems*. Trudy 3-ij mezhdunarodnoj konferencii “Funkcionalnyje prostranstva. Differencialnyje operatory. Obshchaja topologija”, Moskovskij fizicheskotechnicheskij institut, Moskva, pp. 299–314, 2008.
- [5] KRUPKA, D.: *The trace decomposition of tensor spaces*. Linear and Multilinear Algebra, Vol. 54, No. 4, pp. 235–263, 2006.
- [6] LAKOMÁ, L.: *Projections of tensor spaces*. Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica, Vol. 38, pp. 87–93, 1999.
- [7] LAKOMÁ, L., JUKL, M.: *The decomposition of tensor spaces with almost complex structure*. Rendiconti del circolo matematico di Palermo, Ser. II, Suppl., Vol. 72, pp. 145–150, 2004.
- [8] LAKOMÁ, L., MIKEŠ, J.: *On the special trace decomposition problem on quaternionic structure*. Proc. of the Third Internat. Workshop on Diff. Geom. and its Appl.; The First German-Romanian Seminar on Geom., Sibiu, Romania, pp. 225–229, 1997.

- [9] LAKOMÁ, L., MIKEŠ, J., MIKUŠOVÁ, L.: *Decomposition of tensor spaces*. Diff. Geom. and its Appl., Proc. of 7th Int. Conf. DGA 98, Brno, Czech Republic, pp. 371–378, 1998.
- [10] MIKEŠ, J.: *Geodesic and holomorphically projective mappings of special Riemannian space*. PhD. Thesis, Odessa Univ., 107 p., 1979.
- [11] MIKEŠ, J.: *On general trace decomposition problem*. Diff. Geom. and its Appl., Proc. of 6th Int. Conf., Brno, Czech Republic, pp. 45–50, 1995.
- [12] MIKEŠ, J., JUKL, M., JUKLOVÁ, L.: *Some results on traceless decomposition of tensors*. J. Math. Sci. Vol. 174, Issue 5, pp. 627–640, 2011.
- [13] MIKEŠ, J., RACHŮNEK, L.: *Torse-forming vector fields in T-semisymmetric Riemannian spaces*. Steps in differential geometry. Proc. of Colloq. on Diff. Geometry, Debrecen, Hungary, July 25-30, 2000. Debrecen: Univ. Debrecen, Institute of Mathematics and Informatics, pp. 219–229, 2001.
- [14] MIKEŠ, J., RACHŮNEK, L.: *T-semisymmetric spaces and concircular vector fields*. The proceedings of the 21th winter school “Geometry and physics”, Srní, Czech Republic, January 13-20, 2001. Rend. Circ. Mat. Palermo, Suppl., Ser. II., Vol. 69, pp. 187-193, 2002.
- [15] MIKEŠ, J., RADULOVIC, Ž., HADDAD, M.: *Geodesic and holomorphically projective mappings of m-pseudo- and m-quasisymmetric Riemannian spaces*. (English. Russian original) Russ. Math., Vol. 40, No. 10, pp. 28–32, 1996; translation from Izv. Vyssh. Uchebn., Mat., Vol. 413, No. 10, pp. 30–35, 1996.
- [16] MIKEŠ, J., VANŽUROVÁ, A., HINTERLEITNER, I.: *Geodesic mappings and some generalizations*. Palacky University Press., Olomouc, 2009.
- [17] RACHŮNEK, L., MIKEŠ, J.: *On tensor fields semiconjugated with torse-forming vector fields*. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math., Vol. 44, pp. 151–160, 2005.
- [18] VISHNEVSKIJ, V. V., SHIROKOV, A. P., SHURYGIN, V. V.: *Spaces over algebras. (Prostranstva nad algebrami)*. Izd. Kazan. Univ., Kazan’: 1985.
- [19] WEYL, H.: *The classical groups*. Princenton University Press, Princenton, 316 p., 1946; fifteenth printing 1997.
- [20] YANO, K.: *Differential geometry on complex and almost complex spaces*. Pergamon Press, Oxford-London-New York-Paris-Frankfurt, Vol. XII, 323 p., 1965.

Current address

JUKL Marek, RNDr. PhD.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University,
17. listopadu 12, 771 46 Olomouc, Czech Rep., marek.jukl@upol.cz.

JUKLOVÁ Lenka, RNDr. PhD.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University,
17. listopadu 12, 771 46 Olomouc, Czech Rep., lenka.juklova@upol.cz.

MIKEŠ Josef, Prof. RNDr. DrSc.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University,
17. listopadu 12, 771 46 Olomouc, Czech Rep., josef.mikes@upol.cz.

THE CURVATURES OF SPECIAL FUNCTIONS IN ECONOMY APPLICATION OF CARTAN'S MOVING FRAME METHOD

KAŇKA Miloš, (CZ), ELIÁŠOVÁ Lada, (CZ)

Abstract. The aim of this article is to give geometrical analysis of a special type of Cobb-Douglas surface, especially the formula of Gauss curvature

$$\gamma(x, y) = (x, y, Ax^\alpha y^\beta), \text{ where } A = 1, x > 0, y > 0, \alpha = 1, 2, \beta = 1.$$

For this purpose we use the Cartan's moving frame method.

Key words and phrases. Orthogonal frame, orthonormal frame, Cartan's forms, Gaussian curvature, Mean curvature, Maurer-Cartan equations.

1 Introduction

Let $U \subset \mathcal{R}^2$ and $x : U \rightarrow \mathcal{R}^3$ is a map. We say that this map is regular if the Jacobian matrix $J(x)(u, v)$ has rank 2 for all $(u, v) \in U$. Let us suppose that for every point $p \in M \subset \mathcal{R}^3$ exist an open set $U \subset \mathcal{R}^2$, an open set $V \subset \mathcal{R}^3$, $p \in V$, and a regular differentiable homeomorphism $x : U \rightarrow V \cap M$. A subset $M \subset \mathcal{R}^3$ is called a two-dimensional regular surface in \mathcal{R}^3 . Let $x(U) \subset V \cap M \subset \mathcal{R}^3$ be a neighbourhood of $p \in M$ such that the restriction $x|_U$ is an differentiable homeomorphism into $x(U) \subset V \cap M$ and that it is possible to choose in $x(U)$ an orthonormal moving frame $\{E_1, E_2, E_3\}$ in such a way that E_1, E_2 are tangent to $x(U)$ and E_3 is a non-vanishing normal to $x(U)$. We first discuss the Cartan structural equations for a two-dimensional surface in \mathcal{R}^3 .

2 Structural equations

We first discuss the Cartan structural equations for a two-dimensional surface in \mathcal{R}^3 . Differentiating a patch $x(u, v)$ we obtain

$$dx = x_u du + x_v dv,$$

where x_u, x_v are tangent vector fields. Let us denote

$$n(u, v) = x_u \times x_v$$

the normal vector field. With respect to the orthonormal moving frame $\{E_1, E_2, E_3\}$ we define forms

$$\theta_i = E_i dx = E_i x_u du + E_i x_v dv, \quad i = 1, 2,$$

Since x_u and x_v are tangent to $x(U)$ we have $dx \cdot E_3 = 0$ which implies $\theta_3 = 0$ so we have

$$\begin{aligned} \theta_1 &= E_1 x_u du + E_1 x_v dv, \\ \theta_2 &= E_2 x_u du + E_2 x_v dv. \end{aligned} \tag{1}$$

Each vector $E_i : U \subset \mathcal{R}^3 \rightarrow \mathcal{R}^3$ is a differentiable map and the differential

$$dE_i : \mathcal{R}^3 \rightarrow \mathcal{R}^3$$

is a linear map. So we may write (using Einstein's notation)

$$dE_i = \omega_{ij} E_j$$

where ω_{ij} are linear forms on \mathcal{R}^3 and since E_i are differentiable, ω_{ij} are $3^2 = 9$ differentiable forms. So we have

$$\begin{aligned} dE_1 &= \omega_{11} E_1 + \omega_{12} E_2 + \omega_{13} E_3, \\ dE_2 &= \omega_{21} E_1 + \omega_{22} E_2 + \omega_{23} E_3, \\ dE_3 &= \omega_{31} E_1 + \omega_{32} E_2 + \omega_{33} E_3. \end{aligned} \tag{2}$$

Differentiating the equation $E_i \cdot E_j = \delta_{ij}$ we obtain

$$dE_i E_j + E_i dE_j = \omega_{ij} + \omega_{ji} = 0.$$

Forms ω_{ij} are antisymmetric

$$\omega_{ii} = 0, \quad \omega_{ij} = -\omega_{ji}. \tag{3}$$

From (2) and (3) we have

$$\begin{aligned} dE_1 &= \omega_{12} E_2 + \omega_{13} E_3, \\ dE_2 &= -\omega_{12} E_1 + \omega_{23} E_3, \\ dE_3 &= -\omega_{13} E_1 - \omega_{23} E_2. \end{aligned} \tag{4}$$

Forms dx and dE_i have vanishing exterior derivatives

$$0 = d^2 x = dE_1 \wedge \theta_1 + E_1 d\theta_1 + dE_2 \wedge \theta_2 + E_2 d\theta_2. \tag{5}$$

Substituting (4) for (5) we obtain

$$(\omega_{12} E_2 + \omega_{13} E_3) \wedge \theta_1 + E_1 d\theta_1 + (\omega_{21} E_1 + \omega_{23} E_3) \wedge \theta_2 + E_2 d\theta_2 = 0. \tag{6}$$

From (6) there immediately follows

$$(d\theta_1 + \omega_{21} \wedge \theta_2)E_1 + (d\theta_2 + \omega_{12} \wedge \theta_1)E_2 + (\omega_{13} \wedge \theta_1 + \omega_{23} \wedge \theta_2)E_3 = 0. \quad (7)$$

The linear independence of vectors E_1, E_2, E_3 and equation (7) gives the following equations:

$$d\theta_1 = \omega_{12} \wedge \theta_2, \quad (8)$$

$$d\theta_2 = \omega_{21} \wedge \theta_1, \quad (9)$$

$$0 = \omega_{13} \wedge \theta_1 + \omega_{23} \wedge \theta_2. \quad (10)$$

Differentiating (4) gives:

$$\begin{aligned} 0 &= d^2 E_1 = d\omega_{12}E_2 - \omega_{12} \wedge dE_2 + d\omega_{13}E_3 - \omega_{13} \wedge dE_3, \\ d\omega_{12}E_2 - \omega_{12} \wedge (\omega_{21}E_1 + \omega_{23}E_3) + d\omega_{13}E_3 - \omega_{13} \wedge (\omega_{31}E_1 + \omega_{32}E_2) &= 0, \\ (d\omega_{12} - \omega_{13} \wedge \omega_{32})E_2 + (d\omega_{13} - \omega_{12} \wedge \omega_{23})E_3 &= 0. \end{aligned} \quad (11)$$

From (11) we have

$$\begin{aligned} d\omega_{12} &= \omega_{13} \wedge \omega_{32}, \\ d\omega_{13} &= \omega_{12} \wedge \omega_{23}. \end{aligned} \quad (12)$$

Analogically:

$$\begin{aligned} d^2 E_2 &= d\omega_{21}E_1 - \omega_{21} \wedge dE_1 + d\omega_{23}E_3 - \omega_{23} \wedge dE_3 = 0, \\ d\omega_{21}E_1 - \omega_{21} \wedge (\omega_{12}E_2 + \omega_{13}E_3) + d\omega_{23}E_3 - \omega_{23} \wedge (\omega_{31}E_1 + \omega_{32}E_2) &= 0, \\ (d\omega_{23} - \omega_{21} \wedge \omega_{13})E_3 + (d\omega_{21} - \omega_{23} \wedge \omega_{31})E_1 &= 0. \end{aligned} \quad (13)$$

From (13) we have

$$d\omega_{23} = \omega_{21} \wedge \omega_{13}. \quad (14)$$

Equations (8), (9), (10), (12) and (14) are called Maurer-Cartan structural equations. From equation (9) and Cartan's lemma we have

$$\begin{aligned} \omega_{13} &= \alpha_{11}\theta_1 + \alpha_{12}\theta_2, \\ \omega_{23} &= \alpha_{12}\theta_1 + \alpha_{22}\theta_2. \end{aligned} \quad (15)$$

From (15) and (12) we have

$$d\omega_{12} = \omega_{13} \wedge \omega_{32} = -\omega_{13} \wedge \omega_{23} = -(\alpha_{11}\theta_1 + \alpha_{12}\theta_2) \wedge (\alpha_{12}\theta_1 + \alpha_{22}\theta_2). \quad (16)$$

Equation (16) gives

$$d\omega_{12} = -(\alpha_{11}\alpha_{22} - \alpha_{12}^2)\theta_1 \wedge \theta_2 = -K\theta_1 \wedge \theta_2, \quad (17)$$

where $K = \alpha_{11}\alpha_{22} - \alpha_{12}^2$ is the Gaussian curvature.

Differentiating the equation $E_3 \cdot E_3 = 1$ we have

$$dE_3 \cdot E_3 = 0,$$

which means that dE_3 is a tangent vector, i.e. $dE_3 \in T_p(M)$. The mapping

$$W(\alpha x_u + \beta x_v) = -\alpha \frac{\partial E_3}{\partial u} - \beta \frac{\partial E_3}{\partial v}$$

is a linear mapping

$$W : T_p(M) \rightarrow T_p(M).$$

3 Example 1

Let $x(u, v) = (u, v, u \cdot v)$ be a parametrized utility surface in \mathcal{R}^3 . Moving frame is

$$\begin{aligned}x_u &= (1, 0, v), \\x_v &= (0, 1, u), \\n &= (-v, -u, 1).\end{aligned}$$

Orthonormal frame is

$$\begin{aligned}E_1 &= \frac{1}{\sqrt{1+v^2}}(1, 0, v), \\E_2 &= \frac{1}{\sqrt{1+v^2} \cdot \sqrt{1+u^2+v^2}}(-uv, 1+v^2, u), \\E_3 &= \frac{1}{\sqrt{1+u^2+v^2}}(-v, -u, 1).\end{aligned}$$

From (1) follows

$$\theta_1 = \sqrt{1+v^2} du + \frac{uv}{\sqrt{1+v^2}} dv, \quad (18)$$

$$\theta_2 = \frac{\sqrt{1+u^2+v^2}}{\sqrt{1+v^2}} dv. \quad (19)$$

Further we have

$$\begin{aligned}dE_1 &= \left(\frac{-v}{(1+v^2)^{\frac{3}{2}}}, 0, \frac{1}{(1+v^2)^{\frac{3}{2}}} \right) dv, \\ \omega_{12} = dE_1 \cdot E_2 &= \frac{u}{(1+v^2)\sqrt{1+u^2+v^2}} dv.\end{aligned}$$

Analogically we have

$$\omega_{13} = dE_1 \cdot E_3 = \frac{1}{\sqrt{1+v^2}\sqrt{1+u^2+v^2}} dv.$$

Further we have

$$\partial_u E_2 = \frac{\sqrt{1+v^2}}{(1+u^2+v^2)^{\frac{3}{2}}} \cdot (-v, -u, 1), \quad (20)$$

$$\partial_v E_2 = \frac{1}{(1+v^2)^{\frac{3}{2}} \cdot (1+u^2+v^2)^{\frac{3}{2}}} (E_{2v}^1, E_{2v}^2, E_{2v}^3), \quad (21)$$

where

$$\begin{aligned}E_{2v}^1 &= -u(1+v^2)(1+u^2) + uv^2(1+u^2+v^2), \\E_{2v}^2 &= u^2v(1+v^2), \\E_{2v}^3 &= -uv[(1+u^2+v^2) + (1+v^2)].\end{aligned}$$

From (20) and (21) follows

$$\partial_u E_2 \cdot E_3 = \frac{\sqrt{1+v^2}}{1+u^2+v^2}, \quad \partial_v E_2 \cdot E_3 = \frac{-uv}{\sqrt{1+v^2}(1+u^2+v^2)},$$

and

$$\omega_{23} = dE_2 \cdot E_3 = \frac{\sqrt{1+v^2}}{1+u^2+v^2} du - \frac{uv}{\sqrt{1+v^2}(1+u^2+v^2)} dv.$$

Summarizing the previous results, we have

$$\begin{aligned}\omega_{12} &= -\omega_{21} = \frac{u}{(1+v^2)\sqrt{1+u^2+v^2}} dv, \\ \omega_{13} &= -\omega_{31} = \frac{1}{\sqrt{1+v^2}\sqrt{1+u^2+v^2}} dv, \\ \omega_{23} &= -\omega_{32} = \frac{\sqrt{1+v^2}}{1+u^2+v^2} du - \frac{uv}{\sqrt{1+v^2}(1+u^2+v^2)} dv, \\ \theta_1 &= \sqrt{1+v^2} du + \frac{uv}{\sqrt{1+v^2}} dv, \\ \theta_2 &= \frac{\sqrt{1+u^2+v^2}}{\sqrt{1+v^2}} dv.\end{aligned}$$

From equations (18) and (19) follows

$$d\theta_1 = 0, \quad d\theta_2 = \frac{u}{\sqrt{1+v^2}\sqrt{1+u^2+v^2}} du \wedge dv$$

and

$$\theta_1 \wedge \theta_2 = \sqrt{1+u^2+v^2} du \wedge dv. \quad (22)$$

From (12) we have

$$d\omega_{12} = \omega_{13} \wedge \omega_{32} = \frac{1}{(1+u^2+v^2)^{\frac{3}{2}}} du \wedge dv.$$

Thanks to (22) we have

$$du \wedge dv = \frac{1}{\sqrt{1+u^2+v^2}} \theta_1 \wedge \theta_2$$

and

$$d\omega_{12} = \frac{1}{(1+u^2+v^2)^2} \theta_1 \wedge \theta_2. \quad (23)$$

From (23) and (17) immediately follows that

$$K = -\frac{1}{(1+u^2+v^2)^2}, \quad (24)$$

which means that every point of studied surface is hyperbolic. The equation (17) gives

$$W(x_u) = -\partial_u E_3 \quad \text{and} \quad W(x_v) = -\partial_v E_3,$$

where

$$\begin{aligned}\partial_u E_3 &= \frac{1}{(1+u^2+v^2)^{\frac{3}{2}}}(uv, -v^2-1, -u), \\ \partial_v E_3 &= \frac{1}{(1+u^2+v^2)^{\frac{3}{2}}}(-1-u^2, uv, -v).\end{aligned}$$

From the fact $W : T_p(M) \rightarrow T_p(M)$ follows

$$\partial_u E_3 = \beta_{11}x_u + \beta_{12}x_v, \quad \partial_v E_3 = \beta_{21}x_u + \beta_{22}x_v. \quad (25)$$

After a short calculation we obtain

$$\begin{aligned} \beta_{11} &= \frac{uv}{(1+u^2+v^2)^{\frac{3}{2}}}, \\ \beta_{12} &= -\frac{1+v^2}{(1+u^2+v^2)^{\frac{3}{2}}}, \\ \beta_{21} &= -\frac{1+u^2}{(1+u^2+v^2)^{\frac{3}{2}}}, \\ \beta_{22} &= \frac{uv}{(1+u^2+v^2)^{\frac{3}{2}}}. \end{aligned}$$

From equations (25) follows that the mapping W can be described by the matrix

$$W = \frac{1}{(1+u^2+v^2)^{\frac{3}{2}}} \begin{pmatrix} -uv & 1+v^2 \\ 1+u^2 & -uv \end{pmatrix}$$

Determinant

$$\det W = K = \frac{1}{(1+u^2+v^2)^3} \det \begin{pmatrix} -uv & 1+v^2 \\ 1+u^2 & -uv \end{pmatrix} = -\frac{1}{(1+u^2+v^2)^2},$$

as was given in 24 and the formula for mean curvature

$$H = \frac{1}{2} \operatorname{tr} W = -\frac{uv}{(1+u^2+v^2)^{\frac{3}{2}}}.$$

4 Example 2

Let $x(u, v) = (u, v, u^2v)$ be a parameterized utility function. Orthogonal frame is

$$\begin{aligned} x_u &= (1, 0, 2uv) \\ x_v &= (0, 1, u^2) \\ n &= (-2uv, -u^2, 1). \end{aligned}$$

Orthonormal frame is

$$\begin{aligned} E_1 &= \frac{1}{\sqrt{1+4u^2v^2}}(1, 0, 2uv), \\ E_2 &= \frac{1}{\sqrt{1+4u^2v^2} \cdot \sqrt{1+4u^2v^2+u^4}}(-2u^3v, 1+4u^2v^2, u^2), \\ E_3 &= \frac{1}{\sqrt{1+4u^2v^2+u^4}}(-2uv, -u^2, 1). \end{aligned} \quad (26)$$

The forms θ_1 and θ_2 have the form

$$\theta_1 = \sqrt{1 + 4u^2v^2}du + \frac{2u^3v}{\sqrt{1 + 4u^2v^2}}dv, \quad \theta_2 = \frac{\sqrt{1 + 4u^2v^2 + u^4}}{\sqrt{1 + 4u^2v^2}}dv. \quad (27)$$

Further we have

$$dE_1 = \partial_u E_1 du + \partial_v E_1 dv = \frac{1}{(1 + 4u^2v^2)^{\frac{3}{2}}} [(-4uv^2, 0, 2v)du + (-4u^2v, 0, 2u)dv].$$

After a short calculation we obtain

$$\omega_{12} = dE_1 \cdot E_2 = \frac{1}{(1 + 4u^2v^2)\sqrt{1 + 4u^2v^2 + u^4}}(2u^2v \, du + 2u^3 \, dv).$$

Analogically

$$\omega_{13} = dE_1 \cdot E_3 = \frac{1}{\sqrt{1 + 4u^2v^2}\sqrt{1 + 4u^2v^2 + u^4}}(2v \, du + 2u \, dv).$$

From (26) follows

$$dE_3 = (\partial_u E_3)du + (\partial_v E_3)dv.$$

After a short calculation we obtain

$$\omega_{32} = dE_3 \cdot E_2 = \frac{1}{\sqrt{1 + 4u^2v^2}(1 + 4u^2v^2 + u^4)}[(-4u^3v^2 - 2u)du + 4u^4v \, dv].$$

Summarizing the previous results we obtain

$$\begin{aligned} \omega_{12} &= -\omega_{21} = \frac{1}{(1 + 4u^2v^2)\sqrt{1 + 4u^2v^2 + u^4}}(2u^2v \, du + 2u^3 \, dv), \\ \omega_{13} &= -\omega_{31} = \frac{1}{\sqrt{1 + 4u^2v^2}\sqrt{1 + 4u^2v^2 + u^4}}(2v \, du + 2u \, dv), \\ \omega_{23} &= -\omega_{32} = \frac{1}{\sqrt{1 + 4u^2v^2}(1 + 4u^2v^2 + u^4)}[(4u^3v^2 + 2u)du - 4u^4v \, dv]. \end{aligned}$$

From equations (16) and (27) we obtain

$$d\omega_{12} = \omega_{13} \wedge \omega_{32} = \frac{4u^2}{(1 + 4u^2v^2 + u^4)^{\frac{3}{2}}}du \wedge dv = \frac{4u^2}{(1 + 4u^2v^2 + u^4)^2}\theta_1 \wedge \theta_2,$$

from which follows: Gaussian curvature has the form

$$K = -\frac{4u^2}{(1 + 4u^2v^2 + u^4)^2}.$$

5 Conclusion

Two economical examples served as an illustration of Maurer-Cartan equations and we reached the following results:

1. The Gaussian and mean curvatures of the first surface are

$$K = -\frac{1}{(1+u^2+v^2)^2}, \quad H = \frac{1}{2}\text{tr}W = -\frac{uv}{(1+u^2+v^2)^{\frac{3}{2}}}.$$

2. The Gaussian curvature of the second surface is

$$K = -\frac{4u^2}{(1+4u^2v^2+u^4)^2}.$$

References

- [1] BURE, J., KAKA, M.: *Some Conditions for a Surface in E^4 to be a Part of the Sphere S^2* . Mathematica Bohemica 1994 No. 4, pp. 367–371.
- [2] CARTAN, .: *Œuvres complètes-Partie I*. Gauthier-Villars, Paris, 1952 Volume 2.
- [3] CARTAN, .: *Œuvres complètes-Partie II*. Gauthier-Villars, Paris, 1953 Volume 1.
- [4] CARTAN, .: *Œuvres complètes-Partie II*. Gauthier-Villars, Paris, 1953 Volume 2.
- [5] CARTAN, .: *Œuvres complètes-Partie III*. Gauthier-Villars, Paris, 1955 Volume 1.
- [6] CARTAN, .: *Œuvres complètes-Partie III*. Gauthier-Villars, Paris, 1955 Volume 2.
- [7] KAŇKA, M.: *An Example of Basic Structure Equations for Riemannian Manifolds*. Mundus Symbolicus Ronk 3 (1995), pp. 57–62.
- [8] STERNBERG, S.: *Lectures on Differential Geometry*. Prentice-Hall, INC. (1964).
- [9] KOBAYASHI, S., NOMIZU, K.: *Foundations of Differential Geometry*. New York (1963).
- [10] NOMIZU, K.: *Lie Groups and Differential Geometry*. The Mathematical Society of Japan (1956).
- [11] SPIVAK, M.: *Differential geometry*. Berkeley (1979).

Current address

Doc. RNDr. Miloš Kaňka, CSc.

Department of Mathematics

University of Economics

W. Churchill Sq. 4

130 67 Prague 3

e-mail: kankavse.cz

THE UTILITY OF THE VIRTUAL REALITY IN DEEPER UNDERSTANDING OF PEOPLE'S EXPERIENCES OF INFANT FEEDING

**PETRASOVA Alena, (UK), CZANNER Silvester, (UK)
CHALMERS Alan, (UK), WOLKE Dieter, (UK)**

Abstract. A principal strength of Virtual Reality (VR) technology offered to the field of health is in the creation of simulated environments in which performance can be tested and trained in a systematic manner. The aims of this paper were: (1) to observe the influence of the virtual baby's behaviour on user's feeding performance in a virtual environment, and (2) to compare the effect of positive and negative feeding situation on user's emotional response. Methods: Two studies were conducted. In the first study, 33 subjects with (n=21) and without (n=12) feeding experience fed the virtual baby with different levels of baby's initial happiness. The second study was designed to observe the effect of the virtual baby's behaviour on subjects' emotions (12 experts and 5 parents. After the feeding task the subjects filled a questionnaire for subjective rating of their experience. Results: The participants displayed increased confidence with their ability to feed the baby and to make it happy. The repeated virtual feeding increased the efficiency of feeding. The results also indicated a positive correlation between changes in the baby's behaviour and the emotional state of the subject. Conclusion: The virtual feeding application has positive training effect while creating positive subjective experience and feelings of accomplishments. There is a great potential for the application to be therapeutically beneficial, stimulating and effective.

Key words. infant feeding difficulties, virtual reality, emotional effect

Mathematics Subject Classification: Primary 68N30; Secondary 62-07, 62K10.

1 Introduction

About 20-25% of parents report some feeding difficulties with their infants in the first two years with the most frequent problem being the refusal of solid foods [1,2]. Mealtimes can become stressful, and if children fail to gain weight adequately, it is common for parents to describe them as picky eaters with poor appetites and they fear that their child is not eating

enough [3]. Feeding problems are often symptoms of difficulties in caregiver - infant relationships [4] and strategies used by caregivers during meals to encourage eating might lead to deterioration of child behaviour [5]. Then it is critical to establish an emotionally healthy reciprocal relationship between parent and child to achieve optimal feeding interactions [6].

The existing treatments usually combine specific psycho-educational intervention with parent training. The goal is to change the parent's perception and interpretation of infant behavior through therapists' explanations, discussion and questionnaires, as well as examination of the nature of the parent-child interaction at mealtimes [4]. While necessary knowledge can be acquired through books and courses, the training of a relaxed parent - child interaction still can be only achieved through direct interaction with the child. However, the behavioral feeding problems that are a result of a history of maladaptive interaction at mealtime make treatment implementation difficult. Then a key approach is in activation of a specific emotional experience, its guiding and processing; [7] as the emotions are states of motivational arousal and they urge individuals to behave in particular way trying to achieve a specific goal or outcome [8].

VR has been shown to be an effective tool in many areas of psychotherapy because it facilitates focused collaboration between the patient and therapist [7]. VR technology has enough capabilities to influence user cognitive operations and thus offer a new approach to therapy [9]. The warm, sensitive and personal social interaction within the virtual environment is considered as one of the important elements of sense of presence [14]. However there has not been any research done on the use of VR for parents with baby feeding difficulties. Therefore in our Study 1 we studied the utility of the VR feeding environment for determining the extent to which the users may be affected by interaction with the virtual infant. In Study 2, the emotional impact of the VR baby's food rejection/acceptance on the user was analysed while taking into account individuals' personal differences.

Our research virtual scenario consists of the baby sitting on the highchair in the dining room, a bowl with food and a spoon. Interaction is done with a Nintendo Wii controller, see Figure1.



Figure 1: An example of a participant using the virtual baby feeding application

The user's task is to assess the situation, respond appropriately to the changing mood of the baby and feed the baby using the provided controller. Five parameters in the application define the baby's attention, tiredness and happiness, how hungry the baby is and if he/she likes food and its texture or not. Depending on changes to the spoon and the position controlled by the user and the current

values of parameters, the system decides on the response of the baby and makes changes to the parameter values [10]. For the purpose of the study three parameters were used as follows: tiredness - the value 1 represents fresh and active to 5 as very tired and sleepy, happiness - from very happy (5) to unhappy and sad (1), hunger - from very hungry (5) to not hungry at all (1).

2 Study 1

2.1 Methods

We aimed to recruit at least 30 participants to have sufficient power in our analyses. We recruited 33 subjects, where 13 were females and 20 males with a mean age of 27.8 (20 to 58 years old). 21 of 33 participants reported that they had already fed a baby, i.e. their own children, a friend's or a sister's baby. 7 participants worked in the University Nursery as child care professionals. Most of the participants had not played games at all or only up to one hour per day on average. 6 participants spent more than an hour a day using computer games.

2.2 Instruments

2.2.1 Questionnaire

A questionnaire based on a heuristic evaluation for playability was used [11]. The items relevant to this study were administered and rated on a scale of 1 to 4. The movements and the behaviour of the virtual baby were rated for their naturalness, like-ability and expectation. The extent of agreement was provided for the meaningful feedback from sound and correlation of the experience with a real life. The feeling of personal involvement, the difficulty level of the task and participant's subjective feelings were described.

2.2.2. Tasks

The task was to feed the virtual baby until it is not hungry (hunger=1). One feed was represented by a mini-game with two possible endings; the baby was either not hungry anymore or still hungry and too tired to be fed. Each participant completed five games with alternating levels of baby's initial level of happiness. 17 participants fed initially happy (happiness=5) in trials 1, 3 and 5, and they fed unhappy baby (happiness=2) in trials 2 and 4. Other 16 participants fed initially unhappy baby in trials 1, 3 and 5, and they fed happy baby in trials 2 and 4. For all participants, the baby was initially hungry (hunger=5) and not tired (tiredness=1). Before the first trial all users were demonstrated with a proper control of the spoon and they were allowed to try controlling it by themselves.

2.3 Data analysis

To assess the utility of this VR environment we studied three hypotheses:

1. Baby that eats (refuses to eat) has positive (negative) emotional effect on user,
2. Feeding performance depends on the baby's initial state of happiness,
3. Repeated feeding with the VR application improves the feeding performance.

Considering the first hypothesis, we counted number of users who felt happy (unhappy) when baby was eating (refused to eat) and calculated the probability of this count being beyond chance assuming Binomial distribution with the probability of chance being 0.5. For the second hypothesis the feeding performance between the two initial states of baby's

happiness (happy and unhappy) was compared via Mann-Whitney [12] test for 2-independent-samples using exact p-values. Considering the training effect, we compared users' performance between the first and the last feeding via non-parametric Wilcoxon test [12] of two related samples across two levels of the initial happiness. All tests were performed at the 0.05 level of significance. Means and standard deviations (SD) of individual parameters were calculated.

2.4 Results

2.4.1 Emotional effect

When the virtual baby ate the food the VR environment had tendency to create feelings of happiness in 20 of 33 participants though there was not enough evidence for significance (p-value=0.15). However, only 8 of 33 participants (p-value=0.99) felt stressed when the baby refused food.

2.4.2 Feeding performance

In the first feeding trial, the initially happy baby was significantly less hungry (p-value<0.001), less tired (p-values=0.0013), more happy (p-value=0.002) and fed slower (p-value=0.026) than the initially unhappy baby. In the third trial the differences in feeding performances across two initial states of baby's happiness were smaller than in first trial. Specifically, in third feeding trial, the originally happy baby is roughly equally happy, less hungry (p-value<0.020), equally tired (p-value=0.445) and fed in roughly same time (p-value=0.064) than the initially unhappy baby. In the fifth trial the differences in feeding performance across two initial states of baby's initial happiness were again smaller than in first trial. The initially happy baby was in average slightly happier (p=0.066), less hungry (p-value<0.022), roughly equally tired (p-value=0.088) and fed in roughly equal time (p-value=0.129) than the initially unhappy baby. We conclude that the initially happy baby leads to significantly better feeding performance of the participants making the happy baby happier, fuller and less tired than the initially unhappy baby. These differences in performance appear to decrease in later trials, probably because users felt overconfident due to ease of feeding of the happier baby.

Initial condition	Feeding trial	Happiness		Hungriness		Tiredness		Time (sec)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Happy baby (17 subjects)	1	3.41	0.8	1.29	0.59	3.88	0.78	164.59	65.48
	3	2.29	1.16	1.88	0.86	4.53	0.62	109.53	33.35
	5	3.06	1.14	1.59	0.62	4.35	0.79	113.29	45.25
Unhappy baby (16 subjects)	1	2.88	0.72	3.75	1.73	4.81	0.4	80.63	71.54
	3	3	0.73	3.38	1.78	4.69	0.6	76.31	55.43
	5	2.94	0.93	2.56	1.36	4.81	0.4	86.81	48.86

Table 1: Mean values and Standard deviations (SD) of the baby's parameters and time of the feeding performance through feeding trials.

2.4.3 Training effect

The first and last feeding of the happy baby were statistically the same with respect to the final baby's hunger (see Table 1, p-value=0.313), the final state of tiredness (p-value=1.79)

and the final state of the happiness (p -value=0.319). However, the total feeding time in last trial is significantly shorter than that of the first one (p -value=0.017). The feeding of the initially unhappy baby also improved over task repetition, but in a different way. Participants filled the unhappy baby's tummy much better in the last trial than in the first trial (p -value=0.05), but did not improve in terms of baby's tiredness, final happiness and time of feeding (p -values=1.00, 0.875 and 0.782).

3 Study 2

3.1 Methods

In second study we again aimed to recruit 30 participants to facilitate statistical comparisons. However only 17 of 30 asked subjects took part in the second study. They were all female aged 25 to 57 (40.65 in average). Eight participants worked as speech and language therapists at the Children's hospital in Birmingham working with children and families to manage feeding difficulties, which might be motor, sensory or behavioural, giving advice to parents concerned with feeding issues like developed aversions to textures and tastes. Two nurses, one speech and language therapist and one dietitian were from Gulson Feeding Clinic in Coventry. The other five participants were parents. Three of them regularly attended a special session in the feeding clinic with one of their children.

3.2 Instruments

3.2.1 Questionnaires

Each participant reported their feelings through the positive and negative affect schedule (PANAS) [13]. They filled the questionnaire before the experiment and after each of two interaction with the virtual baby. After completing the VR tasks, the participants were asked to fill in a questionnaire for subjective rating of an applications utility as a training tool for parents and interactive teaching tool for parents or clinicians considering the current state of the application on a scale of 1 to 4 (strongly disagree=1, strongly agree=4). They also rated the naturalness and like-ability of the movements, the behaviour and sounds of the baby on same scale. Participants who worked with children with eating disorders listed all their duties related to their work. Parents were asked to report their feeding difficulties.

3.2.2 Tasks

There were two VR tasks during which respondents were asked to feed the VR baby during two minutes. One task was a positive feeding trial in which the participants fed the baby that was initially very hungry (hunger=5) and happy (happiness=5) hence it was easy to feed. Other task was a negative feeding trial where subjects were feeding the baby that was initially tired (tiredness=5) and almost not hungry (hunger=2). Nine of the participants started with the positive feeding situation. This group consisted of 7 experts, 2 parents with feeding problems and 1 parent with no feeding problems. Remaining 8 participants started with negative task followed with a positive task. This group included 5 experts, 1 parent with feeding problem and 1 parent with no feeding problem.

3.3 Data analysis

Due to insufficient data for formal hypothesis testing, we did descriptive evaluation of the data. We were interested if the VR baby's food rejection/acceptance have emotional impact

on the user. To remove the intra-personal variability the analysis was done on the changes in mood scores across two PANAS reports: differences between the initial test and the test after the positive feeding trial, differences between the initial test and the test after the negative feeding trial; and differences between the test after positive trial and test after negative trial. Differences in affects may take values from -40 to 40. Positive values indicate that the emotional impact increased, the negative differences mean that the intensity of feelings decreased.

3.4 Results

The parents appear to suggest large inter-subject differences in emotional reactions during feeding trials, see Table 2. The parent without feeding difficulties was affected mainly by the negative feeding situation by decreased intensity of PA (except in alertness). The parents with feeding problems appear to have stronger behavioural reactions (especially more excited) to the negative feeding trials than positive trials. The virtual baby's behaviour did not markedly affect a parent with no feeding difficulties. However, the happy baby positively improved his PA values during the second feeding when compared to negative feeding trials. The emotional effect does not have the same magnitude in all parents. Dismissive reactions of the baby appear to be insufficient to produce any related changes in NA for parents accustomed to negative behaviour.

Subject	Changes in PA			Changes in NA		
	Parents experiencing the positive feeding as the first one					
	after positive situation – initial rating	after negative situation – initial rating	Negative situation – positive situation	after positive situation – initial rating	after negative situation – initial rating	Negative situation – positive situation
Parent WO	2	-28	-30	0	8	8
Parent W	3	7	4	-7	-10	-3
	Parents experiencing the negative feeding as the first one					
	after negative situation – initial rating	after positive situation – initial rating	Positive situation – negative situation	after negative situation – initial rating	after positive situation – initial rating	Positive situation – negative situation
Parent WP 1	-4	4	8	-1	-1	0
Parent WP 2	4	5	1	0	0	0
Parent WOP	2	5	3	-1	-2	-1

Table 2: The values of changes in positive affect (PA) and negative affect (NA) in individual parents with (W) and without (WO) feeding difficulties.

The experts appear to show lower inter-subject variability in emotional changes than parents, see Table 3. Seven (out of 12) experts experienced the negative feeding in second feeding trial which caused a decrease in PA. Almost all experts increased in NA after negative feeding experience. Five experts did negative feeding in first trial and it resulted in a reduction of PA values. In their second trial with positive feeding their PA increased. Usually, a positive feeding experience followed by a negative one led to a reduction or no changes in NA.

The psychologists and clinicians expressed see potential in using the application by parents for training (mean rating 2.75 out of 4) and as an interactive educational tool (mean rating 2.83) as it can increase parental awareness. They see VR's feeding environment as a complementary tool alongside behavioural support especially for families who have entrenched maladaptive feeding problems. They also positively rated the naturalness and pleasantness of the movements and the behaviour of the virtual baby (means 2.67 and 2.83). The sounds were considered as realistic and

likable (means 3.08 and 2.92). The ratings of the VR baby behaviour by the parents with the feeding difficulties was notably higher than other parents and experts' rating.

4 Discussion

Here we presented VR application in a novel area: a virtual baby feeding environment for enhancing the therapy of parents whose infants refuse to eat. Our studies were aimed to investigate a usefulness of VR in helping parents to read signs of refusal, to increase parental awareness and to solidify their understanding of the child's behaviour.

Our studies were designed to be first investigations of VR application for treatment of parents whose infants have feeding difficulties. Therefore our main priority was to recruit relevant subjects: parents, care givers and feeding professionals. The number of participants in our two studies was not always high enough to gain power in statistical tests. Therefore the generalizability of our results to the larger population is partly limited, yet they studies bring several important insights and suggestions for further development.

Subject	Changes in PA			Changes in NA		
	Experts experiencing the positive feeding as the first one					
	after positive situation – initial rating	after negative situation – initial rating	Negative situation – positive situation	after positive situation – initial rating	after negative situation – initial rating	Negative situation positive situation
	1	-5	-6	4	13	9
	2	-3	-5	1	4	3
	3	-13	-6	1	6	5
	4	-10	-5	-3	-1	2
	5	-4	-1	-2	-1	1
	6	-6	-2	12	18	6
7	6	-4	-10	5	10	5
	Experts experiencing the negative feeding as the first one					
	after negative situation – initial rating	after positive situation – initial rating	Positive situation – negative situation	after negative situation – initial rating	after positive situation – initial rating	Positive situation negative situation
	8	-7	3	0	0	0
	9	2	7	4	-1	-5
	10	-7	11	12	0	-12
	11	4	7	0	-3	-3
	12	-6	4	19	6	-13

Table 3: The values of changes in positive affect (PA) and negative affect (NA) for individual experts.

Our studies showed that the virtual baby affected the emotional state and the performance of the user in a way that is consistent with real feeding situations. Furthermore there was a strong training effect in VR. Finally, as we anticipated, our studies suggest a much greater variability in responses of the parents when compared to the feeding experts.

We conclude that our study shows great potential in promoting a rational use of VR technology in Health Care with emphasis on the accessibility of the delivered treatment, and human-computer interaction research associated with this. For the purpose of this research, a simplified model of a child's behaviour was created. Hence, emotions can be recognised as ambiguous and overlapping experiences not as discrete values,¹⁵ they were merged to the parameter, which made the baby act negatively or positively changing emotional arousal. However, for the

future development of the application, it is important to model the baby's mood in a greater detail as it can bring further insight in the baby-parent interaction dynamics and their emotions. The refinement of the baby's emotional model and its extension to the model of affective experiences should be considered. This may allow to study parents' behaviour in special provoked and well controlled situation. Another potential improvement can be in monitoring of users's facial expressions and in using them as an input for VR mode to make infant-parent interactions more realistic.

References

- [1] LINDBERG L, OSTBERG M, ISACSON IM, et al. Feeding disorders related to nutrition. *Acta Paediatrica* 95. 2006; 425–429.
- [2] WOLKE D. Frequent problems in infancy and toddler years: Excessive crying, sleeping and feeding difficulties. In: Bergmann KE, Bergmann RL, eds. *Health promotion and disease prevention in the family*. Berlin: Walter de Gruyter; 2003.
- [3] SATTER E. Feeding dynamics: Helping children to eat well. *J Journal of pediatric health care*. 1995; 178–184.
- [4] CHATOOR I, GANIBAN J. Food refusal by infants and young children: Diagnosis and treatment. *Cognitive and behavioral practice* 10. Elsevier Ltd; 2003; 138–146.
- [5] PIAZZA CC, CARROLL-Hernandez TA. Assessment and treatment of pediatric feeding disorders. *J Encyclopedia on early childhood development*. 2004.
- [6] SCHELL J. *The art of game design*. USA: Elsevier; 2008.
- [7] RIVA G. Virtual reality in psychotherapy: review. *J CyberPsychology&Behavior*. 2005;8(3):220-240.
- [8] BREHM JW. The intensity of emotion. *Personality and social psychology review*. 1999;3(1):2–22.
- [9] RIVA G, BACCHETTA M, BARUFFI M, et al. Experiential cognitive therapy: a VR based approach for the assessment and treatment of eating disorders. In: Riva G, Wiederhold BK, Molinari E, eds. *Virtual environments in clinical psychology and neuroscience*. Ios Press; 1998;120–135.
- [10] PETRASOVA A, CZANNER S, CHALMERS A, et al. *The playability evaluation of virtual baby feeding application*. The 2nd international IEEE conference on Serious games and virtual worlds. Braga, Portugal; 2010.
- [11] DESURVIRE H, CAPLAN M, TOTH JA. Using heuristics to evaluate the playability of games. *Proceedings ACM SIGCHI 2004*. Vienna, Austria; 2004.
- [12] SIEGEL S, Castellan N. *Nonparametric statistics for the behavioural sciences*. McGraw-Hill, 2. edition; 1988.
- [13] CRAWFORD JR, HENRY JD. The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *J British journal of clinical psychology*. 2004;43:245–265.
- [14] McMAHAN A. Immersion, Engagement, and Presence. A method for analyzing 3-D video games. In: Mark JP, Perron B, eds. *The video game, theory reader*. Routledge, NY, USA; 2003;67–86.
- [15] POSNER J, RUSSEL JA, PETERSON BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *J Development and psychopathology*. 2005;3(17):715-734

Current address

Dr. Silvester Czanner

School of Computing, Mathematics and Digital Technology

Manchester Metropolitan University

John Dalton Building

Chester Street

Manchester M1 5GD

tel.: +44 (0) 161 247 1643

e-

CIRCLE-PRESERVING SUBDIVISION SCHEME BASED ON APOLLONIUS' CIRCLE

SLABÁ Kristýna, (CZ), BASTL Bohumír, (CZ)

Abstract. In this paper, we introduce a new Hermite subdivision scheme. The main idea of inserting a new point corresponding to an edge and its associated tangent vector is to intersect suitably chosen Apollonius' circle with an axis of the angle between one of the associated tangent vectors and this edge. The scheme is proved to converge to a continuous curve and, moreover, the limit curve is G^1 continuous. One of the main properties is that it is circle-preserving, i.e., if the initial vertices and their associated tangent vectors are sampled from a circle, then the subdivision process converges to this circle.

Key words and phrases. Subdivision scheme, Hermite subdivision scheme, Apollonius' circle, Circle-preserving subdivision scheme

Mathematics Subject Classification. Primary 65D17, 68U05; Secondary 53A04.

1 Introduction

The beginning of subdivision techniques for curves and surfaces dates back to 1970s, when G. Chaikin (see [1]) firstly defined a recursive process described by simple linear rules for refinement of a given control polygon converging to quadratic uniform B-spline curve. The generalization to other types of spline functions and the thorough analysis followed (see e.g. [10]).

One of the first interpolatory subdivision schemes was presented in [4], where refinement rules are derived from cubic interpolation of four consecutive points. The generalization of this method, the well known 4-point scheme (see [6]), introduces an optional tension parameter which significantly influences the shape of a limit curve. Many other generalizations followed, e.g. modification for non-uniform parameterization of a cubic interpolant of four consecutive points leading to modified refinement rules was presented in [5], circle-preserving variant of the

4-point scheme was described in [9] or family of subdivision schemes with cubic precision based on 4-point scheme was given in [7].

In [8], J. L. Merrien firstly defined the so-called Hermite subdivision scheme where input data to subdivision process consist not only of an initial control polygon but also of associated tangent/normal vectors. These schemes are typically non-linear and it is usually much harder to analyse their convergence or continuity of limit curves. Despite this fact, a lot of Hermite subdivision schemes appeared during the last years. Especially big effort was put into studying interpolatory schemes preserving circles which is considered to be a desirable feature of a subdivision scheme due to technical importance of circles. For example, non-linear subdivision scheme preserving circles is described in [2] which is based on a minimization procedure to find best circle fitting two consecutive points and associated normal vectors. Or, paper [3] presents a method where a new point is inserted as an incenter a suitably chosen triangle.

In this paper, we introduce a new Hermite subdivision scheme. The main idea combines two approaches – generating curves with chord-length parameterizations (which uses Apollonius' circle) and incenter subdivision scheme (see [3]). A new inserted point is obtained as an intersection of a suitably chosen Apollonius' circle and an axis of an edge of the initial polygon and its associated tangent vector. The scheme is proved to converge to a continuous curve and, moreover, the limit curve is G^1 continuous. One of the important properties is that the scheme is circle-preserving.

The rest of the paper is organized as follows. Section 2 describes the subdivision scheme. In Section 3, we give proofs of a convergence of the scheme to a continuous curve and of G^1 continuity of the limit curve. Then, Section 4 shows that the scheme is circle-preserving and we demonstrate the subdivision scheme on several examples. In Section 5, we conclude the paper.

2 Description of subdivision scheme

In this section, we introduce a new subdivision scheme that preserves circles. The main idea of inserting a new point is based on combination of two methods – generating curves with a chord-length parameterization (see [9]) and incenter subdivision scheme (see [3]).

Let \mathbf{v}_i^0 be given points forming an initial polygon and let \mathbf{t}_i^0 , $i = 0, 1, \dots, n-1$, be associated unit tangent vectors (top index is the step of subdivision process). A new inserted point lies on Apollonius' circle, where a centre of the circle is on the line defined by points \mathbf{v}_i^k and \mathbf{v}_{i+1}^k . Then, the new vertex is an intersection of this Apollonius' circle and an axis of vectors $\mathbf{v}_{i+1}^k - \mathbf{v}_i^k$ and \mathbf{p} , where $\mathbf{p} = \mathbf{t}_i^k \vee \mathbf{t}_{i+1}^k$ (a choice of \mathbf{p} depends on position of the centre of Apollonius' circle).

The choice of unit tangent vectors is not completely arbitrary. Each tangent vector \mathbf{t}_i^0 has to be between vectors $\mathbf{v}_{i+1}^0 - \mathbf{v}_i^0$ and $\bar{\mathbf{v}}_{i-1}^0 - \mathbf{v}_i^0$, where $\bar{\mathbf{v}}_{i-1}^0 = \mathbf{v}_i^0 + (\mathbf{v}_i^0 - \mathbf{v}_{i-1}^0)$ (see Fig. 2). If tangent vectors are not given, we can always choose suitable ones by the following two methods:

- Computation of tangent vectors can be based on circles passing through three adjacent points.
- Tangent vector can be computed as

$$\mathbf{t}_i^0 = \frac{\mathbf{v}_{i+1}^0 - \mathbf{v}_{i-1}^0}{\|\mathbf{v}_{i+1}^0 - \mathbf{v}_{i-1}^0\|}. \quad (1)$$

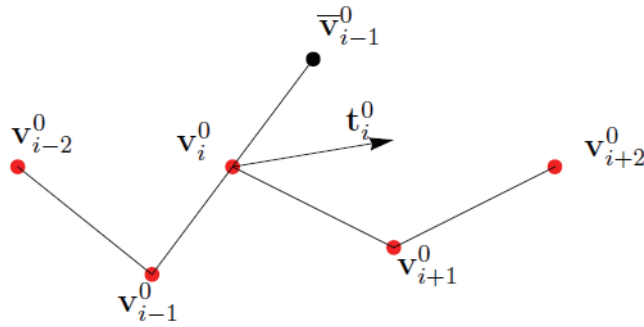


Figure 1: Tangent vector \mathbf{t}_i^0 and initial polygon (red).

We denote the following two oriented angles

$$\alpha_i^k = \angle(\mathbf{t}_i^k, \mathbf{v}_{i+1}^k - \mathbf{v}_i^k), \quad \beta_i^k = \angle(\mathbf{v}_{i+1}^k - \mathbf{v}_i^k, \mathbf{t}_{i+1}^k),$$

see Fig. 2. Moreover, we require $|\alpha_i^0|, |\beta_i^0| < \frac{2}{3}\pi$ and $|\alpha_i^0| + |\beta_i^0| < \pi$, $\forall i$, that can be always ensured with an application of the incenter scheme in the first step of subdivision process (see [3]).

Further, we need to describe a classification of edges that form an initial control polygon. Edges are distinguished into three categories:

- The edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$ is *convex*, if $\alpha_i^k \beta_i^k > 0$ (Fig. 2, left).
- The edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$ is *inflexion*, if $\alpha_i^k \beta_i^k < 0$ (Fig. 2, middle).
- The edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$ is *straight*, if $\alpha_i^k \beta_i^k = 0$ (Fig. 2, right).

If there exists an inflexion edge in the initial control polygon, it is necessary to replace it by two convex edges. We insert a new point $\mathbf{v}_{i+\frac{1}{2}}^0$ between points \mathbf{v}_i^0 and \mathbf{v}_{i+1}^0 , where

$$\mathbf{v}_{i+\frac{1}{2}}^0 = (1 - \omega)\mathbf{v}_i^0 + \omega\mathbf{v}_{i+1}^0 \quad (2)$$

and parameter $\omega \in (0, 1)$ (in the following we choose $\omega = \frac{1}{2}$). The tangent vector $\mathbf{t}_{i+\frac{1}{2}}^0$ in the point $\mathbf{v}_{i+\frac{1}{2}}^0$ is

$$\mathbf{t}_{i+\frac{1}{2}}^0 = \mathbf{R} \left(\frac{\alpha_i^0}{|\alpha_i^0|} \min\{|\alpha_i^0|, |\beta_i^0|\} \right) \frac{\mathbf{v}_{i+1}^0 - \mathbf{v}_i^0}{\|\mathbf{v}_{i+1}^0 - \mathbf{v}_i^0\|}, \quad (3)$$

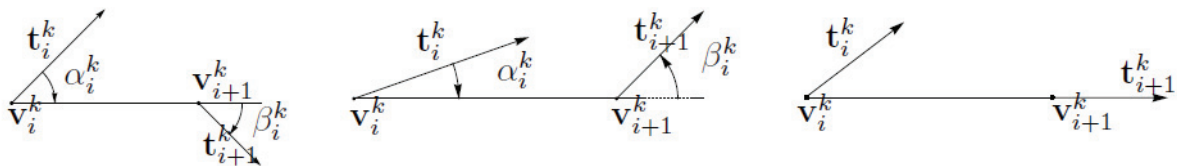


Figure 2: Angles α_i^k and β_i^k . Left: convex edge; Middle: inflexion edge; Right: straight edge.

where $\mathbf{R}(\xi) = \begin{pmatrix} \cos \xi & -\sin \xi \\ \sin \xi & \cos \xi \end{pmatrix}$ is rotation matrix (see Fig. 2). If all inflexion edges are removed before subdivision process starts, then no other inflexion edges can appear during the subdivision process. In the following paragraphs the rules for inserting a new vertex \mathbf{v}_{2i+1}^{k+1} and

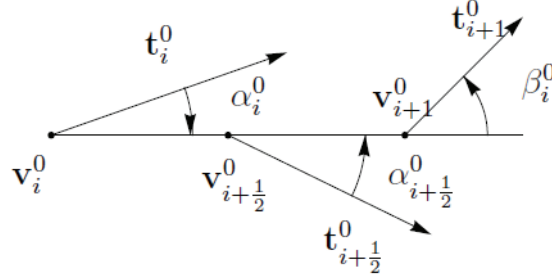


Figure 3: The point $\mathbf{v}_{i+\frac{1}{2}}^0$ and tangent vector $\mathbf{t}_{i+\frac{1}{2}}^0$.

a new tangent vector \mathbf{t}_{2i+1}^{k+1} are presented. We denote

- \mathbf{W} as an intersection of lines $m : \mathbf{x} = \mathbf{v}_i^k + r\mathbf{t}_i^k$ and $n : \mathbf{x} = \mathbf{v}_{i+1}^k + s\mathbf{t}_{i+1}^k$, where $r, s \in \mathbb{R}$,
- $a = \|\mathbf{v}_i^k - \mathbf{W}\|$, $b = \|\mathbf{v}_{i+1}^k - \mathbf{W}\|$.

Now, we need to find the centre \mathbf{S} of Apollonius' circle k . If $a = b$, the circle degenerates to the axis of the edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$. Otherwise, the circle passes through points \mathbf{W} and \mathbf{P} , where $\mathbf{P} = \mathbf{v}_i^k + \frac{a}{a+b}(\mathbf{v}_{i+1}^k - \mathbf{v}_i^k)$. Thus, we obtain \mathbf{S} as

$$\mathbf{S} \in o \cap p, \quad (4)$$

where

$$\begin{aligned} o : \mathbf{x} &= \frac{1}{2}(\mathbf{W} + \mathbf{P}) + g(\mathbf{W} - \mathbf{P})^\perp, \\ p : \mathbf{x} &= \mathbf{v}_i^k + h(\mathbf{v}_{i+1}^k - \mathbf{v}_i^k), \end{aligned} \quad (5)$$

and $(x, y)^\perp = (-y, x)$. Consequently, we solve the system of two linear equations for variables g and h . Finally, the new inserted point depends on value of parameter h . For

- $h < 0$

$$\mathbf{v}_{2i+1}^{k+1} \in k \cap o_1, \quad (6)$$

where $o_1 : \mathbf{x} = \mathbf{v}_i^k + s \left(\mathbf{t}_i^k + \frac{\mathbf{v}_{i+1}^k - \mathbf{v}_i^k}{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\|} \right)$, $s \in \mathbb{R}$. Thus,

$$\mathbf{v}_{2i+1}^{k+1} = \mathbf{v}_i^k + \mathbf{R} \left(-\frac{\alpha_i^k}{2} \right) (\mathbf{P} - \mathbf{v}_i^k). \quad (7)$$

- $h > 1$

$$\mathbf{v}_{2i+1}^{k+1} \in k \cap o_2, \quad (8)$$

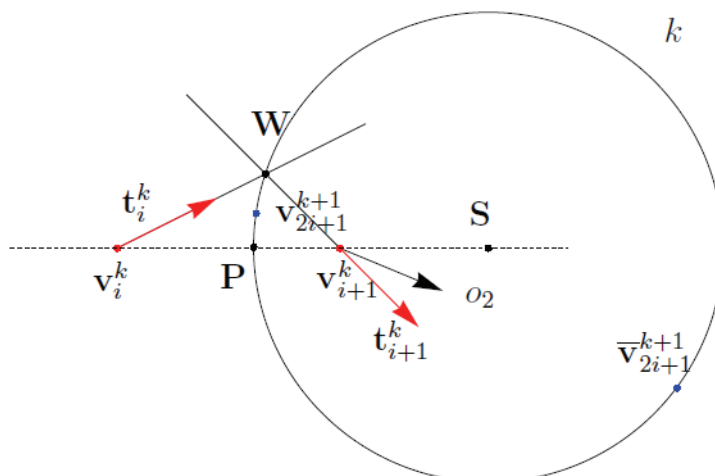


Figure 4: Inserting the new vertex \mathbf{v}_{2i+1}^{k+1} and the tangent vector \mathbf{t}_{2i+1}^{k+1} .

where $o_2 : \mathbf{x} = \mathbf{v}_{i+1}^k + r \left(\mathbf{t}_{i+1}^k + \frac{\mathbf{v}_{i+1}^k - \mathbf{v}_i^k}{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\|} \right)$, $r \in \mathbb{R}$. Thus,

$$\mathbf{v}_{2i+1}^{k+1} = \mathbf{v}_{i+1}^k + \mathbf{R} \left(\frac{\beta_i^k}{2} \right) (\mathbf{P} - \mathbf{v}_{i+1}^k). \quad (9)$$

For $h = 0$ or $h = 1$, Apollonius' circle degenerates to the point \mathbf{v}_i^k or \mathbf{v}_{i+1}^k , respectively, which would coincide also with the new vertex \mathbf{v}_{2i+1}^{k+1} . Thus, there would be no refinement of control polygon. Further, it cannot happen that $h \in (0, 1)$, because the centre of Apollonius' circle cannot lie inside the edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$. If the Apollonius' circle degenerates to the axis, a new vertex is an intersection of o_1 and k . Since there are two intersections \mathbf{v}_{2i+1}^{k+1} , $\bar{\mathbf{v}}_{2i+1}^{k+1}$ of the axis of relevant vectors and of Apollonius' circle, we have to add a condition that the new vertex \mathbf{v}_{2i+1}^{k+1} lies inside the triangle $\triangle \mathbf{v}_i^k \mathbf{v}_{i+1}^k \mathbf{W}$. New tangent vector is defined as

$$\mathbf{t}_{2i+1}^{k+1} = \frac{\mathbf{v}_{i+1}^k - \mathbf{v}_i^k}{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\|}. \quad (10)$$

Then we re-mark vertices and vectors from the previous step k , i.e.,

$$\begin{aligned}\mathbf{v}_{2i}^{k+1} &= \mathbf{v}_i^k, \\ \mathbf{t}_{2i}^{k+1} &= \mathbf{t}_i^k.\end{aligned}\tag{11}$$

Inserting new vertices and vectors is shown in Fig. 2.

3 Analysis of subdivision scheme

In this section, we focus on an analysis of the subdivision scheme described in Section 2. We need to show that our subdivision scheme converges to a continuous curve and determine the continuity of limit curve.

Theorem 3.1 Let $\theta_i^k = \max\{|\alpha_i^k|, |\beta_i^k|\}$ and $\theta_k = \max_i\{\theta_i^k\}$. Then $\lim_{k \rightarrow \infty} \theta_k = 0$.

Proof. Because we insert new point inside the triangle $\Delta \mathbf{v}_i^k \mathbf{v}_{i+1}^k \mathbf{W}$, it holds that

$$\theta_i^k = \xi_i \theta_i^{k-1}, \quad (12)$$

where $\xi_i \in \langle \frac{1}{2}, 1 \rangle$. To be more specific, for $h < 0$ (h determines the position of the centre of Apollonius' circle)

$$|\alpha_{2i}^{k+1}| = \frac{|\alpha_i^k|}{2}, \quad |\beta_{2i}^{k+1}| = \frac{|\alpha_i^k|}{2}, \quad |\alpha_{2i+1}^{k+1}| = (1 - \delta_i)|\beta_i^k|, \quad |\beta_{2i+1}^{k+1}| = \delta_i|\beta_i^k|, \quad \delta_i \in (0, 1)$$

where

$$\begin{aligned} \alpha_{2i}^{k+1} &= \angle(\mathbf{t}_i^k, \mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k), & \beta_{2i}^{k+1} &= \angle(\mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k, \mathbf{t}_{2i+1}^{k+1}), \\ \alpha_{2i+1}^{k+1} &= \angle(\mathbf{t}_{2i+1}^{k+1}, \mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1}), & \beta_{2i+1}^{k+1} &= \angle(\mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1}, \mathbf{t}_{i+1}^k). \end{aligned}$$

Analogously for $h < 0$. If the Apollonius' circle degenerates to axis of edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$, it holds

$$\begin{aligned} |\alpha_{2i}^{k+1}| &= \frac{|\alpha_i^k|}{2}, & |\beta_{2i}^{k+1}| &= \frac{|\alpha_i^k|}{2}, \\ |\alpha_{2i+1}^{k+1}| &= \frac{|\beta_i^k|}{2}, & |\beta_{2i+1}^{k+1}| &= \frac{|\beta_i^k|}{2}. \end{aligned}$$

If we denote $\xi_i = \max\{\delta_i, 1 - \delta_i, \frac{1}{2}\}$, then (12) follows. Let $\xi = \max_i\{\xi_i\}$. Then

$$\theta_k \leq (\xi)^k \theta_0 \quad (13)$$

is geometric sequence with common ratio $q = \xi < 1$ and thus $\lim_{k \rightarrow \infty} \theta_k = 0$. \square

Theorem 3.2 For subdivision scheme described in Section 2, the sequence of polygons converges to a continuous curve.

Proof. We denote Γ_k polygon obtained in the step k of subdivision process. To prove that the scheme converges to a continuous curve, we need to compute an estimation of the distance d_k between Γ_{k+1} and Γ_k . Theorem 3.1 implies that there exists k_0 such that $\theta_k < \frac{\pi}{2}$, if $k > k_0$. Now, for $k > k_0$ and for $h < 0$ with the help of the sine formula we derive the following expression

$$\|\mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k\| = \frac{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\| \sin(|\bar{\beta}_{2i+1}^{k+1}|)}{\sin(|\alpha_{2i}^{k+1}| + |\bar{\beta}_{2i+1}^{k+1}|)} \leq \|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\|, \quad (14)$$

where $\alpha_{2i}^{k+1} = \angle(\mathbf{t}_i^k, \mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k)$ a $\bar{\beta}_{2i+1}^{k+1} = \angle(\mathbf{v}_{i+1}^k - \mathbf{v}_i^k, \mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1})$ and

$$|\alpha_{2i}^{k+1}| = \frac{|\alpha_i^k|}{2}. \quad (15)$$

Analogously, we obtain that $\|\mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1}\| \leq \|\mathbf{v}_i^k - \mathbf{v}_{i+1}^k\|$. Similar procedure can be used for $h > 1$. If Apollonius' circle degenerates to the axis of the edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$, then

$$\|\mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k\| = \frac{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\| \sin\left(\frac{|\beta_i^k|}{2}\right)}{\sin\left(\frac{|\alpha_i^k| + |\beta_i^k|}{2}\right)} \leq \|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\| \quad (16)$$

and analogously $\|\mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1}\| \leq \|\mathbf{v}_i^k - \mathbf{v}_{i+1}^k\|$. Further, we denote

$$L \equiv \max_i \{\|\mathbf{v}_{i+1}^{k_0} - \mathbf{v}_i^{k_0}\|\} \geq \dots \geq \max_i \{\|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\|\} \geq \max_i \{\|\mathbf{v}_{i+1}^{k+1} - \mathbf{v}_i^{k+1}\|\}. \quad (17)$$

Now, the distance between the new vertex \mathbf{v}_{2i+1}^{k+1} and the edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$ for $h < 0$ is

$$d_i^k = \|\mathbf{v}_{2i+1}^{k+1} - \mathbf{v}_i^k\| \sin \left(\frac{|\alpha_i^k|}{2} \right) < \|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\| \frac{|\alpha_i^k|}{2} \leq \frac{L\theta_k}{2} \quad (18)$$

and for $h > 1$

$$d_i^k = \|\mathbf{v}_{i+1}^k - \mathbf{v}_{2i+1}^{k+1}\| \sin \left(\frac{|\beta_i^k|}{2} \right) < \|\mathbf{v}_{i+1}^k - \mathbf{v}_i^k\| \frac{|\beta_i^k|}{2} \leq \frac{L\theta_k}{2}. \quad (19)$$

If the Apollonius' circle degenerates to the axis of the edge $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$, both equations (18) and (19) hold. If we denote $d_k = \max_i \{d_i^k\}$ and due to (13), we arrive at

$$d_k \leq \frac{L\theta_k}{2} \leq \frac{L}{2} \xi^{k-k_0} \theta_{k_0}. \quad (20)$$

This implies that the sequence of polygons $\{\Gamma_k\}$ forms a Cauchy sequence and it converges uniformly. Because each polygon is a piecewise linear curve, the limit curve is continuous. \square

Theorem 3.3 *Limit curve obtained by subdivision scheme described in Section 2 is G^1 continuous.*

Proof. Let $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be a mapping which maps any unit tangent vector \mathbf{t}_i^k to the point $\varphi(\mathbf{t}_i^k) \in \mathbb{S}^1$ such that $\varphi(\mathbf{t}_i^k) - \mathbf{0}$ is parallel to \mathbf{t}_i^k . Let ψ_k be a polygon determined by $\{\varphi(\mathbf{t}_i^k)\}_i$ in the step k of the subdivision process.

Now, we map vector \mathbf{t}_{2i+1}^{k+1} and we obtain the triangle $\triangle \varphi(\mathbf{t}_{2i+1}^{k+1}) \varphi(\mathbf{t}_i^k) \varphi(\mathbf{t}_{i+1}^k)$ (see Fig. 3). It is obvious that

$$\begin{aligned} \|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_i^k)\| &\leq \alpha_i^k, \\ \|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_{i+1}^k)\| &\leq \beta_i^k, \end{aligned} \quad (21)$$

because $\|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_i^k)\|$ is less than the length of the arc that passes through points $\varphi(\mathbf{t}_{2i+1}^{k+1})$ and $\varphi(\mathbf{t}_i^k)$. The new vector \mathbf{t}_{2i+1}^{k+1} is parallel to $\mathbf{v}_{i+1}^k - \mathbf{v}_i^k$ and thus the length of the arc that passes through points $\varphi(\mathbf{t}_{2i+1}^{k+1})$ and $\varphi(\mathbf{t}_i^k)$ is equal to α_i^k (analogously for $\|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_{i+1}^k)\|$).

We denote h_i^k the distance between $\varphi(\mathbf{t}_{2i+1}^{k+1})$ and the edge $\varphi(\mathbf{t}_i^k) \varphi(\mathbf{t}_{i+1}^k)$ in the triangle $\triangle \varphi(\mathbf{t}_{2i+1}^{k+1}) \varphi(\mathbf{t}_i^k) \varphi(\mathbf{t}_{i+1}^k)$. Now, we can estimate the difference $\|\psi_{k+1} - \psi_k\|$, because it holds that

$$\begin{aligned} \|\psi_{k+1} - \psi_k\| &\leq \max_i \{h_i^k\} \leq \max_i \{\|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_i^k)\|, \|\varphi(\mathbf{t}_{2i+1}^{k+1}) - \varphi(\mathbf{t}_{i+1}^k)\|\} \leq \\ &\leq \max_i \{\alpha_i^k, \beta_i^k\} = \theta_k \leq (\xi)^k \theta_0, \end{aligned}$$

where the last inequality is proved in Theorem 3.1. The final estimation $\|\psi_{k+1} - \psi_k\|$ implies that $\{\psi_k\}$ is a Cauchy sequence, which converges uniformly. As ψ_k is a continuous curve, the limit curve is also continuous. Due to the construction of ψ_k the limit curve of the subdivision process is G^1 continuous. \square

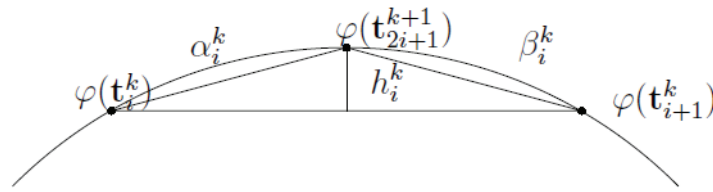


Figure 5: The triangle $\triangle \varphi(\mathbf{t}_{2i+1}^{k+1})\varphi(\mathbf{t}_i^k)\varphi(\mathbf{t}_{i+1}^k)$.

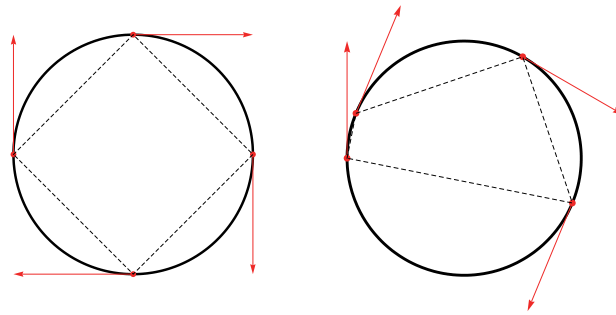


Figure 6: Left: Equidistantly distributed data; Right: Randomly distributed data. Original points and vectors (red), initial polygon (dashed).

4 Properties and examples

In this section, we prove that the proposed subdivision scheme fulfills one of the main desirable properties of subdivision schemes – it preserves circles, i.e., if initial vertices and associated tangent vectors are sampled from a circle k , then the limit curve obtained by our subdivision scheme is the circle k . At the end, we show several examples of an application of our subdivision scheme.

Theorem 4.1 *Subdivision scheme described in Section 2 preserves circles.*

Proof. Let us assume that data $\mathbf{v}_i^k, \mathbf{v}_{i+1}^k$ and $\mathbf{t}_i^k, \mathbf{t}_{i+1}^k$ in the step k are data from a circle k_0 . It is enough to show that the new vertex \mathbf{v}_{2i+1}^{k+1} lies on the circle k_0 and the tangent vector \mathbf{t}_{2i+1}^{k+1}

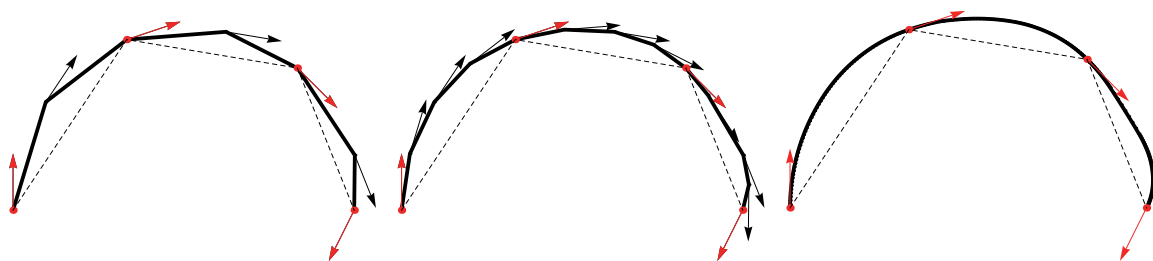


Figure 7: Left: The first step of subdivision; Middle: The second step of subdivision; Right: Limit curve. Initial open polygon (dashed), initial data (red).

is the tangent vector of the circle k_0 in \mathbf{v}_{2i+1}^{k+1} , for $k \in \mathbb{N} \cup 0$. Thus, the triangle $\triangle \mathbf{v}_i^k \mathbf{v}_{i+1}^k \mathbf{W}$ is isosceles triangle with the base-line $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$. Further, $a = b$ and Apollonius' circle degenerates to the axis of the line segment $\mathbf{v}_i^k \mathbf{v}_{i+1}^k$ and, moreover, axis of the angle at the vertex \mathbf{W} . The new vertex \mathbf{v}_{2i+1}^{k+1} is an intersection point of the axis o_1 and k , i.e., the new vertex is the incenter of triangle $\triangle \mathbf{v}_i^k \mathbf{v}_{i+1}^k \mathbf{W}$. Using the Central angle theorem it follows that \mathbf{v}_{2i+1}^{k+1} lies on k_0 and that the tangent vector \mathbf{t}_{2i+1}^{k+1} is the tangent vector of k_0 in \mathbf{v}_{2i+1}^{k+1} . Fig. 6 shows an application of the scheme on data which are sampled from a circle. \square .

Finally, we show examples of using the subdivision scheme to open (see Fig. 7) and closed (see Fig. 8) polygon. Fig. 9 shows a comparison of the subdivision scheme described in Section 2 with incenter subdivision scheme (see [3]).

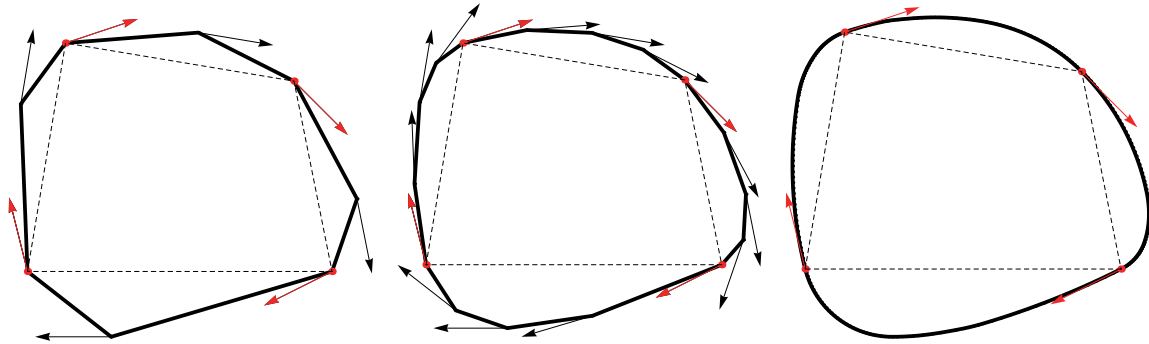


Figure 8: Left: the first step of subdivision, middle: the second step of subdivision, right: limit curve. Initial closed polygon (dashed), initial data (red).

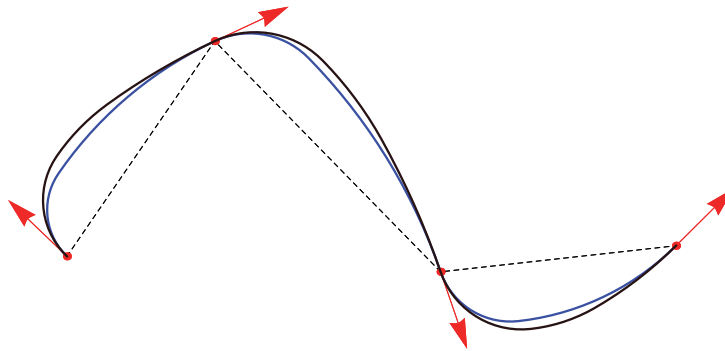


Figure 9: Comparison of incenter subdivision scheme (blue) with the subdivision scheme described in Section 2 (black).

5 Conclusion

We presented a new Hermite subdivision scheme which fulfils one of the most desirable property – it preserves circle. The new point associated to an edge is inserted as an intersection of

suitably chosen Apollonius' circle and a line determined by one of endpoints of this edge and its associated tangent vector. We proved that the scheme converges to a continuous curve and that the limit curve is G^1 continuous. The functionality of the scheme was demonstrated on several examples. In the future work, we want to modify the scheme such that the limit curve is also G^2 continuous.

Acknowledgement

The second author was supported by the Research Plan MSM 4977751301.

References

- [1] CHAIKIN, G.: *An algorithm for high speed curve generation*. Computer Graphics and Image Processing, Vol. 3, pp. 346-349, 1974.
- [2] CHALMOVIANSKÝ, P., JÜTTLER, B.: *A non-linear circle-preserving subdivision scheme*. Advances in Computational Mathematics, pp. 375-400, 2006.
- [3] DENG, Ch., WANG, G.: *Incenter subdivision scheme for curve interpolation*. Computer Aided Geometric Design, Vol. 27, pp. 48-59, 2010.
- [4] DUBUC, S.: *Interpolation through an iterative scheme*. Journal of Mathematical Analysis and Applications, Vol. 114, pp. 185-204, 1986.
- [5] DYN, N., FLOATER, M. S., HORMANN, K.: *Four-point curve subdivision based on chordal and centripetal parametrizations*. Computer Aided Geometric Design, Vol. 26, pp. 279-286, 2009.
- [6] DYN, N., LEVIN, D., GREGORY, J. A.: *A 4-point interpolary scheme for curve design*. Computer Aided Geometric Design, Vol. 4, pp. 257-268, 1987.
- [7] HORMANN, K., SABIN, M. A.: *A family of subdivision schemes with cubic precision*. Computer Aided Geometric Design, Vol. 25, pp. 41-52, 2008.
- [8] MERRIEN, J. L.: *A family of Hermite interpolants by bisection algorithms*. Numerical Algorithms 2, pp. 187-200, 1992.
- [9] SABIN, M. A., DODGSON N. E.: *A circle preserving variant of the four-point subdivision scheme*. In: DÆHLEN, M; MØRKEN, K.; SCHUMAKER, L. (Eds.) Mathematical Methods for Curve and Surfaces. Tromsø, 2004.
- [10] WARREN, J.: *Subdivision Methods For Geometric Design: A Constructive Approach*. Morgan Kaufmann, 2003.

Current address

Slabá Kristýna, Ing.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic, tel. ++420 377 632670, e-mail: kslaba@kma.zcu.cz

Bastl Bohumír, Ing. Ph.D.

Dept. of Mathematics, Univ. of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic, tel. ++420 377 632655, e-mail: bastl@kma.zcu.cz

PARALLELOGRAM SPACES AND MEDIAL QUASIGROUPS

VANŽUROVÁ Alena, (CZ), JUKL Marek (CZ)

Abstract. The concept of an affine space and affine transformation are well known. Our aim is to show that a more general class of the so-called parallelogram spaces can be viewed as a generalization of affine spaces. Natural examples arise from medial quasigroups. The apparatus of parallelograms also enables us to prove a version of Toyoda's Theorem that relates a given medial quasigroup to a particular commutative group.

Key words and phrases. Parallelogram space, affine space, parallelogram, vector, medial quasigroup.

Mathematics Subject Classification. Primary 20N05, 05B15; Secondary 12E20.

1 Parallelogram spaces

So-called parallelogram spaces can be introduced in several ways which are more or less equivalent. We present here some of these possibilities and show advantages of the particular view-points.

1.1 The concept of parallelogram space

Let us start with a view-point that reminds the Weyl approach to analytic geometry.

Definition 1.1 A *parallelogram space* is a pair of non-empty sets \mathcal{P} , \mathcal{V} endowed with a mapping $*$: $\mathcal{P} \times \mathcal{V} \rightarrow \mathcal{P}$, $(p, v) \mapsto p * v$ such that

- (i) for any $P, Q \in \mathcal{P}$ there exists exactly one $v \in \mathcal{V}$ such that $P * v = Q$
(*transitivity* of the star action);
- (ii) $(P * v) * w = (P * w) * v$ for all $P \in \mathcal{P}$ and $v, w \in \mathcal{V}$
(*parallelogram behavior* of the star action).

Denote the parallelogram space as a triplet $(\mathcal{P}, \mathcal{V}, *)$ and use geometrical terminology in what follows: elements of \mathcal{P} are called *points*, elements of \mathcal{V} are *vectors*, though we do not suppose any linear operations introduced in \mathcal{V} . Let us introduce a mapping $\vec{\cdot}: \mathcal{P} \times \mathcal{P} \rightarrow \mathcal{V}$, $(P, Q) \rightarrow \overrightarrow{PQ}$ if and only if $P * \overrightarrow{PQ} = Q$. The mapping is well-defined by (i), satisfies

$$(iii) \text{ for any } P, Q \in \mathcal{P}, P * \overrightarrow{PQ} = Q,$$

and can be called a *translation* on \mathcal{P} determined by (P, Q) . Hence we may consider a parallelogram space as a quadruple $(\mathcal{P}, \mathcal{V}, *, \vec{\cdot})$ consisting of nonempty sets \mathcal{P}, \mathcal{V} and mappings $*$: $\mathcal{P} \times \mathcal{V} \rightarrow \mathcal{P}$, $\vec{\cdot}$: $\mathcal{P} \times \mathcal{P} \rightarrow \mathcal{V}$ satisfying (ii) and (iii).

Equivalently, we may view a parallelogram space also as a triplet $(\mathcal{P}, \mathcal{V}, \vec{\cdot})$ where $\mathcal{P}, \mathcal{V} \neq \emptyset$, $\vec{\cdot}$: $\mathcal{P} \times \mathcal{P} \rightarrow \mathcal{V}$ and the following holds

$$(i') \text{ for any } P \in \mathcal{P}, v \in \mathcal{V} \text{ there is just one } Q \in \mathcal{P} \text{ such that } \overrightarrow{PQ} = v.$$

$$(ii') \text{ for any } P, Q, R, S \in \mathcal{P}, \overrightarrow{PQ} = \overrightarrow{RS} \iff \overrightarrow{PR} = \overrightarrow{QS}.$$

To check equivalence of this new definition to the previous one, let us introduce a mapping $*$: $\mathcal{P} \times \mathcal{V} \rightarrow \mathcal{P}$, $(P, v) \mapsto Q = P * v$ if and only if $v = \overrightarrow{PQ}$. According to (i'), the mapping is well-defined, satisfies (iii), and (ii') turns out into (ii) if we put $\overrightarrow{PQ} = v$, $\overrightarrow{PR} = w$. On the other hand, (ii) becomes (ii') if we take $P * v = Q$, $P * w = R$.

1.2 Structural properties

In what follows, we show that the star operation of the parallelogram space induces a binary operation on \mathcal{V} , written additively here, which turns \mathcal{V} into a commutative group.

Theorem 1.2 *For a parallelogram space $(\mathcal{P}, \mathcal{V}, *)$, there is exactly one map $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ such that for any $P \in \mathcal{P}$ and $v, w \in \mathcal{V}$,*

$$P * (v + w) = (P * v) * w. \quad (1)$$

*On the other hand, the operation of a parallelogram space $(\mathcal{P}, \mathcal{V}, *)$ and of a groupoid $(\mathcal{V}, +)$ are related by (1) if and only if there is an identity element 0 in \mathcal{V} such that $(\mathcal{V}, +; 0)$ is a commutative group.*

Proof. Let $P, Q \in \mathcal{P}$ and $v, w \in \mathcal{V}$. Further let $\bar{v}, \bar{w} \in \mathcal{V}$ be such that $(P * v) * w = P * \bar{v}$, $P * \bar{w} = Q$. Using (ii) we get $(Q * v) * w = ((P * \bar{w}) * v) * w = ((P * v) * w) * \bar{w} = (P * \bar{v}) * \bar{w} = (P * \bar{w}) * \bar{v} = Q * \bar{v}$ which shows that for given v and w , the vector \bar{v} satisfying $(P * v) * w = P * \bar{v}$ is independent of the choice of the point P . Hence a binary operation $+$ on \mathcal{V} , $(v, w) \mapsto \bar{v}$ is by (1) well-defined. If we apply (i') and (ii'), the equality (1) reads

$$\overrightarrow{PQ} + \overrightarrow{QS} = \overrightarrow{PS}. \quad (2)$$

Let $(\mathcal{P}, \mathcal{V}, *)$ be a parallelogram space and $+$ the addition from (1). To deduce commutativity of $+$ we use (1), (i) and (ii), and for associativity of $+$ we apply (1) repeatedly.

Now let $P \in \mathcal{P}$. Denote by 0 the vector uniquely determined by the equality $P * 0 = P$ according to (i). Let us check that, independently of the choice of P , the vector 0 plays the role

of the identity element for the operation $+$. Indeed, by (1), we have $P*v = (P*0)*v = P*(0+v)$ for all $v \in \mathcal{V}$. From (i) it follows $v = 0 + v$, so that 0 the identity element for $+$.

By (i), any $v \in \mathcal{V}$ uniquely determines $\bar{v} \in \mathcal{V}$ that satisfies $(P*v)*\bar{v} = P$. But by (1), $P*(v+\bar{v}) = P = P*0$, hence $v+\bar{v} = 0$, again by (i). Thus \bar{v} is the inverse (opposite vector) of v with respect to $+$, and $(\mathcal{V}, +; 0)$ is a commutative group.

On the other hand, if we start now from a commutative group $(\mathcal{V}, +; 0)$, the triple $(\mathcal{V}, \mathcal{V}, +)$ satisfies the conditions of a parallelogram space. Indeed, (i) holds since the equations of the form $v+x=w$ for $v, w \in \mathcal{V}$ are uniquely solvable in $x \in \mathcal{V}$. From associativity and commutativity of $+$ we can prove (ii), and (1) is guaranteed by associativity.

According to the second part of the above proof, the following holds:

Corollary 1.3 *For all $P \in \mathcal{P}$, the relations $P*0 = P$ and $\overrightarrow{P\bar{P}} = 0$ are satisfied.*

1.3 Morphisms

A *homomorphism* of a parallelogram space $(\mathcal{P}, \mathcal{V}, *)$ to a parallelogram space $(\mathcal{P}', \mathcal{V}', *)'$ is a pair of mappings $\alpha: \mathcal{P} \rightarrow \mathcal{P}'$, $\beta: \mathcal{V} \rightarrow \mathcal{V}'$ such that

$$(iv) \quad \alpha(P) *' \beta(v) = \alpha(P*v) \text{ for all } P \in \mathcal{P}, v \in \mathcal{V}.$$

When α and β are bijections we get an *isomorphism*.

A parallelogram space $(\mathcal{P}, \mathcal{V}, *)$ with a distinguished point $O \in \mathcal{P}$ is here called *pointed*, and O is said to be its *origin*, or *reference point*.

Let us show that parallelogram spaces of the form $(\mathcal{V}, \mathcal{V}, +)$ described in the Theorem 1.2 are, up to isomorphism, typical parallelogram spaces.

Theorem 1.4 *If a pointed parallelogram space $(\mathcal{P}, \mathcal{V}, *)$ with a reference point $O \in \mathcal{P}$ is given, then the pair of mappings $\mathcal{P} \rightarrow \mathcal{V}$, $P \mapsto \overrightarrow{OP}$; $id_{\mathcal{V}}: \mathcal{V} \rightarrow \mathcal{V}$, $v \mapsto v$ is an isomorphism of $(\mathcal{P}, \mathcal{V}, *)$ onto $(\mathcal{V}, \mathcal{V}, +)$ where $+$ is determined by (1).*

We describe the relationship between homomorphisms of parallelogram spaces and homomorphisms of commutative groups.

Theorem 1.5 *Let (α, β) be a homomorphism of a parallelogram space $(\mathcal{P}, \mathcal{V}, *)$ to a parallelogram space $(\mathcal{P}', \mathcal{V}', *)'$, and let $+$ and $+'$ be the corresponding binary operations introduced by (1) in \mathcal{V} and \mathcal{V}' , respectively. Then β is a group homomorphism of the group $(\mathcal{V}, +)$ to the group $(\mathcal{V}', +')$ such that*

$$\alpha(P) = \alpha(O) *' \beta(\overrightarrow{OP}) \quad \text{for all } O, P \in \mathcal{P}. \quad (3)$$

Moreover, the choice of a homomorphism β of $(\mathcal{V}, +)$ to $(\mathcal{V}', +')$, together with a choice of points $O \in \mathcal{P}$ and $O' \in \mathcal{P}'$, determine a mapping $\alpha: \mathcal{P} \rightarrow \mathcal{P}'$, $P \mapsto O' *' \beta(\overrightarrow{OP})$ such that $O' = \alpha(O)$, and (α, β) is a homomorphism of $(\mathcal{P}, \mathcal{V}, *)$ to $(\mathcal{P}', \mathcal{V}', *)'$.

Proof. The mapping β from (iv) is uniquely determined by α . Indeed, if $Q \in \mathcal{P}$ is given then there is unique $v' \in \mathcal{V}'$ such that $\alpha(Q) *' v' = \alpha(Q * v)$; it is sufficient to use (i) for $(\mathcal{P}', \mathcal{V}', *')$. Using (iv) and (1) we get $\alpha(Q) *' \beta(v + w) = (\alpha(Q) *' \beta(v)) *' \beta(w) = \alpha(Q) *' (\beta(v) +' \beta(w))$. Finally by (i), $\beta(v + w) = \beta(v) +' \beta(w)$ for all $v, w \in \mathcal{V}$. Hence β is a group homomorphism of $(\mathcal{V}, +)$ to $(\mathcal{V}', +')$. Now let us verify the condition (3). $O * \overrightarrow{OQ} = Q$ holds by (iii), and if we use (iv) for $P = O$ and $v = \overrightarrow{OQ}$ we get $\alpha(O) *' \beta(\overrightarrow{OQ}) = \alpha(Q)$ for all $Q \in \mathcal{P}$.

On the other hand, if a group homomorphism β of the group $(\mathcal{V}, +)$ to $(\mathcal{V}', +')$ is given and $P, Q \in \mathcal{P}$ let us use (1), (2) and (3) to obtain $\alpha(P) *' \beta(\overrightarrow{PQ}) = (\alpha(O) *' \beta(\overrightarrow{OP})) *' \beta(\overrightarrow{PQ}) = (\alpha(O) *' (\beta(\overrightarrow{OP}) +' \beta(\overrightarrow{PQ}))) = \alpha(Q) = \alpha(P * \overrightarrow{PQ})$. Hence (iv) holds for $v = \overrightarrow{PQ}$. Since \overrightarrow{OO} is the identity element for $+$ and consequently $\beta(\overrightarrow{OO})$ is the identity element for the binary operation $+'$, we get $\alpha(O) = O'$ by Corollary 1.3.

Note that the proof of Theorem 1.5 requires only the conditions (i'), (ii'), and the existence of the right-identity element with respect to $+$. As a consequence of the second part of the proof of Theorem 1.2, $(\mathcal{V}, +)$ is then a group. Indeed, consider a quadruple $(\mathcal{P}, \mathcal{V}, *, +)$ such that (i') and (1) are satisfied and there is a right-identity element 0 for $+$. Then $(\mathcal{V}, +; 0)$ is a group which acts sharply transitively on the set \mathcal{P} . The condition (ii') guarantees commutativity of $+$, which however was not needed in Theorem 1.5. Hence for any group $(\mathcal{V}, +)$, Theorem 1.5 gives a description of homomorphisms (α, β) of $(\mathcal{V}, \mathcal{V}, +)$ to $(\mathcal{V}, \mathcal{V}, +)$ by means of an action of the group on itself. The formula (iv), $\alpha(P) *' \beta(v) = \alpha(P * v)$, $P \in \mathcal{P}$, $v \in \mathcal{V}$, reads $\alpha(v) + \beta(w) = \alpha(v + w)$ for all $v, w \in \mathcal{V}$. Here β becomes an endomorphism of $(\mathcal{V}, +)$, and there exists a vector $v_0 \in \mathcal{V}$ such that $\alpha(v) = v_0 + \beta(v)$ for all $v \in \mathcal{V}$. Particularly, α is a left translation if and only if $\alpha = id_{\mathcal{V}}$, and α is an inner automorphism of $(\mathcal{V}, +)$ just in the same case.

Now let us consider a parallelogram space $(\mathcal{V}, \mathcal{V}, +)$ derived from a commutative group $(\mathcal{V}, +)$. Note that by Theorem 1.2, the group under consideration must be commutative when the operations $*$ and $+$ in (1) coincide.

Theorem 1.6 $(\mathcal{V}, \mathcal{V}, *)$ is a parallelogram space if and only if there is a commutative group $(\mathcal{V}, +')$ and a permutation $f : \mathcal{V} \rightarrow \mathcal{V}$ such that $v * w = v +' f(w)$ for all $v, w \in \mathcal{V}$.

Proof. A triple $(\mathcal{V}, \mathcal{V}, *)$, $\mathcal{V} \neq \emptyset$, $* : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ gives a parallelogram space if and only if the following conditions hold:

(i $^\diamond$) to any $u, v \in \mathcal{V}$ there is just one $w \in \mathcal{V}$ such that $v * w = u$,

(ii $^\diamond$) $(w * u) * v = (w * v) * u$ for all $u, v, w \in \mathcal{V}$.

Given $a, b \in \mathcal{V}$ let $e, c \in \mathcal{V}$ be elements, determined by (i $^\diamond$), such that $a * e = a$, $a * c = b$. Using (ii $^\diamond$) we evaluate $b * e = (a * c) * e = (a * e) * c = a * c = b$. Hence e is the right-neutral element for $*$. By Theorem 1.2, first part, there exists a binary operation $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ satisfying

$$v * (a + b) = (v * a) * b \quad \text{for all } a, b, v \in \mathcal{V}. \quad (4)$$

By Theorem 1.2, second part, $(\mathcal{V}, +)$ is a commutative group with neutral element e . From (i $^\diamond$) it follows that $f : \mathcal{V} \rightarrow \mathcal{V}$, $v \mapsto e * v$ is a permutation of \mathcal{V} . Setting $v = e$ in (4) we get $f(a + b) = f(a) * b$. Introducing a binary operation $+'$ by $f(a) +' f(b) = f(a + b)$ for $a, b \in \mathcal{V}$ we

obtain an isomorphism of $(\mathcal{V}, +)$ onto $(\mathcal{V}, +')$. Hence $(\mathcal{V}, +')$ is a commutative group, too, and $f(a) * b = f(a) +' f(b)$, $a, b \in \mathcal{V}$. So we have obtained an isotopy of $(\mathcal{V}, *)$ onto a commutative group.

Conversely, given a commutative group and a permutation $f : \mathcal{V} \rightarrow \mathcal{V}$, the isotopy mentioned above guarantees (i^\diamond) and (ii^\diamond) . It can be seen from the following: $(c * a) * b = (c +' f(a)) +' f(b) = (c +' f(b)) +' f(a) = (c * b) * a$ where $a * c = b$ is valid if and only if $a +' f(c) = b$.

1.4 Geometric approach

Finally let us mention another approach to parallelogram spaces, which seems to be more “geometric” on the first sight, due to the geometric terminology used (but can in fact be explained in a purely algebraic way, by means of Mal’cev operation, [9]):

Definition 1.7 A *parallelogram space* is a pair $(\mathcal{P}, \mathbf{P})$ such that \mathcal{P} is a nonempty set, \mathbf{P} is a quaternary relation on \mathcal{P} , and the following axioms hold:

- 1° For any $a, b, c \in \mathcal{P}$ there is just one $d \in \mathcal{P}$ such that $\mathbf{P}(a, b, c, d)$ is satisfied.
- 2° If (e, f, g, h) is any cyclic permutation of (a, b, c, d) or (d, c, b, a) , respectively, $a, b, c, d \in \mathcal{P}$, then $\mathbf{P}(a, b, c, d)$ implies $\mathbf{P}(e, f, g, h)$.
- 3° For any $a, b, c, d, e, f \in \mathcal{P}$, if $\mathbf{P}(a, b, c, d)$ and $\mathbf{P}(c, d, e, f)$ then $\mathbf{P}(a, b, f, e)$.

Elements of \mathcal{P} are called *points* of a parallelogram space. Each quadruple $(a, b, c, d) \in \mathbf{P}$ will be called a *parallelogram*, and $(a, b, c, d) \in \mathbf{P}$ will be also written as $\mathbf{P}(a, b, c, d)$.

If a parallelogram space $(\mathcal{P}, \mathbf{P})$ in this sense is given let us define a binary relation \simeq on $\mathcal{P} \times \mathcal{P}$ by

$$(a, b) \simeq (d, c) \Leftrightarrow \mathbf{P}(a, b, c, d). \quad (5)$$

We can check easily that \simeq is an equivalence relation. Elements (classes) of the factor set $\mathcal{P} \times \mathcal{P} / \simeq$ are referred to as *vectors*. Denote by $\overrightarrow{(a, b)}$ the vector containing a pair $(a, b) \in \mathcal{P} \times \mathcal{P}$.

If we take $\mathcal{V} = \mathcal{P} \times \mathcal{P} / \simeq$, $\vec{\cdot} : \mathcal{P} \times \mathcal{P} \rightarrow \mathcal{V}$, $(a, b) \mapsto \overrightarrow{(a, b)}$ then $(\mathcal{P}, \mathcal{V}, \vec{\cdot})$ is a parallelogram space in the sense of the previous (second) definition. Indeed, the condition (i') follows from the definition of $\vec{\cdot}$ and from 1°, the condition (ii') is a consequence of 2°, 3° and of the definition of $\vec{\cdot}$. Conversely, given a parallelogram space $(\mathcal{P}, \mathcal{V}, \vec{\cdot})$ according to the second definition, introduce a parallelogram (in a new sense) as a quadruple (a, b, c, d) , $a, b, c, d \in \mathcal{P}$ such that $\overrightarrow{(a, b)} = \overrightarrow{(d, c)}$. Conditions (i') and (ii') allow to verify all the conditions 1°, 2°, 3° so that \mathcal{P} together with the set of all “new” parallelograms form a parallelogram space in the last sense.

Remark that in the above approach, 1° may be substituted by

- 4° For any $a, b, c \in \mathcal{P}$, there is just one $d \in \mathcal{P}$ such that $\mathbf{P}(a, b, d, c)$ holds.

This can be verified if we pass from $\mathbf{P}(a, b, c, d)$ to $\mathbf{P}(b, a, d, c)$ using suitable cyclic permutations and 2°: $\mathbf{P}(a, b, c, d) \Leftrightarrow \mathbf{P}(c, d, a, b) \Leftrightarrow \mathbf{P}(b, a, d, c)$.

2 Parallelogram space of a medial quasigroup

Let $\mathcal{Q} = (Q, \cdot)$ be a medial quasigroup, i.e. a groupoid such that the equations $xa = b$ and $ay = b$ are uniquely solvable in Q for any $a, b \in Q$ and the identity called *mediality* is satisfied:

$$(xy)(uv) = (xu)(yv). \quad (6)$$

The solutions of the equations are denoted by $x = b/a$, $y = a \backslash b$, and additional accompanying operations $/$, \backslash arise on the support Q . Any quasigroup is left and right cancellative.

We show how a medial quasigroup determines a parallelogram space in the “geometric” sense of Definition 1.7.

2.1 Parallelograms in medial quasigroups

Let $\mathcal{Q} = (Q, \cdot)$ be a medial quasigroup. For $a, b \in Q$, let us introduce a *left transfer* $L_{a,b}$ as the composition (from the right to the left) $L_b^{-1}L_a$, [6]. That is,

$$L_{a,b}(c) = d \iff ac = bd \text{ for all } a, b, c, d \in Q. \quad (7)$$

We say that an ordered quadruple $(a, b, c, d) \in Q \times Q \times Q \times Q$ is a *parallelogram* in \mathcal{Q} if $L_{a,b} = L_{d,c}$. Denote by $\mathbf{P}_{\mathcal{Q}}$ the set of all parallelograms in \mathcal{Q} ; we write $(a, b, c, d) \in \mathbf{P}_{\mathcal{Q}}$ as above. It can be checked that $(Q, \mathbf{P}_{\mathcal{Q}})$ is a parallelogram space.

We can check that $(a, b) \simeq (d, c) \iff \mathbf{P}_{\mathcal{Q}}(a, b, c, d)$, determines an equivalence relation \simeq on Q . The equivalence classes of the relation \simeq are vectors in $(Q, \mathbf{P}_{\mathcal{Q}})$. Denote by \overrightarrow{ab} the vector containing $(a, b) \in Q \times Q$.

2.2 Properties of vector addition

Let us introduce a *vector addition* with respect to a fixed element $o \in Q$, which can be called the *origin*, by $\overrightarrow{oa} +_o \overrightarrow{ob} := \overrightarrow{oc} \iff \mathbf{P}_{\mathcal{Q}}(o, a, c, b)$. It appears that the choice of the origin is not essential since for two different choices, we obtain isomorphic commutative groups. The following can be checked:

$$\overrightarrow{ab} + \overrightarrow{bc} = \overrightarrow{ac} \text{ for all } a, b, c \in Q. \quad (8)$$

$$\mathbf{P}_{\mathcal{Q}}(a, a, b, c) \iff b = c \text{ for } a, b, c \in Q, \quad (9)$$

For any $a, b, c, d, e, f \in Q$,

$$\mathbf{P}_{\mathcal{Q}}(a, b, d, e), \mathbf{P}_{\mathcal{Q}}(b, c, e, f) \implies \mathbf{P}_{\mathcal{Q}}(c, d, f, a). \quad (10)$$

For any $a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2 \in Q$,

$$\mathbf{P}_{\mathcal{Q}}(a_1, b_1, c_1, d_1), \mathbf{P}_{\mathcal{Q}}(a_2, b_2, c_2, d_2) \implies \mathbf{P}_{\mathcal{Q}}(a_1a_2, b_1b_2, c_1c_2, d_1d_2). \quad (11)$$

For any $a, b, c, d, q \in Q$,

$$\mathbf{P}_{\mathcal{Q}}(a, b, c, d) \iff \mathbf{P}_{\mathcal{Q}}(qa, qb, qc, qd) \iff \mathbf{P}_{\mathcal{Q}}(aq, bq, cq, dq), \quad (12)$$

Moreover, for any $a, b, c, d \in Q$, if $\mathbf{P}_{\mathcal{Q}}(a, b, c, d)$ then $\mathbf{P}_{\mathcal{Q}}(ab, bc, ca, db)$,

$$\mathbf{P}_{\mathcal{Q}}(ac, bd, ca, db), \mathbf{P}_{\mathcal{Q}}(ad, ba, cb, dc), \mathbf{P}_{\mathcal{Q}}(ad, bc, cb, da), \quad (13)$$

and

$$\mathbf{P}_{\mathcal{Q}}(ab, ad, cd, cb) \text{ holds for any } a, b, c, d \in Q. \quad (14)$$

$$\mathbf{P}_{\mathcal{Q}}(q, (q/q)a, ba, bq) \text{ and } \mathbf{P}_{\mathcal{Q}}(q, a(q \setminus q), ab, qb) \text{ hold for all } a, b, q \in Q. \quad (15)$$

We take now vectors as position vectors of elements from Q with respect to the origin $o \in Q$, and pass from position vectors to their “end points”. That is, for any $a, b \in Q$, define $a +_o b$ as an element of Q determined uniquely by $\mathbf{P}_{\mathcal{Q}}(o, a, a +_o b, b)$. We obtain a commutative group $(Q, +_o; o)$, [6].

Lemma 2.1 *Let $\mathcal{Q} = (Q, \cdot)$ be a medial quasigroup and $o, a, b, c, d \in Q$. Then*

$$\mathbf{P}_{\mathcal{Q}}(a, b, c, d) \iff a +_o c = b +_o d.$$

Proof. If $\mathbf{P}_{\mathcal{Q}}(o, a, a +_o c, c)$ and $\mathbf{P}_{\mathcal{Q}}(a, b, c, d)$ then $\mathbf{P}_{\mathcal{Q}}(b, a +_o c, d, o)$, by (10). Consequently $\mathbf{P}_{\mathcal{Q}}(b, o, d, a +_o c)$. Similarly, from $\mathbf{P}_{\mathcal{Q}}(o, b, b +_o d, d)$ we obtain $\mathbf{P}_{\mathcal{Q}}(b, o, d, b +_o d)$, that is, $a +_o c = b +_o d$. Vice versa, the equality $a +_o c = b +_o d$ means that $\mathbf{P}_{\mathcal{Q}}(a, o, c, a +_o c)$, and at the same time $\mathbf{P}_{\mathcal{Q}}(o, b, b +_o d, d)$ holds. Hence $\mathbf{P}_{\mathcal{Q}}(b, c, d, a)$ by (10), or equivalently, $\mathbf{P}_{\mathcal{Q}}(a, b, c, d)$.

Lemma 2.2 *Let (\mathcal{Q}, o) be a pointed medial quasigroup where $\mathcal{Q} = (Q, \cdot)$. Then the mappings $L_{o/o}$ and $R_{o \setminus o}$ are commuting automorphisms of the group $(Q, +_o; o)$, i.e. $L_{o/o} \circ R_{o \setminus o} = R_{o \setminus o} \circ L_{o/o}$.*

Proof. Let $q \in Q$. Put $l_o = o/o$ and $r_o = o \setminus o$. Step by step, we obtain $(l_o \cdot (qr_o)) \cdot (o \cdot o) = (l_o o)((qr_o) \cdot o) = o((qr_o) \cdot o) = (or_o) \cdot ((qr_o) \cdot o) = (o \cdot (qr_o)) \cdot (r_o \cdot o) = ((l_o o)(qr_o)) \cdot (r_o o) = ((l_o q)(or_o)) \cdot (r_o o) = ((l_o q) \cdot r_o) \cdot (o \cdot o)$. By right cancellation, $l_o \cdot (qr_o) = (l_o q) \cdot r_o$, which means $L_{l_o} \circ R_{r_o} = R_{r_o} \circ L_{l_o}$. To prove that the mappings are automorphisms let us use (12) and $l_o o = o$: $\mathbf{P}_{\mathcal{Q}}(o, a, a +_o b, b)$ is equivalent with $\mathbf{P}_{\mathcal{Q}}(o, l_o a, l_o(a +_o b), l_o b)$, which reads also $l_o(a +_o b) = l_o a +_o l_o b$. Hence $L_{l_o}(a +_o b) = L_{l_o}(a) +_o L_{l_o}(b)$ for all $a, b \in Q$. Similarly, $R_{r_o}(a +_o b) = R_{r_o}(a) +_o R_{r_o}(b)$ holds for all $a, b \in Q$.

By means of parallelograms, we can formulate and prove a version of the so-called Toyoda’s theorem as follows:

Theorem 2.3 (“Toyoda’s theorem”) *Let $\mathcal{Q} = (Q, \cdot)$ be a medial quasigroup, and $o \in Q$. Then, for all $a, b \in Q$, the equality $a \cdot b = a(o \setminus o) +_o (o/o)b +_o (o \cdot o)$ holds where the binary operation $+_o$ on Q is introduced above (before Lemma 2.1).*

Proof. According to (15), we have $\mathbf{P}_{\mathcal{Q}}(o, ar_o, ab, ob)$ and $\mathbf{P}_{\mathcal{Q}}(o, l_o b, ob, o \cdot o)$ where, again, $r_o = o \setminus o$, $l_o = o/o$. Hence by the definition of addition $+_o$ on Q , $a \cdot b = a \cdot r_o +_o o \cdot b$ and $o \cdot b = l_o b +_o o \cdot o$ holds, and consequently $a \cdot b = a \cdot r_o +_o l_o \cdot b +_o o \cdot o$. That is, $a \cdot b = R_{o \setminus o}(a) +_o L_{o/o}(b) +_o o \cdot o$ holds.

Acknowledgement

The authors are supported by the grant from Grant Agency of Czech Republic GAČR no. P201/11/0356 with the title: “Riemannian, pseudo-Riemannian and affine differential geometry”. Moreover, the first author is supported by the project of specific university research of the Brno University of Technology, No. FAST-S-11-47.

References

- [1] BARTOŠKOVÁ, Z.: *Medial Quasigroups and Geometry*. Mgr. Thesis, Palacký University Olomouc, Fac. Sci., 2011 (in Czech).
- [2] BARTOŠKOVÁ, Z.: *Commutative Groups and Medial Quasigroups*. Bc. Thesis, Palacký University Olomouc, Fac. Sci., 2009 (in Czech).
- [3] BELOUSOV, V.D.: *Foundations of the theory of quasigroups and loops*. Moscow, Nauka, 1967.
- [4] RUCK, R.H.: *A Survey of Binary Systems*. Berlin, Springer, 1958.
- [5] JEŽEK, J., KEPKA, T.: *Medial Groupoids*. Academia Praha, 1983.
- [6] HAVEL, V.J., VANŽUROVÁ, A.: *Medial Quasigroups and Geometry*. Palacký University in Olomouc, Olomouc, 2006.
- [7] PFLUGFELDER, H.O.: *Quasigroups and Loops, Introduction*. Heldermann Verlag, Berlin, 1990.
- [8] PICKERT, G.: *Projektive Ebenen*. Springer, Berlin, Heidelberg, New York, 1975.
- [9] ROMANOVSKA, A., SMITH, J.D.H.: *Modes*. World Scientific, New Jersey, London, Singapore, Hong Kong, 2002.
- [10] VANŽUROVÁ, A.: *Golden Section Quasigroups, Finite Examples*. In Proc. 10th International Conf. APLIMAT 2011, Part IV, pp. 183-190, 2011.
- [11] VANŽUROVÁ, A.: *Algebraic systems in cryptology and the security of information*. In International Conference in Military Technology Proceeding, Section 12 Security technology (ed. Z. Vránová and V. Plátěnka), ICMT'11, Brno : University of Defence, pp. 1251-1256, 2011. ISBN 978-80-7231-787-5 (ISBN 978-80-7231-788-2 for CD).
- [12] VANŽUROVÁ, A., JUKL, M.: *Parallelogram spaces and affine spaces*. Preprint, 2011.
- [13] VOLENEC, V.: *Geometry of medial quasigroups*. Rad. Jugosl. Akad. Znam. Umjet 421, pp. 79-91, 1986.
- [14] VOLENEC, V.: *GS-quasigroups*. Časop. pěst. mat. 115, pp. 307-318, 1990.

Current address

VANŽUROVÁ Alena, Doc. RNDr. CSc.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University, Tř. 17. listopadu 12, 771 46 Olomouc, alena.vanzurova@upol.cz;

Department of Mathematics, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic, vanzurova.a@fce.vutbr.cz.

JUKL Marek, RNDr. PhD.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University, Tř. 17. listopadu 12, 771 46 Olomouc, marek.jukl@upol.cz.

METRIZABLE CONNECTIONS AND RESTRICTIVELY VARIATIONAL CONNECTIONS IN AFFINE MANIFOLDS

VANŽUROVÁ Alena, (CZ), PIRKLOVÁ Petra, (CZ)

Abstract. We discuss the relationship between metrizable and restrictively variational connections and give some partial answers in the case of two-dimensional manifolds.

Key words and phrases. Manifold; connection; metric; Riemannian geometry; calculus of variations; inverse problem; Helmholtz conditions.

Mathematics Subject Classification. Primary 53C05; Secondary 53C60, 53B05, 53B20.

1 Variational connections

1.1 Introduction

In some branches of physics it has been discovered, [8], that the solution of many problems is considerably simplified if the basic equations can be expressed in the form of variational principle.

We consider second order differential equations, in short SODE's. For geometric description of equations under consideration, we use fibred manifolds. Let M be an n -dimensional differentiable manifold. Then $T_s^r(M)$ denotes tensor fields of type (r, s) on M , $p: TM \rightarrow M$ is the tangent bundle of M , and $p^2: T^2M \rightarrow M$ the bundle of two-velocities. Note that the vector bundle $\mathbf{R} \times TM \rightarrow \mathbf{R}$ is canonically identified with the first jet prolongation $J^1\pi = \pi_1: J^1(\mathbf{R} \times M) \rightarrow \mathbf{R}$ of the fibred manifold $\pi: \mathbf{R} \times M \rightarrow \mathbf{R}$, and $\mathbf{R} \times T^2M \rightarrow \mathbf{R}$ is canonically identified with the second jet prolongation $\pi_2: J^2(\mathbf{R} \times M) \rightarrow \mathbf{R}$ of π .

If (x^i) are local coordinates on M and t denote a (global) coordinate on \mathbf{R} we use fibre coordinates (x^i, u^i) on TM (adapted to the projection p_M), (x^i, u^i, z^i) on T^2M (adapted to the projection p_M^2), $1 \leq i \leq n$, where we denoted $u^i = \dot{x}^i$, $z^i = \ddot{x}^i$. Similarly, we use fibre coordinates (t, x^i, u^i) on $\mathbf{R} \times TM \rightarrow \mathbf{R}$, and (t, x^i, u^i, z^i) on $\mathbf{R} \times T^2M \rightarrow \mathbf{R}$.

A smooth, or at least C^2 -differentiable, function $L: \mathbf{R} \times TM \rightarrow \mathbf{R}$ is called here a (first order) *Lagrangian function* or *Lagrangian* in short. In local fibre coordinates, $L = L(t, x, u)$. Often, under a first order *Lagrangian* in $\mathbf{R} \times TM$ we mean a π_1 -horizontal 1-form λ on $\mathbf{R} \times TM$. In adapted coordinates on $\mathbf{R} \times TM$, its expression reads $\lambda = L dt$ where $L(t, x^i, u^i)$ is a Lagrangian function. Given a Lagrangian L , under the *Euler-Lagrange operators*, or *expressions* of L we understand the functions $E_i(L): \mathbf{R} \times T^2M \approx J^2(\mathbf{R} \times M) \rightarrow \mathbf{R}$,

$$E_i(L) = \frac{\partial L}{\partial x^i} - \frac{d}{dt} \frac{\partial L}{\partial u^i}, \quad i = 1, \dots, n. \quad (1)$$

In what follows, $\frac{d}{dt} := \frac{\partial}{\partial t} + x^j \frac{\partial}{\partial x^j} + \dot{x}^j \frac{\partial}{\partial \dot{x}^j}$ is the “total derivation operator”. The *Euler*, or *Euler-Lagrange equations* corresponding to (1) read $E_i(L) = 0$, that is,

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt} \frac{\partial L}{\partial u^i} = 0. \quad (2)$$

The expressions $E_i(L)$ are local components of the Euler-Lagrange form $E_\lambda = E_i dx^i \wedge dt$ of the first-order Lagrangian $\lambda = L dt$.

1.2 Variational Problem

Let M be an n -dimensional manifold and let $\gamma: I \rightarrow M$ be a regular curve on an open interval $I \subset \mathbf{R}$ defined (in local coordinates) by $x(t) = (x^1(t), \dots, x^n(t))$, $t \in I$. Let

$$\frac{d\gamma}{dt} = \dot{x}(t) = dx(t)/dt = (\dot{x}^1(t), \dots, \dot{x}^n(t))$$

be the corresponding tangent vector field along γ , we suppose $\dot{x}(t) \neq 0$ for $t \in I$. In what follows, let all indices vary from 1 to n . Let $A = \gamma(a)$, $B = \gamma(b)$ be two points on γ corresponding to parameters a and $b \in I$, respectively. Given a (differentiable) functions $\omega^i: M \rightarrow \mathbf{R}$ such that $\omega^i(A) = \omega^i(B) = 0$; in local coordinates, $\omega^i(x^1, \dots, x^n)$; and a real parameter $\varepsilon \in \mathbf{R}$, the formulas

$$\bar{x}^i(t) = x^i(t) + \varepsilon \cdot \omega^i(x^1(t), \dots, x^n(t))$$

define a new curve $\bar{\gamma}: \bar{x}(t) = (\bar{x}^1(t), \dots, \bar{x}^n(t))$ which is, for small ε , “close” to the original one and passes through the given points A and B as well.

Let us consider the integral

$$I(\gamma) = \int_a^b L(t, x^1(t), x^2(t), \dots, x^n(t), \dot{x}^1(t), \dot{x}^2(t), \dots, \dot{x}^n(t)) dt \quad (3)$$

where L is an analytic function of the given arguments (Lagrangian function).

The *variational problem associated with L* is to find those curves $x(t)$ which minimize the fundamental integral (3) subject to the prescribed boundary conditions $A = \gamma(a)$, $B = \gamma(b)$, [15].

If $\bar{I} = I(\bar{\gamma})$ is a similar integral for $\bar{\gamma}$ then expanding L as a Taylor series in powers of ε we get

$$\bar{I} = I + \varepsilon \cdot \int_a^b \left(\frac{\partial L}{\partial x^i} \omega^i + \frac{\partial L}{\partial \dot{x}^i} \frac{\partial \omega^i}{\partial x^j} \dot{x}^j \right) dt + \dots$$

where “dots” stand for the members of order ε^2 and higher. Coefficients at ε , ε^2 etc. in the above expansion are denoted by δI , $\delta^2 I$ etc., and are called the *first*, *second* etc. *variation* of the integral I . Particularly, the first variation

$$\delta I = \int_a^b \left(\frac{\partial L}{\partial x^i} \omega^i + \frac{\partial L}{\partial \dot{x}^i} \frac{\partial \omega^i}{\partial x^j} \dot{x}^j \right) dt$$

is often written in the following form (we integrate in parts in the second summand and use properties of the functions $\omega^i(x)$)

$$\delta I = \int_a^b \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) \right) \omega^i dt. \quad (4)$$

The integral (3) is called *stationary* if its first variation vanishes, $\delta I = 0$, for an arbitrary choice of functions ω^i (the so-called Hamilton’s variational principle, $\delta \int L(t, x^i(t), \dot{x}^i(t)) dt = 0$). The curve for which it holds is called the *extremal* of the integral under consideration. As it follows from (4), the integral (3) is stationary if and only if the Euler-Lagrange equations (2) are satisfied.

Functions $x^i(t)$ for which the integral reaches its minimum or maximum must satisfy the Euler-Lagrange equations, and any solution $x = x(t)$ of the equations (2) is an extremal curve of the integral (3). Note that H. Cartan precisely formulated this variational problem for Lagrangian functions L in the class C^2 [2]. In this case, the formula for first variation reads $\delta I = dI(\bar{\gamma})/d\epsilon|_{\epsilon=0}$ and its extremals $\gamma \in C^2$ are solutions of the Euler-Lagrange equations (2). The problem can be considered in a more general setting [16].

1.3 SODEs and the Inverse Problem

The most general form of a system of second-order ordinary differential equations, which may be time-dependent in general, for functions $t \mapsto x^k(t)$; $k = 1, \dots, n$ with the definition domain in \mathbf{R}^n or in a coordinate neighborhood $U \subset M$ of some differentiable n -manifold M , reads

$$E_i(t, x^k, \dot{x}^k, \ddot{x}^k) = 0, \quad i = 1, \dots, n, \quad (5)$$

and is called the “Helmholtz set”, according to H. Helmholtz who did the first investigation (although actually, Helmholtz did not consider any explicit time dependence). To solve the system means to find local curves $\gamma : I \rightarrow M$ on M where $I \subset \mathbf{R}$ is an open interval, $\gamma(t) = (x^1(t), \dots, x^n(t))$, such that

$$E_i \left(t, \gamma(t), \frac{d\gamma}{dt}, \frac{d^2\gamma}{dt^2} \right) = 0, \quad i = 1, \dots, n. \quad (6)$$

If second derivatives are put in evidence, the equations read

$$a_{ij}(t, x^k, \dot{x}^k) \cdot \ddot{x}^j + b_i = 0, \quad i = 1, \dots, n \quad (7)$$

where the Einstein summation convention is used. According to [8, p. 367], E_i may contain second derivatives only linearly, therefore (6) can be written in the form (7) without loss of generality.

Given a manifold (M, ∇) endowed with linear connection ∇ , the equations of canonically parametrized geodesics in M ,

$$\ddot{x}^i + \Gamma_{jk}^i(x) \dot{x}^j \dot{x}^k = 0, \quad (8)$$

yield a natural example of such a system, which is in a special form, solved to second derivatives, i.e. in the form

$$\ddot{x}^i - f^i(t, x^k, \dot{x}^k) = 0, \quad i, k = 1, \dots, n. \quad (9)$$

On the other hand, whenever the functions f^i in (9) are quadratic forms in first derivatives, “components of velocities”, with coefficients depending on coordinates x^k , “components of positions” only, (9) represent geodesic paths.

It might be useful to know whether the system (6), or (7), can be expressed in the form of a variational principle, or at least, some linear combination of the original equations takes the form of Lagrange equations for some Lagrangian. More precisely [8], the so-called Strong Inverse Problem of the Calculus of Variations means:

Is there a function $L(t, x^k, \dot{x}^k)$, sufficiently differentiable, such that $E_i = 0$ are just Euler-Lagrange equations of a variational principle $\delta \int L(t, x, \dot{x}) dt = 0$, $E_i = E_i(L)$? Another speaking, we ask if there are function L such that the following equations hold:

$$E_i = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i}. \quad (10)$$

If the answer is affirmative we want to find all such L .

In dimension $n = 2$, a complete answer for the inverse problem in the real analytic class was given by Jesse Douglas in [3].

The range of applicability of Lagrange’s formalism can be extended, [8]. E.g. we try to substitute linear combinations of the original equations: Helmholtz set (5) is replaced by

$$\sum_j g_{ij}(t, x, u) E_j = 0. \quad (11)$$

The so-called Weak Inverse Problem of the Calculus of Variations, or the multiplier problem, means:

Find all pairs $((g_{ij}), L)$ where $L(t, x^k, \dot{x}^k)$ is a Lagrangian and $(g_{ij}(t, x^k, \dot{x}^k))$ is a non-degenerate functional matrix (all objects sufficiently differentiable) such that

$$\sum_j g_{ij} E_j = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i}. \quad (12)$$

If the answer is affirmative find all such L and g_{ij} .

Recall that g_{ij} are usually called (Lagrange) *multipliers*. Although some necessary and sufficient conditions have been formulated, a lot of progress have been made recently, and some partial answers have been given, the problem is difficult and still open in a way. Only particular cases can be answered easily; e.g. equations for geodesics of a pseudo-Riemannian space are always variational in this weak sense.

1.4 Helmholtz conditions

The classical result states that necessary and sufficient conditions are as follows.

Lemma 1.1 *A family of functions $E_i : J^2(\mathbf{R} \times M) \rightarrow \mathbf{R}$ (depending on t, x^i, u^i, z^i in general), $i = 1, \dots, n$, can be regarded as Euler-Lagrange expressions of some Lagrange function $L : J^1(\mathbf{R} \times M) \rightarrow \mathbf{R}$ if and only if the so-called Helmholtz conditions hold, [8]: with respect to any adapted chart, the following relations hold identically for all $i, k = 1, \dots, n$*

$$\frac{\partial E_i}{\partial z^k} - \frac{\partial E_k}{\partial z^i} = 0, \quad (13)$$

$$\frac{\partial E_i}{\partial u^k} + \frac{\partial E_k}{\partial u^i} = \frac{d}{dt} \left(\frac{\partial E_i}{\partial z^k} + \frac{\partial E_k}{\partial z^i} \right), \quad (14)$$

$$\frac{\partial E_i}{\partial x^k} - \frac{\partial E_k}{\partial x^i} = \frac{1}{2} \frac{d}{dt} \left(\frac{\partial E_i}{\partial u^k} - \frac{\partial E_k}{\partial u^i} \right). \quad (15)$$

Of course, if the starting equations take some particular shape the conditions are appropriately modified.

Recall that the first attempt to investigate this question is due to Helmholtz; in [9], necessary and sufficient conditions were stated, but only necessity was proven. The proof of sufficiency was given a bit later by Mayer [13].

2 Restrictively variational connections

The Fundamental Lemma of (pseudo-) Riemannian geometry states that given a (non-degenerate) metric g on M there is a unique connection ∇ on M that is symmetric and “compatible” with the metric in the sense that the metric is covariantly constant with respect to the connection, $\nabla g = 0$ (geometrically speaking, the scalar product defined by the quadratic metric tensor on each tangent space translates paralelly along any curve). Such a connection is called Riemannian, or Levi-Civita.

The metrizability problem is the “reverse” question: given a connection ∇ on M , find necessary and sufficient conditions (formulated in terms of the given connection) for ∇ to be just the Levi-Civita connection of some metric, eventually find all such metrics. A system of integrability conditions has been given by Eisenhart and Veblen, many particular answers are known, an equivalent formulation in terms of geodesic mappings can be realized, and a system of differential equations that controls this question was in fact found, [14]. But even this problem is far from being completely solved.

We say that a linear connection on M is *restrictively variational* [11], [12], [26] if there exists a (smooth, or of the class at least two) function $L : \mathbf{R} \times TM \rightarrow \mathbf{R}$, in local fibre coordinates $L = L(t, x, u)$, and a non-singular type $(0, 2)$ tensor field $g : M \rightarrow T_2^0(M)$ on M such that with respect to any local fibre chart, the functions

$$-E_i = g_{ik}(x)(z^k + \Gamma_{rs}^k(x)u^r u^s), \quad i = 1, \dots, n, \quad (16)$$

coincide with Euler-Lagrange expressions $E_i(L)$ of some Lagrangian L ; here Γ_{jk}^i are components of ∇ .

The relationship between restrictive variationality and metrizability for a linear connection on M can be expressed as follows [26]):

Theorem 2.1 *Given a manifold (M, ∇) with linear connection, the following conditions are equivalent:*

- (i) ∇ is restrictively variational;
- (ii) the symmetric part $\tilde{\nabla}$ of ∇ is metrizable;
- (iii) there is a non-singular symmetric type $(0, 2)$ tensor field g on M such that

$$\tilde{\Gamma}_{jk}^i = \frac{1}{2} g^{i\ell} \left(\frac{\partial g_{j\ell}}{\partial x^k} + \frac{\partial g_{k\ell}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^\ell} \right); \quad (17)$$

$\tilde{\Gamma}_{jk}^i$ are components of the (symmetric) connection $\tilde{\nabla}$, and $g^{i\ell}$ are components of the tensor g^* dual to g (in a natural pairing, i.e. $g^{i\ell}(x)$ is the inverse matrix of $g_{i\ell}(x)$, $x \in M$).

- (iv) there is a non-singular symmetric type $(0, 2)$ tensor field g on M such that the equations

$$g_{ik}(x)(z^k + \Gamma_{rs}^k(x)u^r u^s) = 0 \quad (18)$$

are variational.

Proof. The equivalence (i) \Leftrightarrow (iv) is a matter of definition. The equivalence (ii) \Leftrightarrow (iii) is classical: if a symmetric connection $\tilde{\nabla}$ on M is metrizable and g is a suitable metric, then $\tilde{\nabla}$ is exactly the Riemannian (Levi-Civita) connection of the pseudo-Riemannian manifold (M, g) ; $\tilde{\nabla}g = 0$ is equivalent to (17) for any symmetric connection. Now let ∇ be a linear connection on M , and let $\tilde{\Gamma}_{jk}^i = \tilde{\Gamma}_{kj}^i$ denote components of its symmetric part $\tilde{\nabla}$. Assume that $\tilde{\nabla}$ is metrizable, g being a metric tensor compatible with $\tilde{\nabla}$. Let us introduce a (global) function L , $L(u) = \frac{1}{2}g_x(u, u)$, $u \in T_x M$; in local fibre coordinates, $L = \frac{1}{2}g_{rs}(x)u^r u^s$. To check that L is a Lagrangian function we must evaluate the corresponding expressions $E_i(L)$ and verify that they obey the Helmholtz conditions. We get

$$\frac{\partial L}{\partial x^i} = \frac{1}{2} \partial_i g_{rs} u^r u^s, \quad \frac{d}{dt} \left(\frac{\partial L}{\partial u^i} \right) = u^j \partial_j g_{is} u^s + z^j g_{is} \delta_j^s, \quad (19)$$

$$E_i(L) = \frac{1}{2} \partial_i g_{rs} u^r u^s - g_{is} z^s - \partial_r g_{is} u^r u^s - \left[g_{is} z^s + \frac{1}{2} (\partial_s g_{ir} + \partial_r g_{is} - \partial_i g_{rs}) \right] u^r u^s. \quad (20)$$

In short, $E_i(L) = - \left(g_{is} z^s + \tilde{\Gamma}_{irs} u^r u^s \right) = - \left(g_{is} z^s + g_{i\ell} \tilde{\Gamma}_{rs}^\ell u^r u^s \right)$ where $2\tilde{\Gamma}_{irs} = 2g_{i\ell} \tilde{\Gamma}_{rs}^\ell = \partial_s g_{ir} + \partial_r g_{is} - \partial_i g_{rs}$. To check that L is a Lagrangian function, let us verify (13)–(15). In fact, according to symmetry of g , (13) is satisfied for the expressions (16):

$$\frac{\partial E_i}{\partial z^k} - \frac{\partial E_k}{\partial z^i} = -g_{is} \delta_k^s + g_{ks} \delta_i^s = -g_{ik} + g_{ki} = 0. \quad (21)$$

To verify (14) we use the condition $\tilde{\nabla}g = 0$ as well as symmetry of g : $\frac{\partial E_i}{\partial u^k} + \frac{\partial E_k}{\partial u^i} - \frac{d}{dt} \left(\frac{\partial E_i}{\partial z^k} + \frac{\partial E_k}{\partial z^i} \right) = 2 \left(\tilde{\Gamma}_{iks} + \tilde{\Gamma}_{kis} + \frac{\partial g_{ki}}{\partial x^s} \right) u^s = 2(\tilde{\nabla}g) \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}; u^s \frac{\partial}{\partial x^j} \right) = 0$. After some calculations, using the formula (20), we get

$$\frac{\partial E_i}{\partial x^k} = \frac{\partial g_{is}}{\partial x^k} z^s + \left(\frac{\partial^2 g_{is}}{\partial x^k \partial x^r} - \frac{1}{2} \frac{\partial^2 g_{rs}}{\partial x^k \partial x^r} \right) u^r u^s, \quad (22)$$

and finally (15) which proves (ii) \Rightarrow (i). Vice versa, if the connection is restrictively variational, that is, if the Helmholtz conditions are satisfied then symmetry of g follows according to (13);

(14) together with (13) give $\tilde{\nabla}g = 0$; if g is symmetric and covariantly constant with respect to $\tilde{\nabla}g$ then (15) gives no new condition. Hence (i) \Rightarrow (ii), which completes the proof.

Given a system of SODE's of a particular form in which second derivatives are expressible as quadratic forms in first derivatives,

$$\ddot{x}^k + \Gamma_{rs}^k(x) \dot{x}^r \dot{x}^s = 0, \quad k = 1, \dots, n, \quad (23)$$

we can use Theorem 2.1 for deciding whether the system is derivable from a Lagrangian. Namely, we can assume that the functions $\Gamma_{rs}^k(x)$ are components of a symmetric linear connection ∇ on some neighborhood $U \subset \mathbf{R}^n$. If ∇ is (locally) metrizable, $g_{ij}(x)$ (with $\det(g_{ij}(x)) \neq 0$ at any $x \in U$) being components of some non-degenerate metric g compatible with ∇ on U then the system of equations (23) is equivalent to the system

$$g_{ik} (\ddot{x}^k + \Gamma_{rs}^k(x) \dot{x}^r \dot{x}^s) = 0, \quad k = 1, \dots, n, \quad (24)$$

hence the functions $g_{ik}(x)$ are the desired variational multipliers. On the other hand, given a system of SODE's (23), if there are Lagrange multipliers independent on time and velocities then they are just components of a metric with geodesics given exactly by (23).

Acknowledgement

The paper was supported by the grant from Grant Agency of Czech Republic GAČR no. P201/11/0356 with the title: "Riemannian, pseudo-Riemannian and affine differential geometry" and by the project of specific university research of the Brno University of Technology, No. FAST-S-11-47.

References

- [1] ANDERSON, I., THOMPSON, G.: *The inverse problem of the calculus of variations for ordinary differential equations*. Memoires AMS 98, No 473, 1992.
- [2] CARTAN, H.: *Theorie Elementaire des Fonctions Analytiques D'une ou Plusieurs Variabl.* Hermann, 1975.
- [3] DOUGLAS, J.: *Solution of the inverse problem of the calculus of variations*. Trans. Amer. Math. Soc., Vol. 50, pp. 71-128, 1941.
- [4] EISENHART, L.P., VEULEN, O.: *The Riemann geometry and its generalization*. Proc. London Math. Soc. Vol. 8, pp. 19-23, 1922.
- [5] GODBILLON, C.: *Géométrie Différentielle et Mécanique Analytique*. Hermann, Paris, 1969.
- [6] GRIFONE, J.: *Structure presque-tangente et connexions, I*. Ann. Inst. Fourier (Grenoble), Vol. 22, pp. 287-334, 1972.
- [7] GRIFONE, J., MUSZNAY, Z.: *Variational Principles of Second-order Differential Equations*. World Scientific, Singapore, 2000.
- [8] HAVAS, P.: *The range of application of the Lagrange formalism I*. Nuovo Cimento Suppl., Vol. 3, pp. 363-388, 1957.
- [9] HELMHOLTZ, H.: *Ueber die physikalische Bedeutung des Princips der kleinsten Wirkung*. Jour. f. d. reine u. angew. Math., Vol. 100, pp. 137-166, 1887.

- [10] KOLÁŘ, I., SLOVÁK, J., MICHOR, P.W.: *Natural Operations in Differential Geometry*. Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [11] KLEIN, J.: *Geometry of sprays. Lagrangian case. Principle of least curvature*. Proc. IUTAM-ISIMM Symp. on Modern Developements in Analytical Mechanics, Vol. I, Torino, pp. 177-196, 1982.
- [12] KLEIN, J.: *On variational second order variational equations; Polynomial case*. Diff. Geom. and Its Appl. Proc. Conf. Opava 1992. Silesian Univ. Opava 1993, pp. 449-456.
- [13] MAYER, A.: Ber. Kon. Ges. Wiss. Leipzig, Math.-Phys. Kl., Vol. 48, p. 519, 1896.
- [14] MIKEŠ, J.: *Geodesic mappings of affine-connected and Riemannian spaces*. Jour. Math. Sci. Vol. 78, pp. 311-333, 1996.
- [15] MIKEŠ, J., KIOSAK, V., VANŽUROVÁ A.: *Geodesic mappings of manifolds with affine connection*. Palacký University, Olomouc, 2008.
- [16] MIKEŠ, J., HINTERLEITNER, I., VANŽUROVÁ A.: *One remark on variational properties of geodesics in pseudo-Riemannian and generalized Finsler spaces*. IN: Ninth Int. Conf. On Geometry, Integrability and Quantization. June 8-13, 2007. Varna. SOFTEX, pp. 261-264, Sofia 2008.
- [17] MIKEŠ, J., HINTERLEITNER, I., VANŽUROVÁ A.: *Geodesic Mappings and Some Generalizations*. Palack University, Olomouc, Faculty of Science, 2009.
- [18] VANŽUROVÁ, A.: *Linear connections on two-manifolds and SODEs*. Proc. Conf. Aplimat 2007 (Bratislava, Slov. Rep.), Part II, p. 325-332, 2007.
- [19] VANŽUROVÁ, A.: *Metrization of linear connections, holonomy groups and holonomy algebras*. Acta Physica Debrecina, Vol. 42, pp. 39-48, 2008.
- [20] VANŽUROVÁ, A.: *Metrization problem for linear connections and holonomy algebras*. Archivum Mathematicum (Brno), Vol. 44, pp. 339-348, 2008.
- [21] VANŽUROVÁ, A., ŽÁČKOVÁ, P.: *Metrization of linear connections*. Journal of Applied Mathematics (Bratislava, SR) Vol. II, No. I, pp. 151-163, 2009.
- [22] VANŽUROVÁ, A., ŽÁČKOVÁ, P.: *Metrization of linear connections*. In Aplimat 2009: 8th International Conference Proceedings, Fakulty of Mechanical Ingeneering Slovak University of Technology in Bratislava, pp. 453-464, 2009.
- [23] VANŽUROVÁ, A.: *Metrization of connections with regular curvature*. Archivum Mathematicum (Brno), Vol. 45, No. 4, pp. 325-333, 2009.
- [24] VANŽUROVÁ, A., ŽÁČKOVÁ, P.: *Metrizability of connections on two-manifolds*. ACTA UPO, Math. Vol. 48, 157-170, 2009.
- [25] VANŽUROVÁ, A., ŽÁČKOVÁ, P.: *The metrization problem for linear connections*. Proc. Internat. Conf. Presentation of Mathematics, ICPM09 Liberec, pp. 137-143, 2009.
- [26] VANŽUROVÁ, A.: *A note on variational and metrizable connections*. Acta Mathematicae Academiae Nyrenghziensis, Vol. 20, 2010, www.emis.de/journals.

Current address

VANŽUROVÁ Alena, Doc. RNDr. CSc.

Faculty of Science, Dept. of Algebra and Geometry, Palacký University,

Tř. 17. listopadu 12,

771 46 Olomouc,

e-mail: alena.vanzurova@upol.cz;

Department of Mathematics,

Faculty of Civil Engineering,

Brno University of Technology,

Veveří 331/95,

602 00 Brno, Czech Republic,

e-mail: vanzurova.a@fce.vutbr.cz.

PIRKLOVÁ Petra, Mgr.

Faculty of Science, Dept. of Algebra and Geometry,

Palacký University,

Tř. 17. listopadu 12,

771 46 Olomouc,

e-mail: petra.zacek@tul.cz

MULTIDIMENSIONAL RIEMANNIAN MANIFOLDS AS MINKOWSKI PRODUCTS

VELICHOVÁ Daniela, (SK)

Abstract. The aim of this paper is to bring ideas on utilisation of the set operation of Minkowski product of point sets in the n dimensional Euclidean space \mathbf{E}^n for $n > 3$ in modelling Riemannian manifolds. Some examples are introduced on modelling manifolds in \mathbf{E}^6 and their visualisations by means of orthographic views in the three-dimensional coordinate subspaces of the basic 6 dimensional space. The comparison of different 3D views of manifolds determined as Minkowski sum, difference and product of two curve segments defined by vector representations is provided on particular examples.

Key words. Minkowski set operations, modelling Riemannian manifolds, multidimensional visualisation, orthographic projections in more dimensional spaces

Mathematics Subject Classification: Primary 51N25, Secondary 53A056.

1 Minkowski product of Riemannian manifolds

Minkowski product of two point sets is a binary geometric operation defined on point sets in the n -dimensional Euclidean space \mathbf{E}^n , and it can be determined and interpreted in various ways, as e.g. in [2], [3]. Most commonly appearing definition is based on the concept of the outer vector product well defined in the associated vector space over the \mathbf{E}^n , see in [1], [2].

Let V^n be the associated vector space to the Euclidean space \mathbf{E}^n with the Cartesian orthogonal coordinate system defined by the direction unit vectors of the coordinate axes

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \mathbf{e}_2 = (0, 1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, 0, 0, \dots, 1). \quad (1.1)$$

Let the vectors $\mathbf{u}, \mathbf{v} \in V^n$ be given by their coordinates, while

$$\mathbf{u} = u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2 + \dots + u_n \mathbf{e}_n, \mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \dots + v_n \mathbf{e}_n. \quad (1.2)$$

Outer (wedge) product of vectors \mathbf{u} and \mathbf{v} is vector $\mathbf{u} \wedge \mathbf{v} \in V^p$ in the vector space associated to the space $\Lambda^2(\mathbf{E}^p)$ of dimension $p = n(n-1)/2$, which can be determined as follows

$$\begin{aligned}
\mathbf{u} \wedge \mathbf{v} &= (u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2 + \dots + u_n \mathbf{e}_n) \wedge (v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \dots + v_n \mathbf{e}_n) = \\
&= u_1 v_1 (\mathbf{e}_1 \wedge \mathbf{e}_1) + u_1 v_2 (\mathbf{e}_1 \wedge \mathbf{e}_2) + \dots + u_1 v_n (\mathbf{e}_1 \wedge \mathbf{e}_n) + \\
&+ u_2 v_1 (\mathbf{e}_2 \wedge \mathbf{e}_1) + u_2 v_2 (\mathbf{e}_2 \wedge \mathbf{e}_2) + \dots + u_2 v_n (\mathbf{e}_2 \wedge \mathbf{e}_n) + \dots + \\
&+ u_n v_1 (\mathbf{e}_n \wedge \mathbf{e}_1) + u_n v_2 (\mathbf{e}_n \wedge \mathbf{e}_2) + \dots + u_n v_n (\mathbf{e}_n \wedge \mathbf{e}_n) = \\
&= (u_1 v_2 - u_2 v_1) (\mathbf{e}_1 \wedge \mathbf{e}_2) + (u_1 v_3 - u_3 v_1) (\mathbf{e}_1 \wedge \mathbf{e}_3) + \dots + (u_1 v_n - u_n v_1) (\mathbf{e}_1 \wedge \mathbf{e}_n) + \\
&+ (u_2 v_3 - u_3 v_2) (\mathbf{e}_2 \wedge \mathbf{e}_3) + \dots + (u_2 v_n - u_n v_2) (\mathbf{e}_2 \wedge \mathbf{e}_n) + \dots + (u_{n-1} v_n - u_n v_{n-1}) (\mathbf{e}_{n-1} \wedge \mathbf{e}_n)
\end{aligned} \tag{1.3}$$

whereas $\{\mathbf{e}_1 \wedge \mathbf{e}_2, \mathbf{e}_1 \wedge \mathbf{e}_3, \dots, \mathbf{e}_1 \wedge \mathbf{e}_n, \mathbf{e}_2 \wedge \mathbf{e}_3, \dots, \mathbf{e}_2 \wedge \mathbf{e}_n, \dots, \mathbf{e}_{n-1} \wedge \mathbf{e}_n\}$ is the basis of the associated vector space V^p . The following basic properties of the outer vector product were used to determine the resulting vector:

$$1. \mathbf{u} \wedge \mathbf{u} = 0 \quad (1.4)$$

$$2. k(\mathbf{u} \wedge \mathbf{w}) + l(\mathbf{u} \wedge \mathbf{w}) = (k + l)(\mathbf{u} \wedge \mathbf{w}) \quad (1.5)$$

$$3. \mathbf{u} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{u} \quad (1.6)$$

4. Outer product of two unit vectors is again a unit vector,

$$\|\mathbf{e}_i \wedge \mathbf{e}_j\| = 1, i, j = 1, 2, \dots, n, i \neq j. \quad (1.7)$$

Considering the two point sets in the form of Riemannian manifolds A and B determined by vector maps defined on the simply connected elementary regions in \mathbf{R}^n , the following definition of a Minkowski product of A and B can be adopted (see also in [4], [5]).

Definition 1. Let A and B be two Riemannian manifolds of dimensions k and l respectively, $k, l \leq n$, in the n -dimensional Euclidean space \mathbf{E}^n with the related vector space V^n and Cartesian orthogonal coordinate system $\langle 0; \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n \rangle$ defined by the origin and direction unit vectors of the coordinate axes. Let the vector maps of the respective point sets A and B be in the form

$$\mathbf{a} = ({}^1xa(u_i), {}^2xa(u_i), \dots, {}^nxa(u_i)) = {}^1xa(u_i)\mathbf{e}_1 + {}^2xa(u_i)\mathbf{e}_2 + \dots + {}^nxa(u_i)\mathbf{e}_n, i = 1, 2, \dots, k \quad (1.8)$$

$$\mathbf{b} = ({}^1xb(v_j), {}^2xb(v_j), \dots, {}^nxb(v_j)) = {}^1xb(v_j)\mathbf{e}_1 + {}^2xb(v_j)\mathbf{e}_2 + \dots + {}^nxb(v_j)\mathbf{e}_n, j = 1, 2, \dots, l \quad (1.9)$$

defined on the regions $\Omega \subset R^k, \Pi \subset R^l$. Minkowski product of point sets A and B in the Euclidean space \mathbf{E}^n is the Riemannian manifold $A \otimes B$ in the space \mathbf{E}^p of dimension $p = n(n - 1)/2$, which is determined as the outer product of vector maps \mathbf{a} and \mathbf{b} of sets A and B

$$\mathbf{a} \wedge \mathbf{b} = \begin{pmatrix} {}^1x(u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_j) \\ {}^2x(u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_j) \\ \dots \\ x(u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_j) \end{pmatrix}^T = \begin{pmatrix} {}^1xa(u_i){}^2xb(v_j) - {}^2xa(u_i){}^1xb(v_j) \\ {}^1xa(u_i){}^3xb(v_j) - {}^3xa(u_i){}^1xb(v_j) \\ \dots \\ {}^{n-1}xa(u_i){}^nxb(v_j) - {}^nxa(u_i){}^{n-1}xb(v_j) \end{pmatrix}^T \tag{1.10}$$

for $i = 1, 2, \dots, k, j = 1, 2, \dots, l$ and defined on the region $\Omega \times \Pi \subset R^{k+l}$.

2 Minkowski product of two curve segments in E^4

Let k and h be two curve segments in E^4 defined by their vector representations parameterized on the unit intervals

$$\mathbf{r}(u) = (xr(u), yr(u), zr(u), wr(u)), u \in \langle 0, 1 \rangle \subset R \quad (2.1)$$

$$\mathbf{s}(v) = (xs(v), ys(v), zs(v), ws(v)), v \in \langle 0, 1 \rangle \subset R. \quad (2.2)$$

Minkowski product of curves k and h according to [4] is a surface patch χ^\otimes in the space E^6 with the Cartesian coordinate system determined by unit vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_6$ defined on the unit square in R by the vector function defined in the following form

$$\mathbf{p}^\otimes(u, v) = \mathbf{r}(u) \wedge \mathbf{s}(v) = \begin{pmatrix} xr(u)ys(v) - yr(u)xs(v) \\ xr(u)zs(v) - zr(u)xs(v) \\ xr(u)ws(v) - wr(u)xs(v) \\ yr(u)zs(v) - zr(u)ys(v) \\ yr(u)ws(v) - wr(u)ys(v) \\ zr(u)ws(v) - wr(u)zs(v) \end{pmatrix}^T, (u, v) \in \langle 0, 1 \rangle^2 \quad (2.3)$$

Using consecutive orthographic projections to subspaces E^5, E^4 and E^3 we can achieve visualisation of the surface view in one from twenty possible three-dimensional coordinate subspaces $O\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k$, for $i, j, k = 1, \dots, 6$.

Let two segments of conical helices be determined in E^4 by their vector representations

$$\mathbf{r}(u) = (a_1(1-u)\cos 2\pi u, a_1(1-u)\sin 2\pi u, 0, b_1u), u \in \langle 0, 1 \rangle, a_1, b_1 \in R \quad (2.4)$$

$$\mathbf{s}(v) = (0, a_2(1-v)\cos 2\pi v, a_2(1-v)\sin 2\pi v, b_2v), v \in \langle 0, 1 \rangle, a_2, b_2 \in R \quad (2.5)$$

Minkowski product of these curve segments – one-dimensional Riemannian manifolds in E^4 is a surface patch – two-dimensional Riemannian manifold in the 6 dimensional space E^6 determined by vector representation

$$\mathbf{p}^\otimes(u, v) = \begin{pmatrix} a_1a_2(1-u)(1-v)\cos 2\pi u \cos 2\pi v \\ a_1a_2(1-u)(1-v)\cos 2\pi u \sin 2\pi v \\ a_1b_2v(1-u)\cos 2\pi u \\ a_1a_2(1-u)(1-v)\sin 2\pi u \sin 2\pi v \\ a_1b_2v(1-u)\sin 2\pi u - b_1a_2u(1-v)\cos 2\pi v \\ b_1a_2u(1-v)\sin 2\pi v \end{pmatrix}^T, (u, v) \in \langle 0, 1 \rangle^2 \quad (2.6)$$

Orthographic views of this surface patch onto several 3-dimensional coordinate subspaces of E^6 are illustrated in the Fig. 1, while these views can be regarded as special surfaces in E^6 created by means of a special modelling tool, Minkowski product of 2 curve segments.

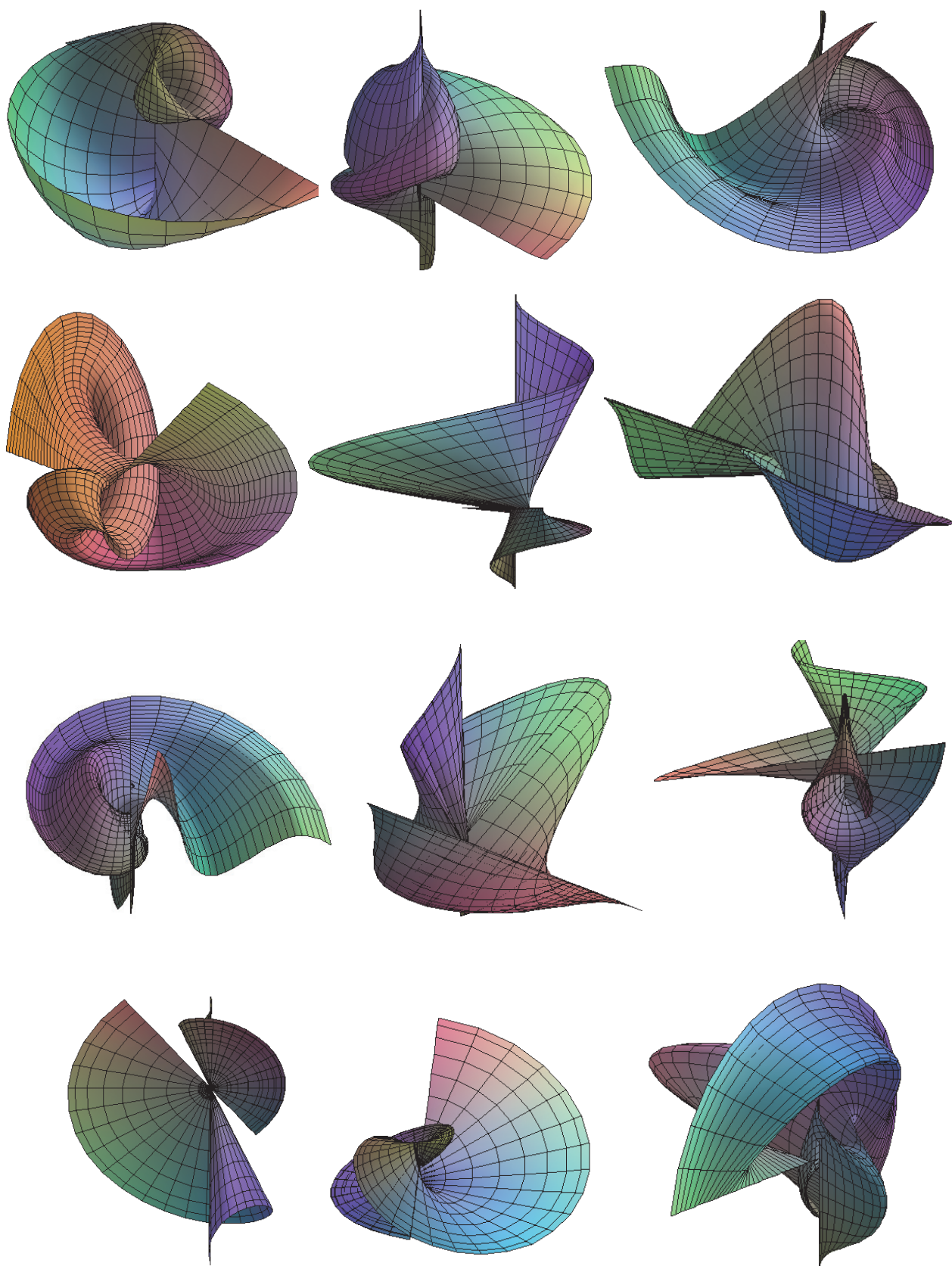


Fig. 1. Orthographic views of Minkowski product of 2 conical helices in E^6

Considering two segments of conical and cylindrical helices determined in \mathbf{E}^4 by their vector representations

$$\mathbf{r}(u) = (a_1 \cos 2\pi u, a_1 \sin 2\pi u, 0, b_1 u), u \in \langle 0, 1 \rangle \quad (2.7)$$

$$\mathbf{s}(v) = (0, a_2(1-v)\cos 2\pi v, a_2(1-v)\sin 2\pi v, b_2(1-v)), v \in \langle 0, 1 \rangle \quad (2.8)$$

Minkowski product of these one dimensional Riemannian manifolds is a surface patch in \mathbf{E}^6 determined by vector representation

$$\mathbf{p}^{\otimes}(u, v) = \begin{pmatrix} a_1 a_2 (1-v) \cos 2\pi u \cos 2\pi v \\ a_1 a_2 (1-v) \cos 2\pi u \sin 2\pi v \\ a_1 b_2 (1-v) \\ a_1 a_2 (1-v) \sin 2\pi u \sin 2\pi v \\ a_1 b_2 (1-v) \sin 2\pi u - b_1 a_2 u (1-v) \cos 2\pi v \\ b_1 a_2 u (1-v) \sin 2\pi v \end{pmatrix}^T, (u, v) \in \langle 0, 1 \rangle^2 \quad (2.9)$$

Orthographic views of this surface into several 3-dimensional coordinate subspaces of \mathbf{E}^6 are illustrated in the Fig. 2.

In the last example illustrated in the Fig. 3 we consider simple Euler orbit (that is a spherical curve generated as trajectory of a point revolving simultaneously about 3 intersecting orthogonal axes in the space) and line segment determined in \mathbf{E}^4 by vector representations

$$\mathbf{r}(u) = a_1 (\cos^2 2\pi u, \sin 2\pi u \cos 2\pi u (1 + \sin 2\pi u), \sin 2\pi u (\sin 2\pi u - \cos^2 2\pi u), 0), u \in \langle 0, 1 \rangle \quad (2.10)$$

$$\mathbf{s}(v) = (a_2(1-v), b_2 v, c_2 v, d_2(1-v)), v \in \langle 0, 1 \rangle \quad (2.11)$$

whose Minkowski product is a surface patch in \mathbf{E}^6 determined on unit square by vector representation

$$\mathbf{p}^{\otimes}(u, v) = \begin{pmatrix} a_1 b_2 v \cos^2 2\pi u - a_1 a_2 (1-v) \sin 2\pi u \cos 2\pi u (1 + \sin 2\pi u) \\ a_1 c_2 v \cos^2 2\pi u - a_1 a_2 (1-v) \sin 2\pi u (\sin 2\pi u - \cos^2 2\pi u) \\ a_1 d_2 (1-v) \cos^2 2\pi u \\ a_1 c_2 v \sin 2\pi u \cos 2\pi u (1 + \sin 2\pi u) - a_1 b_2 v \sin 2\pi u (\sin 2\pi u - \cos^2 2\pi u) \\ a_1 d_2 (1-v) \sin 2\pi u \cos 2\pi u (1 + \sin 2\pi u) \\ a_1 d_2 (1-v) \sin 2\pi u (\sin 2\pi u - \cos^2 2\pi u) \end{pmatrix}^T \quad (2.12)$$

Orthographic views of generated surfaces into the space \mathbf{E}^3 present unusually interesting forms and structures, which might be considered interesting for design and architectural purposes as new forms generated on the idea of an abstract algebra set operation with the additional value of being views of object from more dimensional spaces.

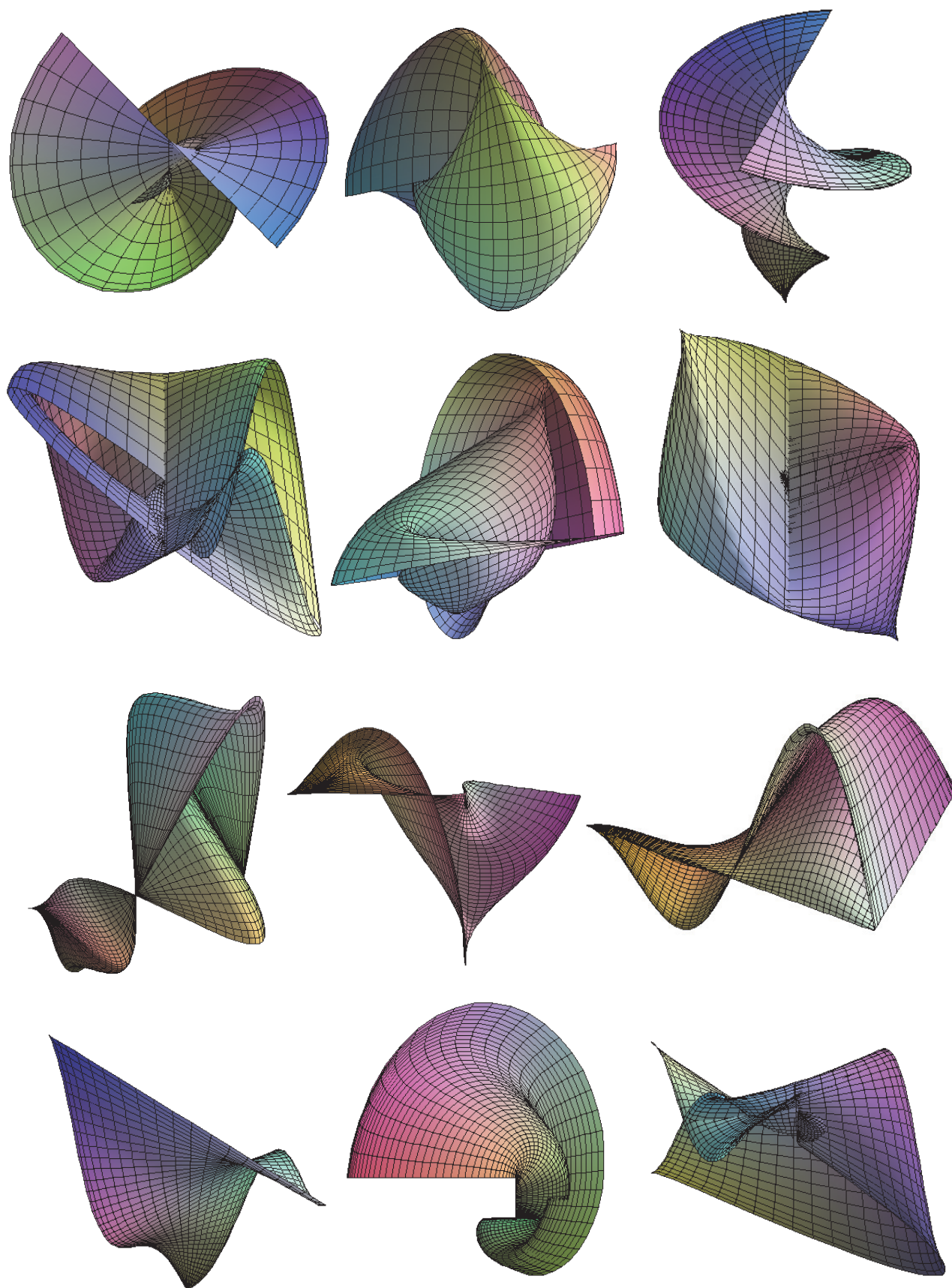


Fig. 2. Orthographic views of Minkowski product of conical and cylindrical helices in E^6

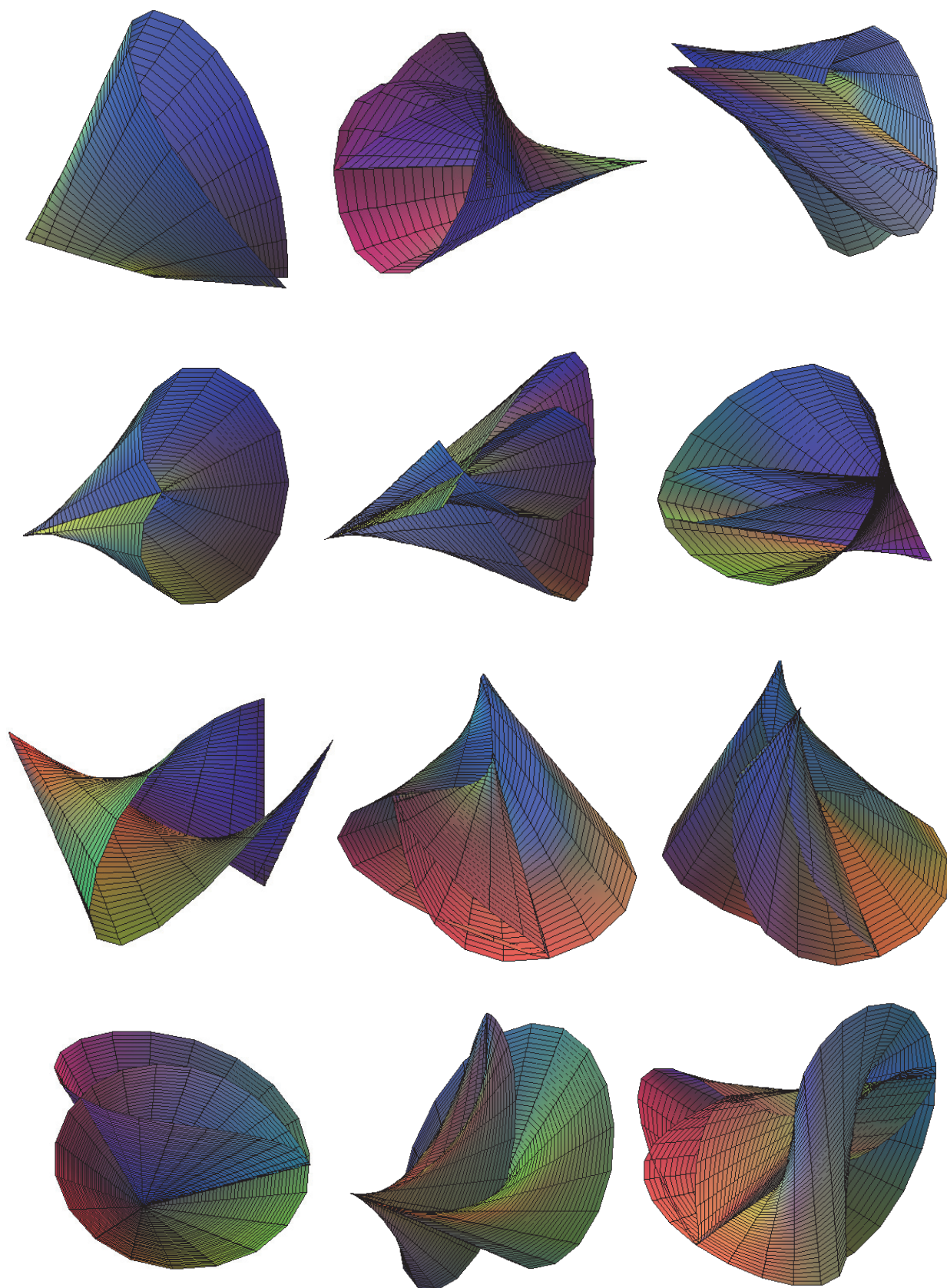


Fig. 3. Orthographic views of Minkowski product of Euler curve and line segment in E^6

4 Conclusions

Minkowski set operations are interesting geometric tools suitable for modelling of surfaces in the three dimensional Euclidean space, which are characteristic by unusual forms, complex shapes and which posses many self-intersections and singularities. These modelling tools provide a large scale of possibilities for creative applications in graphical and visual design, in different areas of computer graphics and in architectonical design. Intrinsic geometric properties of surfaces generated by means of these relatively simple generating principles can be exactly calculated by means of well-known methods in differential geometry. Last but not least important and interesting feature of this modelling tool is the possibility to visualise partially manifolds from the higher dimensional spaces by their orthographic views into 3D subspaces, which gives an opportunity to realise and better understand the geometry of manifolds in the higher dimensions.

Acknowledgement

The paper was supported by grant VEGA no. 1/0230/11.

References

- [1.] MacLANE, S.; BIRKHOFF, G.: *Algebra*. AMS Chelsea, 1999, ISBN 0-8218-1646-2.
- [2.] SMUKLER, M.: *Geometry, Topology and aplications of the Minkowski Product and Action*. Harvey Mudd College. Senior thesis, 2003.
- [3.] TOMIČKOVÁ, S.: *Minkowského operace a jejich aplikace*. Plzeň, ZČU 2006.
- [4.] VELICHOVÁ, D. *Minkowski product in surface modelling*. In Proceedings of Symposium on Computer Geometry SCG'2009. Vol. 18, 2009, STU Bratislava. pp.107-112.
- [5.] VELICHOVÁ D.: *Minkowski product in Surface Modelling*. In Aplimat - Journal of Applied Mathematics, Vol. 3, No.1/2010, pp. 277-286, 2010.

Current address

Daniela Velichová, doc. RNDr. CSc., mim. prof.

Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, Nám. slobody 17, 8132 31 Bratislava, Slovakia, tel. +4212 5729 6115, e-mail: daniela.velichova@stuba.sk

FINSLERIAN CONNECTIONS AND THE EQUATIONS OF SPINNING CHARGED PARTICLES IN GENERAL RELATIVITY

VOICU Nicoleta, (RO)

Abstract. In several previous articles, we proposed a unified description of the main equations of gravity and electromagnetism in terms of a 1-parameter family of Finslerian connections on the tangent bundle of the (4-dimensional) space-time manifold. In the present paper, we apply this construction in the description of the worldline equations of spinning charged particles in General Relativity.

Key words and phrases. tangent bundle, spray, Ehresmann connection, Einstein-Maxwell equations, Dixon-Souriau equations.

Mathematics Subject Classification. **MSC 2000:** 53Z05, 53B05, 53B40, 53C60, 83C22

1 Introduction

In [11], [12], we built a 1-parameter family of affine connections $\overset{\alpha}{D}$ on the space-time tangent bundle with the following properties:

- 1) worldlines of charged particles subject to gravitational and electromagnetic fields are autoparallel curves;
- 2) deviation equations of the above worldlines have the simplest form;
- 3) for a conveniently chosen α , the Ricci tensor of $\overset{\alpha}{D}$ is dynamically equivalent to the Lagrangian which provides the usual Einstein-Maxwell equations;
- 4) Maxwell equations can be expressed directly in terms of tidal tensors attached to these connections.

This description starts from two ideas. One is the use of tidal tensors, [4], in describing Einstein and Maxwell equations and the other is the idea proposed by Miron and collaborators, [7], [9], of encoding information regarding gravity in a Lorentzian metric on the (4-dimensional)

base manifold and the information regarding the electromagnetic field, in connections on its tangent bundle. Thus, we included the action $F_{ij}F^{ij}$ for the electromagnetic field in the Ricci tensor, without adding any supplementary dimensions to the space-time manifold (as in the case of Kaluza-Klein theory).

In the present work, we present in brief these geometric structures and their use, [11], [12], in the description of Einstein and Maxwell equations and afterwards, we express the Dixon-Souriau equations governing the motion of spinning charged particles under the action of gravitational and electromagnetic fields, in terms of these connections. The spin precession equations achieve a very simple form.

2 Geometric structures

Consider a 4-dimensional Lorentzian manifold (M, g) , with signature $(+, -, -, -)$, regarded as space-time manifold, with local coordinates $x = (x^i)_{i=0,3}$ and Levi-Civita connection ∇ (having the coefficients γ^i_{jk} and curvature tensor r); on the tangent bundle (TM, π, M) , we denote the local coordinates by $(x \circ \pi, y) =: (x^i, y^i)_{i=0,3}$ and by \cdot_i and \cdot_i , partial differentiation with respect to x^i and y^i respectively. Throughout the paper, we will assume that light speed in vacuum c and the gravitational constant G are both equal to 1.

A Finslerian tensor field on TM , [2], is a tensor field on TM , whose local components transform, with respect to coordinate changes on TM , by the same rule as the components of a tensor field on M (ex.: $X^{i'}_{j'}(x', y') = \frac{\partial x^{i'}}{\partial x^k} \frac{\partial x^l}{\partial x^{j'}} X^k_l(x, y)$).

For an Ehresmann connection N on TM , [7], we denote the adapted basis by

$$(\delta_i = \frac{\partial}{\partial x^i} - N^j_i(x, y) \frac{\partial}{\partial y^j}, \quad \dot{\delta}_i = \frac{\partial}{\partial y^i}), \quad (1)$$

and its dual, by $(dx^i, \delta y^i = dy^i + N^i_j dx^j)$. Any vector field $X = X^i \delta_i + \tilde{X}^i \dot{\delta}_i$ on TM is decomposed into a horizontal component $hX = X^i \delta_i$ and a vertical component $vX = \tilde{X}^i \dot{\delta}_i$, which are both Finslerian tensor fields.

We will focus in the following upon a 1-parameter family of Lagrangians depending on α :

$$\overset{\alpha}{L} = \sqrt{g_{ij}(x) \dot{x}^i \dot{x}^j} + \alpha A_i \dot{x}^i; \quad (2)$$

where $A = A_i(x) dx^i$ is a 1-form on M and $\alpha \in \mathbb{R}$ is a parameter.

Extremal curves $x = x(t)$ (where $t = \text{const} \cdot s$ is proportional to the arclength) for the action $\int \overset{\alpha}{L} dt$ are given by:

$$\frac{dy^i}{dt} + \gamma^i_{jk} y^j y^k - \alpha \|y\| F^i_j y^j = 0, \quad y^i = \dot{x}^i, \quad (3)$$

with

$$F^i_j := g^{ih}(A_{j,h} - A_{h,j}), \quad \|y\| = \sqrt{g_{ij} y^i y^j}; \quad (4)$$

The functions:

$$2\overset{\alpha}{G}^i(x, y) = \gamma^i_{jk} y^j y^k + 2\overset{\alpha}{B}^i, \quad 2\overset{\alpha}{B}^i = -\alpha \|y\| F^i_j y^j, \quad (5)$$

(with $\alpha \in \mathbb{R}$) have the property that their y -derivatives

$$\overset{\alpha}{G}{}^i{}_j := \overset{\alpha}{G}{}^i{}_{\cdot j} = \gamma^i{}_{jk} y^j y^k + \overset{\alpha}{B}{}^i{}_{\cdot j};$$

are the coefficients of an Ehresmann connection¹ $\overset{\alpha}{N}$; extremal curves of $\overset{\alpha}{L}$, (3), are autoparallel curves for the connection $\overset{\alpha}{N}$, i.e.,

$$\frac{dy^i}{dt} + \overset{\alpha}{N}{}^i{}_j(x, y) y^j = 0, \quad y = \frac{dx}{dt}.$$

We denote: $F^i = F^i{}_j y^j$, i.e., $2\overset{\alpha}{B}{}^i := -\alpha \|y\| F^i$. If there is no risk of confusion, we will omit the parameter α in the notation of the connections $\overset{\alpha}{N}$ and of the related quantities, i.e., we will write simply $G^i, B^i, G^i{}_j, B^i{}_{\cdot j}, \delta_i \dots$ instead of $\overset{\alpha}{G}{}^i, \overset{\alpha}{B}{}^i, \overset{\alpha}{G}{}^i{}_j, \overset{\alpha}{B}{}^i{}_{\cdot j}, \overset{\alpha}{\delta}_i$ etc.

Further, we define the affine connections $D := \overset{\alpha}{D}$ (the *Berwald connections attached to $\overset{\alpha}{N}$*) given in the adapted basis to $N = \overset{\alpha}{N}$, by:

$$D_{\delta_k} \delta_j = G^i{}_{jk} \delta_i, \quad D_{\delta_k} \dot{\delta}_j = G^i{}_{jk} \dot{\delta}_i, \quad D_{\dot{\delta}_k} \delta_j = 0, \quad D_{\dot{\delta}_k} \dot{\delta}_j = 0, \quad D_{\dot{\delta}_k} \ddot{\delta}_j = 0,$$

where:

$$G^i{}_{jk} = \gamma^i{}_{jk} + B^i{}_{\cdot jk}.$$

The functions B^i are the components of a Finslerian vector field $B = B^i \delta_i$ on TM . With $l_i := \frac{y_i}{\|y\|} = \frac{g_{ij} y^j}{\|y\|}$, we have:

$$\begin{aligned} B^i{}_j &= B^i{}_{\cdot j} = -\frac{\alpha}{2} (F^i l_j + \|y\| F^i{}_j), \\ B^i{}_{jk} &:= B^i{}_{\cdot jk} = -\frac{\alpha}{2} (l_{jk} F^i + l_j F^i{}_k + l_k F^i{}_j). \end{aligned} \quad (6)$$

From the homogeneity of degree 2 of G^i and B^i in y , it follows:

$$G^i{}_j y^j = 2G^i, \quad G^i{}_{jk} y^k = G^i{}_j, \quad B^i{}_j y^j = 2B^i, \quad B^i{}_{jk} y^k = B^i{}_j.$$

Connections D are generally non-metrical. They also have generally nonvanishing torsion, given by:

$$\mathbb{T} = R^i{}_{jk} \dot{\delta}_i \otimes dx^j \otimes dx^k, \quad R^i{}_{jk} = \delta_k N^i{}_j - \delta_j N^i{}_k. \quad (7)$$

The curvature of D is:

$$\mathbb{R} = R^i{}_{kl} \delta_i \otimes dx^j \otimes dx^k \otimes dx^l + R^i{}_{kl} \dot{\delta}_i \otimes \delta y^j \otimes dx^k \otimes dx^l + \quad (8)$$

$$+ B^i{}_{\cdot jkl} \delta_i \otimes dx^j \otimes dx^k \otimes \delta y^l, \quad (9)$$

where

$$R^i{}_{jkl} = R^i{}_{klj}. \quad (10)$$

¹ $\overset{\alpha}{N}$ is the *spray connection*, [1], [2], attached to the spray with coefficients $\overset{\alpha}{G}{}^i$.

Let us fix a curve $c : t \mapsto x^i(t)$ on the base manifold M and denote by $c' : t \mapsto (x^i(t), \dot{x}^i(t))$ its lift to the tangent bundle. For a Finsler vector field $X^i = X^i(t)$ along c' , it makes sense the *Berwald covariant derivative*, [1], [2]:

$$DX^i := D_{\frac{dc'}{dt}} X^i = \frac{dX^i}{dt} + G^i_j X^j. \quad (11)$$

In terms of (11), the equations of extremal curves of $L = \overset{\alpha}{L}$ become:

$$Dy^i = 0, \quad y^i = \dot{x}^i \quad (12)$$

and geodesic deviation equations are written, [11], as $D^2w = E(w)$, or locally:

$$D^2w^i = E^i_j(x, y)w^j, \quad y = \dot{x}, \quad (13)$$

where $w = w^i(t)\delta_i$ is the horizontal lift to TM of the deviation vector field and

$$E^i_j = R^i_{jk}y^k = R^i_{jl}y^l y^k \quad (14)$$

define the *tidal tensor* $E = E^i_j \delta_i \otimes dx^j$ attached to the connection N .

3 Basic equations of gravitational and electromagnetic fields

In the following, we will assume that g_{ij} describes the gravitational field and $A = A_i dx^i$ is physically interpreted as the 4-potential of the electromagnetic field. We will denote the horizontal lifts to TM of the differential forms A and $F := dA = \frac{1}{2}F_{ij}dx^i \wedge dx^j$ by the same letters A and F . Unless elsewhere specified, the parameter $\alpha \neq 0$ is considered arbitrary.

The components of the electromagnetic 2-form F are expressed, [11], as:

$$\frac{\alpha}{2}F_{ij} = D_{\delta_j}l_i. \quad (15)$$

Consider the *angular metric*, $h = h_{ij}(x, y)dx^i \otimes dx^j$, given by:

$$h_{ij} = g_{ij} - l_i l_j.$$

Then:

- *Homogeneous Maxwell equations* $\nabla_{\partial_i}F_{jk} + \nabla_{\partial_k}F_{ij} + \nabla_{\partial_j}F_{ki} = 0$, [8], are written in terms of tidal tensors, [11], as:

$$\tilde{E}_{[ij]} = 0, \quad (16)$$

where $\tilde{E}_{ij} = h_{ik}E^k_j$ and square brackets denote antisymmetrization.

- *Inhomogeneous Maxwell equations* $\nabla_{\partial_i}F^{ij} = 4\pi J^i$, are expressed as:

$$E^i_i = e^i_i - 4\pi\alpha\rho_c \|y\|^2 + B^l_i B^i_l, \quad (17)$$

where $\rho_c = -J^i l_i$ and $e^i_i = \overset{0}{E^i_i}$.

- The Ricci tensor of $D = \overset{\alpha}{D}$ is

$$R_{ij} = -\frac{1}{2}(E^l_l)_{.ij}.$$

We have proved, [12], that for $\frac{3\alpha^2}{2} = 1$, the Ricci scalar $R = g^{ij}\overset{\alpha}{R}_{ij}$ is dynamically equivalent to:

$$\tilde{R} = r + F_{ij}F^{ij}$$

that is, to the Lagrangian (on M) which leads to the usual Einstein-Maxwell equations, [8].

Here, $\tilde{R} = R_{ij} - \overset{0}{D}_{\delta_k}(B^k_{ij})$ and $\tilde{R} = g^{ij}\tilde{R}_{ij}$. Then, *Einstein field equations*

$$r_{ij} - \frac{1}{2}rg_{ij} = 8\pi(T^{em}_{ij} + T^m_{ij})$$

(where T^{em}_{ij} is the stress-energy tensor of the electromagnetic field and T^m_{ij} , the stress-energy tensor of matter) are equivalent to:

$$\mathcal{G}_{ij} = 8\pi T^m_{ij}, \quad (18)$$

where

$$\mathcal{G}_{ij} = \tilde{R}_{ij} - \frac{1}{2}\tilde{R}g_{ij} + \mathcal{B}_{.ij}$$

and $\mathcal{B} := \frac{3}{2}\frac{B^l B_l}{\|y\|^2} + \frac{1}{2}B^i_h B^h_i$ (the term $\mathcal{B}_{.ij}$ comes from the non-metricity of D); thus, the

electromagnetic stress-energy tensor T^{em}_{ij} is included in the generalized Einstein tensor \mathcal{G}_{ij} .

- *Equations of motion of a (non-spinning) charged particle*, [8], are (3):

$$\overset{\alpha}{D}y^i = 0, \quad y = \dot{x}, \quad (19)$$

where $\alpha = \frac{q}{m}$. For particles having the same ratio $\frac{q}{m}$, worldline deviation equations are:

$$\overset{\alpha}{D}^2 w^i = E^i_j w^j, \quad \alpha = \frac{q}{m}. \quad (20)$$

4 Equations of motion of spinning charged particles

In General Relativity, worldlines of particles subjected only to the gravitational field, are geodesics of the Levi-Civita connection of the space-time manifold (M, g) ; accordingly, equations of motion of charged particles under the action of the gravitational and electromagnetic fields, are (3) (where it is usually considered that $t = s$, i.e., $\|\dot{x}\| = 1$). But, if the considered particles are also spinning, then their equations of motion become more complicated. More in detail:

1) Worldlines of spinning particles subject to the gravitational field are given by the *Mathisson-Papapetrou equations*, [3], [6],[10]:

$$\frac{\nabla p^i}{ds} = \frac{1}{2}r^i_{jkl}\dot{x}^j S^{kl}, \quad (21)$$

$$\frac{\nabla S^{ij}}{ds} = p^i \dot{x}^j - p^j \dot{x}^i, \quad (22)$$

where $r_j^i{}_{kl} = \gamma^i_{jk,l} - \gamma^i_{jl,k} + \gamma^h_{jk}\gamma^i_{hl} - \gamma^h_{jl}\gamma^i_{hk}$, p^i are the components of the 4-momentum of the particle and S^{ij} is the *spin tensor*. It is noteworthy that, if the particle is spinning, the 4-momentum and the 4-velocity of the moving object are generally non-collinear. From (22), it follows that $\frac{\nabla S^{ij}}{ds}\dot{x}_j = p^i - m\dot{x}^i$, where $m := p_j\dot{x}^j$, or, [10]:

$$p^i = m\dot{x}^i + \frac{\nabla S^{ij}}{ds}\dot{x}_j. \quad (23)$$

In order that the system be closed, some supplementary conditions are needed. Usually, one assumes either $S^{ij}\dot{x}_j = 0$ (the *Pirani condition*) or $S^{ij}p_j = 0$ (the *Tulczyjew, or Dixon condition*).

2) Worldlines of spinning particles subject to both gravitational and electromagnetic fields are given by the *Dixon-Souriau equations*, [3]:

$$\frac{\nabla p^i}{ds} = \frac{1}{2}r_j^i{}_{kl}\dot{x}^j S^{kl} + qF^i{}_j\dot{x}^j + \frac{\mathbf{k}}{2}S^{kl}\nabla^i F_{kl}, \quad (24)$$

$$\frac{\nabla S^{ij}}{ds} = p^i\dot{x}^j - p^j\dot{x}^i - \mathbf{k}(S^{ik}F_k{}^j - S^{jk}F_k{}^i), \quad (25)$$

where \mathbf{k} is a constant.

In the following, we will adopt the Pirani condition:

$$S^{ij}\dot{x}_j = 0. \quad (26)$$

Let us assume that the trajectories are parametrized by the arc length $t = s$; thus, setting $y^i := \dot{x}^i$, we will also have $\dot{x}^i = l^i$. The horizontal lifts:

$$p := p^i\delta_i, \quad S := \frac{1}{2}S_{ij}dx^i \wedge dx^j$$

of p and S are Finslerian tensors, for which we can write:

$$\frac{\nabla p^i}{ds} = \overset{0}{D}p^i, \quad \frac{\nabla S^{ij}}{ds} = \overset{0}{D}S^{ij}.$$

In terms of other connections $D = \overset{\alpha}{D}$ of the family, we will have:

$$Dp^i = \overset{0}{D}p^i + B^i{}_jp^j, \quad DS^{ij} = \overset{0}{D}S^{ij} + B^i{}_hS^{hj} + B^j{}_hS^{ih}.$$

Choosing

$$\frac{\alpha}{2} := \mathbf{k},$$

and under the condition (26), the *spin precession* equations (25) take the form:

$$DS^{ij} = p^iy^j - p^jy^i, \quad y^i = \frac{dx^i}{ds}; \quad (27)$$

i.e., they become formally similar to (22).

Contracting the above relation with S_{ij} , we find that $S_{ij}DS^{ij} = 0$, which implies the already known relation, [3], [5]:

$$S_{ij}S^{ij} = \text{const.}$$

Also, contracting (27) with $y_j = \dot{x}_j$, we find for the momentum p and the velocity \dot{x} a relation which is formally similar to (23):

$$p^i = my^i + y_j DS^{ij}, \quad y^i = \dot{x}^i. \quad (28)$$

In the following, we will determine the form of equations (24) in terms of D . In terms of $\gamma^i_{jk} = \overset{0}{G}^i_{jk}$ and B^i_j , the torsion R^i_{jk} , (7), is expressed as:

$$R^i_{jk} = y^l r_l{}^i{}_{jk} + \overset{0}{D}_k B^i_j - \overset{0}{D}_j B^i_k + B^l_j B^i_{kl} - B^l_k B^i_{jl}, \quad (29)$$

with $\overset{0}{D}_k = \overset{0}{D}_{\delta_k}$. Further, taking into account that, [11], $\overset{0}{D}_k l^i = 0$, the Pirani condition and (6), we get:

$$S^{jk} R^i_{jk} = S^{jk} \{y^l r_l{}^i{}_{jk} - \frac{\alpha}{2} (\overset{0}{D}_k F^i_j - \overset{0}{D}_j F^i_k) + M^i_{jk}\}, \quad (30)$$

where $M^i_{jk} = \frac{\alpha^2}{4} \{2F^i F_{kj} + F^i_{[j} F_{k]}\}$.

The derivatives $\overset{0}{D}_k F^i_j$ coincide with their Levi-Civita counterparts on M . Hence, from the Maxwell equations, we have: $\overset{0}{D}_k F^i_j - \overset{0}{D}_j F^i_k = -\overset{0}{D}^i F_{jk} = -\nabla^i F_{jk}$. We thus get:

$$S^{jk} (r_l{}^i{}_{jk} y^l + \frac{\alpha}{2} \nabla^i F_{jk}) = S^{jk} (R^i_{jk} - M^i_{jk}).$$

With $\frac{\alpha}{2} = \mathbf{k}$ as above, we recognize the terms in the right hand side of (24). We can thus write:

$$\overset{0}{D} p^i = \frac{1}{2} S^{jk} (R^i_{jk} - M^i_{jk}) + q F^i_j \dot{x}^j.$$

Passing to the D -covariant Dp^i , equations (24) take the form:

$$Dp^i = \frac{1}{2} S^{jk} (R^i_{jk} - M^i_{jk}) + q F^i_j y^j + B^i_h p^h. \quad (31)$$

Acknowledgement

The work was supported by the Sectorial Operational Program Human Resources Development (SOP HRD), financed from the European Social Fund and by Romanian Government under the Project number POSDRU/89/1.5/S/59323.

References

- [1] P.L. ANTONELLI, I. BUCATARU, *New results about the geometric invariants in KCC-theory*, An. St. Univ. "Al. I. Cuza", Iasi, XLVII (s. I) (2011) 405-414.

- [2] I. BUCATARU, O. CONSTANTINESCU, M. DAHL, A geometric setting for systems of ordinary differential equations, *Int. J. of Geom. Methods in Physics*, 8(6) (2011), 1291-1327.
- [3] F. CIANFRANI, I. MILILLO, G. MONTANI, Dixon-Souriau equations from a 5-dimensional spinning particle in a Kaluza-Klein framework, *Phys. Lett. A*, 366 (7) (2007).
- [4] L.F.O. COSTA, C.A.R. HERDEIRO, Gravitoelectromagnetic analogy based on tidal tensors, *Phys. Rev. D* 78(2) (2008), id: 024021.
- [5] M. HEYDARI-FARD, M. MOHSENI, H. R. SEPANGI, *Worldline deviations of charged spinning particles*, arXiv: gr-qc/0509111 (2011).
- [6] O.B. KARPOV, The Papapetrou equations and supplementary conditions, arXiv:gr-qc/0406002v2 (2004).
- [7] R. MIRON, M. ANASTASIEI, *The Geometry of Lagrange Spaces: Theory and Applications*, Kluwer Acad. Publ., Dordrecht, Boston, London, 1994.
- [8] L.D. LANDAU, E.M. LIFSCHIZ, *The Classical Theory of Fields*, 4th edn. (Elsevier, 1975).
- [9] R. MIRON, The geometry of Ingarden spaces, *Rep. on Math. Phys.* 54(2) (2004) 131-147.
- [10] R M PLYATSKO, O B STEFANYSHYN, M T FENYK, Mathisson-Papapetrou-Dixon equations in the Schwarzschild and Kerr backgrounds, *Class. Quantum Grav.* 28 195025 (2011).
- [11] N. VOICU, Tidal tensors in the description of gravity and electromagnetism, *arXiv:1111.1435v2 [math-ph]*, 2011.
- [12] N. VOICU, On a new unified geometric description of gravity and electromagnetism, *arXiv:1111.5270v1 [math-ph]*, 2011.

Current address

Voicu Nicoleta Dr.

Faculty of Mathematics and Computer Science,
"Transilvania" University,
Str. Iuliu Maniu, nr. 50
500091 Brasov, Romania
e-mail: nico.brinzei@unitbv.ro

CRIME SCENE INVESTIGATION THROUGH DNA TRACES USING BAYESIAN NETWORKS

ANDRADE Marina, (PT), FERREIRA Manuel Alberto M., (PT)

Abstract. The use of biological information in crime scene identification problems is becoming more and more common. In this work, a crime scene is presented and used to exemplify the use of Bayesian networks, to analyse the information contained in a mixture DNA trace, referring to a crime in which there are two victims' and two suspect's involved. It is also made some discussion about the hypotheses to be considered in order to make them compatible with the possible information extracted from the mixture trace, that constitutes the evidence.

Key words: Mixture traces, forensic identification, Bayesian networks, DNA profiling.

Mathematics Subject Classification: 62C10.

1 Introduction

A crime has been committed and two persons, V_1 and V_2 , were murdered. One mixture trace was found and S_1 and S_2 are potential suspects. Their DNA profiles were measured and considered to be compatible with the mixture trace.

Possibly a fight occurred during the assault and some material was produced. It is acceptable that the individuals who perpetrated the crime could have left some of their genetic material in the trace.

The crime scene is analysed in section 2. It will be presented the evidence, E , and explained the hypotheses to be considered.

In section 3 the Bayesian network – example of a probability expert system - built expressly to perform the necessary calculations is shown¹. Then are discussed the number, quite large, and the significance of the results that is possible to obtain.

¹ To perform the calculations it is mandatory to apply repeatedly the Bayes' Law. This leads to very complicated computations impossible to perform algebraically. So the use of a probabilistic expert system is needed.

In section 4 the numerical results will be seen. In section 5 a brief discussion is outlined.

With the geneticist Sewall Wright, in the beginning of the 20th century, began the use of networks transporting probabilities. In (Dawid *et al.*, 2002) it is described this new approach to problems of the kind of the one described above. The conception and use of Bayesian networks to analyse problems in forensic identification inference, was initially done there, followed by (Evetts *et al.*, 2002), (Mortera, 2003) and (Mortera *et al.*, 2003). The analysis of a crime scene analogous to the considered in this work, but with two victims' and one perpetrator and two mixture traces was presented in (Andrade and Ferreira, 2009). About this subject see (Andrade and Ferreira, 2011a).

2 Evidence and Hypotheses

To summarize the evidence it is presented in Table 1 the DNA profiles of the victims' and the suspect's, V_1 , V_2 , S_1 , S_2 , and the trace found at the crime scene, E .

	V_1	V_2	S_1	S_2	E
VWA	14,15	16,17	14,14	14,17	14,15,16,17
D21S11	29,31.2	28,28	30,31.2	28,32.2	28,29,30,31.2,32.2
SE33	19,30.2	17,30.2	17,18	16,19	16,17,18,19,30.2

Table 1: **Two Victims' and Two Suspect's DNA Profiles and Evidence.**

In Table 2 the allele frequencies, for each marker found in the trace, are presented.

	p_{14}	p_{15}	p_{16}	p_{17}	
VWA	0.1101	0.1197	0.1827	0.2753	
	p_{28}	p_{29}	p_{30}	$p_{31.2}$	$p_{32.2}$
D21S11	0.1674	0.2136	0.2437	0.1138	0.0894
	p_{16}	p_{17}	p_{18}	p_{19}	$p_{30.2}$
SE33	0.0590	0.0660	0.0833	0.0868	0.0140

Table 2: **Allele frequencies.**

The allele frequencies in Table 2 were collected in the database "The Distribution of Human DNA-PCR Polymorphisms".

The crime trace can contain DNA from up to four unknown contributors, in addition to the victims and/or the suspects.

If the DNA of S_i with $i = 1, 2$ is presented in the trace this will place him/her at the crime scene and consequently as one of the possible perpetrators.

The court has to determine if each suspect is or is not guilty. The hypotheses to be evaluated are:

H_1 : S_1 is a contributor to the trace but S_2 is not, given the evidence.

H_2 : S_2 is a contributor to the trace but S_1 is not, given the evidence.

H_3 : S_1 and S_2 are both contributors to the trace, given the evidence.

H_4 : Neither S_1 nor S_2 are contributors to the trace, given the evidence.

The respective events probabilities are called $p_{10}, p_{02}, p_{12}, p_{00}$, where 0 mentions the absence of the respective, in order, individual DNA in the trace. So:

If $p_{00} > p_{10} + p_{02} + p_{12}$ the two suspects are acquitted. If not it must be seen if $p_{12} > p_{10} + p_{02}$ case at which the two suspects are both placed at the crime scene. If not p_{10} must be compared with p_{02} . If $p_{10} > p_{02}$ the evidence favours the presence of S_1 at the crime scene and the acquaintance of S_2 . The contrary happens when $p_{02} > p_{10}$.

3 Bayesian Network for One Marker

The probabilities referred above will be computed using the Bayesian network of Figure 1.

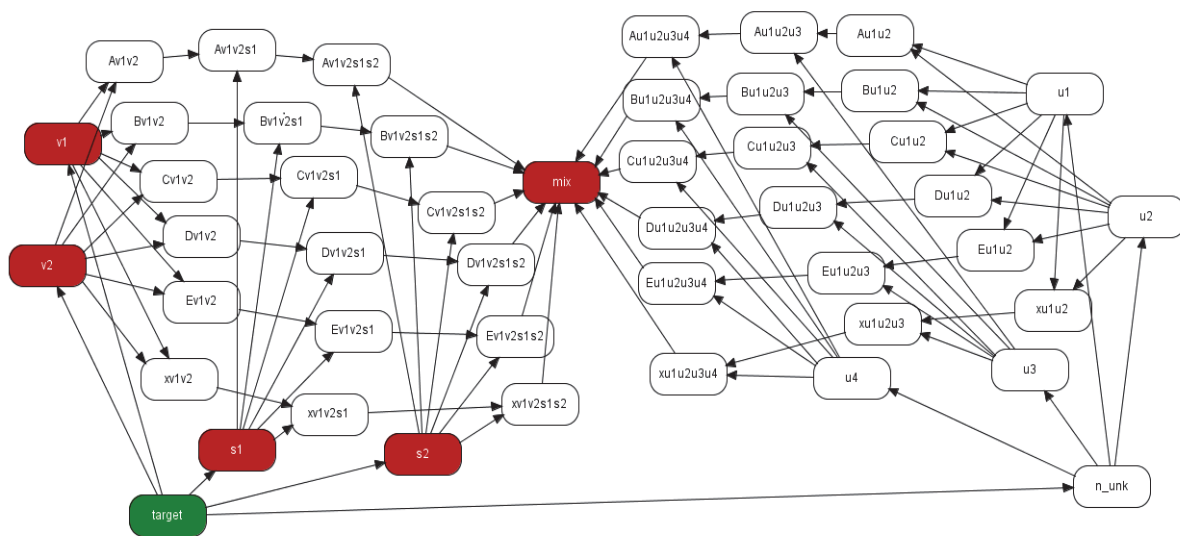


Figure 1: Marker network.

Nodes $v_i, i = 1, 2, s_j, j = 1, 2$ and $u_k, k = 1, 2, 3, 4$, in Figure 1 are themselves Bayesian networks that represent the genetic structure and inheritance of each individual - the victims, the suspects and the unknowns, respectively - and have all the same structure. The $v_i, i = 1, 2$ and $s_j, j = 1, 2$ constitute data of the problem. The nodes in white, at the left of the node mix, that represents the mixture and it is also comprised by known data (E), represent the relations in which the nodes $v_i, i = 1, 2$ or $s_j, j = 1, 2$ may contribute to the mixture. The nodes in white, at the right of the node mix, except the $u_k, k = 1, 2, 3, 4$ and n_unk - that is a counter for the number of unknowns in the mixture - represent the relations in which the $u_k, k = 1, 2, 3, 4$ may contribute to the mixture. Node target collects the states and the respective probabilities.

As it is mandatory to consider the possible contribution of till four unknown individuals to the mixture, the number of admissible states jumps to 80, numbered from 0 - no one in the mixture - to 79 - the two victims, the two suspects and the four unknowns are all in the mixture. Of course these two states are unrealistic and there are other ones also unrealistic because are incompatible with the minimum number of contributors to the mixture, according to the evidence inserted. These unrealistic states are discarded by the network but have to be considered conceptually in its building.

Among the realistic states only a few ones are interesting to the problem: the corresponding to the hypotheses events defined above. The network was implemented using Hugin² software³.

4 Results

For marker VWA, alleles 14, 15, 16, 17 are considered, Table 1. And so they are represented in the Figure 1 Bayesian network by A, B, C, D, respectively. E is considered with 0 frequency. When considering marker D21S11, the alleles are 28, 29, 30, 31.2 and 32.2 corresponding to A, B, C, D, E. In marker SE33 the alleles are 16, 17, 18, 19, 30.2 corresponding to A, B, C, D, E. In the whole cases x accumulates the remaining frequencies of the non considered alleles for each marker.

The results obtained are presented in Table 3. The values in line rescale are constituted by the ratios of the products of the values in the respective column⁴ by the total sum of the four products. The values in this line are the used ones in the tests described in section 2.

	p_{00}	p_{12}	p_{10}	p_{02}
VWA	0.2134	0.2699	0.2473	0.2699
D21S11	0.0752	0.5354	0.1335	0.2554
SE33	0.0091	0.8868	0.0454	0.0585
Rescale	0.0011	0.9576	0.0112	0.0301

Table 3: **Results.**

Following the procedure recommended in section 2 the conclusion is that both suspects are placed at the crime scene – note the great value of $p_{12} = 0.9576$. For marker VWA, alone, S_2 is placed at the crime scene but S_1 is not. But note that the probability is not very convincing: p_{02} is only slightly greater than p_{10} . For markers D21S11 and SE33, each of them alone, both suspects are placed together at the crime scene, in a much more convincing way for SE33.

5 Discussion

The problems and the difficulties posed in the interpretation and evaluation of DNA evidence are very well outlined, for instance, in (Andrade and Ferreira, 2011) and (Lauritzen, 2003). As in general they are stated in probabilistic terms leads to some confusion to the judges, when they have to issue a decision, because of its difficulty in interpreting the meaning of the measure of

² www.hugin.com

³ To compute the interesting probabilities there must be considered the following states probabilities:

- p_{00} : 1, 2, 3, 16, 17, 18, 19, 32, 33, 34, 35, 48, 49, 50, 51, 64, 65, 66 and 67,
- p_{12} : 12, 13, 14, 15, 28, 29, 30, 31, 44, 45, 46, 47, 60, 61, 62, 63, 76, 77, 78 and 79,
- p_{10} : 8, 9, 10, 11, 24, 25, 26, 27, 40, 41, 42, 43, 56, 57, 58, 59, 72, 73, 74 and 75,
- p_{02} : 4, 5, 6, 7, 20, 21, 22, 23, 36, 37, 38, 39, 52, 53, 54, 55, 68, 69, 70 and 71

from the output given by Hugin after the inserted evidence.

⁴ It is possible to multiply the respective probabilities, for each marker, because it is assumed independence between and across marker, i.e., linkage and Hardy-Weinberg Equilibrium (Andrade, 2007).

probability⁵. In this situation the Bayesian approach is the most clear to explain the significance of the evidence through the comparison of the hypotheses likelihood ratios see (Ferreira and Andrade, 2009). So the use of Bayesian networks to compute the interesting probabilities is a natural option, due to the intractable algebraic manipulation, when attempting to use successively the Bayes' Law in very complicated situations.

When the inference tool is the hypotheses tests, in these problems, they must be defined for each type of problem. This was exemplified with the four hypotheses suggested for this crime scene investigation. Only after this definition results clear which probabilities must be computed among a huge of possible ones.

Note finally that this methodology allows to conclude for the absolution of a suspect but not for the conviction. Eventually it only places the suspect in the crime scene. In this case further police work must be made.

References

- [1] ANDRADE, M.: *A Estatística Bayesiana na Identificação Forense: Análise e avaliação de vestígios de DNA com redes Bayesianas*. Phd Thesis, ISCTE, 2007.
- [2] ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol. 1 (1), pp 96-98, 2010.
- [3] ANDRADE, M. and FERREIRA, M. A. M.: *Bayesian networks in forensic identification problems*. Aplimat - Journal of Applied Mathematics, vol. 2 (3), pp. 13-30, 2009.
- [4] ANDRADE, M. and FERREIRA, M. A. M.: *Some considerations about forensic DNA evidences*. International Journal of Academic Research, vol. 3, (1, Part I), pp 7-10, 2011.
- [5] ANDRADE, M. and FERREIRA, M. A. M.: *Evidence evaluation in DNA mixture traces possibly resulting from two victims and two suspects*. Portuguese Journal of Quantitative Methods, vol 2 (1), pp 99-103, 2011a
- [6] ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with probabilistic expert systems*. International Journal of Academic Research, vol. 3 (3, Part I), pp 7-15, 2011b.
- [7] ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with two victims and a perpetrator*. 4 th IISMES Conference, Dalian, P. R. China. 2011c. Forthcoming.
- [8] DAWID, A. P., MORTERA, J., PASCALI, V. L. and BOXEL, D. W.: Probabilistic expert systems for forensic inference from genetic markers. Scandinavian Journal of Statistics vol. 29, pp 577-595, 2002.
- [9] EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. and CHAMPOD, C.: *Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks*. Journal of Forensic Science, vol. 47, pp 520-539, 2002.
- [10] FERREIRA, M. A. M. and ANDRADE, M.: *A note on Dawnie Wolfe Steadman, Bradley J. Adams, and Lyle W. Konigsberg, Statistical Basis for Positive Identification in Forensic Anthropology. American Journal of Physical Anthropology 131: 15-26 (2006)*. International Journal of Academic Research, vol. 1 (2), pp 23-26, 2009.
- [11] LAURITZEN, S. L.: *Bayesian networks for forensic identification Problems*. Tutorial 19th Conference on Uncertainty in Artificial Intelligence, Mexico, 2003.
- [12] MORTERA, J.: *Analysis of DNA mixtures using probabilistic expert systems*. In: P. J. Green, N. L. Hjort and S. Richardson (Eds.), Highly Structured Stochastic Systems. Oxford University Press, 2003.

⁵ This confusion is originated by the conflict: probability frequentist interpretation vs subjectivist probability interpretation (Andrade and Ferreira, 2009) and (Andrade, 2010).

- [13] MORTERA, J., DAWID, A. P. and LAURITZEN, S. L.: *Probabilistic expert systems for DNA mixture profiling*. Theoretical Population Biology, vol. 63, pp 191-205, 2003.

Current address

Marina Andrade, Professor Auxiliar

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

Telefone: + 351 21 790 34 05

Fax: + 351 21 790 39 41

e-mail: marina.andrade@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

TELEFONE: + 351 21 790 37 03

FAX: + 351 21 790 39 41

e-mail: manuel.ferreira@iscte.pt

CIVIL AND CRIMINAL IDENTIFICATION WITH BAYESIAN NETWORKS

ANDRADE Marina, (PT), FERREIRA Manuel Alberto M., (PT)

Abstract. The use of DNA evidence in problems of civil and criminal identification is becoming greater and greater. Being necessary to evaluate the weight of that evidence one of the most powerful tools to help in it is the Bayesian networks. This is exemplified with the presentation of a civil identification problem and of a criminal identification problem in this work.

Key words: Bayesian networks, DNA database profiles, civil and criminal identification problems.

Mathematics Subject Classification: 62C10.

1 Introduction

In this work it is intended to exemplify how to apply Bayesian networks in civil and criminal identification problems. The first objective is to give a methodology, with the appropriate tools, to use in a correct way a DNA profiles database in the problem of civil identification when there is a partial match between the genetic characteristic of an individual whose body was found, one volunteer who claimed a family member disappearance and one sample belonging to the DNA database.

In section 2 the civil identification case to be studied is presented and discussed. The Bayesian network, that allows the efficient probabilities computation, determinant to evaluate the hypothesis in comparison, is presented. In section 3, real examples clarifying the application are exhibited. And in section 4 a brief discussion is outlined related with this objective.

The second objective is to illustrate the use of biological information in crime scene identification problems, example of a criminal identification problem. In section 5 a crime scene, the correspondent evidence, E , and the hypotheses to be considered are presented. In section 6 the Bayesian network built expressly to perform the calculations is shown. And in section 7 the numerical results will be seen. Finally in section 8 a brief discussion, related with this second objective is presented.

General conclusions and references are presented at the end of the paper.

2 The Civil Identification Problem

Frequent examples of civil identification problems are the case of a body identification, jointly with information of a missing person belonging to a known family, or the identification of more than one body resultant of a disaster or an attempt. And even immigration cases in which it is important to establish family relations.

The establishment and use of DNA database files for a great number of European countries was an incentive to study the mentioned problems and the use of these database files for identification, (Corte-Real, 2004), (Martin, 2004) and (Andrade and Ferreira, 2009a, 2010). In this context it may be useful when unidentified corpses appear and may be identified by comparison of their DNA profiles with family volunteer's profiles.

The Portuguese law nº 5/2008 establishes the principles for creating and maintaining a database of DNA profiles for identification purposes, and regulates the collection, processing and conservation of samples of human cells, their analysis and collection of DNA profiles, the methodology for comparison of DNA profiles taken from the samples, and the processing and storage of information in a computer file¹.

So the database is, in general terms, composed of a file containing information of samples from convicted offenders with 3 years of imprisonment or more - α ; a file containing the information of samples of volunteers - β ; a file containing information on the “problem samples” or “reference samples” from corpses, or parts of corpses, or things in places where the authorities collect samples - γ .

It matters to study civil identification problems, mainly if there is a partial match between the genetic characteristic of an individual whose body was found and one volunteer who claimed a family member disappearance and one sample in the database file γ .

2.1 The Case of Partial Match with the Volunteer and one γ – sample

When there is an individual claiming for a disappeared person who gives his/her genetic characteristic, C_{Vol} , to be compared with the genetic characteristic of a body found, the first action to do is to check if there is a match between the genetic characteristic of the individual whose body was found, C_{BF} , and any sample of the DNA file, γ - sample, which is named “problem samples”.

If there is a partial match between the genetic profile of the individual whose body was found and one sample in the file γ , the evidence now is $E = (C_{BF}, \gamma - sample, C_{Vol})$.

Then it follows the establishment of the hypotheses of interest: the identification hypothesis (H_{ID}) versus the non identification hypothesis ($H_{not ID}$):

H_{ID} : It is possible to reach an identification of the individual whose body was found.
vs

¹ The implementation of this process is not going very well. In the moment the number of samples in database is not significant.

$H_{not\ ID}$: It is not possible to reach an identification of the individual whose body was found.

After checking the possibility of a partial match between the profile of the individual whose body was found, C_{BF} , the sample in the file γ , γ - sample, and the volunteer, C_{Vol} , two different comparisons are made in order to obtain a measure either of the possible genetic relation between the individual whose body was found with the γ - sample (*bf_match_gs?*), or of the possible genetic relation between the individual whose body was found and the volunteer (*bf_match_vol?*). The possible answers are: yes or no.

So the resulting states are:

- A (yes, no) - defines the possibility of genetic relationship between the individual whose body was found and the γ - sample but not the volunteer;
- B (no, yes) - defines the possibility of genetic relationship between the individual whose body was found and the volunteer but not the γ - sample;
- C (yes, yes) - defines the possibility of genetic relationship between the individual whose body was found and both the volunteer and the γ - sample;
- D (no, no) - defines the possibility of genetic relationship between the individual whose body was found neither with the volunteer nor with the γ - sample;

A , B define the identification hypothesis, H_{ID} , and C , D define the non identification hypothesis, $H_{not\ ID}$. B is a particular case: the simple problem studied in (Andrade and Ferreira, 2009a). Each of the four possible states probabilities provides a measure for each event, and the four are pairwise incompatible.

Following the probabilities computation it is important to compare state D versus A , B , C ; i.e., to evaluate the event “the individual whose body was found is not genetically related either with the γ - sample or the volunteer”. This comparison intends to evaluate the situation “the genetic information of the individual whose body was found is not compatible with the other genetic information available” and “the genetic information of the individual whose body was found is compatible with at least one of the remaining genetic information”.

If D is accepted the process ends. And the body genetic information joins the file γ in the database. If D is discarded then it is necessary to perform a comparison between A , B and C events. If C is accepted the process ends and police intelligence investigations must be done. If C is discarded, finally A and B must be compared. If A is accepted the individual whose body was found is related with the γ - sample. If B is accepted the conclusion is that the individual whose body was found is a volunteer relative.

3 The Bayesian Network for the Civil Identification Problem

The comparisons described above are performed through the respective probabilities events ratios: the likelihood ratios, (Ferreira and Andrade, 2009). The hypothesis with the greatest probability is the accepted one. Thus the probabilities associated to the states A , B , C and D must be computed. In

order to do so a lot of intermediary conditional probabilities computation, that are impossible to do with algebraic manipulations, must be done.

To overcome this situation those probabilities will be computed using the Bayesian network, see (Andrade and Ferreira, 2009) and (Andrade *et al.*, 2010), in the Figure 1².

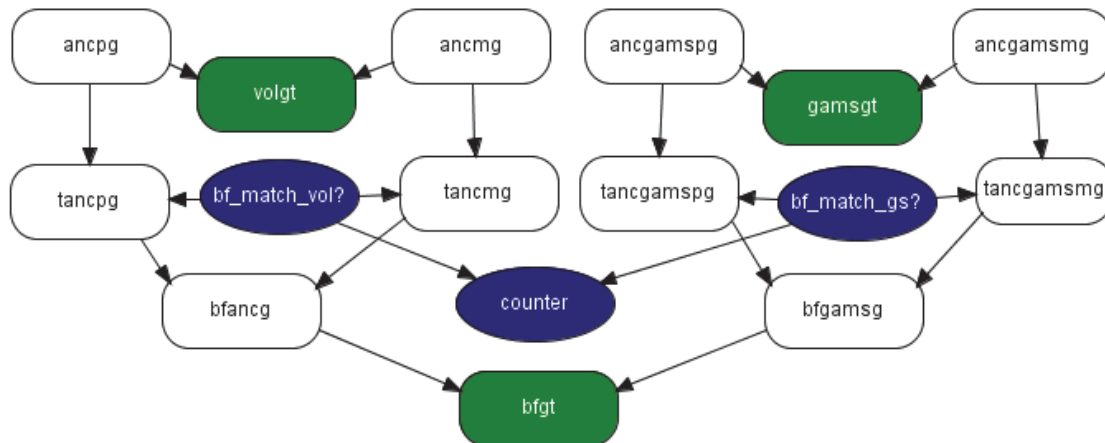


Figure 1: Network for civil identification with one volunteer and one γ – sample

The nodes **ancpg**, **ancmg**, **ancgamspg** and **ancgamsmg** are of class founder: a network with only one node which states are the alleles in the problem where the respective frequencies in the population are specified, and represent the volunteer's ancient paternal and maternal inheritance.

The nodes **volgt**, **gamsgt** and **bfgt** are of class genotype: the volunteer, the γ - sample and the body found genotypes.

Nodes **tancmg**, **tancpg**, **tancgamspg** and **tancgamsmg** specify whether the correspondent allele is or is not the same as the volunteer and the same as the γ - sample.

If **bf_match_vol?** is true then the volunteer's allele will be identical with the body found allele, otherwise the allele is randomly chosen in the population and if **bf_match_gs?** is true then the γ - sample's allele will be identical with the body found allele, otherwise the allele is randomly chosen in the population.

The nodes **bfancg** and **bfgamsg** define the Mendel inheritance in which the allele of the individual whose body was found is chosen at random from the ancient's paternal and maternal gene.

Node counter counts the number of true states of the preceding nodes, accounting the results for the A, B, C, D possible events.

3.1 Examples

To exemplify the described methodology, in Table 1 the allele frequencies, real ones, for some genetic markers and, for each marker, possible evidence profiles for the body found C_{BF} , the γ - sample and the volunteer C_{Vol} are presented.

²The networks mentioned in this work were implemented using Hugin software: www.hugin.com

Marker	Allele Frequencies				$\{C_{BF}, \gamma - sample, C_{Vol}\}$
D21S11	p_{28} 0.1647	p_{29} 0.2136	p_{30} 0.2437	$p_{31.2}$ 0.1138	$\{(29,30), (28,30), (29,31.2)\}$
F13A1	p_5 0.1985	p_6 0.2890	p_7 0.3377	p_8 0.0112	$\{(6,7), (7,8), (5,6)\}$
TH01	p_6 0.2044	p_7 0.1696	p_9 0.1984	$p_{9.3}$ 0.2748	$\{(7,9), (9,9.3), (6,7)\}$
TPOX	p_8 0.5053	p_9 0.0974	p_{10} 0.0647	p_{11} 0.2893	$\{(8,11), (8,10), (9,11)\}$
VWA31	p_{15} 0.1216	p_{16} 0.2300	p_{17} 0.2649	p_{18} 0.1859	$\{(16,17), (15,17), (16,18)\}$

Table 1: Allele frequencies and genetic profiles.

In Table 2 the state probabilities, the node counter states, see Figure 1, are presented.

States	D21S11	F13A1	TH01	TPOX	VWA31
A	0.5322	0.3296	0.4987	0.2661	0.4548
B	0.1296	0.2226	0.1978	0.2688	0.2251
C	0.2274	0.1904	0.1692	0.1539	0.2092
D	0.1108	0.2574	0.1343	0.3112	0.1109

Table 2: State probabilities.

And in Table 3 the decisions, consequence of the procedures proposed in section 2.1, are presented for each example evidence profile.

Evidence Profiles	Decision
$\{(29,30), (28,30), (29,31.2)\}$	Police intelligence investigations must be done
$\{(6,7), (7,8), (5,6)\}$	The individual whose body was found is a volunteer relative
$\{(7,9), (9,9.3), (6,7)\}$	Police intelligence investigations must be done
$\{(8,11), (8,10), (9,11)\}$	The individual whose body was found is a volunteer relative
$\{(16,17), (15,17), (16,18)\}$	Police intelligence investigations must be done

Table 3: Decisions for each evidence profile.

4 First Discussion

Using the Bayesian network built expressly for civil identification problem, in which there is a partial match between an individual whose body was found, a volunteer who claimed a relative disappearance supplying his/her own genetic information and a DNA database file sample existent, it is possible to perform the sequence of three hypothesis tests described above. Thus it is possible

to decide first if an identification is possible or not; second if an effective identification is possible or not; third to make the identification. So with a procedure technically simple, it is possible to make an adequate and correct use of a DNA database.

As the examples illustrate, the procedure leads almost surely to a decision: whether it is to close the case identifying the individual, or concluding that it is not possible any identification, or to go on with the police investigations.

5 Crime Scene Investigation

A crime has been committed. Two persons, V_1 and V_2 , were murdered. One mixture trace was found. S_1 and S_2 are potential suspects. S_1 and S_2 DNA profiles were measured and considered to be compatible with the mixture trace.

Being possible that a fight occurred during the assault, producing some material, it is acceptable that the individuals who perpetrated the crime could have left some of their material in the trace.

To analyse the crime scene, in this section, it will be presented the evidence, E , and the hypotheses to be considered.

To summarize the evidence it is presented in Table 4 the DNA profiles of the victims' and the suspect's, V_1 , V_2 , S_1 , S_2 , and the trace found at the crime scene, E .

	V_1	V_2	S_1	S_2	E
TH01	9,9.3	9,9.3	7,8	6,9	6,7,8,9,9.3
F13A1	5,7	5,6	3.2,5	6,7	3.2,5,6,7
FGA	22,26	22,23	24,24	19,24	19,22,23, 24,26

Table 4: Two victim's and two suspect's DNA profiles and evidence.

In Table 5 the allele frequencies, for each marker found in the trace, are presented.

	p_6	p_7	p_8	p_9	$p_{9.3}$
TH01	0.2044	0.1696	0.1386	0.1984	0.2748
	$p_{3.2}$	p_5	p_6	p_7	
F13A1	0.0806	0.1985	0.2890	0.3377	
	p_{19}	p_{22}	p_{23}	p_{24}	p_{26}
FGA	0.0684	0.1740	0.1606	0.1325	0.0321

Table 5: Allele frequencies.

The allele frequencies in Table 5 are the Portuguese population frequencies collected in the database "The Distribution of Human DNA-PCR Polymorphisms", since the mentioned case is supposed to have occurred in Portugal.

The crime trace can contain DNA from up to four unknown contributors, in addition to the victims and/or the suspects.

If the DNA of S_i with $i = 1, 2$ is presented in the trace this will place him/her at the crime scene and consequently as one of the possible perpetrators.

The court has to determine if each suspect is or is not guilty. The hypotheses to be evaluated are:

H_1 : S_1 is a contributor to the trace but S_2 is not, given the evidence.

H_2 : S_2 is a contributor to the trace but S_1 is not, given the evidence.

H_3 : S_1 and S_2 are both contributors to the trace, given the evidence.

H_4 : Neither S_1 nor S_2 are contributors to the trace, given the evidence.

The respective events probabilities are called $p_{10}, p_{02}, p_{12}, p_{00}$, where 0 mentions the absence of the respective, in order, individual DNA in the trace. So:

If $p_{00} > p_{10} + p_{02} + p_{12}$ the two suspects are acquitted. If not it must be seen if $p_{12} > p_{10} + p_{02}$ case at which the two suspects are both placed at the crime scene. If not p_{10} must be compared with p_{02} . If $p_{10} > p_{02}$ the evidence favours the presence of S_1 at the crime scene and the acquaintance of S_2 . The contrary happens when $p_{02} > p_{10}$.

6 Marker Bayesian Network

The probabilities referred above are very hard to compute algebraically, demanding a great use of Bayes' Law because of the number of the dependencies to be considered. So they will be computed using the Bayesian network of Figure 2.

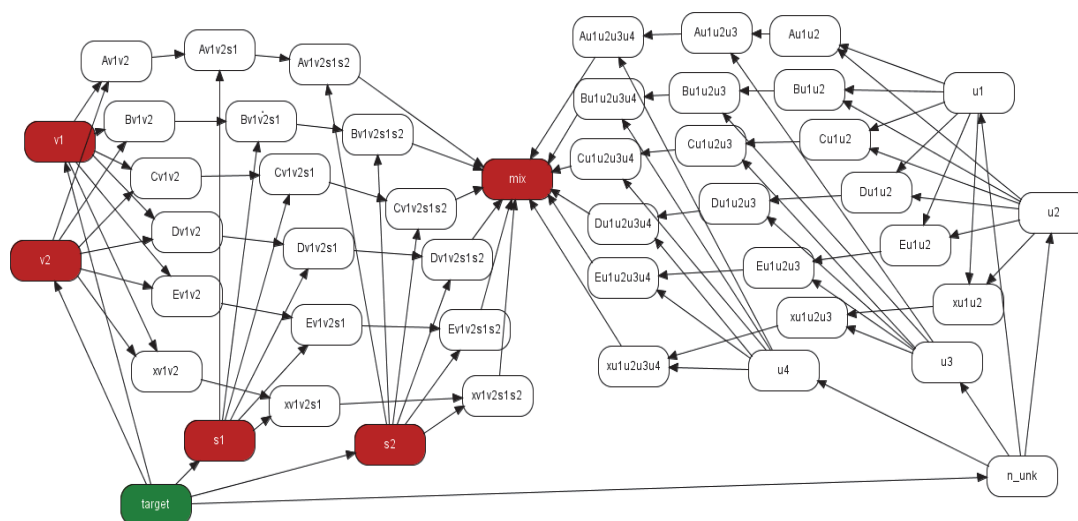


Figure 2: Marker network

Nodes $\mathbf{v}_i, i = 1, 2, \mathbf{s}_j, j = 1, 2$ and $\mathbf{u}_k, k = 1, 2, 3, 4$, in Figure 2 are themselves Bayesian networks that represent the genetic structure and inheritance of each individual - the victims, the suspects and the unknowns, respectively - and have all the same structure. The $\mathbf{v}_i, i = 1, 2$ and $\mathbf{s}_j, j = 1, 2$ are represented in red colour meaning that the respective profiles are known and constitute data of the problem. The nodes in white, below the node mix, that represents the mixture and is also in red colour because it is comprised by known data (E), represent the relations in which the nodes in red may contribute to the mixture. The nodes in white, above the node mix, except the $\mathbf{u}_k, k = 1, 2, 3, 4$ and $\mathbf{n_unk}$ - that is a counter for the number of unknowns in the mixture - represent the relations in

which the $u_k, k = 1, 2, 3, 4$ may contribute to the mixture. Node **target**, in green colour, collects the states and the respective probabilities.

As it is mandatory to consider the possible contribution of till four unknown individuals to the mixture, the number of admissible states jumps to 80, numbered from 0 - no one in the mixture - to 79 - the two victims, the two suspects and the four unknowns are all in the mixture. Of course these two states are unrealistic and there are other ones also unrealistic because are incompatible with the minimum number of contributors to the mixture, according to the evidence inserted. These unrealistic states are discarded by the network but have to be considered conceptually in its building.

Among the realistic states only a few ones are interesting to the problem: the corresponding to the hypotheses events defined above.

7 Numerical Results

For marker TH01, alleles 6, 7, 8, 9, 9.3 are considered, Table 4, and so they are represented in the Figure 1 Bayesian network by A, B, C, D, E, respectively. When considering marker F13A1, the alleles are 3.2, 5, 6, 7, corresponding to A, B, C, D. E is considered with 0 frequency. In marker FGA the alleles are 19, 22, 23, 24, 26 corresponding to A, B, C, D, E. In any case x accumulates the remaining frequencies of the non considered alleles for each marker.

The results obtained using Table 4 data together with Table 5 frequencies are in Table 6, where the values in line rescale are constituted by the ratios of the products of the values in the respective column³ by the total sum of the four products. The values in this line are the used ones in the tests described in section 5.

	p_{00}	p_{12}	p_{10}	p_{02}
TH01	0.0830	0.5029	0.2773	0.1367
F13A1	0.0986	0.4544	0.3279	0.1187
FGA	0.0378	0.4398	0.0820	0.4398
Rescale	0.0027	0.8709	0.0646	0.0618

Table 6: **Results.**

Following the procedure outlined in section 5 the conclusion is that both suspects are placed at the crime scene – note the great value of $p_{12} = 0.8709$. For TH01 and F13A1, alone, the conclusion is the same. But for FGA this does not happens. Note that, $p_{12} = p_{02}$. This is justified by the fact that in marker FGA there are two rare alleles, $p_{19} = 0.0684$ and $p_{26} = 0.0321$, that are in consequence “good identifiers”. Each one is present in V_1 and S_2 . Besides S_1 is homozygote for this marker and this genotype may be hidden by S_2 ’s genotype. In consequence it is natural that p_{12} and p_{02} are of the same magnitude.

To compute the interesting probabilities there must be considered the following states probabilities:

³ It is possible to multiply the respective probabilities, for each marker, because it is assumed independence between and across marker, i.e., linkage and Hardy-Weinberg Equilibrium [8].

- p_{00} : 1, 2, 3, 16, 17, 18, 19, 32, 33, 34, 35, 48, 49, 50, 51, 64, 65, 66 and 67,
- p_{12} : 12, 13, 14, 15, 28, 29, 30, 31, 44, 45, 46, 47, 60, 61, 62, 63, 76, 77, 78 and 79,
- p_{10} : 8, 9, 10, 11, 24, 25, 26, 27, 40, 41, 42, 43, 56, 57, 58, 59, 72, 73, 74 and 75,
- p_{02} : 4, 5, 6, 7, 20, 21, 22, 23, 36, 37, 38, 39, 52, 53, 54, 55, 68, 69, 70 and 71

from the output given by Hugin after the inserted evidence.

8 Second Discussion

Criminal identification problems are examples of situations, in which forensic approach, the DNA profiles study is usual. But the interpretation and evaluation of DNA evidences is not an easy task, see for instance, (Andrade and Ferreira, 2011) and (Lauritzen, 2003). Also the fact that in general they are posed in probabilistic terms leads to some confusion to the judges when they have to issue a decision. In this situation the Bayesian approach is maybe the most clear to explain the significance of the evidence, see (Ferreira and Andrade, 2009). And for it the use of Bayesian networks to compute the interesting probabilities is a natural option, as it was exemplified in this paper.

It is important to define which probabilities, among the possible ones to compute, interest to the problem. And in consequence to define, for each case, which hypotheses tests to implement. Of course they are Bayesian tests.

Note finally, as this example shows, that this methodology may conclude for the absolution of a suspect but not for the conviction. It only can place the suspect in the crime scene. Further work of the police must be made to conclude by the conviction or absolution.

9 General Conclusions

The use of networks transporting probabilities began with the geneticist Sewall Wright in the beginning of the 20th century (1921). (Dawid *et al.*, 2002) describes this new approach to problems of the kind of the one described above. The construction and use of Bayesian networks to analyse problems in forensic identification inference, was initially done there, followed by (Evetts *et al.*, 2002), (Mortera, 2003) and (Mortera *et al.*, 2003).

The civil identification problem presented obviously may occur in situations of catastrophes or accidents at which it is possible to have unidentified victims. The use of DNA evidence is quite recent in helping to solve this situations. It was shown in this work how the use of Bayesian networks is useful to evaluate that kind of evidence.

The analysis of a crime scene analogous to the considered in this work, but with two victims' and one perpetrator and two mixture traces was presented in (Andrade and Ferreira, 2009, 2011b, 2011c). A problem dealing with a crime scene analogous to the one considered in this work may be seen at (Andrade and Ferreira, 2011a). Also was shown in this work how useful are the Bayesian networks in the evaluation of DNA evidence in problems of criminal identification.

Acknowledgement

This work was financially supported by FCT through the Strategic Project PEST-OE/EGE/UI0315/2011.

References

- ANDRADE, M.: *A Estatística Bayesiana na Identificação Forense: Análise e avaliação de vestígios de DNA com redes Bayesianas*. Phd Thesis, ISCTE, 2007.
- ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol. 1 (1), pp 96-98, 2010.
- ANDRADE, M. and FERREIRA, M. A. M.: *Bayesian networks in forensic identification problems*. Aplimat - Journal of Applied Mathematics, vol. 2 (3), pp. 13-30, 2009.
- ANDRADE, M. and FERREIRA, M. A. M.: *Civil identification problems with DNA databases using Bayesian networks*. International Journal of Security-CSC Journals, vol 3 (4), pp 65-74, 2009a.
- ANDRADE, M. and FERREIRA, M. A. M.: *Civil identification problems with Bayesian networks using official DNA databases*. Aplimat-Journal of Applied Mathematics, vol. 3 (3), pp 155-162, 2010.
- ANDRADE, M. and FERREIRA, M. A. M.: *Some considerations about forensic DNA evidences*. International Journal of Academic Research, vol. 3, (1, Part I), pp 7-10, 2011.
- ANDRADE, M. and FERREIRA, M. A. M.: *Evidence evaluation in DNA mixture traces possibly resulting from two victims and two suspects*. Portuguese Journal of Quantitative Methods, vol 2 (1), pp 99-103, 2011a
- ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with probabilistic expert systems*. International Journal of Academic Research, vol. 3 (3, Part I), pp 7-15, 2011b.
- ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with two victims and a perpetrator*. 4 th IISMES Conference, Dalian, P. R. China. 2011c. Forthcoming.
- ANDRADE, M., FERREIRA, M. A. M., ABRANTES, D., PONTES, M. L. and PINHEIRO, M. F.: *Object-oriented Bayesian Networks in the evaluation of paternities in less usual environments*. Journal of Mathematics and Technology, vol. 1 (1), pp 161-164, 2010.
- Corte-REAL, F.: *Forensic DNA databases*. Forensic Science International. 146s:s143-s144, 2004.
- DAWID, A. P., MORTERA, J., PASCALI, V. L. and BOXEL, D. W.: *Probabilistic expert systems for forensic inference from genetic markers*. Scandinavian Journal of Statistics vol. 29, pp 577-595, 2002.
- EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. and CHAMPOD, C.: *Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesain networks*. Journal of Forensic Science, vol. 47, pp 520-539, 2002.
- FERREIRA, M. A. M. and ANDRADE, M.: *A note on Dawnie Wolfe Steadman, Bradley J. Adams, and Lyle W. Konigsberg, Statistical Basis for Positive Identification in Forensic Anthropology. American Journal of Physical Anthropology 131: 15-26 (2006)*. International Journal of Academic Research, vol. 1 (2), pp 23-26, 2009.
- LAURITZEN, S. L.: *Bayesian networks for forensic identification Problems*. Tutorial 19th Conference on Uncertainty in Artificial Intelligence, Mexico, 2003.
- MARTIN, P.: *National DNA databases - practice and practability. A forum for discussion*. In International Congress Series 1261, pp 1-8, 2004.

MORTERA, J.: *Analysis of DNA mixtures using probabilistic expert systems*. In: P. J. Green, N. L. Hjort and S. Richardson (Eds.), *Highly Structured Stochastic Systems*. Oxford University Press, 2003.

MORTERA, J., DAWID, A. P. and LAURITZEN, S. L.: *Probabilistic expert systems for DNA mixture profiling*. *Theoretical Population Biology*, vol. 63, pp 191-205, 2003.

Current address

Marina Andrade, Professor Auxiliar

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

Telefone: + 351 21 790 34 05

Fax: + 351 21 790 39 41

e-mail: marina.andrade@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

TELEFONE: + 351 21 790 37 03

FAX: + 351 21 790 39 41

e-mail: manuel.ferreira@iscte.pt

RISK MANAGEMENT OF EQUITY PORTFOLIO CONSTRUCTION ON THE BASIS OF DATA ENVELOPMENT ANALYSIS APPROACH

ARSHINOVA Tatyana, (LV)

Abstract. The research focus of the scientific paper is on the problem of equity portfolio construction. The author recommends applying frontier analysis technique such as Data Envelopment Analysis to the performance measurement of emitters. Using modern computer technologies, the author has calculated efficiency score of twenty Baltic companies which are quoted at NASDAQ OMX Riga and NASDAQ OMX Tallinn stock exchanges on the basis of DEA CCR approach and elaborated proposals for effective asset allocation.

Keywords: Data Envelopment Analysis, Decision Making Units (DMUs) portfolio construction, DEA, performance measurement

Introduction

Actual macroeconomic events are indicative of Eurozone crisis threats. In late October 2011 European leaders obtained an agreement from banks to take 50% loss on the face value of their Greek debt, decreasing its value by 100 billion euro. They also neared agreement on boosting the firepower of the Continent's bailout fund to around €1 trillion to help it protect larger economies like Italy and Spain from the sort of market pressures that pushed Greece to need a rescue. In September 2011 debt rating agency Standard & Poor's downgraded Italy one level, from A +/A+1 to A/A-1, assessing the prospects for economic growth in Italy in the category of "negative." The remaining uncertainty in Europe's recovery and the future of euro impacts activities of all economical subjects. Thus, European companies and potential investors become especially vulnerable.

The construction of profitable and effective equity portfolio is among the most important investment problems. Traditionally the portfolio construction process includes four steps: creation of risk profile, asset allocation, correction of the portfolio structure corresponding to the investor's requirements and regular control over the portfolio structure to avoid the overweight risk in a particular asset class. Currently risk measurement and asset allocation stages are completed using methods of technical and fundamental analysis, Modern Portfolio Theory by H.Markovitz, Sharpe ratio analysis etc. The main principle of technical analysis is an assumption that the market price of

an asset includes information on all influencing factors; the development strategy of the company, balance data and perspectives of development are not estimated. However, fundamental analysis is based on the estimation of financial statements and competitive advantages. The Modern Portfolio Theory is the theory of investment which attempts to maximize portfolio expected return for a given amount of portfolio risk, or equivalently minimize risk for a given level of expected return, by carefully choosing the proportions of various assets. Nevertheless, efforts to translate the theoretical foundation into a viable portfolio construction algorithm have been plagued by technical difficulties stemming from the instability of the original optimization problem with respect to the available data. The results which are obtained on the basis of the mentioned approaches often provide inconsistent conclusions concerning potential investment opportunities and do not provide the possibility to evaluate the enterprise activity of emitters as a process.

Methods of frontier analysis ensure a principally different approach to the problem of equity performance measurement, estimating the performance of production process of each company. They provide an opportunity of complex analysis of company's efficiency level for a certain period of time and comparison of it among investigated objects. The objective of the author's research is to improve and supplement the methodology of risk measurement before the equity portfolio construction on the basis of the Data Envelopment Analysis approach.

In the circumstances of unstable macroeconomic environment and competition, profitability and market capitalization are among the most important indicators of stability and development of companies for the potential investor. Total operating revenue is a measure of the market value of company's production and the demand for it. Market capitalization is a parameter that reflects market value of all of a company's outstanding shares. It is a basic determinant of asset allocation and risk-return parameters. In this connection, the author analyzed the performance of a set of Baltic companies, assuming total operating revenue and market capitalization value as outputs. The objects of the research are Baltic companies which are quoted at NASDAQ OMX Riga and NASDAQ OMX Tallinn stock exchanges; their efficiency level is analyzed using data for the second quarter 2011. Evaluating the performance on the basis of the Data Envelopment Analysis approach, the author included into the set of investigated objects companies that are considered to be liquid at the Baltic stock markets (according to the amount of operations): JSC "Latvijas Balzāms", JSC "Grindeks", JSC "Latvijas gāze", JSC "Liepājas metalurģs", JSC "Latvijas Kuģniecība", JSC "Olainfarm", JSC "Rīgas kuģu būvētava", JSC "SAF Tehnika", JSC "Ventspils nafta", JSC "Valmieras stikla šķiedra", JSC "Arco Vara", JSC "Baltika", JSC "Ekspress Grupp", JSC "Harju Elekter", JSC "Olympic Entertainment Group", JSC "Silvano Fashion Group", JSC "Tallink Grupp", JSC "Tallina Kaubamaja", JSC "Tallina Vesi", JSC "Viisnurk".

1 Methods of frontier data analysis

The progress of production technology and increase of production volumes have stimulated the development of performance measurement methodology. In the second part of the 20th century there were introduced methods of frontier data analysis that provided a qualitatively different approach to the problem. According to the methodology of methods of frontier data analysis, the efficiency score of investigated DMUs is calculated as a distance from the point that defines the production process of a Decision Making Unit (DMU) to the certain efficiency frontier. Entities that are functioning on the efficiency frontier are considered to be absolutely technically efficient; inefficiency of other DMUs is increasing together with extension of the distance to the efficiency frontier. The value of efficiency score is fluctuating from zero to one.

Methods of frontier analysis may be divided into two groups: parametric (Stochastic Frontier Approach (SFA), Distribution-Free Approach (DFA), Thick Frontier Approach (TFA)) and non-parametric (Data Envelopment Analysis (DEA), Free Disposal Hull (FDH)) methods.

In accordance with parametric approaches, the efficiency frontier is constructed on the basis of econometric modelling, usually in form of Cobb-Douglas (log-linear) production function. Econometric analyses include two error components: an error term that captures inefficiency (u_i) and a random error (v_i). Parametric methods have significant advantages – they provide the possibilities to use panel data, to distinguish the random noise from inefficiency and to calculate the standard error of efficiency measurement results. Nevertheless, the stochastic approaches of performance measurement presume the comparison of investigated DMUs efficiency to the theoretically developed benchmark frontier; therefore the optimal combinations of inputs and outputs sometimes are not achievable practically. The application of parametric methods also requires observance of the restrictions imposed on the distributional assumptions on the inefficiencies and random error.[6]

In contrast to the econometric approaches, non-parametric methods are based on the hypothesis that the efficiency frontier is generated from the empirical results of the most efficient DMUs i.e. benchmarks that „float” on the piecewise linear frontier. The level of technical efficiency of these DMUs is 100%. However, the level of scale efficiency that defines the optimality of output and input proportions may have different values even among absolutely technically efficient DMUs. While mathematical, non-parametric methods require few assumptions when specifying the best-practice frontier, they generally do not account for random errors [7].

2 The CCR DEA Model

The CCR DEA model was developed by Charnes, Cooper and Rhodes in 1978 to evaluate the performance of Decision Making Units (DMUs). To allow for applications to a wide variety of activities, the term DMU might be used to refer to any entity that is to be evaluated in terms of its abilities to convert inputs into outputs. These evaluations can involve governmental agencies and non-profit organizations as well as business firms, hospitals and educational institutions.

The production process might be aimed either at minimization of resources or maximization of production volumes. The orientation of the model should be aimed at controllable variables. Volumes of resources are usually over control of management; therefore only input-oriented model will be examined in the paper.

The measurement of comparative efficiency is based on the assumption that the performance of each DMU is calculated in comparison to n investigated DMUs. Each DMU consumes varying amounts of m different inputs to produce s different outputs. Specifically, DMU_j consumes amount x_{ij} of input i and produces amount y_{rj} of output r . It is necessary to assume that $x_{ij} \geq 0$ and $y_{rj} \geq 0$ and further to assume that each DMU has at least one positive input and one positive output value. Primarily the DEA model was expressed in fractional, i.e. ratio-form. In this form the ratio of outputs to inputs is used to measure the relative efficiency of the $DMU_j = DMU_0$ to be evaluated relative to the ratios of all of the $j = 1, 2, \dots, n$ DMU_j . The CCR construction can be interpreted as the reduction of the multiple-output/multiple-input situation (for each DMU) to that of a single 'virtual' output and 'virtual' input. For a particular DMU the ratio of this single virtual output to single virtual input provides a measure of efficiency that is a function of the multipliers. In mathematical programming parlance, this ratio, which is to be maximized, forms the objective function for the particular DMU being evaluated. A set of normalizing constraints (one for each DMU) reflects the condition that the virtual output to virtual input ratio of every DMU, including

$DMU_j = DMU_0$, must be less than or equal to unity. [4] The mathematical programming problem may thus be stated as (1):

$$\begin{aligned} \max h_0(u, v) &= \sum_r u_r y_{r0} / \sum_i v_i x_{i0} \\ \text{subject to} \\ \sum_r u_r y_{rj} / \sum_i v_i x_{ij} &\leq 1 \text{ for } j = 1, \dots, n, \\ u_r, v_i &\geq 0 \text{ for all } i \text{ and } r, \end{aligned} \quad (1)$$

where

- h_0 – the function of virtual output and virtual input ratio of DMU_0 ;
- u_r – the output multiplier of DMU_0 ;
- v_i – the input multiplier of DMU_0 ;
- y_{r0} – the output of DMU_0 ;
- x_{i0} – the input of DMU_0 ;
- y_{rj} – outputs of 1,2...n DMUs;
- x_{ij} – inputs of 1,2...n DMUs.

The above ratio form yields an infinite number of solutions; if (u^*, v^*) is optimal, then $(\alpha u^*, \alpha v^*)$ is also optimal for $\alpha > 0$. However, the transformation developed by Charnes and Cooper (1962) for linear fractional programming selects a representative solution (u, v) $\sum_{i=1}^m v_i x_{i0} = 1$ and yields the equivalent linear programming problem in which the change of variables from (u, v) to (μ, v) is a result of the Charnes-Cooper transformation (2):

$$\begin{aligned} \max z &= \sum_{r=1}^s \mu_r y_{r0} \\ \text{subject to} \\ \sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0 \\ \sum_{i=1}^m v_i x_{i0} &= 1 \\ \mu_r, v_i &\geq 0, \end{aligned} \quad (2)$$

where

- z – the CCR input-oriented function of DMU_0 (multiplier form);
- μ_r – the output multiplier of DMU_0 ;
- v_i – the input multiplier of DMU_0 ;
- y_{r0} – the output of DMU_0 ;
- x_{i0} – the input of DMU_0 ;
- y_{rj} – outputs of 1,2...n DMUs;
- x_{ij} – inputs of 1,2...n DMUs;
- s – number of outputs;
- m – number of inputs.

Model that is expressed by (2) can be solved by its dual problem (3):

$$\begin{aligned}
 &\theta^* = \min \theta \\
 &\text{subject to} \\
 &\sum_{j=1}^n x_{ij} \lambda_j \leq \theta x_{io} \quad i = 1, 2, \dots, m; \\
 &\sum_{j=1}^n y_{rj} \lambda_j \geq y_{ro} \quad r = 1, 2, \dots, s; \\
 &\lambda_j \geq 0 \quad j = 1, 2, \dots, n,
 \end{aligned} \tag{3}$$

where

θ^* – the optimal value of dual variable θ of DMU₀;
 θ, λ_j – dual variables of DMU₀;
 y_{r0} – the output of DMU₀;
 x_{i0} – the input of DMU₀;
 y_{rj} – outputs of 1,2...n DMUs;
 x_{ij} – inputs of 1,2...n DMUs;
 s – number of outputs;
 m – number of inputs.

This last model is sometimes referred to as the "Farrell model" because it is the one used in Farrell (1957). By virtue of the dual theorem of linear programming we have $z^* = \theta$. Hence either problem may be used. One can solve the dual linear program, to obtain an efficiency score. Setting $\theta = 1$ and $\lambda_k^* = 1$ with $\lambda_k = \lambda_o^*$ and all other $\lambda_k^* = 0$, a solution of dual problem (see Formula 3) always exists. Moreover this solution implies $\theta^* \leq 1$. The optimal solution, θ^* , yields an efficiency score for a particular DMU. [3]

The process is repeated for each DMU. i.e., solving the model, expressed by Formula 3, with $(X_o, Y_o) = (X_k, Y_k)$, where (X_k, Y_k) represent vectors with components x_{ik}, y_{rk} and, similarly (X_o, Y_o) has components x_{ok}, y_{ok} . DMUs for which $\theta^* < 1$ are inefficient, while DMUs for which $\theta^* = 1$ are boundary points. Some boundary points may be "weakly efficient" because we have non-zero slacks. This may appear because alternate optima may have non-zero slacks in some solutions, but not in others. However, we can avoid this effect by invoking the following linear program in which the slacks are taken to their maximal values (4).

$$\begin{aligned}
 &\max \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \\
 &\text{subject to} \\
 &\sum_{j=1}^n x_{ij} \lambda_j + s_i^- = \theta^* x_{io} \quad i = 1, 2, \dots, m; \\
 &\sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{ro} \quad r = 1, 2, \dots, s; \\
 &\lambda_j, s_i^-, s_r^+ \geq 0 \quad \forall i, j, r,
 \end{aligned} \tag{4}$$

where

s_i^- – input slacks;
 s_r^+ – output slacks;
 θ^* – the optimal value of dual variable θ of DMU₀;
 λ_j – the dual variable of DMU₀;
 y_{r0} – the output of DMU₀;

x_{i0} – the input of DMU₀;
 y_{rj} – outputs of 1,2...n DMUs;
 x_{ij} – inputs of 1,2...n DMUs;
 s – number of outputs;
 m – number of inputs.

It shall be noted that the choices of s_i^- and s_r^+ do not affect the optimal θ^* which is determined from model expressed by (3). These developments lead to the following definitions of DEA efficiency:

DEA Efficiency: The performance of DMU₀ is fully (100%) efficient if and only if both (i) $\theta^* = 1$ and (ii) all slacks $s_i^{-*} = s_r^{+*} = 0$.

Weakly DEA Efficiency: The performance of DMU₀ is weakly efficient if and only if both (i) $\theta^* = 1$ and (ii) $s_i^{-*} \neq 0$ and/or $s_r^{+*} \neq 0$ for some i and r in some alternate optima [1].

The CCR efficiency score is indicative of the overall efficiency level of investigated DMUs. [5]

3 The application of data envelopment analysis approach to the equity portfolio construction

3.1 Methodology of the research

Due to the methodology, the Data Envelopment Analysis approach of comparative performance measurement does not require the specific functional form of the model. Therefore choice of outputs and inputs that are corresponding to the objectives of the research is among significant conditions for the achievement of plausible results.

The problem of keeping profitability is especially topical and important in the circumstances of unstable macroeconomic environment. The market capitalization value reflects the risk-return parameters that are indicative of company's stability and development opportunities. In this connection, there is developed a concept of efficiency measurement of companies which are quoted at the NASDAQ OMX Riga and NASDAQ OMX Tallinn in the research, assuming total operational revenue to be outputs, while equity, operating expenses and finance (interest) expenses are defined as inputs.

The performance evaluation will be completed on the basis of DEA CCR approach that allows calculating overall efficiency score of investigated companies.

3.2 Efficiency measurement results of Baltic companies on the basis of CCR DEA approach

The application of the DEA approach requires the determination of assumptions, concerning orientation measures of the model and the concept of returns to scale (RTS). The production process may be aimed either at minimization of resources (input-oriented) or maximization of production volumes (output-oriented). It is emphasized in the international researches that the orientation of the model should be aimed at controllable variables. Usually volumes of resources are considered to be over control of management, therefore there is applied the assumption of input orientation in the research. Since the constant returns to scale CRS approach represents the total (overall) efficiency level, CCR DEA model is considered to be the basic concept of the research.[2]

The results of companies' performance evaluation on the basis of CCR input-oriented model, assuming total operating revenue and market capitalization values as outputs, are represented in Figure 1.

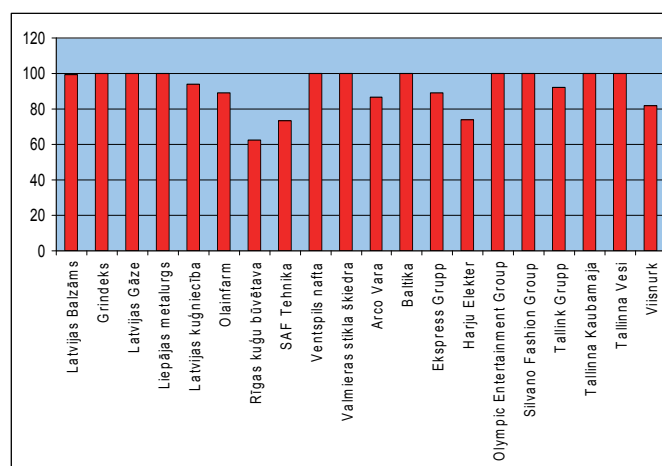


Fig. 1. DEA CCR efficiency score of Baltic stock exchange quoted companies, (%)

According to the obtained results, investigated companies might be separated into three groups. The first group includes 100% DEA CCR efficient companies: JSC “Latvijas Balzāms”, JSC “Grindeks”, JSC “Latvijas gāze”, JSC “Liepājas metalurģs”, JSC “Ventspils nafta”, JSC “Valmieras stikla šķiedra”, JSC “Baltika”, JSC “Olympic Entertainment Group”, JSC “Silvano Fashion Group”, JSC “Tallina Kaubamaja”, JSC “Tallina Vesi”. The above-mentioned companies have demonstrated the best result, operating on the efficiency frontier at the observation period. High efficiency level of these emitters is indicative of their ability to maximize the volume of outputs using minimal volumes of inputs and to ensure optimal proportions of output and inputs in the process of production, thus of both 100% technical and scale efficiency in comparison to the set of investigated objects. For example, the state-owned company JSC “Latvijas gāze” ensuring 375.9 million euro market capitalization value and 278.9 million euro total revenue, is operating using only equity capital and having no interest expenses. Despite of high volatility of the share price, JSC “Liepājas metalurģs” has the total revenue value 183 million euro at the second quarter 2011. According to the latest company’s announcement, JSC “Liepājas metalurģs” is investing into the equipment modernization project; the commercial pledge of 72.19 million lats is guaranteed by the Ministry of Finance of the Republic of Latvia. Due to this fact, potential investors might expect the reduction of company’s operational costs. JSC “Olympic Entertainment Group” and JSC “Tallina Kaubamaja” are among the leading companies on the Tallinn stock exchange according to their output values. The enterprise activity of JSC “Olympic Entertainment Group” is oriented at casino and hotel business segments, ensuring the total revenue of 60.8 million euro by the end of the second quarter 2011. JSC “Tallina Kaubamaja” is the largest department store in Estonia that is listed since 1996 on the Tallinn stock exchange. This fact makes securities of the emitter attractive for potential investors.

The second group consists of companies that are having the performance above the 80% level: JSC “Latvijas Kuģniecība”, JSC “Olainfarm”, JSC “Arco Vara”, JSC “Ekspress Grupp”, JSC “Tallink Grupp”, JSC “Viisnurk”. JSC “Olainfarm” is one of the most rapidly growing Baltic companies. According to the latest company’s announcement, preliminary sales results of JSC “Olainfarm” for September 2011 show that sales have increased by 102% compared to the same period last year and have reached 3.71 million lats (5.28 million euro). The most rapid sales increase has been experienced in Canada, where sales have increased 442 times, in Ukraine they increased nearly 4 times, in Belarus, a country heavily hit by its currency crisis, the sales have

grown by 71%, in Russia by 63%. Main sale markets of AS “Olainfarm” during September 2011 were Ukraine, Russia, Belarus and Latvia. Nevertheless, the company has lower production volumes, higher finance expenses (228 thousand euro) than its nearest competitor JSC “Grindeks”, having the 88.89% performance level. Having the highest market capitalization value 467.6 million euro and total revenue of 400.8 million euro, JSC “Tallink Grupp” is only 92.31% DEA CCR efficient. Among possible reasons are high finance (26.7 million euro) and operational expenses (391.5 million euro). This fact is indicative of inefficient organization of company’s operational activity.

The third group includes companies that are having the efficiency below the 80% level: JSC “Rīgas kuģu būvētava”, JSC “SAF Tehnika”, JSC “Harju Elekter”.

According to the JSC “SAF Tehnika” interim report data in August 2011, the company’s non-audited net sales for 12 months of the financial year 2010/11 were 10.9 million LVL (15.5 million EUR) representing a year-on-year increase of 7%. Sales in the Asia Pacific, Middle East and Africa region formed the largest sales proportion (37%) comprising 4.05 million LVL (5.76 million EUR) although it was by 32% less than in previous financial year 2009/10. The net profit of JSC “SAF Tehnika” for the 12 months of financial year 2010/11 was 780 thousand LVL (1.1 million EUR) representing 52% of the net profit of previous financial year 2009/10. JSC “SAF Tehnika’s” non-audited net sales for the fourth quarter of financial year 2010/11 were 1.99 million LVL (2.84 million EUR), representing 53% of the fourth quarter of the previous financial year. [8] Reporting quarter was the weakest in this financial year unlike from last financial year 2009/10 when the fourth quarter was the best. This information had negative impact on the share price of JSC “SAF Tehnika” that decreased by 27.27% since January 2011.

Operating with loss in both 2010 and 2011 financial years (572.6 thousand euro in the second quarter 2011), JSC “Rīgas kuģu būvētava” has the lowest performance level among all investigated companies. Nevertheless, on October 19th 2011 JSC „Rīgas kuģu būvētava” received an official announcement from SJSC «Черноморнефтегаз» tenders trade commission about the approval of the proposed price by JSC „Rīgas kuģu būvētava. The proposal of JSC „Rīgas kuģu būvētava „to delivery gas platform 35.11.4. for USD 399 800 000 was considered the most economically beneficial and was accepted as a result of evaluation. [8] This corporative event caused the increase of equity price by 77.6%, demonstrating that shares of JSC „Rīgas kuģu būvētava” are a good investment opportunity for a risk-tolerant investor.

Conclusions

The scientific paper is devoted to the equity portfolio construction problem. The most important stages of this process are choice of potential assets, risk evaluation and asset allocation. Traditionally potential investors are using methods of technical and fundamental analysis, Modern Portfolio Theory for this purpose. However, the results which are obtained on the basis of the mentioned approaches often provide inconsistent conclusions concerning investment opportunities and do not provide the possibility to evaluate the enterprise activity of emitters as a process.

The methodology of Data Envelopment Analysis is considered to be a sophisticated tool for performance measurement that allows the investigation of complex production processes among a set of Decision Making Units (DMUs).

The author has implemented the DEA CCR approach, analyzing efficiency scores of a set of companies which are quoted at NASDAQ OMX Riga and Tallinn. According to the results, emitters might be divided into three groups: 100% DEA CCR efficient, having the performance above the 80% level and having the performance below the 80% level. Equities of fully CCR DEA efficient companies could be included into the portfolio with conservative investment strategy. Companies

which are included into the second and the third group, have lower level of performance. However, securities of these emitters might be attractive for investors with higher level of risk tolerance.

To sum up, the author recommends using the Data Envelopment Analysis approach methodology as an additional tool for analysis of investment opportunities and equity portfolio creation.

References

- [1] Data Envelopment Analysis: History, Models and Interpretations. Cooper W.W., SEIFORD L.M., J. ZHU. Handbook on Data Envelopment Analysis. – Boston: Kluwer, 2004, pp.1-39.
- [2] Returns to Scale in DEA. Banker R.D., Cooper W.W., SEIFORD L.M., J. ZHU. Handbook on Data Envelopment Analysis. – Boston: Kluwer, 2004, pp. 41-74.
- [3] FARRELL, M. J. (1957): The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, Vol. 120, pp. 253-281.
- [4] CHARNES, A., COOPER, W., LEWIN, A.Y., and SEIFORD, L.M. (eds.). Data Envelopment Analysis - Theory, Methodology and Applications. – Dordrecht: Kluwer, 1997, 513 pp.
- [5] COOPER, W.W., SEIFORD, L.M., and TONE, K. (2000). Data Envelopment Analysis - A Comprehensive Text with Models, Applications, References and DEA-Solver Software. – New York: Springer, 2007, 490 pp.
- [6] FARE, R., GROSSKOPF, S., and Lovell, C.A.K. Production Frontiers. – Cambridge: University of Cambridge Press, 1994, 296 pp.
- [7] T. ARSHINOVA, “The Application of Data Envelopment Analysis Approach to the Efficiency Measurement of Latvian Banks”, *Scientific Journal of Riga Technical University, Computer Science*, Vol.5, no. 39, p.92-104, 2009
- [8] Information about Baltic equity listed companies. Available: <http://www.nasdaqomxbaltic.com/market/?pg=mainlist&lang=lv>

Current address

Tatyana Arshinova, Mg.oec.

Department of Probability Theory and Mathematical Statistics,

Faculty of Computer Science and Information Technologies,

Riga Technical University, 1/4 Meza Street, Riga, LV-1048

Phone: (+371)29855850

e-mail: tatjana.arsinova@rtu.lv

GLMM MODEL OF AT-RISK-OF-POVERTY CZECH HOUSEHOLDS DEPENDING ON THE AGE AND SEX OF THE HOUSEHOLDER (EU-SILC 2005-2009)

BARTOŠOVÁ Jitka, (CZ), FORBELSKÁ Marie, (CZ)

Abstract. The paper presents a quantitative study of at-risk-of-poverty Czech households depending on the age and sex of the householder. We focus mainly on two age categories: juniors and seniors. At-risk-of-poverty rates of age categories are calculated as the proportion of households with an equivalised income below the poverty threshold, which is set at 60% of the national median equivalised income. We use Generalized Linear Mixed Models (GLMM) to model at-risk-of-poverty rates between 2005 and 2009. GLMM have become a very powerful and widely used statistical tool. The R environment (R Development Core Team, 2010) is used for GLMM analysis.

Keywords. Survey on income and living conditions, equivalised household income, generalized linear mixed models, poverty rate.

Mathematics Subject Classification: Primary 62H30, Secondary 30C40.

1 Introduction

The current economic crisis is having an important impact on social surveys. There is a pressure on statistics to provide updated information to monitor the extent of the crisis in the social field. It is the case of EU-SILC, the main source of comparable information on income and living conditions across Europe. As a consequence of the crisis, there has been an increase in unemployment and hence of poverty in the Czech Republic and elsewhere. The crisis has mostly worsened the financial situation of juniors and seniors. In these two categories, we can observe an increase in unemployment the most and therefore we can also expect the highest at-risk-of-poverty rates.

The European Union Statistics on Income and Living Conditions (EU-SILC) collects information about income, age and social structure, level of living and other characteristics of households and individuals yearly in all countries of EU. In the present paper, we use data EU-SILC 2005 – 2009. The **equivalised household income** is used to allow comparisons between households of different sizes and composition.

1.1 Poverty rate

The poverty definition adopted in this paper is the relative country-specific poverty measure: this views poverty in a nationally defined social and economic context. It is commonly measured as the percentage of population with cash income less than some fixed proportion (say, 60%) of national median income. Such relative poverty measures are now commonly used as the official poverty risk rate in EU-countries. The measurements are usually based on a household's yearly cash income and frequently take no account of household wealth, or inequality of resource distribution that may exist within a household. Household income includes earnings, transfers and income from capital, as well as the imputed rent for owner-occupied households, and is measured here net of direct taxes, social security contributions and interest on mortgages paid by households.

The data reported here are collected in the EU-SILC surveys that apply common conventions and definitions to collect unit record data. EUROSTAT supply detailed cross-tabulations of these results in their statistical database.

The **poverty rates** discussed here are defined as the percentage of those having less than 60% of the median income. At-risk-of-poverty rates are calculated as the proportion of households with an equivalised income below the poverty threshold, which is set at 60% of the national median equivalised income.

2 Model Specification

2.1 Generalized Linear Mixed Model

The Generalized Linear Mixed Models (GLMM) extends the Generalized Linear Model (GLM) by incorporating random effects into the linear predictor to accommodate random variations and correlations from different sources (McCulloch & Searle 2001).

Generalized linear models (GLMs) represent a class of fixed effects regression models for several types of dependent variables (i.e., continuous, dichotomous, counts). Common Generalized linear models (GLMs) include linear regression, logistic regression, and Poisson regression.

There are three specifications in a GLM. First, the linear predictor, denoted as η_i , of a GLM is of the form $\eta_i = \mathbf{x}_i\boldsymbol{\beta}$, where \mathbf{x}_i is the vector of regressors for unit i with fixed effects $\boldsymbol{\beta}$. Then, a link function g is specified which converts the expected value μ_i of the outcome variable Y_i (i.e., $\mu_i = EY_i$) to the linear predictor η_i , i.e., $g(\mu_i) = \eta_i$. Finally, a specification for the form of the variance in terms of the mean μ_i is made. The latter two specifications usually depend on the distribution of the outcome Y_i , which is assumed to fall within the exponential family of distributions.

Fixed effects models, which assume that all observations are independent of each other, are not appropriate for analysis of several types of correlated data structures, in particular, for clustered data. In clustered designs, subjects are observed nested within larger units. For analysis of such data, random cluster effects can be added into the regression model to account for the correlation of the data. The resulting model is a mixed model including the usual fixed effects for the regressors plus the random effects.

Let i denote the number of clusters and let j denote the nested observation. Assume there are $i = 1, \dots, N$ clusters and $j = 1, \dots, n_i$ repeated observations nested within each cluster. A random - intercept model, which is the simplest mixed model, augments the linear predictor with a single random effect for subject i , $\eta_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + v_i$, where v_i is the random effect (one for each cluster). These random effects represent the influence of cluster i , on its repeated observations that is not captured by the observed covariates. These are treated as random effects because the sampled clusters are thought to represent a population of clusters, and they are usually assumed to be distributed as $N(0, \sigma_v^2)$. The parameter σ_v^2 indicates the variance in the population distribution, and therefore the degree of heterogeneity of clusters. Including the random effects, the expected value of the outcome variable, which is related to the linear predictor via the link function, is given as $\mu_{ij} = E(Y_{ij} | v_i, \mathbf{x}_{ij})$. This is the expectation of the conditional distribution of the outcome given the random effects. As a result, GLMMs are often referred to as conditional models in contrast to the marginal generalized estimating equations (GEE) models.

The model can be easily extended to include multiple random effects. The model is now written as

$$\eta_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij} \mathbf{v}_i. \quad (1)$$

The vector of random effects \mathbf{v}_i is assumed to follow a multivariate normal distribution with mean vector $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}_v$. Note that the conditional mean μ_{ij} is now specified as $E(Y_{ij} | \mathbf{v}_i, \mathbf{x}_{ij})$, namely, in terms of the vector of random effects.

With N independent sampling units and conditionally on the random effects, assume that the responses are independent with density function that is a member of the exponential family $f(Y_{ij} | \mathbf{v}_i) = \exp[\{Y_{ij} \theta_{ij} - b(\theta_{ij})\} / a(\tau) + c(Y_{ij}, \tau)]$ for some functions a , b , and c .

Parameter estimation in GLMMs typically involves maximum likelihood (ML) or variants of ML. Additionally, the solutions are usually iterative ones that can be numerically quite intensive.

Several methods have been proposed for inference and estimation in the GLMM. Existing estimation methods for the GLMM include: (a) analytically simplifying the problem, for example, by the use of Laplace approximation to the integrated likelihood, including the penalized quasi-likelihood (PQL) estimation (Breslow and Clayton, 1993) and the hierarchical generalized linear models (HGLM) procedure (Lee and Nelder, 2001); (b) using computation-intensive techniques such as the MCEM algorithm (Booth and Hobert, 1999), Markov Chain Monte Carlo (MCMC) (Zeger and Karim, 1991) and Gauss–Hermite quadrature (GHQ) approaches (Pan and Thompson, 2003). The applications of these methods to date the EU-SILC, see for example in Bartošová, Forbelská, 2010 and 2011, Forbelská, 2010, Forbelská, Bartošová, 2010).

2.2 Mixed-Effects Logistic Regression Model

The mixed-effects logistic regression model is a common choice for analysis of multilevel dichotomous data and is arguably the most popular GLMM. In the GLMM context, this model utilizes the logit link, namely

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) = \log(\mu_{ij} / (1 - \mu_{ij})) = \eta_{ij}. \quad (2)$$

Here, the conditional expectation $\mu_{ij} = E(Y_{ij} | \mathbf{v}_i, \mathbf{x}_{ij})$ equals $P(Y_{ij} = 1 | \mathbf{v}_i, \mathbf{x}_{ij})$, namely, the conditional probability of a response given the random effects (and covariate values). This model can also be written as

$$P(Y_{ij} = 1 | \mathbf{v}_i, \mathbf{x}_{ij}, \mathbf{z}_{ij}) = g^{-1}(\eta_{ij}) = F_{\text{logist}}(\eta_{ij}), \quad (3)$$

where the inverse link function $F_{\text{logist}}(\eta_{ij})$ is the logistic cumulative distribution function (cdf), namely

$$F_{\text{logist}}(\eta_{ij}) = [1 + \exp(-\eta_{ij})]^{-1}. \quad (4)$$

A nicety of the logistic distribution, that simplifies parameter estimation, is that the probability density function (pdf) is related to the cdf in a simple way, as

$$f_{\text{logist}}(\eta_{ij}) = F_{\text{logist}}(\eta_{ij})[1 - F_{\text{logist}}(\eta_{ij})]. \quad (5)$$

3 Fitting GLMM model

A binary mixed logit model was used to analyse the risk monetary poverty for different age groups.

We consider a two factorial ANOVA model with a factor *year* measured at 5 levels (years 2005 – 2009) and a factor *sex* (*M* male, *F* female), i.e.

$$\eta_{ij} = (\alpha_{\text{year}} + u_i) \text{year}_{ij} + (\beta_{\text{sex}} + v_i) \text{sex}_{ij}, \quad (6)$$

where u_i, v_i are random parameters and $\alpha_{\text{year}}, \beta_{\text{sex}}$ are fixed parameters, $i = 1, \dots, 5$ (age groups).

3.1 Results of construction of GLMM model

Results for the mentioned model are presented in Table 1. The fitted parameters of the GLMM model are given in Tables 1 and 2.

Estimates of fixed effects α_{year} and β_{sex} , their variability and tests of importance of both factor, i.e. years and sex of the householder are shown in Table 1. The group considered in the model consists of households with a male householder in 2005. It is obvious that the chances that the household drops below the poverty line in this period have increased with the lowest value in 2006 and the highest in 2008. At the same time, we can see that standard errors of estimates of effects in individual years are high, which shows that these coefficients are statistically insignificant. This is also confirmed by tests of significance. On the contrary, the effect of sex of the householder is statistically significant. When the householder is female, the chances that the household drops below the poverty line significantly increases.

Table 1. Final estimation of the fixed effects

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.62148	0.335183	-7.82104	5.24E-15
year2006	-0.03157	0.056689	-0.55698	0.577539
year2007	0.12352	0.081366	0.51806	0.87934
year2008	0.05993	0.146056	0.410319	0.681572
year2009	0.014536	0.193081	0.075283	0.93999
SEX F	1.537917	0.184257	8.346599	7.03E-17

Table 2 containing prediction of random effects shows that the impact of age in 2005-2009 gradually changed. In the whole reported period, the chances of drop below the poverty line was the highest when the householder was younger than 25. At the other edge of the spectrum, at-risk-of-poverty rate decreased the most for households with householders over 60 years old. Random effects for households in the middle of the age spectrum (i.e. age categories (25,30], (30,55] and (55,60]) oscillated around zero.

Table 2. Prediction of the random effects

	(0,25]	(25,30]	(30,55]	(55,60]	(60,100]
(Intercept)	0.981	0.232	0.409	-0.416	-1.205
year2006	-0.032	-0.179	-0.076	0.140	0.147
year2007	0.163	-0.236	-0.187	0.215	0.045
year2008	0.132	-0.481	-0.285	0.349	0.285
year2009	0.095	-0.594	-0.403	0.457	0.446
SEX F	-0.530	0.301	-0.326	-0.052	0.607

AIC	BIC	logLik	deviance
737433.1	737556	-368690	737379.1

Table 3. Correlation components of the random-effects terms

	(Intercept)	year2006	year2007	year2008	year2009	SEX F
(Intercept)	1.000	-0.718	-0.169	-0.480	-0.580	-0.838
year2006	-0.718	1.000	0.762	0.940	0.964	0.270
year2007	-0.169	0.762	1.000	0.929	0.889	-0.214
year2008	-0.480	0.940	0.929	1.000	0.993	0.050
year2009	-0.580	0.964	0.889	0.993	1.000	0.168
SEX F	-0.838	0.270	-0.214	0.050	0.168	1.000

Correlation matrix (see Table 3) shows high correlation between years which is solved using GLMM.

Table 4. Poverty rates estimates (in %) by means of GLMM model

SEX	AGE	2005	2006	2007	2008	2009
Male	(0,25]	16.23	15.38	18.76	19.02	17.77
	(25,30]	8.40	6.91	6.83	5.68	4.88
	(30,55]	9.86	8.95	8.42	8.03	6.91
	(55,60]	4.57	5.07	5.68	6.73	7.13
	(60,100]	2.13	2.39	2.25	2.98	3.34
Female	(0,25]	34.69	33.25	38.75	39.16	37.20
	(25,30]	36.57	31.84	31.56	27.46	24.41
	(30,55]	26.89	24.82	23.60	22.69	19.97
	(55,60]	17.49	19.11	21.01	24.19	25.34
	(60,100]	15.69	17.28	16.46	20.80	22.77

In Table 4, there are poverty rates estimates (in %) by means of GLMM model. It is obvious that between 2005 and 2009 the households of seniors were the least endangered with relative poverty. The proportion of households below the poverty line with a male householder was 2.13%, 2.39%, 2.25%, 2.98% and 3.34%. In case of female householder, the proportion was about eight times higher - 15.69%, 17.28%, 16.46%, 20.80% and 22.77%.

On the contrary, households of juniors were the most endangered with relative poverty. When the householder was male, the proportion of households below the poverty line was 16.23%, 15.38%, 18.76%, 19.02%, 17.77%. In case of a female householder, the proportion was about twice as large - 34.69%, 33.25%, 38.75%, 39.16% and 37.20%.

3.2 Graphing the results

Figure 1 highlights the variations observed across 5 age-household groups.

Figure 1 shows, that the values of at-risk-of-poverty-rate modelled with GLMM with fixed and random effects was relatively stable over the monitored period and with respect to individual age categories with both male and female householders. Trend in at-risk-of-poverty rate in the junior category (0,25] was close to being constant. The following two age categories, (25,30] and (30,55] show decreasing trend, as opposed to the last two age categories (55,60] and (60,100]. We can see that the situation of seniors deteriorates.

The usual diagnostic plots are shown in Figures 2, 3 and indicate no particular problems (except M:(30,55]).

Calculations were performed in an R. Software environment R is a free, cooperatively developed, open-source implementation of the S statistical programming language and computing environment, a language that has become a de-facto standard among statisticians for the development of statistical software. In R one of the most popular functions for fitting GLMMs is called glmer and is found within the package lme4 which is written and maintained by Douglas Bates. Methodology can be found at [1].

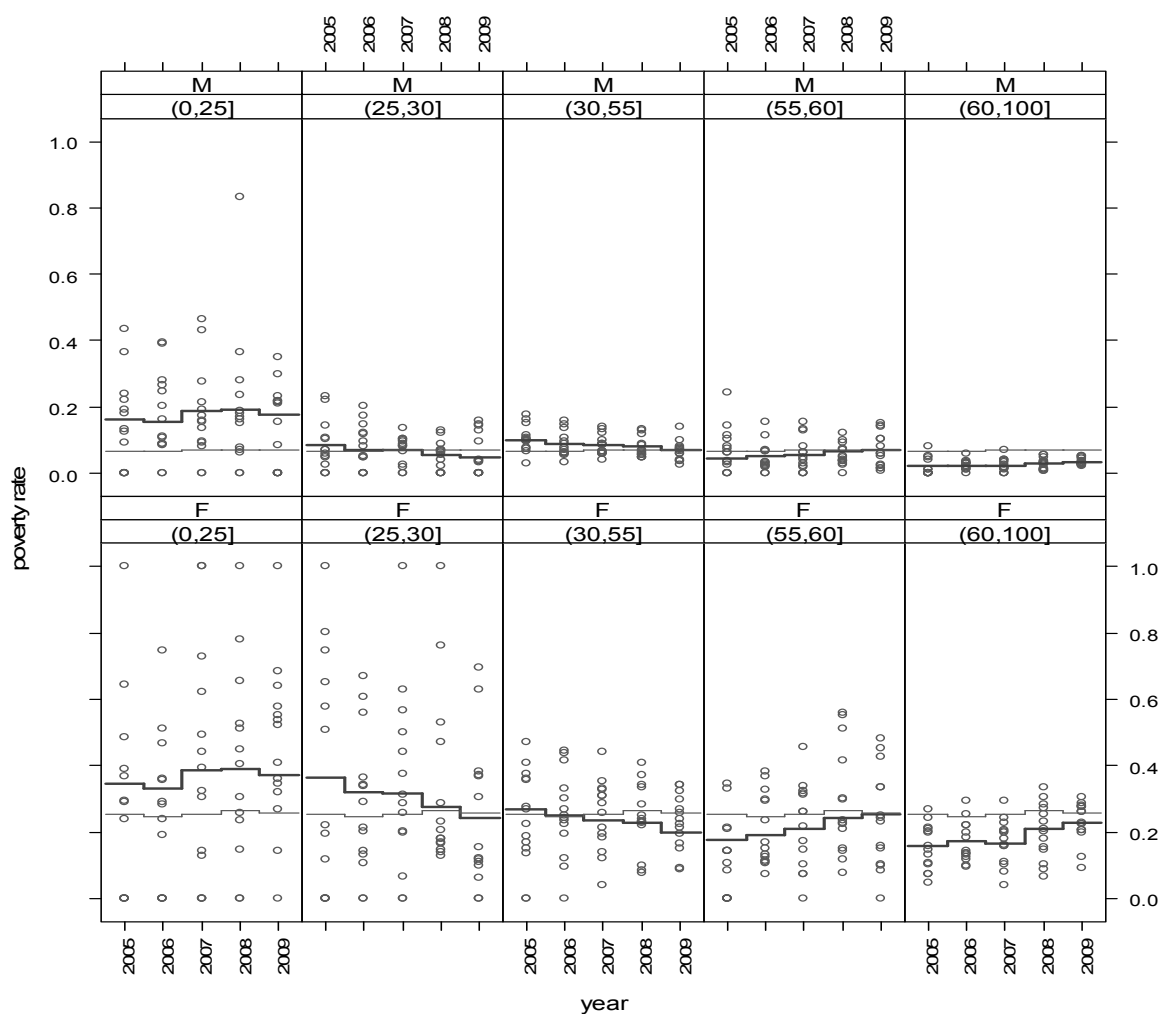


Figure 1. Trends in at-risk-of-poverty rates (black lines: poverty rates using fixed and random parameters, gray lines: poverty rates by means of fixed parameters)

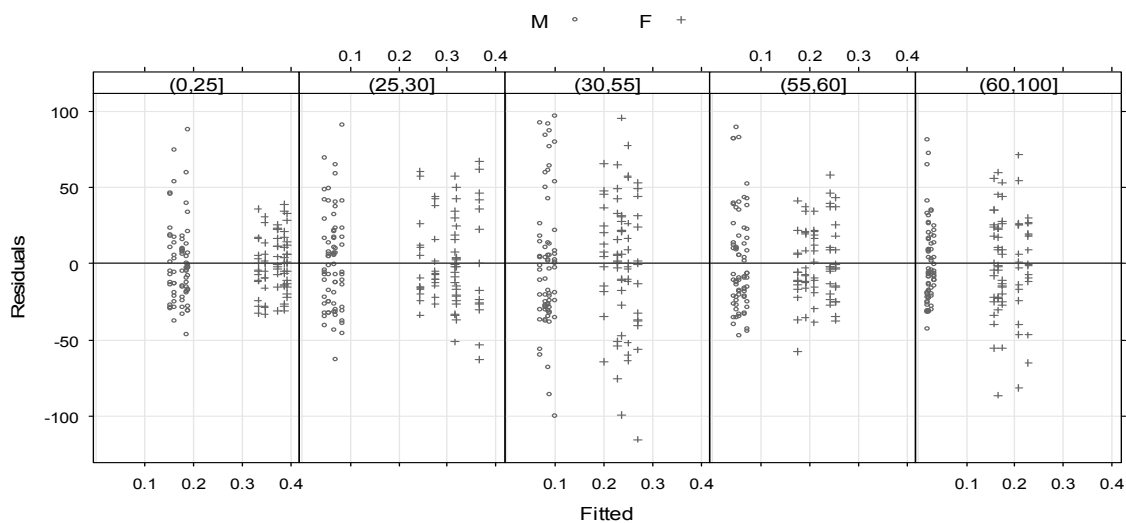


Figure 2: Residuals vs. Fitted plots.

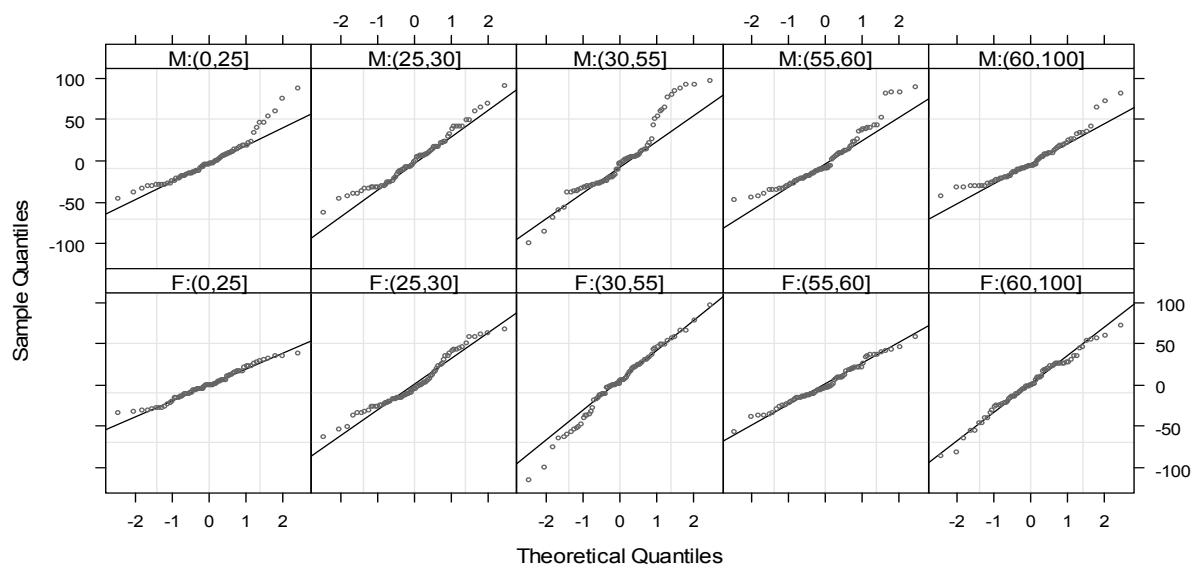


Figure 3: QQ plots.

4 Conclusions

Generalized Linear Mixed Model (GLMM) was used to model at-risk-of-poverty rates between 2005 and 2009. Our results show that households with householders at both ends of age spectrum (juniors up to 25 and seniors over 60) had extremely (i.e. minimal and maximal) at-risk-of-poverty rates between 2005 and 2009, regardless of sex. The proportion of junior households below the poverty line was about seven times higher than the proportion of senior households. Sex of the householder has significant impact on the proportion of households below the poverty line in all age categories and over the whole monitored period. A female householder doubles the risk at the junior category and raises the risk eight times at the senior category in comparison to a male householder.

Trend of at-risk-of-poverty-rate over the monitored period was quite stable in all age categories regardless of sex of the householder. The trend in the junior category is almost constant, the two following age categories show decrease and the two oldest categories increase. The situation of junior households is still unfavourable, the situation of the two middle-age categories (25-55] improves, while the situation of the oldest (55-100] gradually deteriorates.

Acknowledgement

The research was supported by project of Grant Agency of the Czech Republic no. 402/09/0515 with title: “Analysis and modelling of financial power of Czech and Slovak Households”.

References

- [1] BARTOŠOVÁ, J., FORBELSKÁ, M.: Mixture Model Clustering for Household Incomes. *APLIMAT - Journal of Applied Mathematics*, **3**: 163-172, 2010.

- [2] BARTOŠOVÁ, J., FORBELSKÁ, M.: Differentiation and dynamics of household incomes in the Czech EU-SILC survey in the years 2005 - 2008. *APLIMAT - Journal of Applied Mathematics*, 4: 145-460, 2011.
- [3] BATES, D.: lme4: *Mixed-effects modeling with r*. URL <http://lme4.r-forge.r-project.org/>
- [4] [book](#), 2010.
- [5] BATES, D., MAECHLER, M.: *lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-37*. <http://CRAN.R-project.org/package=lme4>, 2010
- [6] BOOTH, J.G., HOBERT, J.H.: Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm, *Journal of the Royal Statistical Society B* **62**: 265-285, 1999.
- [7] BRESLOW, N.E., CLAYTON, D.G.: Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9-25, 1993.
- [8] FORBELSKÁ, M.: Exploring the Regional Czech Household Income Dynamics Via Regression Mixtures. *APLIMAT - Journal of Applied Mathematics*, **3**: 1-10, 2010.
- [9] FORBELSKÁ, M., BARTOŠOVÁ, J.: Clustering of Czech Household Incomes Over Very Short Time Period. Proceedings in *COMPSTAT 2010*. Physica-Verlag, 1015-1022, 2010.
- [10] LEE, Y., NELDER, J.A., PAWITAN, Y.: *Generalized Linear Models with Random Effects*. London: Chapman & Hall, 2006.
- [11] [10] MCCULLAGH, C.E. , NELDER, J. A.: *Generalized linear models*. Boca Raton Chapman & Hall, 1989.
- [12] [11] MCCULLOCH, C.E. , SEARLE, S.R.: *Generalized, linear, and mixed models*. Wiley-Interscience, 2001.
- [13] [12] PAN, J., THOMPSON, R.: Quasi-Monte Carlo estimation in generalized linear mixed models . *Computational Statistics & Data Analysis*, **51**: 5765-5775. 2007.
- [14] [13] R Development Core Team: *A language and environment for statistical computing. R*. Foundation for Statistical Computing. Vienna, Austria, 2010. URL <http://www.R-project.org>
- [15] [14] ZEGER S.L., KARIM M.R.: Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*, **86**: 79-86, 1991.

Current address

Jitka Bartošová, RNDr., PhD.

University of Economics Prague, Department of Management of Information of the Faculty of Management, Jarošovská 1117/II, Jindřichův Hradec, 377 01, Czech Republic,
tel.: +420 384 417 221,
email: bartosov@fm.vse.cz

Marie Forbelská, RNDr., PhD.

Masaryk University, Department of Mathematics and Statistics of the Faculty of Science, Kotlářská 2, Brno, 611 37, Czech Republic,
tel.: +420 549 493 811
email: forbels@math.muni.cz

LATVIAN GDP: TIME SERIES FORECASTING USING VECTOR AUTO REGRESSION

BEZRUCKO Aleksandrs, (LV)

Abstract: The target goal of this work is to develop a methodology of forecasting Latvian GDP using ARMA (AutoRegressive-Moving-Average) and VAR methods. The paper follows up with the papers published in the proceedings of the APLIMAT journal in 2011 – see [1]. The algorithm is developed for finding optimal time series model for GDP forecasting. Latvian GDP data with quarterly observation frequency is taken as time series. ARMA and VAR Analysis of Latvian GDP, M2X and inflation indicators time series is performed. The set of models has been constructed. In order to check the accuracy of models, different residual tests are performed: autocorrelation, Portmaneteau, heteroscedasticity and normality of residual distribution. Models are compared in their forecast quality.

Keywords and phrases: time series, Gross Domestic Product, Inflation, VAR (Vector Auto Regression), Residual tests, Serial Correlation

1 Introduction

The urgency of use of VAR models for an estimation of macroeconomic indicators doesn't cause today any doubts. It has been proved by And Sims (one of the main promoters of the use of vector autoregression in empirical macroeconomics, and contributed to the development of Bayes estimators for vector autoregression) in the works on use of VAR-models, which have received the The Nobel Prize in Economic Sciences in 2011 "*for their empirical research on cause and effect in the macroeconomy*".

The analysis and forecast of GDP for any time and any country is important task for economists, policy makers and entrepreneurs. These researches are consisting of many objective and subjective factors.

The base of this work is taken from previous paper [1] and a research has continued taken into account Vector Auto Regression. Different methods of econometrical modelling have been analyzed. For example, analysis methods for German GDP forecast that are described by Lutkepohl in "Applied Time Series Analysis" [2]. Lutkepohl described different ways of ARMA and Residual analysis of time series. In this paper author uses familiar methods of statistical analysis of time series for forecasting Latvian GDP. Computer software enabled the author to perform the search for

the best models for certain time series. Based on the analysis of these models, a search algorithm of optimal model is created. A lot of attention is given to use VAR-models.

In order to find an optimal model of forecasting Latvian Gross Domestic Product, three different cases of Latvian GDP series with quarterly observation frequency are taken. The first case is quarterly Latvian GDP series in levels (Latvian lats), the second case is same data in percentage growth and in the third case different VAR models are build using Latvian GDP, Inflation and M2X cash mass data. The GDP series are given in Figure 1, Comparison with M2X and inflation data are given in Figure 2. The time series length is $T = 57$. The time series is taken from the first quarter of year 1995 till the first quarter of year 2009. All searches and forecasts are made using econometrical software EViews 6.0.

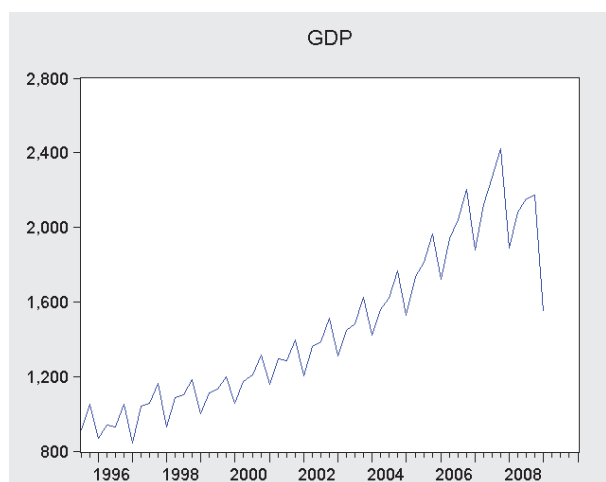


Figure 1 Latvian GDP (Lats) 1995Q1-2009Q1

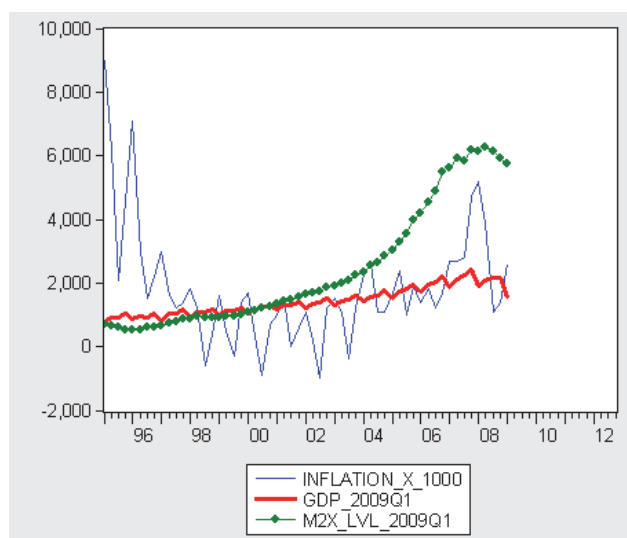


Figure 2 Inflation (%) / GDP (Lats) / M2X (Lats) 1995Q1-2009Q1

2 Analysis description

2.1. The analysis of criteria

ARMA case: At the first stage of choosing the best model, 3 criteria are analyzed: Akaike info, Schwarz and Hannan-Quinn.

VAR case: 2 criteria are analyzed: Akaike info and Schwarz

Constructing of ARMA and VAR models occurs in parallel, but separately from each other. The best model has minimal values. At this stage models with best criteria are taken. ARMA and VAR Analysis are performed in EViews program language and statistical criteria represent the result of the program (Fig.3).

Table: CRITS Workfile: GDP_AUTOREGRESSION_1:Untitled\					
View	Proc	Object	Print	Name	Edit+/-
CellFmt	Grid+/-	Title	Comments+/-		
A	B	C	D	E	
1	AR / MA	0.000000	1.000000	2.000000	3.000000
2	0.000000	0.347646	-0.239621	-1.415521	-1.446306
3	1.000000	-1.378436	-1.841098	-2.093701	-2.105267
4	2.000000	-1.662691	-2.019881	-2.076773	-2.204613
5					

Figure 3 ARMA Analysis in EViews 6.0

The R-squared (R^2) statistic measures the success of the regression in predicting the values of the dependent variable within the sample. In standard settings, R^2 may be interpreted as the fraction of the variance of the dependent variable explained by the independent variables. The statistic will equal one if the regression fits perfectly, and zero if it fits no better than the simple mean of the dependent variable. It can be negative for a number of reasons. For example, if the regression does not have an intercept or constant, if the regression contains coefficient restrictions, or if the estimation method is two-stage least squares or ARCH.

The Akaike Information Criterion (AIC) is computed as: $AIC = -2l/T + 2k/T$ where l is the log likelihood. The AIC is often used in model selection for non-nested alternatives-smaller values of the AIC are preferred. For example, you can choose the length of a lag distribution by choosing the specification with the lowest value of the AIC.

The Schwarz Criterion (SC) is an alternative to the AIC that imposes a larger penalty for additional coefficients: $SC = -2l/T + (k \log T)/T$

2.2. Residual tests

The second stage is represented by Residual tests:

ARMA case: Serial Correlation LM test, Histogram – Normality test, Heteroskedasticity ARCH test and Correlogram Square Residual test. Models have passed the test if P Value is higher than 0.1.

VAR case: VAR Residual Serial correlation LM test (Fig. 4.) and VAR Residual Portmanteau Tests for Autocorrelation.

Serial Correlation LM test is an alternative to the Q-statistics for testing serial correlation. The test belongs to the class of asymptotic (large sample) tests known as Lagrange multiplier (LM) tests. Serial Correlation LM test has the higher importance because on this step we are concerning with the possibility that our errors exhibit autocorrelation. LM test check for higher order ARMA errors and is applicable whether or not there are lagged dependent variables.

The null hypothesis of the LM test is that there is no serial correlation up to lag order p , where p is a pre-specified integer. The local alternative is ARMA(r, q) errors, where the number of lag terms $p = \max(r, q)$. Note that this alternative includes both AR(p) and MA(p) error processes, so that the test may have power against a variety of alternative autocorrelation structures.

The test statistic is computed by an auxiliary regression as follows. First, suppose you have estimated the regression:

$$y_t = \beta + e_t$$

where β are the estimated coefficients and e are the errors. The test statistic for lag order p is based on the auxiliary regression for the residuals $e = y - \hat{X}\beta$:

$$e_t = X_t\gamma + \left(\sum_{s=1}^p \alpha_s e_{t-s} \right) + v_t$$

Histogram and normality tests are displays a histogram and descriptive statistics of the residuals, including the Jarque-Bera statistic for testing normality. If the residuals are normally distributed, the histogram should be bell-shaped and the Jarque-Bera statistic should not be significant. The Jarque-Bera statistic has a χ^2 distribution with two degrees of freedom under the null hypothesis of normally distributed errors. [2]

The ARCH test is a Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) in the residuals. This particular heteroskedasticity specification was motivated by the observation that in many financial time series, the magnitude of residuals appeared

to be related to the magnitude of recent residuals. ARCH in itself does not invalidate standard LS inference. However, ignoring ARCH effects may result in loss of efficiency.

The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis that there is no ARCH up to order q in the residuals, we run the regression:

$$e_t^2 = \beta_0 + \left(\sum_{s=1}^q \beta_s e_{t-s}^2 \right) + v_t,$$

where e is the residual. This is a regression of the squared residuals on a constant and lagged squared residuals up to order q . The F-statistic is an omitted variable test for the joint significance of all lagged squared residuals. The Obs*R-squared statistic is Engle's LM test statistic, computed as the number of observations times the R^2 from the test regression. The exact finite sample distribution of the F-statistic under H_0 is not known, but the LM test statistic is asymptotically distributed as a $\chi^2(q)$ under quite general conditions.

Correlogram of squared residuals test displays the autocorrelations and partial autocorrelations of the squared residuals up to any specified number of lags and computes the Ljung-Box Q-statistics for the corresponding lags. The correlograms of the squared residuals can be used to check autoregressive conditional heteroskedasticity (ARCH) in the residuals.

If there is no ARCH in the residuals, the autocorrelations and partial autocorrelations should be zero at all lags and the Q-statistics should not be significant inclusion of ARMA terms. [2]

VAR Residual Serial Correlation LM T...

Null Hypothesis: no serial correlation ...

Date: 12/14/11 Time: 23:24

Sample: 1995Q1 2012Q4

Included observations: 55

Lags	LM-Stat	Prob
1	48.38674	0.0000
2	21.90037	0.0092
3	8.646018	0.4706
4	22.88185	0.0065
5	13.97551	0.1232
6	10.34989	0.3229
7	7.698525	0.5648
8	12.41758	0.1908
9	6.160805	0.7237
10	6.560951	0.6827
11	6.727117	0.6655
12	9.181946	0.4207

Probs from chi-square with 9 df.

Figure 4 VAR Residual Serial Correlation in EViews 6.0

Autocorrelation LM Test (Fig.4.)

Reports the multivariate LM test statistics for residual serial correlation up to the specified order. The test statistic for lag order h is computed by running an auxiliary regression of the residuals u_t on the original right-hand regressors and the lagged residual u_{t-h} , where the missing first h values of u_{t-h} are filled with zeros. See Johansen (1995, p. 22) for the formula of the LM statistic. Under the null hypothesis of no serial correlation of order h , the LM statistic is asymptotically distributed χ^2 with k^2 degrees of freedom.

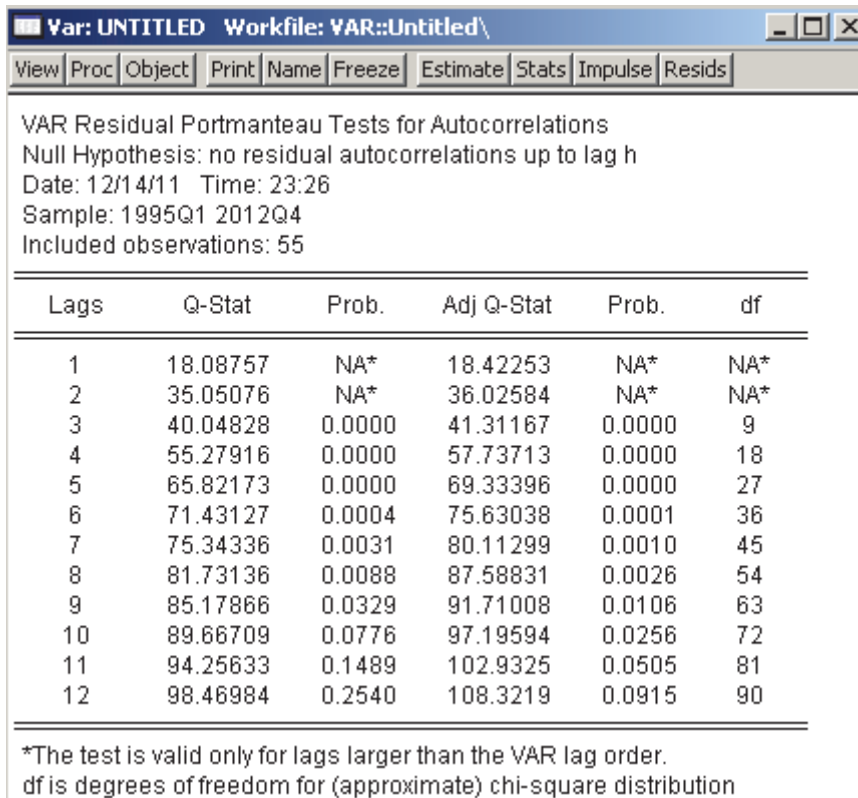


Figure 5 Portmanteau Tests in EViews 6.0

Portmanteau Autocorrelation Test (Fig.5.)

Computes the multivariate Box-Pierce/Ljung-Box Q -statistics for residual serial correlation up to the specified order (see Lütkepohl, 1991, 4.4.21 & 4.4.23 for details). We report both the Q -statistics and the adjusted Q -statistics (with a small sample correction). Under the null hypothesis of no serial correlation up to lag h , both statistics are *approximately* distributed χ^2 with degrees of freedom $k^2(h-p)$ where p is the VAR lag order. The asymptotic distribution is approximate in the sense that it requires the MA coefficients to be zero for lags $i > h-p$. Therefore, this approximation will be poor if the roots of the AR polynomial are close to one and h is small. In fact, the degrees of freedom becomes negative for $h < p$.

2.3. Out-Of-Sample Forecasting

The final evaluation test is “Out-Of-Sample Forecasting”. At this stage forecasts are compared to real data that we have for the period of the last 3 quarters of 2009.

3 The search algorithm

The search algorithm is shown in Figure 7. Step-by-step description looks as follows:
Input data: GDP Time series

1. Construction of ARMA models in levels and in differences and VAR models separately. Starting from this point the model is divided in two branches and the subsequent steps are carried out in parallel for levels and for differences.
2. Model ARMA Analysis.
3. Performing Residual tests.
4. If during residual tests probability value is less than 10%, the model is excluded from further evaluation. This is not a reliable model.
5. If P Value is higher than 10%, the forecast for specified periods of time is performed.
6. Out of Sample forecasting test. Comparing forecast data with real data.
7. The analysis of the results. Output data: Best model for the GDP Forecasts.

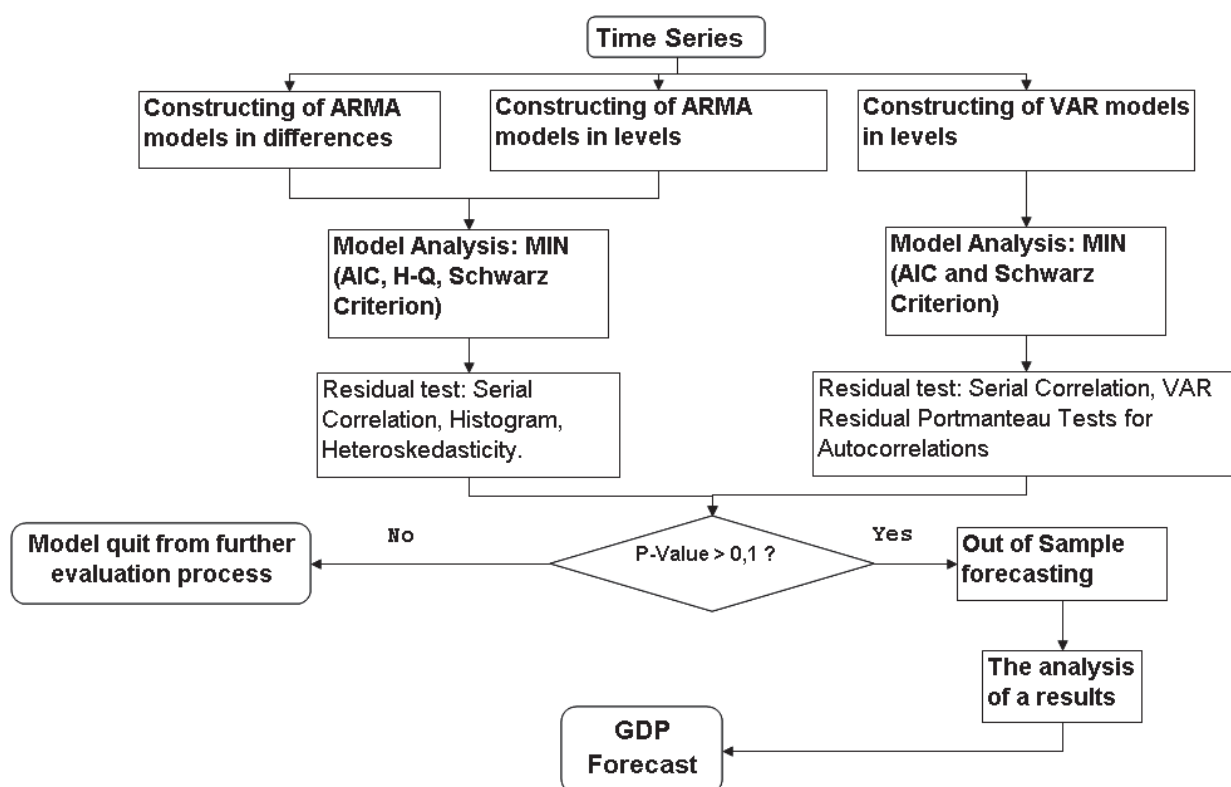


Figure 6 The search algorithm

4 Latvian GDP in levels

The analyzed series consist of seasonally unadjusted Latvian quarterly GDP in levels for the period of 1995Q1 – 2009Q1. It is represented in Figure 1. Constructing a model for the logs is more advantageous because the changes in the log series display a more stable variance than the changes in the original series. Time series in logs are shown in Figure 7.

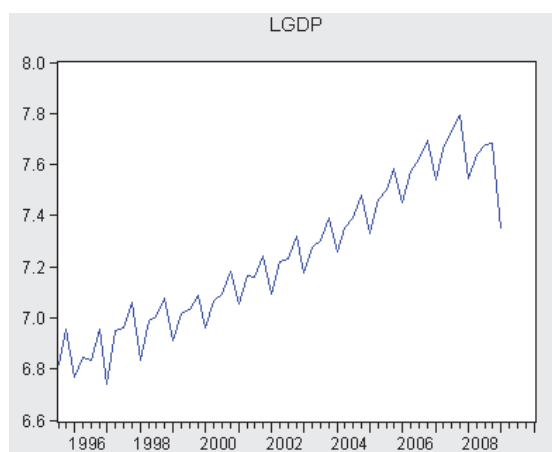


Figure 7 Latvian GDP in logs

Table 1

The analysis of criteria (Levels)

Nr.	LGDP	Akaike	Schwarz	Han-Quinn
1	Ar(1)	-1.326483	-1.290316	-1.312461
2	Ar(2)	-1.487976	-1.451479	-1.473862
3	Ma(1) C	-0.182363	-0.110677	-0.154503
4	Ma(2) C	-0.143123	-0.071437	-0.115263
5	AR(1) MA(1)	-1.878947	-1.806613	-1.850903
6	AR(1) AR(2) MA(1) MA(2)	-2.087592	-1.941604	-2.031137
7	AR(1) AR(2) SAR(4) MA(1)	-3.746621	-3.595105	-3.688722
8	AR(1) AR(2) SAR(4) MA(4)	-4.18207	-4.030555	-4.124172
9	Trend C	-1.914989	-1.843303	-1.887129
10	AR(1) Trend	-1.294331	-1.221997	-1.266287
11	AR(2) SAR(4)	-3.027406	-2.951648	-2.998456
12	AR(1) SAR(4) MA(4)	-3.646746	-3.534175	-3.603589
13	AR(1) AR(2) SAR(4) MA(4) SEAS(1)	-4.224953	-4.035559	-4.15258
14	AR(1) SAR(4) MA(4) D1997Q2	-4.047836	-3.89774	-3.990293

Best models: Nr. 5,7,8,12,13,14. Other models are excluded from further evaluation process. The residual test results are given in Table 2. Models Nr.13 and Nr.14 have undergone all tests. Residuals graph of Model Nr. 13 is given in Figure 4.

Table 2

Residual Test (Levels)			
Nr.	Serial Correlation	Histogram	Heteroskedasticity
5	0.69290	0.00000	0.61360
7	0.69350	0.00000	0.79830
8	0.08070	0.40000	0.53920
12	0.61610	0.00003	0.84840
13	0.09250	0.83015	0.87060
14	0.10670	0.40223	0.85250

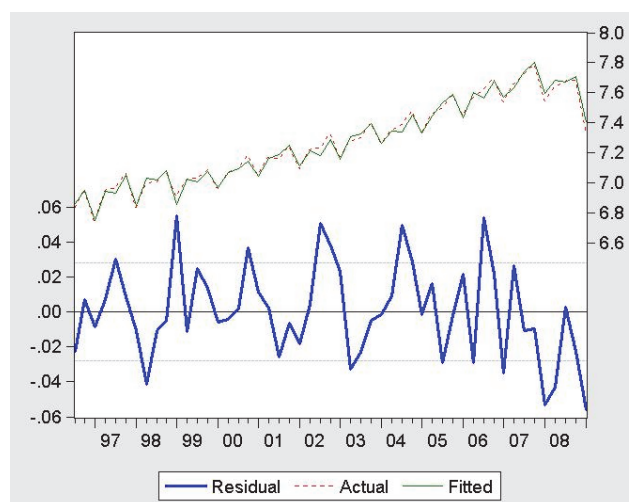


Figure 8 Residuals graph for Model Nr.13

The worst model residuals are given in Figure 5 for comparison.

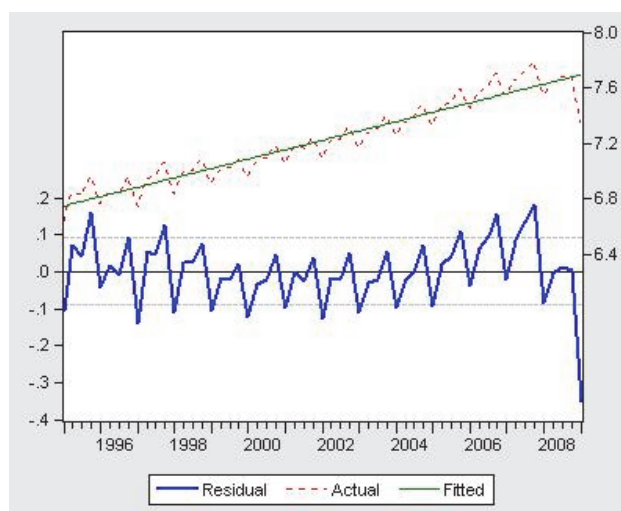


Figure 9 Residuals graph for Model Nr.9

“Out-Of-Sample Forecasting” test (Table 3) shows that the most real forecasts are gained from models Nr.12: AR(1) SAR(4) MA(4). Absolute difference (0.037) is minimal in this case. The best model has passed all Residual tests except Normality. The second and the third results are shown by models Nr. 7 and Nr.13., which have passed almost all residual tests except Nr.7., which also did not pass the Normality test.

Table 3

Out-Of-Sample Forecasting (Levels)

	Forecast - 1 Step - 09q2			Forecast - 2 Step - 09q3			Forecast - 3 Step - 09q4			
Nr.	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	Sum %
5	0.088	0.435	-0.347	0.031	0.022	0.008	0.034	0.018	0.016	0.371
7	0.088	0.082	0.007	0.031	0.017	0.014	0.034	-0.012	0.045	0.066
8	0.088	0.044	0.044	0.031	0.023	0.007	0.034	-0.020	0.054	0.106
12	0.088	0.103	-0.014	0.031	0.042	-0.012	0.034	0.023	0.011	0.037
13	0.088	0.057	0.031	0.031	0.017	0.014	0.034	-0.021	0.055	0.099
14	0.088	0.050	0.039	0.031	0.013	0.018	0.034	-0.019	0.053	0.109

5 Latvian GDP in percentage growth

All evaluations are made with Latvian GDP. The differences and the log difference of time series are shown in Figure 6. The log difference displays a more stable variance than the changes in the original series.

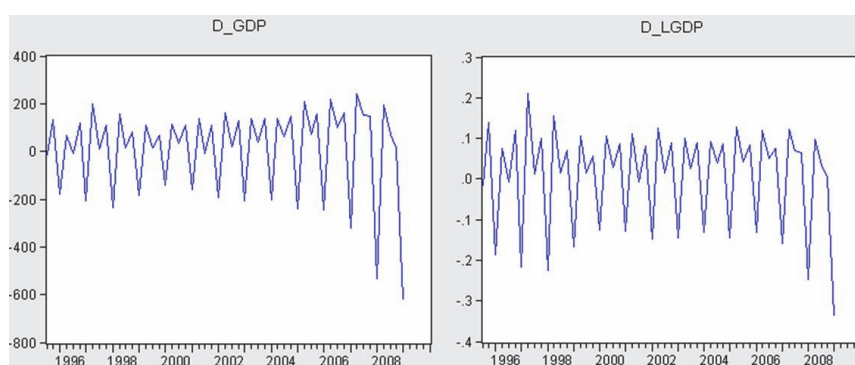


Figure 1 Residuals of time series in differences and in log differences

Table 4

The analysis of criteria (Difference)

Nr.	d (LGDP)	Akaike	Schwarz	Han-Quinn
1	Ar(1)	-1.659793	-1.623296	-1.645679
2	Ar(2)	-1.379102	-1.342269	-1.364897
3	Ma(1) C	-1.876758	-1.804424	-1.848714
4	Ma(2) C	-1.309161	-1.236827	-1.281117
5	AR(1) MA(1)	-1.712348	-1.639354	-1.684121
6	AR(1) AR(2) MA(1) MA(2)	-2.033619	-1.886286	-1.976798
7	AR(1) AR(2) SAR(4) MA(1)	-3.728444	-3.575482	-3.670195
8	AR(1) AR(2) SAR(4) MA(4)	-4.197451	-4.044489	-4.139202
9	Trend C	-1.306934	-1.2346	-1.278891
10	AR(1) Trend	-1.636801	-1.563807	-1.608574
11	AR(2) SAR(4)	-3.803588	-3.727107	-3.774463
12	AR(1) SAR(4) MA(4)	-4.191773	-4.078136	-4.148349
13	AR(1) AR(2) SAR(4) MA(4) SEAS(1)	-4.307962	-4.116759	-4.235151
14	AR(1) SAR(4) MA(4) D1997Q2	-4.273152	-4.121636	-4.215253

Models with the best criteria: Nr. 7, 8, 11, 12, 13, 14. Other models are excluded from further evaluation process.

Table 5

<i>Residual Test (Difference)</i>			
Nr.	Serial Correlation	Histogram	Heteroskedasticity: ARCH
7	0.6938	0	0.825
8	0.6895	0.393373	0.3453
11	0.8852	0	0.017538
12	0.0503	0.421976	0.6305
13	0.64	0.560093	0.2823
14	0.5554	0.997972	0.7499

Models Nr. 8, 13, 14 have completed all tests. Model Nr. 12 also has good statistic. “Out-Of-Sample” forecasting test (Table 6) shows that the most real forecasts are gained from models Nr.11: AR(2) SAR(4). Absolute difference (0.037) is minimal in this case. This model did not complete the histogram test, but passed all other residual tests. The second result belongs to model Nr.7, which has the same problem with Normality test. Third result is shown by model Nr.14 – this model passed all residual tests.

Table 6

<i>Out-Of-Sample Forecasting (Difference)</i>										
Nr.	Forecast - 1 Step -09q2			Forecast - 2 Step -09q3			Forecast - 3 Step -09q4			Sum %
	dLGDP real	dLGDPf	Diff	dLGDP real	dLGDPf	Diff	LGDP real	LGDPf	Diff	
7	0.088	0.084	0.004	0.031	0.022	0.009	0.034	0.004	0.030	0.043
8	0.088	0.034	0.054	0.031	0.019	0.012	0.034	-0.020	0.054	0.120
11	0.088	0.091	-0.003	0.031	0.024	0.007	0.034	0.007	0.027	0.037
12	0.088	0.045	0.043	0.031	0.024	0.007	0.034	-0.020	0.054	0.104
13	0.088	0.022	0.067	0.031	0.009	0.022	0.034	-0.021	0.054	0.143
14	0.088	0.070	0.018	0.031	0.011	0.020	0.034	-0.020	0.054	0.091

6 Vector auto regression

The analyzed series consist of seasonally unadjusted Latvian GDP, Inflation and M2X cash mass data indicators in for the period of 1995Q1 – 2009Q1 with quarterly observation. It is represented in Figure 2. M2X and GDP data are constructed in logs. Constructing a model for the logs is more advantageous because the changes in the log series display a more stable variance than the changes in the original series. Inflation time series is taken as is, because in some points it has negative values. Time series are shown in Figure 2.

Table 7

The analysis of criteria and residual tests for VAR models

Nr.	GDP	Akaike	Schwarz	Serial Correlation	VAR Residual Portmanteau Tests for Autocorrelations
1	C	-3.19764	-2.94216	0.42070	0.09150
2	C Trend	-2.18555	-1.89358	0.29130	0.07320
3	C SEAS(1)	-3.48239	-3.19041	0.99600	0.09250
4	C SEAS(1) SEAS(2) SEAS(3)	-3.67567	-3.3107	0.94820	0.14090
5	C Trend SEAS(1) SEAS(2) SEAS(3)	-3.6687	-3.26723	0.83680	0.09700
6	C Trend SEAS(1) SEAS(2)	-3.54695	-3.18198	0.92760	0.07230
7	C SEAS(1) SEAS(2)	-3.4629	-3.13442	0.9883	0.1457
8	C SEAS(2) SEAS(4)	-3.26908	-2.94061	0.893	0.1733
9	Trend	-1.97041	-1.71493	0.00640	0.00000
10	Trend SEAS(1)	-3.56915	-3.27717	0.60660	0.00790
11	Trend SEAS(1)	-3.56854	-3.24007	0.9648	0.0524

Model analysis and the residual test results are given in Table 7. After carrying out of the analysis of model it is appreciable that best models are Nr. 4,5,10. In the residual tests case best model are Nr. 4,7,8. In Figure 7 we can see residual analysis for model Nr.1 that are performed in EViews software.

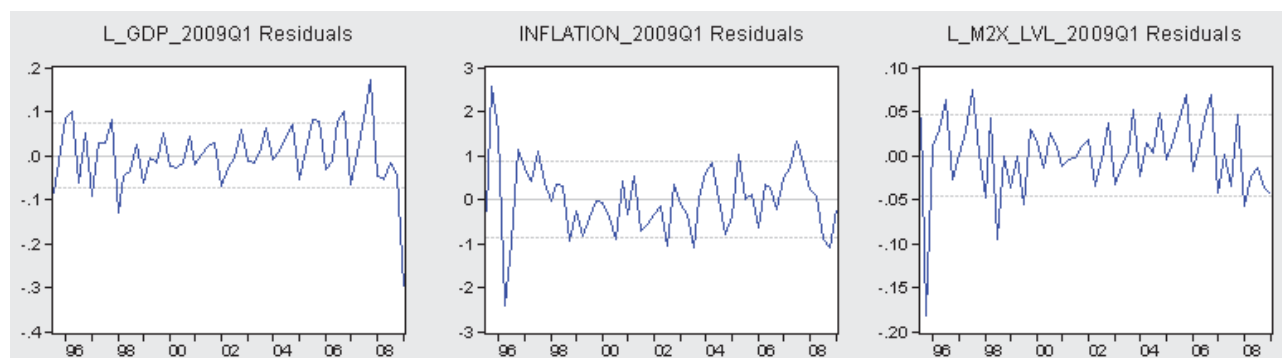


Figure 7 Residuals of time series in VAR model (Nr.1.)

Table 8

Out-Of-Sample Forecasting (VAR)

	Forecast - 1 Step - 2009q2			Forecast - 2 Step - 2009q3			Forecast - 3 Step - 2009q4			
Nr.	LGDP real	LGDPf	Diff	dLGDP re	dLGDPf	Diff	LGDP rea	LGDPf	Diff	Sum %
1	0.08837	0.13140	-0.04303	0.03073	0.07745	-0.04672	0.03364	0.04118	-0.00754	0.09729
2	0.08837	0.36690	-0.27853	0.03073	-0.00229	0.03302	0.03364	0.00154	0.03211	0.34366
3	0.08837	0.14402	-0.05566	0.03073	-0.00366	0.03439	0.03364	0.04070	-0.00706	0.09710
4	0.08837	0.18926	-0.10089	0.03073	0.00038	0.03035	0.03364	0.07325	-0.03961	0.17084
5	0.08837	0.16749	-0.07912	0.03073	-0.01937	0.05010	0.03364	0.07205	-0.03840	0.16763
6	0.08837	0.13501	-0.04664	0.03073	-0.06039	0.09111	0.03364	0.03107	0.00257	0.14032
7	0.08837	0.17461	-0.08624	0.03073	-0.02411	0.05484	0.03364	0.06105	-0.02741	0.16849
8	0.08837	0.29866	-0.21030	0.03073	0.03009	0.00064	0.03364	0.01722	0.01642	0.22735
9	0.08837	0.18843	-0.10006	0.03073	-0.13286	0.16358	0.03364	0.10985	-0.07620	0.33985
10	0.08837	0.14931	-0.06094	0.03073	0.00365	0.02707	0.03364	0.04039	-0.00675	0.09476
11	0.08837	0.10749	-0.01912	0.03073	-0.04182	0.07255	0.03364	0.01343	0.02022	0.11188

“Out-Of-Sample Forecasting” test (Table 8) shows that the most real forecasts are gained from models Nr. 10. The absolute difference is equal to 9,476 percents. The second and the third results are shown by models Nr. 3 and Nr.1. The model Nr.1 has high results in the analysis of models (Akaike and Schwarz Criterion) and in Serial Correlation but did not pass the Portmanteau test.

Conclusion

In the given paper the search algorithm of optimal time series is described. With a help of statistical modelling the econometric analysis of Latvian GDP is done. Different cases of constructing model are made in levels, in percentage growth and constructing a Vector Auto Regression model.

Comparison of 1 and 2 cases shows that case in levels and in differences gave approximately the same result – 3.7% deviation from real data in absolute value for forecasts for 3 steps in future.

Comparison of ARMA and VAR models shows that ARMA models win the competition. If we analyse tables of results of models in all three cases we will see that VAR models have the best dispersion of results.

The conclusion is that very important to analyze all the results and understand how they are calculated and evaluated. Data received as a result of this work will be used for the further researches in a scope of Vector Auto Regress for forecasting of macroeconomic variables.

References

- [1] BEZRUCKO A. Latvian GDP: The Optimal Time Series Forecasting Algorithm. – (APLIMAT Proceeding and Journal (2011), pp 477-486)
- [2] LUTKEPOHL H., KRATZIG M. Applied Time Series Econometrics. – Cambridge: Cambridge University press, 2004, 323 pp.
- [3] EViews software documentation: Quantitative Micro Software– <http://www.eviews.com> – Resource is described 2010, 30 of august.
- [4] NOSKO V., Econometrics: introduction to regression analysis of time series. – Moscow: Moscow Institute of Physics and Technology , 2002, 273 pp.
- [5] MOLCANOV I., ARZENOVSKI S. Statistical methods of forecasting. – Rostov-na-Donu: Rostov State University, 2001, 74 pp.

Current address

Bezrucko Aleksandrs, Mgr.

Chair of the theory of probability and the mathematical statistics

Faculty of Computer Science and Information Technology

Riga Technical University, 1/4 Meza Street, LV-1048 Riga, Latvia

e-mail: bezrucko@gmail.com

A METHOD TO APPROXIMATE FIRST PASSAGE TIMES DISTRIBUTIONS IN DIRECT TIME MARKOV PROCESSES

FERREIRA Manuel Alberto M., (PT), ANDRADE Marina, (PT)

Abstract. A numerical method to approximate first passage times distributions in direct Markov processes will be presented. It is useful to compute sojourn times in queue systems, namely in Jackson queuing networks. Using this method (Kiessler *et al.*, 1988) achieved to clear a problem that arises in the Jackson three node acyclic networks sojourn times.

Key words. randomisation procedure, sojourn time, Jackson three node acyclic networks

1 Introduction

In this work it will be described a general method, which key is the proceeding called, in the English language literature, “randomisation procedure” to approximate “first passage times” distributions in direct time Markov Processes, being the sojourn times in queue systems a particular case.

Call $\mathfrak{X} = \{X(t): t \geq 0\}$ a regular Markov Process, in continuous time with a countable states space E and a bounded matrix infinitesimal generator Q .

The elements of Q are designated by

$$Q(x, y), x, y \in E \text{ and } Q(x) = \sum_{y \in E - \{x\}} Q(x, y).$$

$\psi(t)$ designates the $X(t)$ state probability vector:

$$\psi_t(x) = P\{X(t) = x\}, x \in E \quad (1).$$

X models, for instance, the evolution of a queue system during the sojourn of a given, “marked”, customer in it.

The states of E have two main components:

- i) The queue system state,
- ii) The “marked” customer position.

Call

- A the states subset that describes the system till the departure of the “marked” customer, and
- B the state subset that describes the system after the departure of that customer.

Evidently

- $\{A, B\}$ is a partition of E ,
- If T is the time that the process \aleph spends in A till attaining B , for the first time, T is precisely the sojourn time of the “marked” customer in the queue system.

It is supposed that \aleph will remain in B , with probability 1 after having attained it for the first time. In fact, as the evolution of the system after the departure of the “marked” customer is irrelevant, it may be supposed that B is a closed set. That is, the process \aleph cannot come back to A after reaching B . The quantity of interest is the T distribution function, $\tau(t)$. Note that

$$\tau(t) = P\{T \leq t\} = P\{X(t) \in B\} = 1 - P\{X(t) \in A\}, t \geq 0 \quad (2)$$

since the presented hypotheses guarantee that $\{T \leq t\} = \{X(t) \in B\}$.

After (2) it is concluded that

- The problem of computing $\tau(t)$ is equivalent to the one of the computation of the transient distribution of $X(t)$ in A .

2 The Randomisation Procedure

From section 1 it follows that it is necessary to compute the vector $\psi_t, t \geq 0$. Being $P_t, t \geq 0$, the \aleph n transition matrix,

$$\psi_t = \psi_0 P_t, t \geq 0 \quad (3)$$

and

$$P_t = \exp(Qt) = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n, t \geq 0 \quad (4).$$

The “randomisation procedure” consists in using in (4) an equivalent representation, see (Çınlar, 1975):

$$P_t = \exp(-\alpha t) \exp\left(\alpha t \left(I + \frac{1}{\alpha} Q\right)\right) = \exp(-\alpha t) \sum_{n=0}^{\infty} \frac{\alpha^n t^n}{n!} R^n \quad (5)$$

where

$$R = I + \frac{1}{\alpha} Q \quad (6)$$

is called the “randomised matrix” in English language literature,

- I is the identity matrix, and
- α is a positive majorant of the whole $Q(x), x \in E$.

Note that, see (Melamed and Yadin, 1984, 1984a),

- Although the equation (5) seems more complex than (4), fulfils in fact more favourable computational properties. The most important is that R is a stochastic matrix while Q is not. Consequently, the computation using (5) is stable and using (4) is not,
- The “randomisation procedure” has an interesting probabilistic meaning, useful to determine bounds for $\tau(t)$. In fact, being R a stochastic matrix, it defines a discrete time Markov Process

$$\mathfrak{J} = \{Y_n : n = 0, 1, \dots\} \quad (7)$$

if it is assumed $Y_0 = X(0)$. With this procedure, the relation between the processes \mathfrak{X} and \mathfrak{J} is quite simple as it will be seen next.

Extending the discrete time process \mathfrak{J} to a continuous time Markov Process such that

- The time intervals between jumps are exponential random variables i.i.d. with mean
- The jumps are commanded by R .

In (Melamed and Yadin, 1984) it is shown that the resulting process is precisely the original process \mathfrak{X} ; but when there is a sequence of jumps in \mathfrak{J} from the state $x \in E$, this will be noticed in \mathfrak{X} as a long sojourn in state x .

So, the “randomisation procedure” may be interpreted as a sowing in the process \mathfrak{X} with “fake” random jumps between the true jumps. The resulting process, designated by $\bar{\mathfrak{X}}$, at which the “fake” jumps are visible, has the same probabilistic structure than \mathfrak{X} but with an advantage:

- The sequence of the jump instants in $\bar{\mathfrak{X}}$, “fake” and “true”, is now a Poisson Process. This is not, in general, the case of \mathfrak{X} .

Note that Y_n is the state of $\bar{\mathfrak{X}}$ in the instant of the n^{th} jump, “fake” or “true”.

Suppose that $\bar{\mathfrak{X}}$ reaches the set B in its n^{th} jump. Consequently the $\bar{\mathfrak{X}}$ sojourn time, and so also the \mathfrak{X} , in A is the sum of n exponential independent random variables with mean $\frac{1}{\alpha}$. That is the sojourn time has a n order Erlang distribution with parameter α . Its distribution function will be designated $E_{n,\alpha}(t)$.

Be $h(n)$ the probability that $\bar{\mathfrak{X}}$ reaches B in its n^{th} jump. Call ϕ_n the state probability vector of Y_n :

$$\phi_n = \psi_0 R^n \quad (8).$$

The quantities $h(n)$ are given by the equivalent formulae:

$$h(n) = \begin{cases} \sum_{x \in B} \phi_0(x), n = 0 \\ \sum_{x \in A} \sum_{y \in B} \phi_{n-1}(x) R(x, y), n > 0 \end{cases} \quad (9)$$

or

$$h(n) = \begin{cases} 1 - \sum_{x \in A} \phi_0(x), n = 0 \\ \sum_{x \in A} \phi_{n-1}(x) - \sum_{x \in A} \phi_n(x), n > 0 \end{cases} \quad (10).$$

Given the probabilities $h(n)$ and, noting that $\sum_{n=0}^{\infty} h(n) = 1$, it is obtained

$$\tau(t) = \sum_{n=0}^{\infty} h(n) E_{n,\alpha}(t), t \geq 0 \quad (11),$$

$$E[T^m] = \frac{1}{\alpha^m} \sum_{n=0}^{\infty} n(n+1) \dots (n+m-1) h(n), m = 1, 2, \dots \quad (12).$$

The formula (12) for $m = 1$ is

$$E[T] = \frac{1}{\alpha} E[H] \quad (13)$$

being H the number of \aleph jumps till reaching B . Expression (13) is the Little's Formula in this context.

Equation (11) allows obtaining simple bounds for $\tau(t)$ that may, in principle, to become arbitrarily close. Equation (12) allows to obtain $E[T^k]$, in principle, so close of $E[T^k]$ as wished. So, given any integer $k \geq 0$

$$L_k(t) \leq \tau(t) \leq U_k(t) \quad (14)$$

where

$$L_k(t) = \sum_{n=0}^k h(n) E_{n,\alpha}(t), t \geq 0 \quad (15),$$

$$U_k(t) = 1 - \sum_{n=0}^k h(n) \bar{E}_{n,\alpha}(t), t \geq 0 \quad (16)$$

and

$$E[T^m]_{L,k} \leq E[T^m], m = 1, 2, \dots \quad (17)$$

where

$$E[T^m]_{L,k} = \frac{1}{\alpha^m} \sum_{n=0}^k n(n+1) \dots (n+m-1) h(n), m = 1, 2, \dots \quad (18).$$

It is easy to prove that

Proposition

If, for any $\varepsilon > 0$, k is chosen in accordance with the rule

$$k = \min \left\{ n \geq 0 : \sum_{i=0}^n h(i) \geq 1 - \varepsilon \right\} = k(\varepsilon), \quad (19)$$

or equivalently

$$J = \min \left\{ n \geq 0 : \sum_{x \in A} \phi_n(x) \leq \varepsilon \right\} = J(\varepsilon), \quad (20)$$

$$|\tau(t) - L_{J(\varepsilon)}| \leq \varepsilon \text{ and } |\tau(t) - U_{J(\varepsilon)}| \leq \varepsilon, \text{ uniformly in } t \geq 0. \blacksquare$$

3 Concluding Remarks

The main problem in the application of the method presented, that in principle would solve any computation problems related to the distribution of sojourn times, stays in the difficulty of the $h(n)$ computation. In fact, for it, it is necessary to compute the vectors ϕ_n but only in the subset A of the states space. When states space E is finite, as it happens, for instance in the case of closed queue networks, both $h(n)$ and ϕ_n can, at first glance, be computed exactly, apart the mistakes brought by the approximations.

In practice the states space is often infinite or, although finite prohibitively great. In this situations it is mandatory to truncate E . So, it must be considered a new level of approximation since the $h(n)$, ϕ_n , etc. must also be approximated now.

In fact, what is viable to obtain is $h(n)$ minorants because the E truncation is translated in probability loss (Melamed and Yadin, 1984a). So, with these $h(n)$ approximate values, (14) and (17) go on being valid but

- The uniform convergence property seen above is lost,
- The rules analogous to (19) and (20) are not equivalent. The one generated by (19) may be even unviable and in practice it is used only the one generated by (20) (Melamed and Yadin, 1984a).

Using this method (Kiessler *et al.*, 1988) achieved to show that, in a Jackson three node acyclic network, see Figure 1, the total sojourn time distribution function for a customer that follows

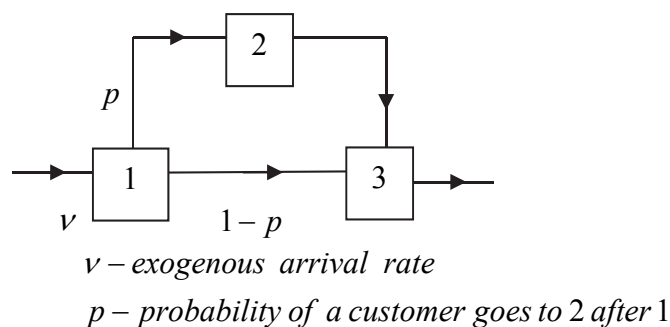


Figure 1: **Jackson Three Node Acyclic Network**

the path integrated by the nodes 1, 2, and 3 is not the same obtained considering that S_1, S_2 and S_3 (the sojourn times at nodes 1, 2 and 3 respectively) are independent although this one, designated by $F(t)$, is a “good” approximation of that one. They show that in some particular cases it was not true that

$$F^L(t) \leq F(t) \leq F^U(t), t \geq 0 \quad (21)$$

being $F^L(t)$ and $F^U(t)$ the minorant and the majorant, respectively, of that customer sojourn time distribution function, obtained through the described method.

This conclusion is important because, in spite of the dependence between S_1 and S_3 , see for instance (Ferreira, 2010), $F(t)$ could be the $S = S_1 + S_2 + S_3$ distribution function. In fact, (Feller, 1966) presents an example of dependent random variables which sum has the same distribution as if the random variables were independent.

Finally note that the formula (12), apparently new, seems to be of great efficiency, although only allows to obtain moments minorants, because its field of application is a broad one.

References

- ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol.. 1 (1), pp 96-98, 2010.
- [2] ÇINLAR, E.: *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- [3] DISNEY, R. L. and KÖNIG, D.: *Queueing Networks: A Survey of their Radom Processes*. SIAM Review, 3, pp. 335-403, 1985.
- [4] FELLER, W.: *An Introduction to Probability Theory and its Applications*. Vol. II, John Wiley & Sons, New York, 1966.
- [5] FERREIRA, M. A. M.: *O Tempo de Permanência em Redes de Jackson*. Revista de Estatística, 2 (2), pp. 25-44, 1997.
- [6] FERREIRA, M. A. M.: Correlation Coefficient of the Sojourn Times in Nodes 1 and 3 of Three Node Acyclic Jackson Network. Proceedings of the Third South China International Business Symposium. Vol. 2. Macau, Hong-Kong, Jiangmen (China), pp. 875-881, 1998.

- [7] FERREIRA, M. A. M.: *A note on Jackson Networks Sojourn Times*. Journal of Mathematics and Technology, vol. 1 (1), pp 91-95, 2010.
- [8] FERREIRA, M. A. M. and ANDRADE, M.: *Fundamentals of Theory of Queues*. International Journal of Academic Research, Vol. 3 (1, part II), pp. 427-429, 2011.
- [9] FERREIRA, M. A. M. and ANDRADE, M.: *Stochastic Processes in Networks of Queues with Losses: A Review*. International Journal of Academic Research, Vol. 3 (2, part IV), pp. 989-991, 2011.
- [10] FERREIRA, M. A. M. and ANDRADE, M.: *Grouping and Reordering in a Server Series*. Journal of Mathematics and Technology, Vol 2 (2), pp. 4-8, 2011a.
- [11] FERREIRA, M. A. M. and ANDRADE, M.: *Non-Homogeneous Networks of Queues*. Journal of Mathematics and Technology, Vol 2 (2), pp. 24-29, 2011b.
- [12] FOLEY, R. D. and KIESSLER, P. C.: *Positive Correlations in a Three-Node Jackson Queueing Network*. Advanced Applied Probability, 21, pp. 241-242, 1989.
- [13] JACKSON, J. R.: *Network of Waiting Lines*. Operations Research, 5, pp. 518-521, 1957.
- [14] JACKSON, J. R.: *Jobshop-like Queueing Systems*. Management Science, 10, pp. 131-142, 1963.
- [15] KELLY, F. P.: *Reversibility and Stochastic Networks*. John Willey & Sons, New York, 1979.
- [16] KIESSLER, P. C., MELAMED, B., YADIN, M. and FOLLEY, R. D.: *Analysis of a Three Node Queueing Network*. Queueing Systems, 3, pp. 53-72, 1988.
- [17] LEMOINE, A. J.: *On Sojourn Time in Jackson Networks of Queues*. Journal of Applied Probability, 24, pp. 495-510, 1987.
- [18] MELAMED, B. and YADIN, M.: *Randomization Procedures in the Computation of Cumulative-Time Distributions over Discrete State Markov Processes*. Operations Research, 4 (32), pp. 926-944, 1984.
- [19] MELAMED, B. and YADIN, M.: *Numerical Computation of Sojourn-Time Distributions in Queueing Networks*. Journal of the Association for Computing Machinery, 4 (31), pp. 839-854, 1984a.

Current address

Marina Andrade, Professor Auxiliar

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

Telefone: + 351 21 790 34 05

Fax: + 351 21 790 39 41

e-mail: marina.andrade@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

TELEFONE: + 351 21 790 37 03

FAX: + 351 21 790 39 41

e-mail: manuel.ferreira@iscte.pt

SOJOURN TIMES IN JACKSON NETWORKS

FERREIRA Manuel Alberto M., (PT), ANDRADE Marina, (PT)

Abstract. Jackson queuing networks have a lot of practical applications, mainly in the modelling of computation and telecommunications networks. Evidently the time that one customer - a person, a job, a message ... – spends in this kind of systems, its sojourn time, is an important measure of its performance. In this work the practical known results about the sojourn time distribution are collected and presented.

Key words: Jackson networks, sojourn time, randomisation procedure.

Mathematics Subject Classification: 60K35.

1 Introduction

In this work it is intended to present some problems and results that arise in the study of the sojourn time in Jackson networks of queues. These networks have many applications, namely in the modelling of computation and telecommunications networks. And a customer sojourn time, in this kind of system, is evidently an important element to be considered in its performance evaluation.

The model of network to be considered in this paper is briefly described in section 2. The main objective of section 3 is the presentation of formula (10) that, in some situations allows the sojourn times moments exact computation. In section 4 it is given a numerical method for the sojourn times distribution function and any order moments computation, adequate to any Jackson network.

2 General Results and Examples

Along this work the sojourn times in a class of Markovian networks of queues, introduced initially by Jackson, see (Jackson, 1957-1963), will be studied. They are called Jackson networks and have only one class of customers. They are composed of J nodes numbered $1, 2, \dots, J$. It is usual to put $U = \{1, 2, \dots, J\}$.

In each node there is

- Only one server,
- A queue discipline “first-come-first-served” (FCFS),
- An infinite waiting capacity.

They are open networks: any customer may enter and leave.

The exogenous arrivals process at node j is a Poisson Process at rate $\nu_j, j \in U$, independent of the exogenous arrivals processes to the other nodes: $\nu = \sum_{j=1}^J \nu_j$.

The service times at node j are independent and identically distributed, having exponential distribution with parameter $\mu_j, j \in U$, and independent from the other nodes service times.

After the completion of a service at node j , a customer is immediately directed to node l with probability p_{jl} , or abandons the network with probability $q_j = 1 - \sum_{l=1}^J p_{jl}, j \in U$.

These probabilities are not influenced by the movements of the other customers in the network. The p_{jl} matrix is called P .

The total arrivals rate, exogenous and endogenous, at node j, θ_j satisfies the network traffic equations:

$$\theta_j = \nu_j + \sum_{l=1}^J \theta_l p_{lj}, j = 1, 2, \dots, J \quad (1).$$

The state of the network at instant t is given by $N(t) = [N_1(t), \dots, N_J(t)]$, where $N_j(t)$ is the number of customers at node j in instant $t, j = 1, 2, \dots, J$.

If $\rho_j = \frac{\theta_j}{\mu_j} < 1, j = 1, 2, \dots, J$ the process $N = \{N(t)\}$ has stationary, or equilibrium, distribution, see for instance (Disney and König, 1985),

$$\pi(n_1, n_2, \dots, n_J) = \prod_{j=1}^J (1 - \rho_j) \rho_j^{n_j}, n_j \geq 0, j = 1, 2, \dots, J \quad (2).$$

Calling S_j, W_j and X_j the sojourn, waiting and service, respectively, times of a customer at node j

$$S_j = W_j + X_j \quad (3).$$

The Jackson networks sojourn times considered in this paper are those of typical customers that, arriving at the network, find the process in an equilibrium state. Call S the sojourn time in the network, that is: the time that goes between the arrival at the network and the departure of one of those customers. If in its path it traverses the nodes $1, 2, \dots, l$, $S = S_1 + S_2 + \dots + S_l$.

To study the sojourn time, the following is important:

- A network has “feedback” if a customer may come back to the same node after the completion of its service, immediately or in a future instant,
- A network without “feedback” is an “acyclic” one.

Then some examples of typical Jackson networks usually considered in the study of sojourn times are presented.

Simple Queues Series

For this Jackson network

$$p_{jl} = \begin{cases} 1, & \text{if } l = j + 1, j = 1, 2, \dots, J - 1 \\ 0 & \text{otherwise} \end{cases}$$

$\nu_1 = \nu$, $\nu_j = 0, j = 2, \dots, J$ and $\theta_j = \nu, j = 1, 2, \dots, J$. Figure 1 is a graphical representation of a simple queues series.

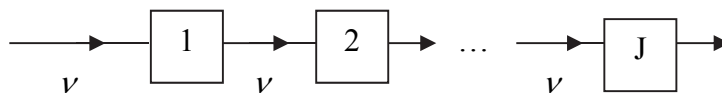


Figure 1: Simple Queues Series

Some important results are:

- All customers' flows, in this network, at stationary state, are Poisson Processes. It is a consequence of it, in stationary state, that the departure process from an M/M/1 queue is a Poisson Process, see for instance (Kelly, 1979),
- The sojourn times in the various nodes are independent random variables. In (Kelly, 1979) it is presented a demonstration of this statement based on the reversibility concept,
- The sojourn time at node j is an exponential random variable with parameter $\mu_j - \nu, j = 1, 2, \dots, J$,
- The waiting times are dependent random variables. See also (Kelly, 1979).

So the sojourn time study in these networks has no difficulty. The same is not true for the waiting time.

M/M/1 Queue with Instantaneous Bernoulli Feedback

It is a network with a single node. $J = 1, p_{11} = p, q_1 = 1 - p$ and $\theta = \frac{\nu}{1 - \rho}$, where $\theta = \theta_1$ and

$\nu = \nu_1$, see Figure 2.

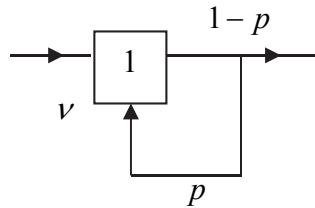


Figure 2: M/M/1 Queue with Instantaneous Bernoulli Feedback

Call S_m the m^{th} customer sojourn time in the network. So, if it is served k times

$$S_m = (t_{m1}^0 - t_{m1}^a) + (t_{m2}^0 - t_{m2}^i) + \dots + (t_{mk-1}^0 - t_{mk-1}^i) + (t_{mk}^d - t_{mk}^i) \quad (4)$$

where

- $t_{ml}^0 - t_{ml}^i$ is the time that the customer spends passing by the service system in the l^{th} time, given by the difference between the l^{th} output (0) instant from the server and the one of the l^{th} junction (i) to the queue,
- $t_{m1}^0 - t_{m1}^a$ is the time that the customer spends passing by the service system for the first time, given by the difference between the first output (0) instant from the server and the one of the arrival (a) to the queue,
- $t_{mk}^d - t_{mk}^i$ is the time that the customer spends passing by the service system for the last time, given by the difference between the departure (d) instant from the network and the one of the k^{th} junction (i) to the queue.

Note that K , the number of times that the customer passes by the server, is a random variable and $P(K = k) = p(1 - p)^{k-1}, k = 1, 2, \dots$.

$\{(t_{ml}^0 - t_{ml}^i) : l = 2, 2, \dots\}$ is not a sequence of independent random variables, see (Disney and König, 1985). So it is not possible to make use of the usual statement to sum independent random variables. But it is possible to get an expression to $P(S_m \leq t)$ that requires the k steps transition probabilities for the delayed Markovian renewal process $\{(N(t^i - 0), (t_l^o - t_l^i)), l = 0, 1, 2, \dots\}$ conditioning to the number of times that the customer returns to the queue.

Calling that transition probabilities matrix $Q_i^k(t)$, see still (Disney and König, 1985),

$$P(S_m \leq t) = \sum_{k=1}^{\infty} \pi Q_i^k(t) p(1-p)V \quad (5)$$

where π is the N^i (embedded version of N in the input instants) stationary distribution, k is the number of times the customer passes by the server and V is a vector which entries are all 1.

So, now, the situation is much more complicated than in the former case owing to the feedback.

The Jackson Three Node Acyclic Network

It is a network with three nodes where $p_{12} = p$, $p_{13} = 1-p$, $p_{23} = 1$, $p_{ji} = 0$ in the other cases, $\nu_1 = \nu$, $\nu_j = 0, j = 2, 3$, $\theta_1 = \nu$, $\theta_2 = p\nu$ and $\theta_3 = \nu$.

In equilibrium, all customers' flows are Poisson Process in this network.

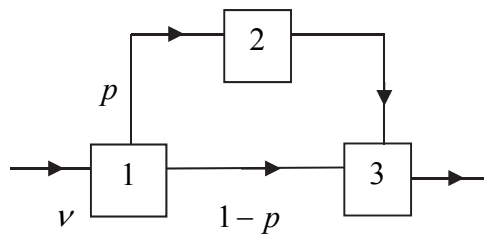


Figure 3: Jackson Three Node Acyclic Network

Consequently,

- The sojourn time at node j is a random variable exponentially distributed with parameter $\mu_j - \theta_j, j = 1, 2, 3$. S_1 and S_2 are independent random variables as well as S_2 and S_3 .

This result is valid for any Jackson acyclic network:

- Suppose that a customer follows a path r in a Jackson acyclic network with only one server at each node. If node j belongs to path r , S_j is such that

$$P(S_j \leq t | \text{the followed path is } r) = 1 - e^{-(\mu_j - \theta_j)t}, t \geq 0 \quad (6)$$

and, if node j is the next to the server after node l , S_j and S_l are independent random variables.

But,

- S_1 and S_3 are dependent random variables: (Foley and Kiessler, 1989) showed that, in fact, S_1 and S_3 are positively correlated. (Ferreira, 1998) showed that if

$$2(\nu + \mu_1 + \mu_2 + \mu_3) \left(\frac{1}{\nu + \mu_1 + \mu_2 + \mu_3} \left(\frac{1}{\mu_1 - \nu} + \frac{1}{\mu_3 - \nu} \right) + \frac{1}{(\mu_1 - \nu)^2} + \frac{1}{(\mu_3 - \nu)^2} \right)^{\frac{1}{2}} < 1 \quad (7)$$

and if

$$\frac{\mu_2 \left(\frac{1}{2(\nu + \mu_1 + \mu_2 + \mu_3)} - \frac{1}{2} \left(\frac{1}{(\nu + \mu_1 + \mu_2 + \mu_3)^2} + \frac{4}{\nu + \mu_1 + \mu_2 + \mu_3} \left(\frac{1}{\mu_1 - \nu} + \frac{1}{\mu_3 - \nu} \right) - 4 \left(\frac{1}{(\mu_1 - \nu)^2} + \frac{1}{(\mu_3 - \nu)^2} \right) \right)^{\frac{1}{2}} \right)}{1 - \nu \left(\frac{1}{2(\nu + \mu_1 + \mu_2 + \mu_3)} - \frac{1}{2} \left(\frac{1}{(\nu + \mu_1 + \mu_2 + \mu_3)^2} + \frac{4}{\nu + \mu_1 + \mu_2 + \mu_3} \left(\frac{1}{\mu_1 - \nu} + \frac{1}{\mu_3 - \nu} \right) - 4 \left(\frac{1}{(\mu_1 - \nu)^2} + \frac{1}{(\mu_3 - \nu)^2} \right) \right)^{\frac{1}{2}} \right)} < p < \frac{\mu_2 \left(\frac{1}{2(\nu + \mu_1 + \mu_2 + \mu_3)} + \frac{1}{2} \left(\frac{1}{(\nu + \mu_1 + \mu_2 + \mu_3)^2} + \frac{4}{\nu + \mu_1 + \mu_2 + \mu_3} \left(\frac{1}{\mu_1 - \nu} + \frac{1}{\mu_3 - \nu} \right) - 4 \left(\frac{1}{(\mu_1 - \nu)^2} + \frac{1}{(\mu_3 - \nu)^2} \right) \right)^{\frac{1}{2}} \right)}{1 - \nu \left(\frac{1}{2(\nu + \mu_1 + \mu_2 + \mu_3)} + \frac{1}{2} \left(\frac{1}{(\nu + \mu_1 + \mu_2 + \mu_3)^2} + \frac{4}{\nu + \mu_1 + \mu_2 + \mu_3} \left(\frac{1}{\mu_1 - \nu} + \frac{1}{\mu_3 - \nu} \right) - 4 \left(\frac{1}{(\mu_1 - \nu)^2} + \frac{1}{(\mu_3 - \nu)^2} \right) \right)^{\frac{1}{2}} \right)} \quad (8)$$

verify both simultaneously it is possible to guarantee that S_1 and S_3 are positively correlated in equilibrium.

There are two alternative paths for a customer to go from node 1 to node 3. And a customer that follows by node 2 may be overtaken by another one that goes directly from node 1 to node 3. So, a customer, when arriving at node 3, may meet there another one that was behind it at node 1 or even that had not arrived when it was there.

These overtaking customers can delay a certain customer, when it arrives at node 3, for a longer time than that if they were not present. The number of these customers depends, partly, on the number of the customers that arrive while the customer that is being followed is in node 1, partly owing to the supposition of a FCFS discipline. Consequently, the time that a customer waits at node 3 depends on how much time it has waited at node 1.

3 Network Flow Equations

The objective of this section is to present the so called “network flow equations” for the Jackson networks, that allow the deduction of formulae to the computation of sojourn times moments of any order, efficient in some situations.

Following the work of (Lemoine, 1987) call τ_j an arrival instant, endogenous or exogenous, at node j and $\tau_j + T_j$ the departure instant from the network of the customer that arrived in τ_j , $j = 1, 2, \dots, J$, so

- T_j is the remaining sojourn time, in the network, for the arrival at node j in the instant $\tau_j, j = 1, 2, \dots, J$.

Call h_j the Laplace Transform of the $T_j, j = 1, 2, \dots, J$ distribution. As N is a strong Markov Process, and the network state process “seen by the arrivals” is in equilibrium, the $T_j, j = 1, 2, \dots, J$ and its Laplace Transforms are uniquely determined.

Dealing with the sojourn time as the life time of a Markov Process ϑ – as it will be seen in section 4 – it is possible to show that the Laplace Transforms $h_j, j = 1, 2, \dots, J$ satisfy an equations system called the “network flow equations”. That is, according with (Lemoine, 1987)

- Being H_j the probability distribution with Laplace Transform h_j , there is a distribution probability with Laplace Transform q_j such as

$$h_j(s) + \frac{sg_j(s)}{\mu_j - \theta_j} = q_j(s) + \sum_{l=1}^J p_{jl} h_l(s), s \geq 0 \text{ and } j = 1, 2, \dots, J \quad (9).$$

In Jackson networks without “overtaking” the Transforms h_j and g_j are identical for each j .

Given $h_j, j = 1, 2, \dots, J$ the Transforms $g_j, j = 1, 2, \dots, J$ are uniquely determined by (9). The converse is also true since $I - P$, being I the identity matrix, is invertible.

After (9), by successive derivations, (Lemoine, 1987) showed that

- Network Flow Equations

For $j = 1, 2, \dots, J$ and $r = 1, 2, \dots$

$$\begin{aligned} E[T_j^r] &= r! (\mu_j - \theta_j)^{-r} + \sum_{l=1}^J p_{jl} E[T_l^r] \\ &+ \sum_{l=1}^J p_{jl} \sum_{n=1}^{r-1} \frac{r!}{n! (r-n)!} \mu_j^{-n} E \left[T_l^{r-n} \prod_{m=1}^n (N_j(T_l^-) + m) \right] \end{aligned} \quad (10).$$

For $r = 1$, (10) assumes the matrix form

$$[E[T_j]] = (I - P)^{-1} [(\mu_j - \theta_j)^{-1}] \quad (11).$$

For $r = 2$ (10) assumes the form

$$E[T_j^2] = 2(\mu_j - \theta_j)^{-2} + \sum_{l=1}^J p_{jl} E[T_l^2] + 2\mu_j^{-1} \sum_{l=1}^J p_{jl} E[T_l (N_j(T_l^-) + 1)], j = 1, 2, \dots, J \quad (12).$$

Equality (12) defines a system of J equations and $J^2 + J$ unknowns. In general, when $r \geq 2$, the product terms involving the variables T_l and $N_j(\tau_l^-)$ prevent the exact computation of the sojourn times r order moments; there are too many unknowns and too few equations. In these cases other independent equations are needed to complement (12) in order to be possible to obtain exact solutions.

When any pair of nodes in the network is connected by, in the maximum, one oriented path and $p_{jj} = 0, j = 1, 2, \dots, J$, T_l and $N_j(\tau_l^-)$ are independent for $j \neq l$. The computation of $E[T_j(N_j(\tau_j^-) + 1)]$ is irrelevant since $p_{jj} = 0, j = 1, 2, \dots, J$. In this case (10) becomes a compact recursive formula that allows the computation of any order moments of the sojourn times, $T_j, j = 1, 2, \dots, J$. For instance, as, in these conditions,

$$E[N_j(\tau_l^-)] = \frac{\theta_j}{\mu_j - \theta_j}, j = 1, 2, \dots, J \quad (13),$$

(12) assumes the form

$$E[T_j^2] = 2(\mu_j - \theta_j)^{-2} + \sum_{l=1}^J p_{jl} E[T_l^2] + 2(\mu_j - \theta_j)^{-1} \sum_{l=1}^J p_{jl} E[T_l], j = 1, 2, \dots, J \quad (14).$$

Applying (14) to the simple queues series

$$\begin{aligned} E[T_j^2] &= 2(\mu_j - \nu)^{-2} + E[T_{l+1}^2] + 2(\mu_j - \nu)^{-1} E[T_{l+1}], \quad j = 1, 2, \dots, J-1 \\ E[T_j^2] &= 2(\mu_j - \nu)^{-2} \end{aligned} \quad (15)$$

Putting together (15) and (11) it may be concluded that

$$VAR[T_j] = \sum_{l=j}^J (\mu_l - \nu)^{-2} \quad (16).$$

For Jackson networks that do not fulfil those conditions, in (Lemoine, 1987) it is suggested to identify adequate Martingale families in N as a process to determine independent equations to complement (10). Applying this proceeding to the M/M/1 queue with instantaneous Bernoulli feedback it was obtained

$$VAR[T] = \frac{1}{((1-p)\mu - \nu)^2} \frac{(1-p^2)\mu + \nu p}{(1-p^2)\mu - \nu p} \quad (17)$$

and

$$COV[N(\tau^-), T] = \frac{\nu(1-p)\mu}{(1-p^2)\mu - \nu p} \quad (18).$$

4 Sojourn Times Distributions and Moments Numerical Computation

Now it will be described a general method, which key is the proceeding called, in the English language literature “randomisation procedure”, to approximate “first passage times” distributions in direct time Markov Processes, being the sojourn times in queue systems a particular case.

Call $\aleph = \{X(t): t \geq 0\}$ a regular Markov Process, in continuous time with a countable states space E and a bounded matrix infinitesimal generator Q .

The elements of Q are designated by $Q(x, y), x, y \in E$ and $Q(x) = \sum_{y \in E - \{x\}} Q(x, y)$. $\psi(t)$ designates the $X(t)$ state probability vector:

$$\psi_t(x) = P\{X(t) = x\}, x \in E \quad (19).$$

X models the evolution of a queue system during the sojourn of a given, “marked”, customer in it.

The states of E have two main components:

- i) The queue system state,
- ii) The “marked” customer position.

Be

- A the states subset that describes the system till the departure of the “marked” customer, and
- B the states subset that describes the system after the departure of that customer.

Evidently

- $\{A, B\}$ is a partition of E ,
- If T is the time that the process \aleph spends in A till attaining B , for the first time, T is precisely the sojourn time of the “marked” customer in the network.

It is supposed that \aleph will remain in B , with probability 1 after having attained it for the first time. In fact, as the evolution of the system after the departure of the “marked” customer is irrelevant, it may be supposed that B is a closed set. That is, the process \aleph cannot come back to A after reaching B . The quantity of interest is the T distribution function, $\tau(t)$. Note that

$$\tau(t) = P\{T \leq t\} = P\{X(t) \in B\} = 1 - P\{X(t) \in A\}, t \geq 0 \quad (20)$$

since the presented hypotheses guarantee that $\{T \leq t\} = \{X(t) \in B\}$.

After (20) it is concluded that

- The problem of computing $\tau(t)$ is equivalent to the one of the computation of the transient distribution of $X(t)$ in A .

So it is necessary to compute the vector $\psi_t, t \geq 0$. Being $P_t, t \geq 0$, the \aleph n transition matrix,

$$\psi_t = \psi_0 P_t, t \geq 0 \quad (21)$$

and

$$P_t = \exp(Qt) = \sum_{i=0}^{\infty} \frac{t^i}{i!} Q^i, t \geq 0 \quad (22).$$

The “randomisation procedure” consists in using in (22) an equivalent representation; see (Çınlar, 1975):

$$P_t = \exp(-\alpha t) \exp\left(\alpha t \left(I + \frac{1}{\alpha} Q\right)\right) = \exp(-\alpha t) \sum_{i=0}^{\infty} \frac{\alpha^i t^i}{i!} R^i \quad (23)$$

where

$$R = I + \frac{1}{\alpha} Q \quad (24)$$

is called the “randomised matrix” in English language literature,

- I is the identity matrix, and
- α is a positive upper bound for the whole $Q(x), x \in E$.

Note that, see (Melamed and Yadin, 1984, 1984a),

- Although the equation (23) seems more complex than (22), it fulfils in fact more favourable computational properties. The most important is that R is a stochastic matrix while Q is not. Consequently, the computation using (23) is stable and using (22) is not,
- The “randomisation procedure” has an interesting probabilistic meaning, useful to determine bounds for $\tau(t)$. In fact, being R a stochastic matrix, it defines a discrete time Markov Process

$$\mathfrak{S} = \{Y_n: n = 0, 1, \dots\} \quad (25)$$

if it is assumed $Y_0 = X(0)$. With this procedure, the relation between the processes \mathfrak{N} and \mathfrak{S} is quite simple as it will be seen next.

Extend the discrete time process \mathfrak{S} to a continuous time Markov Process such that

- i) The time intervals between jumps are exponential random variables i.i.d. with mean $1/\alpha$
- ii) The jumps are commanded by R .

In (Melamed and Yadin, 1984) it is shown that the resulting process is precisely the original process \mathfrak{N} ; but when there is a sequence of jumps in \mathfrak{S} from the state $x \in E$ to itself, this will be noticed in \mathfrak{N} as a long sojourn in state x .

So, the “randomisation procedure” may be interpreted as a sowing in the process \mathfrak{N} with “fake” random jumps between the true jumps. The resulting process, designated by $\bar{\mathfrak{N}}$, at which the “fake” jumps are visible, has the same probabilistic structure than \mathfrak{N} but with an advantage:

- The sequence of the jump instants in $\bar{\mathfrak{N}}$, “fake” and “true”, is now a Poisson Process. This is not, in general, the case of \mathfrak{N} .

Note that Y_n is the state of $\bar{\aleph}$ in the instant of the n^{th} jump, “fake” or “true”.

Suppose that $\bar{\aleph}$ reaches the set B in its n^{th} jump. Consequently the $\bar{\aleph}$ sojourn time, and so also the \aleph , in A is the sum of n exponential independent random variables with mean $1/\alpha$. That is, the sojourn time has a n order Erlang distribution with parameter α . Its distribution function will be designated $E_{n,\alpha}(t)$.

Be $h(n)$ the probability that $\bar{\aleph}$ reaches B in its n^{th} jump. Call ϕ_n the state probability vector of Y_n :

$$\phi_n = \psi_0 R^n \quad (26).$$

The quantities $h(n)$ are given by the equivalent formulae:

$$h(n) = \begin{cases} \sum_{x \in B} \phi_0(x), n = 0 \\ \sum_{x \in A} \sum_{y \in B} \phi_{n-1}(x) R(x, y), n > 0 \end{cases} \quad (27)$$

or

$$h(n) = \begin{cases} 1 - \sum_{x \in A} \phi_0(x), n = 0 \\ \sum_{x \in A} \phi_{n-1}(x) - \sum_{x \in A} \phi_n(x), n > 0 \end{cases} \quad (28).$$

Given the probabilities $h(n)$ and, noting that $\sum_{n=0}^{\infty} h(n) = 1$, it is obtained

$$\tau(t) = \sum_{n=0}^{\infty} h(n) E_{n,\alpha}(t), t \geq 0 \quad (29),$$

$$E[T^m] = \frac{1}{\alpha^m} \sum_{n=0}^{\infty} n(n+1) \dots (n+m-1) h(n), m = 1, 2, \dots \quad (30).$$

The formula (30) for $m = 1$ is

$$E[T] = \frac{1}{\alpha} E[H] \quad (31)$$

being H the number of \aleph jumps till reaching B . Expression (31) is the Little's Formula in this queues context.

Equation (29) allows obtaining simple bounds for $\tau(t)$ that may, in principle, to become arbitrarily close. Equation (30) allows obtaining $E[T^k]$, in principle, so close of $E[T^k]$ as wished. So, given any integer $k \geq 0$

$$L_k(t) \leq \tau(t) \leq U_k(t) \quad (32)$$

where

$$L_k(t) = \sum_{n=0}^k h(n) E_{n,\alpha}(t), t \geq 0 \quad (33),$$

$$U_k(t) = 1 - \sum_{n=0}^k h(n) \bar{E}_{n,\alpha}(t), t \geq 0 \quad (34)$$

and

$$E[T^m]_{L,k} \leq E[T^m], \quad m = 1, 2, \dots \quad (35)$$

where

$$E[T^m]_{L,k} = \frac{1}{\alpha^m} \sum_{n=0}^k n(n+1) \dots (n+m-1) h(n), m = 1, 2, \dots \quad (36).$$

It is easy to prove that

Proposition

If, for any $\varepsilon > 0$, k is chosen in accordance with the rule

$$k = \min \left\{ n \geq 0: \sum_{i=0}^n h(i) \geq 1 - \varepsilon \right\} = k(\varepsilon), \quad (37)$$

or equivalently

$$J = \min \left\{ n \geq 0: \sum_{x \in A} \phi_n(x) \leq \varepsilon \right\} = J(\varepsilon), \quad (38)$$

$$|\tau(t) - L_{J(\varepsilon)}| \leq \varepsilon \text{ and } |\tau(t) - U_{J(\varepsilon)}| \leq \varepsilon, \text{ uniformly in } t \geq 0. \blacksquare$$

The main problem in the application of the method presented, that in principle would solve any computation problems related to the distribution of sojourn times, stays in the difficulty of the $h(n)$ computation. In fact, for it, it is necessary to compute the vectors ϕ_n but only in the subset A of the states space. When states space E is finite, as it happens in the case of closed networks, both $h(n)$ and ϕ_n can, at first glance, be computed exactly, apart the mistakes brought by the approximations.

In practice the states space is often infinite or, although finite, prohibitively great. In this situations it is mandatory to truncate E . So, it must be considered a new level of approximation since the $h(n)$, ϕ_n , etc. must also be approximated now.

In fact, what are viable to obtain is $h(n)$ lower bounds because the E truncation is translated in probability loss (Melamed and Yadin, 1984a). So, with these $h(n)$ approximate values, (32) and (35) go on being valid but

- The uniform convergence property seen above is lost,
- The rules analogous to (37) and (38) are not equivalent. The one generated by (37) may be even unviable and in practice it is used only the one generated by (38) (Melamed and Yadin, 1984a).

Using this method (Kiessler *et al.*, 1988) achieved to show that, in a Jackson three node acyclic network, the total sojourn time distribution function for a customer that follows the path integrated

by the nodes 1, 2, and 3 is not the same obtained considering that S_1, S_2 and S_3 are independent although this one, designated by $F(t)$, is a “good” approximation of that one. They show that in some particular cases it was not true that

$$F^L(t) \leq F(t) \leq F^U(t), t \geq 0 \quad (39)$$

being $F^L(t)$ and $F^U(t)$ the lower bound and the upper bound, respectively, of that customer sojourn time distribution function, obtained through the described method.

This conclusion is important because, in spite of the dependence between S_1 and S_3 , $F(t)$ could be the S distribution function. In fact, (Feller, 1966) presents an example of dependent random variables which sum has the same distribution as if the random variables were independent.

Finally note that the formula (30), apparently new, seems to be more efficient than (10), although only allows to obtain moments lower bounds, because its field of application is much greater.

5 Conclusions

The sojourn time has an evident practice interest. And is and has been intensively studied. Evidently the problem of the computation of the sojourn times in networks of queues is one of the most difficult in these networks study. In fact, analytic solutions are the exception and not the rule. And, when existing, are quite rough.

The most of the known works only present results on sojourn time distributions for only one customer in paths without overtaking with FCFS disciplines in the nodes. It seems that still there are not results for simultaneous distributions of various customers sojourn times.

It follows, from the examples seen in section 2, that the sojourn times, at Jackson networks computations, difficulties occur when there are feedback and overtaking. In the first case the input server process is not a Poisson Process, becoming everything more complex. In the second case dependencies exist among a customer sojourn times in the various nodes, simultaneously complicated and subtle, that make the total sojourn time computation difficult even if the sojourn times in each node are easy to compute.

From all this it results the interest of the methods presented in sections 3 and 4 to compute exactly and approximately the quantities related with the Jackson networks sojourn times.

Acknowledgement

This work was financially supported by FCT through the Strategic Project PEst-OE/EGE/UI0315/2011.

References

- [1] ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol. 1 (1), pp 96-98, 2010.

- [2] ÇINLAR, E.: *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- [3] DISNEY, R. L. and KÖNIG, D.: *Queueing Networks: A Survey of their Random Processes*. SIAM Review, 3, pp. 335-403, 1985.
- [4] FELLER, W.: *An Introduction to Probability Theory and its Applications*. Vol. II, John Wiley & Sons, New York, 1966.
- [5] FERREIRA, M. A. M.: *O Tempo de Permanência em Redes de Jackson*. Revista de Estatística, 2 (2), pp. 25-44, 1997.
- [6] FERREIRA, M. A. M.: Correlation Coefficient of the Sojourn Times in Nodes 1 and 3 of Three Node Acyclic Jackson Network. Proceedings of the Third South China International Business Symposium. Vol. 2. Macau, Hong-Kong, Jiangmen (China), pp. 875-881, 1998.
- [7] FERREIRA, M. A. M.: *A note on Jackson Networks Sojourn Times*. Journal of Mathematics and Technology, vol. 1 (1), pp 91-95, 2010.
- [8] FERREIRA, M. A. M. and ANDRADE, M.: *Fundaments of Theory of Queues*. International Journal of Academic Research, Vol. 3 (1, part II), pp. 427-429, 2011.
- [9] FERREIRA, M. A. M. and Andrade, M.: *Stochastic Processes in Networks of Queues with Losses: A Review*. International Journal of Academic Research, Vol. 3 (2, part IV), pp. 989-991, 2011.
- [10] FERREIRA, M. A. M. and ANDRADE, M.: *Grouping and Reordering in a Server Series*. Journal of Mathematics and Technology, Vol. 2 (2), pp. 4-8, 2011a.
- [11] FERREIRA, M. A. M. and ANDRADE, M.: *Non-Homogeneous Networks of Queues*. Journal of Mathematics and Technology, Vol. 2 (2), pp. 24-29, 2011b.
- [12] FOLEY, R. D. and KIESSLER, P. C.: *Positive Correlations in a Three-Node Jackson Queueing Network*. Advanced Applied Probability, 21, pp. 241-242, 1989.
- [13] JACKSON, J. R.: *Network of Waiting Lines*. Operations Research, 5, pp. 518-521, 1957.
- [14] JACKSON, J. R.: *Jobshop-like Queueing Systems*. Management Science, 10, pp. 131-142, 1963.
- [15] KELLY, F. P.: *Reversibility and Stochastic Networks*. John Willey & Sons, New York, 1979.
- [16] KIESSLER, P. C., MELAMED, B., YADIN, M. and FOLLEY, R. D.: *Analysis of a Three Node Queueing Network*. Queueing Systems, 3, pp. 53-72, 1988.
- [17] LEMOINE, A. J.: *On Sojourn Time in Jackson Networks of Queues*. Journal of Applied Probability, 24, pp. 495-510, 1987.
- [18] MELAMED, B. and YADIN, M.: *Randomization Procedures in the Computation of Cumulative-Time Distributions over Discrete State Markov Processes*. Operations Research, 4 (32), pp. 926-944, 1984.
- [19] MELAMED, B. and YADIN, M.: *Numerical Computation of Sojourn-Time Distributions in Queueing Networks*. Journal of the Association for Computing Machinery, 4 (31), pp. 839-854, 1984a.

Current address

Marina Andrade, Professor Auxiliar

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

Telefone: + 351 21 790 34 05

Fax: + 351 21 790 39 41
e-mail: marina.andrade@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

TELEFONE: + 351 21 790 37 03

FAX: + 351 21 790 39 41

e-mail: manuel.ferreira@iscte.pt

COPULA BASED SEMIPARAMETRIC REGRESSIVE MODELS

FJODOROVŠ Jegors, (LV), MATVEJEVS Andrejs, (LV)

Abstract. This paper studies the estimation of copula-based semi parametric stationary Markov models. Described models allow us evaluate the parameters of copula, which has the best fit to previously selected model (simple estimators of the marginal distribution and the copula parameter are provided). These copula-based models are characterized by nonparametric marginal distributions and parametric copula functions, while the copulas capture all the scale-free temporal dependence of the processes. In our copula dependence study we used MatLab, which help to evaluate copula parameters and choose the best copula class, based on log likelihood estimation, for the selected financial market data. Also, using this MatLab we made VIX option index simulation - found the best copula fit under our condition and show the evaluation steps for copula based semi parametric autoregression.

Key words: copula, diffusion processes, time series, semi parametric regressions, VIX index.

Mathematics Subject Classification: 60J70, 62M10, 47D07

1. Introduction

The possibility of identifying nonlinear time series using nonparametric estimates of the conditional mean and conditional variance were studied in many papers (see, for example, [1], and references there). As a rule analyzing the dependence structure of stationary time series $\{x_t, t \in \mathbb{Z}\}$ regressive models defined by invariant marginal distributions and copula functions that capture the temporal dependence of the processes. As it indicated in [1] this permits to separate out the temporal dependence (such as tail dependence) from the marginal behavior (such as fat tails) of a time series. One more advantage of this type regressive approach is a possibility to apply probabilistic limit theorems for transition from difference equations to continuous time stochastic differential equations ([2], [3]). In our paper we also study a class of copula-based semi parametric stationary Markov models in a form of scalar difference equation

$$t \in \mathbb{Z} : X_t = X_{t-1} + \mathcal{E}(X_{t-1}, \varepsilon) + \mathcal{E}(X_{t-1}, \varepsilon) \xi_t \quad (1)$$

where $\{\xi_t, t \in \mathbb{Z}\}$ is i.i.d., $N(0; 1)$, and ε is a small positive parameter, which will be used for diffusion approximation of (1). Regressions (1) are high-usage equations for simulation and

parameter estimation of stochastic volatility models ([2]). But unfortunately defined by (1) Markov chain has incompact phase space that complicates an application of probabilistic limit theorem. Copula approach helps to simplify asymptotic analysis of (1). Let us remember that to construct a copula $C(u; v)$ for pair $\{X_{t-1}, X_t\}$ from (1) one should find a marginal invariant distribution $F(x)$ for X_t and to substitute this in joint distribution function $H(x, y) = P(X_{t-1} \leq x, X_t \leq y)$, that is, $C(u, v) = H(F^{-1}(u), F^{-1}(v))$ and $H(x, y) = C(F^{-1}(x), F^{-1}(y))$. Due to persistence of small parameter ε after a substitution $U_t = F(X_t)$ in equation (1) for a further diffusion approximation one can write a difference equation in a same form like (1):

$$t \in Z : U_t = U_{t-1} + \varepsilon f(U_{t-1}, \varepsilon) + \varepsilon g(U_{t-1}, \varepsilon) \xi_t \quad (2)$$

But now this equation defines Markov chain on the compact $[0, 1]$. This makes easier formulate construction for transition probability and further estimators of functions $\hat{f}(u)$ and $\hat{g}(u)$. After diffusion approximation of (2) one can make inverse substitution and derive stochastic differential equation as diffusion approximation for (1).

In the copula approach to univariate time series modeling, the finite dimensional distributions of the time series are generated by copulas. By coupling different marginal distributions with different copula functions, copula-based time series models are able to model a wide variety of marginal behaviors (such as skewness and fat tails) and dependence properties (such as clusters, positive or negative tail dependence). (see [Darsow et al. \(1992\) \[4\]](#) and [Joe \(1997\)\[5\]](#)).

Described algorithm allow us evaluate the parameters of copula, which has the best fit to previously selected model. In our copula dependence study we used MatLab, which help to evaluate copula parameters and choose the best copula class, based on log likelihood estimation, for the selected financial market data. These copula based models are easy to simulate, and can be expressed as semi parametric regression transformation models. Also, using this MatLab we made VIX option index simulation - found the best copula fit under our condition and built semi parametric autoregression.

The paper is structured as follows. Section 2 gives a brief review of the copula functions definition. Section 3 describes our approach. In Section 4 we report our results for the VIX index data Section 5 concludes and discusses several possible avenues of future research.

2. Copula functions

Copulas became popular in the finance and insurance community in the past years, where modeling and estimating the dependence structure between several univariate time series are of great interest; see [Frees and Valdez \(1998\) \[6\]](#) and [Embrechts et al. \(2002\) \[7\]](#) for reviews.

A copula function is a multivariate distribution function with standard uniform marginals. By [Sklar's \(1959\) \[8\]](#) theorem, one can always model any multivariate distribution by modeling its marginal distributions and its copula function separately, where the copula captures all the scale-free dependence in the multivariate distribution.

The central result of this theorem, which states that any continuous N -dimensional cumulative distribution function F , evaluated at point $x = (x_1, \dots, x_n)$ can be represented as

$$F(x) = C(F_1(x_1), \dots, F_n(x_n)),$$

where C is called a copula function and $F_i(x_i)$, $i = 1, \dots, n$ are the marginal distributions. The use of copulas therefore splits a complicated problem (finding a multivariate distribution) into two simpler tasks. The first task is to model the univariate marginal distributions and the second task is finding a copula that summarises the dependence structure between them.

It is also useful to represent of copulas as joint distribution functions of standard uniform random variables:

$$U = F(X_1) \quad \text{and} \quad V = F(X_2)$$

$$C(u, v) = P(U \leq u, V \leq v)$$

The outcome of uniform random variables falls into the interval $[0, 1]$, therefore the domain of a copula must be the N -dimensional unit cube. Similarly, because the mapping represents a probability, the range of the copula must also be the unit interval. Also, it is easy to determine the value of a copula on the border of its domain. When one argument equals zero, the probability of any joint event must also be zero. Similarly, when all but one of the inputs are equal to one the joint probability must be equal to the (marginal) probability of the argument that does not equal one. Finally, the function must be increasing in all its arguments.

Besides the standard distribution functions, copulas have associated densities:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

which permit the bivariate density $f(u, v)$ as the product of the copula density and the density functions of the margins

$$f(u, v) = c(F_1(u), F_2(v))f_1(u)f_2(v)$$

This expression indicates how the simple product of two marginal distributions will fail to properly measure the joint distribution of two asset prices unless they are in fact independent and the dependence information captured by the copula density, $c(F_1(u), F_2(v))$; is normalized to unity.

3. Evaluation of parameters for the semi parametric regression model

Copula based semi parametric models are characterized by conditional heteroscedasticity and have been often used in modeling the variability of statistical data. In paper [1] the basic idea was to apply a local linear regression to the squared residuals for finding the unknown functions f and g .

Our methodology builds on the finding conditional expectation of the first and second order.

Let $\{Y_t\}$ be a stationary Markov process of order 1 with continuous state space. Then its probabilistic properties are completely determined by the joint distribution function of $\{Y_{t-1}\}$ and $\{Y_t\}$. For the determination of the copula based model we should use Markov model in the scalar difference equation form:

$$t \in Z : X_t = X_{t-1} + \varepsilon f(X_{t-1}, \varepsilon) + \varepsilon g(X_{t-1}, \varepsilon) \xi_t$$

Where

$$E\xi_t = 0, DE\xi_t = 1, \{\xi_t, t \in Z\} \text{ is i.i.d. } N(0,1)$$

And our goal reduced to the estimation of conditional moments, which will be our base regression model parameters:

$$g(X_{t-1}, \varepsilon) - ? \text{ and } f(X_{t-1}, \varepsilon) - ?$$

As was mentioned above it is not easy task, especially this representation complicates an application of probabilistic limit theorem. That is why; if we have stationary distribution our suggestion is to find parameters through Markov chain using copula approach.

Firstly, let's show that copula distribution density equals Markov chain transition density:

$$P(X_{t-1} \leq y) = F(y) = \int_{-\infty}^y p(z) dz$$

$$P(X_{t+1} \in A / X_t) = P(X, A) = \int_A p(x, y) dy$$

$$\begin{aligned} P(X_{t+1} \in A, X_t \in B) &= \int_A p(z, A) dz = \iint_{AB} p(z, y) p(z) dz dy = \int P(x, A) dF(x) = \\ &= \iint_{BA} p(x, y) p(z) dF(x) dy \end{aligned}$$

$$C(F(y), F(x)) = P(X_{t+1} \leq y, X_t \leq x) = \iint p(z, u) p(z) dz du$$

$$p(x, y) = \frac{\partial^2 C(F(y), F(x))}{\partial x \partial y} p(y)$$

As the result we see that Markov transition density is copula density.

Secondly, we should expand our semi parametric regression into Taylor series:

$$F(X_t) = F(X_{t-1}) + F'(X_t) \mathcal{E}f(X_{t-1}, \varepsilon) + F'(X_t) \mathcal{E}g(X_{t-1}, \varepsilon) \xi_t \dots$$

And due to persistence of the small parameter ε , we can rewrite our expression is:

$$\begin{aligned} t \in Z : U_t &= U_{t-1} + \mathcal{E}f(U_{t-1}, \varepsilon) + \mathcal{E}g(U_{t-1}, \varepsilon) \xi_t \\ f(U_{t-1}, \varepsilon) &= E(U_t | U_{t-1} = u) - ? \end{aligned} \tag{3}$$

$$g(U_{t-1}, \varepsilon) = E((U_{t-1} - f(U_{t-1}, \varepsilon))^2 | U_{t-1} = u) - ? \tag{4}$$

After conditional expectations of (3) and (4) evaluation one can make inverse substitution and derive stochastic differential equation as diffusion approximation for the base semi parametric model (1). Of course, our algorithm works only if inverse function exists. For example, Gamble copula, which don't have standard inverse function.

Now we derived a tool for model (1) parameters evaluation. For describing our idea briefly, let's take a look in the next section how works our algorithm with the true market data.

4. Practical approach of the proposed algorithm

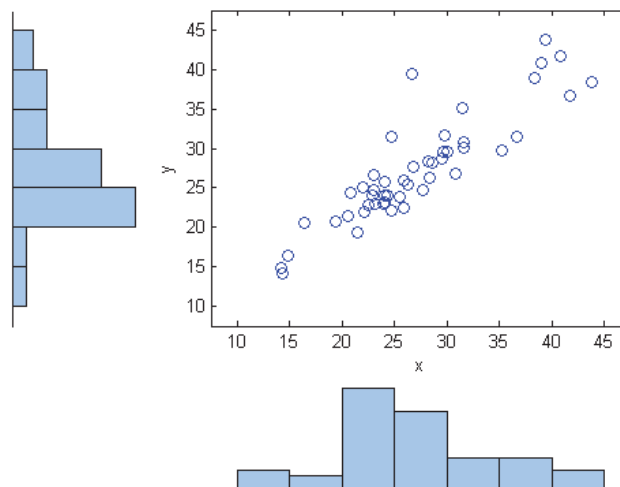
We'll analyze the VIX - **Market Volatility Index** – daily data from 31.03.2007. to 10.12.2011. The VIX is a market mechanism that measures the 30-day forward implied volatility of the underlying

index, the S&P 500. Being able to meaningfully interpret movements in the VIX and its reaction to market events can give investors an edge in managing the risk and profitability of their trading book and in designing portfolio strategies using VIX derivatives to capitalize on the dynamic and time-varying correlation of the VIX with its underlying S&P 500 Index. Let's built for this option index semi parametric copula based model, using AIC and BIC criteria.

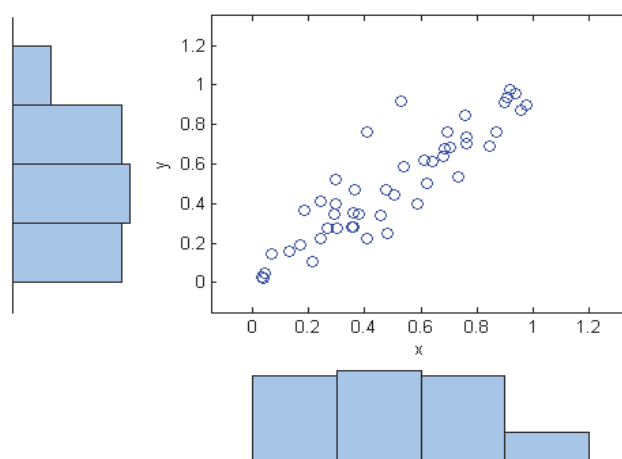
An easiest way of parameters estimating of the semi regressive model for the VIX index would be to hold the algorithm:

- Simulate U_t points which is $R[0,1]$ (uniform) or transform the existing sample into $R[0,1]$;
- Build scatter plot for (U_{t-1}, U_t) ;
- Make several statistical tests to find the suited distribution of data;
- Taking into account scatter plot and distribution of data try to choose copula from existing class or build your own copula, if you know marginal distributions;
- Test copula consistency to data (for example AIB and BIC);
- Find regression parameters.

Using Matlab program we have built scatter plots for VIX index transformed into uniform distribution ($R[0,1]$) and non transformed data.



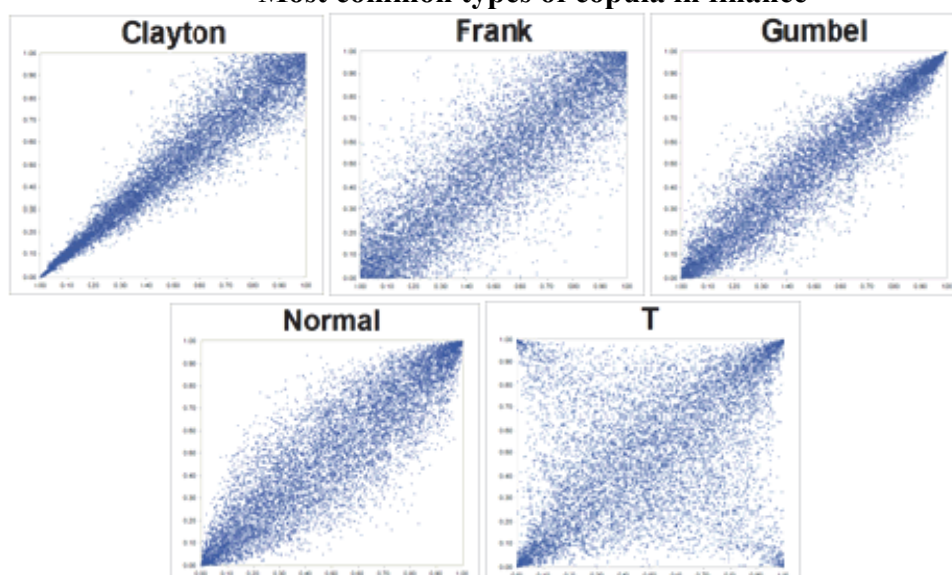
Graph1. Scatter plot for non transformed into $R[0,1]$ VIX index data



Graph 2. Scatter plot for transformed into $R[0,1]$ VIX index data

An important issue faced by an applied researcher interested in using the class of semi parametric copula-based time series models is the choice of an appropriate parametric copula. In different papers [Chen et al. \(2003\)](#) [9] propose two simple tests for the correct specification of a parametric copula in the context of modeling the contemporaneous dependence between several univariate time series and of the innovations of univariate GARCH models used to filter each univariate time series; (2) [Chen and Fan \(2004b\)](#) [10] establish pseudo-likelihood ratio tests for selection of parametric copula models for multivariate i.i.d. observations under copula misspecification [1]. But our suggestion is simpler – we can choose the best copula fit using AIC and BIC criteria or using χ^2 test for data distribution. We take for different copula comparisons AIC and BIC (see Table 1).

Most common types of copula in finance



Graph 3. Scatter plot for transformed into $R[0,1]$ VIX index data

For the first copula choosing step it is reasonable to compare graphical parametric copulas with VIX data scatter plot (Graph 2). As we can see the most suitable copulas for our data are Gumbel, Frank and Normal. For this sample of copulas is useful to calculate AIC and BIC criteria.

Copula	AIC	BIC
Gumbel copula	-124,1	-119,3
Frank copula	-267,4	-261,5
Normal copula	-230,3	-227,9

Table 1. AIC and BIC criteria for VIX index data.

Taking into account AIC and BIC criteria we should choose Frank copula for further model estimation. Let see how to derive semi parametric regression parameters using Frank copula representation:

$$C(u_{t+1}, u_t) = -a^{-1} \ln \left(1 + \frac{g_{u_{t+1}} g_{u_t}}{g_1} \right) \quad g_{u_t} = e^{-au_t} - 1 \quad (5)$$

And insert expression (5) into conditional expectation, we get our parameters:

$$\begin{aligned} E(U_{t+1} | U_t = u) &= \int_0^1 u_{t+1} dF_{u_{t+1}|u_t}(u) = \int_0^1 u_{t+1} p(u_{t+1} | u_t) du_{t+1} = \int_0^1 u_{t+1} \frac{\partial C(u_{t+1}, u_t)}{\partial u_{t+1} \partial u_t} du_{t+1} = \int_0^1 u_{t+1} c(u_{t+1}, u_t) du_{t+1} \\ &= -a \int_0^1 u_{t+1} \frac{g_1 (1 + g_{u_t + u_{t+1}})}{(g_{u_t} g_{u_{t+1}} + g_1)^2} du \end{aligned} \quad (6)$$

$$g((U_{t+1} | U_t = u)) = E((U_{t+1} - f(U_t))^2 | U_t = u) = \int_0^1 (U_{t+1} - f(U_t))^2 c(U_{t+1}, U_t) dU_{t+1} \quad (7)$$

It is impossible to solve analytically (5) and (6) expressions. But numerically it is doable for example in the Matlab. For the Frank copula we can use inverse function with the aim to return to our base equation (1). Of course, if we want use this model in practice, it is crucial to compare different class models which could be suitable for this data. This can give applied added value for this method.

But if we deal with copulas we should not skip some facts. For example, it is not easy to say which parametric copula best fits a given dataset, since some copulas may fit better near the center and other near the tails and many copulas do not have moments that are directly related to the Pearson correlation, it is difficult to compare financial models based on correlation.

Conclusions and further work

The algorithm for copula simulation and semi parametric regression coefficients finding through Markov chain have been presented. For Option VIX index data was found via MatLab the best fitted copula model, which is Frank copula. According to this copula, were shown principals of the semi parametric regression model coefficients evaluation. The next step in this research can be evaluating of the applied characteristic of the copula based semi parametric model as well as studying efficient estimation of conditional variance function in stochastic regression and to build continuous stochastic model using limit theorems.

References

- [1] X.CHEN, Y.FAN. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 2006.
- [2] D.B. NELSON. ARCH models as diffusion approximations. *Journal of Econometrics*, 441 1-2: 7|38, 1990.
- [3] Y. AIT-SAHALIA, R. KIMMEL. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 2007.
- [4] DARSOW, W., NGUYEN, B., OLSEN, E., 1992. Copulas and Markov processes. *Illinois Journal of Mathematics* 36, 600–642.
- [5] JOE, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC;
- [6] FREES, E.W., VALDEZ, E.A., 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2, 1–25.
- [7] EMBRECHTS, P., MCNEIL, A., Straumann, D., 2002. Correlation and dependence properties in risk management: properties and pitfalls. In: Dempster, M. (Ed.), *Risk Management: Value at Risk and Beyond*. Cambridge University Press, Cambridge, pp. 176–223.
- [8] SKLAR, A., 1959. Fonctions de répartition à n dimensionset leurs marges. *Publ. Inst. Statist. Univ. Paris* 8,229–231.
- [9] CHEN, X., HANSEN, L.P., CARRASCO, M., 1998. Nonlinearity and temporal dependence. Working Paper, University of Chicago.
- [10] CHEN, X., FAN, Y., 2004b. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *Canadian Journal of Statistics*.

Current address

Jegors Fjodorovs, Phd student

Affiliation: Department of Probability Theory and Mathematical Statistics, Riga Technical University

Address: 1 Kalku Street, LV-1658, Riga, Latvia

e-mail: Jegors.Fjodorovs@rtu.lv

Andrejs Matvejevs, Dr.sc.ing., Professor

Affiliation: Department of Probability Theory and Mathematical Statistics, Riga Technical University

Address: 1 Kalku Street, LV-1658, Riga, Latvia

e-mail: Andrejs.Matvejevs@rtu.lv

NONSMOOTH FUNCTION APPROXIMATION IN PRACTICAL CHANGE POINT PROBLEM

HECKENBERGEROVÁ Jana, (CZ), MAREK Jaroslav, (CZ),
SOUČKOVÁ Jitka, (CZ), TUČEK Pavel, (CZ)

Abstract. Indirect method for evaluation of critical micelle concentration, owing to their wide application potential, are in the great interest of the chemist community. In this article we propose regression models based on explicitly defined nonsmooth functions for approximation dependence of micelle concentration and voltage potential. The main goal is to find the change point of approximation function and determine the value of critical micelle concentration including its uncertainty.

Key words and phrases. Critical micelle concentration, Nonlinear regression model, Linearization, BLUE, Least Squares Method.

Mathematics Subject Classification. Primary 62J02, 62J05; Secondary 62B15

1 Introduction

Paper presents the solution of actual problem solved within chemistry community. They are currently looking for general algorithm that determinates the change point of micelle concentration known as critical micelle concentration (CMC).

The critical micelle concentration is defined as relatively small range of concentration of surfactant above which micelles are spontaneously formed. The value of the CMC is important parameter for characterization of each micelle-forming compound/surface active compound. Electrochemical techniques are not least important group used to observation of micelle aggregation. Conductivity measurement is the most widely used electrochemical method for CMC determination. Formation of micelle can be indirectly measured by a cyclic voltametric method on a hanging mercury drop electrode without electrochemical active probe [1].

Previously, values of CMC were evaluated from concentration dependence as the intersection of two straight lines. The approximation lines were intuitively interlaid through concentration data, therefore CMC value was not accompanied with its precision or its uncertainty.

In this paper, we suggest to use statistical regression models to solve CMC evaluation problem. In statistical point of view, CMC can be defined as a change point in the behavior of the micelle. For estimation of CMC value augmented with its variance, it is required to find/design suitable regression models either with or without additional constraints [2], [3].

First CMC algorithm, introduced in this paper, is based on two straight lines approximation, same as standard CMC evaluation methods. Next proposed approximations is using two quadratic functions with one intersection. Last but definitely not least CMC algorithm is combining advantages of previous algorithms and it is based on assumption of concavity of both quadratic approximation functions. The presumption of concavity comes from data character and surprising failure of previous algorithm.

All designed algorithms are tested on real data-sets and compared by the variance of change point, coefficient of determination and residual variance. Employment of these brand-new algorithms to chemical research will benefit whole chemistry community dealing with CMC evaluation problem. After all, they will gain simple and general tool for determination of critical micelle concentration value including its uncertainty.

This paper consists 4 sections. Section 2 describes statistical theory in background of CMC algorithms, especially the model of incomplete measurement without constrains and linearization of nonlinear models. Three designed CMC algorithms are summarized and evaluated in following Section 3. Last section contains major conclusions and directions of future work.

2 Statistical background

In the first section, parts of the theory of nonlinear regression models are presented. See [2], [3] or [4]. Major accent is given to the model of incomplete measurement without constrains and to the properties of unknown parameter estimator. Best Linear Unbiased Estimator (BLUE) is derivated with the help of Least Squares Method and its variance follows from matrix theory. In the following part, description of linearization process is presented. It shows us the procedure how to transform the nonlinear regression model to the linear one. The linearization process is mainly based on the Taylor series expansion and neglecting of the elements with second order and higher.

2.1 Model of incomplete measurement without constrains

Definition 2.1

The model of incomplete measurement without constrains is given in the form of

$$\mathbf{Y} \sim (\mathbf{F}\boldsymbol{\beta}, \sigma^2\mathbf{V}), \quad (1)$$

where \mathbf{Y} is n -dimensional random vector of measurements, $\sigma^2\mathbf{V}$ is known symmetric covariance matrix of type $(n \times n)$, \mathbf{F} is $(n \times k)$ -matrix and $\boldsymbol{\beta} \in R^k$ is vector of unknown parameters.

If \mathbf{F} is full-ranked in columns so $r(\mathbf{F}) = k < n$ and \mathbf{V} is a positively definite matrix then the model (1) is regular.

Theorem 2.2

Let

$$\mathbf{Y} \sim N_n(\mathbf{F}\boldsymbol{\beta}, \sigma^2\mathbf{V}) \quad (2)$$

be regular regression model of incomplete measurement without constraints, then BLUE (Best Linear Unbiased Estimator) of unknown first order parameter β in the model (2) is given by

$$\hat{\beta} = (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}^{-1}\mathbf{Y}. \quad (3)$$

The covariance matrix $\text{Var}(\hat{\beta})$ is in the form of

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}. \quad (4)$$

and

$$\hat{\beta} \sim N_k(\beta, \text{Var}(\hat{\beta})). \quad (5)$$

Proof. The derivation of the relations for the estimators is based on the least-squares method. For details see [2] or [3].

2.2 Linearization of nonlinear model

Let us assume that measurements are forming together the random vector \mathbf{Y} satisfying the nonlinear regression model

$$\mathbf{Y} = \Phi(\beta) + \varepsilon, \quad \varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{V}), \quad (6)$$

where $\Phi: R^k \rightarrow R^n$ is known nonlinear function, $\beta \in R^k$ is the unknown parameter, R^k is the parametric space and \mathbf{V} is the known positive definite matrix. Further, we consider the point $\beta^0 \in R^k$ and its neighbourhood $\mathcal{O}(\beta^0)$ in the parametric space R^k such that the true value of the parameter β is lying inside the $\mathcal{O}(\beta^0)$.

Additional assumptions about the model (6) are:

1. model is regular at the point β^0 , so $r(\mathbf{F}) = k$ where $\mathbf{F} = \frac{\partial \Phi(\beta)}{\partial \beta'}|_{\beta^0}$;
2. for arbitrary $\beta \in \mathcal{O}(\beta^0)$ and $\forall i, j, l \in 1, \dots, k: \frac{\partial^3 \Phi(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} = 0$.

Above-mentioned assumptions about the model (6) imply that the parameter space R^k can be restricted to the set $\mathcal{O}(\beta^0)$, and furthermore the model can be approximated by the quadratic model, i.e.

$$\mathbf{Y} - \Phi_0 \sim N_n(\mathbf{F}\delta\beta + \frac{1}{2}\kappa(\delta\beta), \sigma^2\mathbf{V}), \quad \beta \in \mathcal{O}(\beta^0) \quad (7)$$

where

$$\Phi_0 = \Phi(\beta^0), \quad \kappa(\delta\beta) = \begin{pmatrix} \kappa_1(\delta\beta) \\ \vdots \\ \kappa_n(\delta\beta) \end{pmatrix},$$

$$\kappa_i(\delta\beta) = (\beta - \beta^0)' h_i (\beta - \beta^0), \quad h_i = \frac{\partial^2 \Phi_i(\beta^0)}{\partial \beta' \partial \beta'}, \quad i = 1, \dots, n.$$

Linearization of the quadratic model (7) is made by neglecting the elements of the Taylor series expansion with second order and higher. So resulting linearized model is in form

$$\mathbf{Y} \sim N_n(\Phi_0 + \mathbf{F}\delta\beta, \sigma^2\mathbf{V}), \quad (8)$$

where $\Phi_0 = \Phi(\beta^0)$, $\mathbf{F} = \frac{\partial\Phi(\beta)}{\partial\beta}|_{\beta^0}$ and $\delta\beta = \beta - \beta^0$. Under special circumstances such as $\beta^0 = \mathbf{0}$ and $\Phi(\beta^0) = \mathbf{0}$, the linear model can be written as

$$\mathbf{Y} \sim N_n(\mathbf{F}\beta, \sigma^2\mathbf{V}). \quad (9)$$

It is important to mention that selection of initial parameter β^0 play the significant role in whole linearization process. Inappropriate estimation of parameter β can be caused by the low-quality initial solutions. For such cases, it is necessary to evaluate the linearization domain of differences $\delta\beta$ that are acceptable for linearization of the model. Problem of linearization domains is described and solved in full detail in [3].

3 Example and Algorithms

Example

Formation of sodium dodecyl sulfate (SDS) micelle in phosphate buffer system was observed by a cyclic voltametric method on a hanging mercury drop electrode without electrochemical active probe. All measurements of potential for different micelle concentrations were arranged to the column vector. Critical micelle concentration (CMC) of SDS have to be determined as the SDS micelle concentration in the change point of desorption peak potentials.

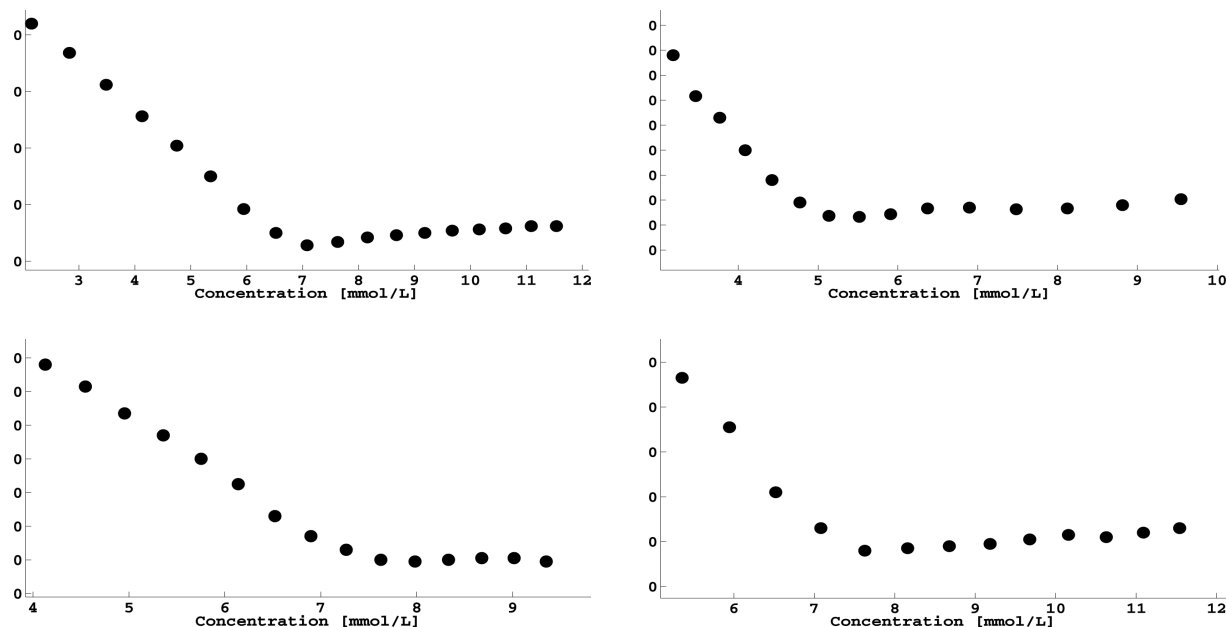


Figure 1: Dependence of potential and SDS micelle concentration - measurements

Figure 1 illustrates a few examples of source data catching the measurements of potential for different concentrations. One can recognize, that critical micelle concentration can be intuitively determined by the data change point in all pictured examples. However, it is also easy to see that the relation between concentration and potential is nonlinear and moreover nonsmooth. The aim of following section is to describe general algorithms for CMC determination. Main target of these algorithms is to find or design advisable mathematical model for evaluation of CMC including its uncertainty. Proposed relationships of potential and concentration are nonsmooth functions widely recognized as the most appropriate ones used for the approximation of the measured CMC.

3.1 Algorithm A:

First presented algorithm is based on standard way of critical micelle concentration evaluation. Let us assume, that dependence between potential and concentration is given by two straight lines and CMC is given by their intersection, resp. by their change point. So the measured potentials can be approximated by just one nonsmooth curve of absolute value function which follows,

$$f_A(\boldsymbol{\beta}, x) = \beta_1 x + \beta_2 + \beta_3 |x - \beta_4|, \quad (10)$$

where variable x represents micelle concentrations and $\beta_1, \beta_2, \beta_3, \beta_4$ are the unknown parameters. Parameter β_4 has a special meaning, it determinates the CMC value. It is important to mention, that values of micelle concentration are assumed to be deterministic.

Let all measurements of potential be forming n -dimensional random vector \mathbf{Y} , then we can design the nonlinear regression model

$$\mathbf{Y} = \boldsymbol{\Phi}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{V}), \quad (11)$$

where $\boldsymbol{\Phi}_i(\boldsymbol{\beta}) = f_A(\boldsymbol{\beta}, x_i) \quad \forall i = 1, \dots, n$, x_i is micelle concentration corresponding to the measurement Y_i and uncertainty of source data is described by standard deviation 0.5 [mV], so covariance matrix of measurements $\sigma^2 \mathbf{V} = 0.25 \cdot \mathbf{I}_n$.

In the next step, linearization of the model (11), we will follow the procedure described in Section 2.2. Let us assume that the initial solution $\boldsymbol{\beta}^0 = \mathbf{0}$ and $\boldsymbol{\Phi}(\boldsymbol{\beta}^0) = \mathbf{0}$, then linearized model for CMC Algorithm A is in form

$$\mathbf{Y} \sim N_n(\mathbf{F}\boldsymbol{\beta}, \sigma^2 \mathbf{V}),$$

where

$$\begin{aligned} \mathbf{F}_i &= \frac{\partial f_A(x_i, \boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}'} = \left(\frac{\partial f_A(x_i, \boldsymbol{\beta}^0)}{\partial \beta_1}, \frac{\partial f_A(x_i, \boldsymbol{\beta}^0)}{\partial \beta_2}, \frac{\partial f_A(x_i, \boldsymbol{\beta}^0)}{\partial \beta_3}, \frac{\partial f_A(x_i, \boldsymbol{\beta}^0)}{\partial \beta_4} \right) = \\ &= (x_i, 1, |x_i - \beta_4^0|, |x_i - \beta_4^0|(-\beta_3^0)), \quad i = 1, \dots, n. \end{aligned} \quad (12)$$

Evaluation of unknown parameter estimator $\hat{\boldsymbol{\beta}}$ is based on formulas from Section 2.1, especially on (3) and (4). Source data from two different measurements together with their approximation functions are illustrated in the Figure 2, it is supplemented with numerical results as well. In the first case, CMC is evaluated as 6.7492 with variance equal to 0.040683², in the other case the CMC corresponds to 4.8555 with standard deviation of 0.05959. Index of determination and residual variance are determined for each example to see propriety and quality of chosen approximation.

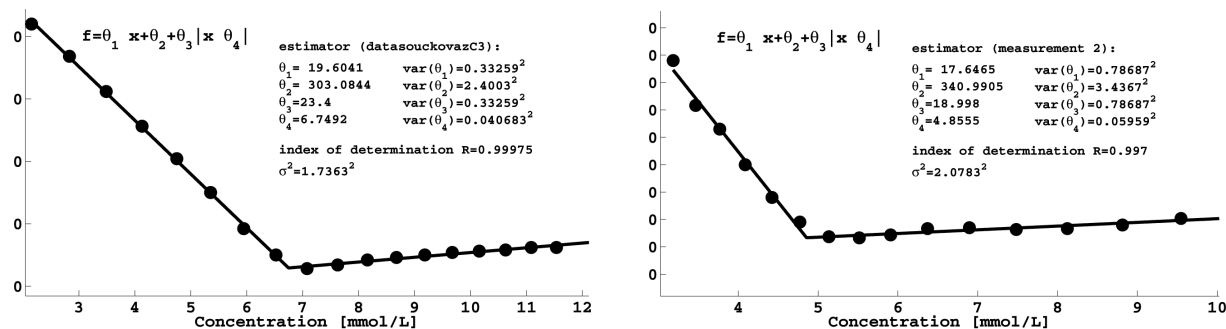


Figure 2: Resulting potential approximation and CMC evaluation with Algorithm A

3.2 Algorithm B:

In the following algorithm, we have decided to extend piecewise linear function (10) from previous Algorithm A to the quadratic one. Curve approximating the measurements of potential is then defined by

$$f_B(\beta, x) = \beta_1 x^2 + \beta_2 x + \beta_3 + \beta_4 |x - \beta_5|, \quad (13)$$

where β_i , $i = 1, \dots, 5$, are the unknown parameters and variable x represents micelle concentrations. Value of parameter β_5 corresponds to change point of concentration and therefore defines CMC value.

When we build the regression model for random vector of potential measurements, we have to consider the nonlinearity and nonsmoothness of approximating function (13). Resulting nonlinear model is in form

$$\mathbf{Y} = \Phi(\beta) + \varepsilon, \quad \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{V}), \quad (14)$$

where $\Phi_i(\beta) = f_B(\beta, x_i) \quad \forall i = 1, \dots, n$, x_i is corresponding concentration and covariance matrix $\sigma^2 \mathbf{V} = 0.25 \cdot \mathbf{I}_n$.

Linearization process of the model (14) is again based on the procedure described in Section 2.2. Let us use the estimator from previous CMC Algorithm A as initial parameter β^0 of linearization, then

$$\mathbf{Y} \sim N_n(\Phi_0 + \mathbf{F} \delta \beta, \sigma^2 \mathbf{V}),$$

where $\Phi_0 = \Phi(\beta^0)$, $\delta \beta = \beta - \beta^0$ and

$$\begin{aligned} \mathbf{F}_i &= \frac{\partial f_B(x_i, \beta^0)}{\partial \beta'} = \left(\frac{\partial f_B(x_i, \beta^0)}{\partial \beta_1}, \frac{\partial f_B(x_i, \beta^0)}{\partial \beta_2}, \frac{\partial f_B(x_i, \beta^0)}{\partial \beta_3}, \frac{\partial f_B(x_i, \beta^0)}{\partial \beta_4}, \frac{\partial f_B(x_i, \beta^0)}{\partial \beta_5} \right) = \\ &= (x_i^2, x_i, 1, |x_i - \beta_5^0|, |x_i - \beta_5^0|(-\beta_4^0)), \quad i = 1, \dots, n \end{aligned} \quad (15)$$

is linearized form of the model (14).

Last but not least, it is necessary to estimate the values of unknown parameter β . As the linearized model (15) is regular model of incomplete measurement without constraints, we can

use formulas (3) and (4) from Section 2.1. That leads us to

$$\begin{aligned}\delta\hat{\beta} &= (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}^{-1}(\mathbf{Y} - \Phi_0), \\ \hat{\beta} &= \beta^0 + \delta\hat{\beta} \text{ and} \\ \text{Var}(\hat{\beta}) &= \sigma^2 (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}.\end{aligned}\quad (16)$$

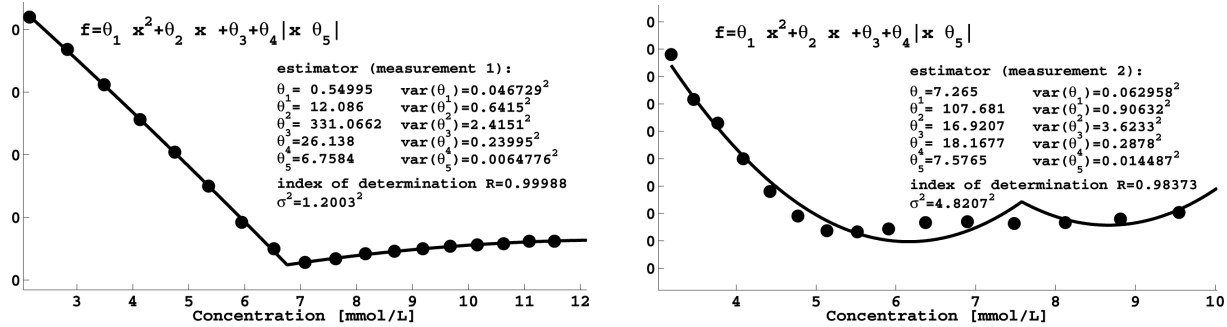


Figure 3: Resulting potential approximation and CMC evaluation with Algorithm B

Figure 3 shows the resulting approximation functions augmented with numerical results for two different datasets. For the first sample, the result looks very promising; CMC is 6.7584 with standard deviation of 0.0065 and index of determination resp. residual variance shows propriety of the chosen model. However, everything is negated in the second sample, where the quadratic model (15) completely failed in CMC value determination. Inappropriate estimation arises due to the low-quality initial solutions. Dealing with this issue by linearization domain determination will be part of our future work.

3.3 Algorithm C:

After the failure of Algorithm B, we investigated all test samples to find out what happened. Inappropriate estimation of CMC has been observed in 1% of experimental data, where resulting approximation curve became convex. For all other 99% test samples, Algorithm B worked correctly; moreover, indexes of determination were close to unit and approximation functions were concave. Results of this study leads us to the next Algorithm C, where concavity of approximation curve is presumed.

Let us assume, that concave piecewise quadratic function approximating voltage potential is given by relationship

$$f_C(\beta, x) = -\beta_1^2 x^2 + \beta_2 x + \beta_3 + \beta_4 |x - \beta_5|, \quad (17)$$

where $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the unknown parameters and x represents the micelle concentrations. Value of $-\beta_1^2$ is always non-positive, therefore $f_C(\beta, x)$ is concave function regardless of β_1 . Parameter β_5 indicates the critical micelle concentration, same as in previous algorithm.

The nonlinear, nonsmooth and concave regression model corresponding to chosen approximation is given by

$$\mathbf{Y} = \Phi(\beta) + \varepsilon, \quad \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{V}), \quad (18)$$

where \mathbf{Y} is random vector of potential measurements, $\Phi_i(\beta) = f_C(\beta, x_i) \quad \forall i = 1, \dots, n$, x_i is corresponding concentration and covariance matrix $\sigma^2 \mathbf{V} = 0.25 \cdot \mathbf{I}_n$.

After the linearization process (see Section 2.2), we are getting to

$$\mathbf{Y} \sim N_n(\Phi_0 + \mathbf{F}\delta\beta, \sigma^2 \mathbf{V}),$$

where initial parameter β^0 is evaluated from resulting function of Algorithm A, $\Phi_0 = \Phi(\beta^0)$, $\delta\beta = \beta - \beta^0$ and

$$\begin{aligned} \mathbf{F}_i &= \frac{\partial f_C(x_i, \beta^0)}{\partial \beta'} = \left(\frac{\partial f_C(x_i, \beta^0)}{\partial \beta_1}, \frac{\partial f_C(x_i, \beta^0)}{\partial \beta_2}, \frac{\partial f_C(x_i, \beta^0)}{\partial \beta_3}, \frac{\partial f_C(x_i, \beta^0)}{\partial \beta_4}, \frac{\partial f_C(x_i, \beta^0)}{\partial \beta_5} \right) = \\ &= (-2\beta_1^0 x_i^2, x_i, 1, |x_i - \beta_5^0|, |x_i - \beta_5^0|(-\beta_4^0)), \quad i = 1, \dots, n. \end{aligned} \quad (19)$$

Evaluation of estimator $\hat{\beta}$ and CMC value follows from equations in (16) and results for two chosen samples are illustrated in the Figure 4. This time Algorithm C worked correctly for problematic samples as well as for the normal ones; CMC is 4.8615 with standard deviation 0.011 for the second sample.

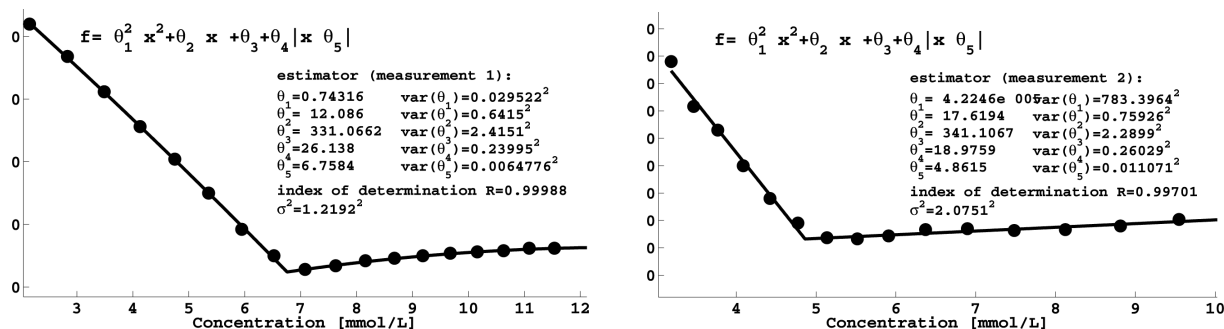


Figure 4: Resulting potential approximation and CMC evaluation with Algorithm C

There is another way how to deal with convexity-concavity approximation problem in this example. We can add the constrain to the regression model such as $f''(x) = -K^2$, $K \in R^+$ that guarantees the concavity of approximation function $f(x)$. Regression model is then in form of the model of incomplete measurement with the condition type II. Unfortunately, description of following algorithm is beyond the range of this paper. Details about constrained regression model can be found in [2] and [3].

3.4 Numerical results

To conclude this main section of our paper, all numerical results are summarized in Table 1. Residual sum of squares (S_e) and residual variance (s^2) are defined commonly as

$$S_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad s^2 = \frac{S_e}{n-p}, \quad (20)$$

where Y_i and \hat{Y}_i are observed values and estimated values of dependent variable \mathbf{Y}_i , n is the number of measurements and p is the number of regression parameters. As an additional

Algorithm	Sample No.	Figure	Estimator of change point	Uncertainty of estimator	S_e	s^2	R
A	1	2 left	6.7492	0.0407^2	12.0590	1.7363^2	0.99975
B	1	3 left	6.7584	0.0065^2	7.2036	1.2003^2	0.99988
C	1	4 left	6.7584	0.0065^2	7.4322	1.2192^2	0.99988
A	2	2 right	4.8555	0.0596^2	17.2773	2.0783^2	0.99700
B	2	3 right	7.5765	0.0145^2	116.1957	4.8207^2	0.98373
C	2	4 right	4.8615	0.0111^2	21.5302	2.0751^2	0.99701

Table 1: Estimator of change point and the criteria of the quality of regression model applied to the measured data

criterion for comparison of the models index of determination is used, which is defined in (21) and ranges from 0 to 1;

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}}. \quad (21)$$

It is easy to see from Table 1, that critical micelle concentration of two selected samples is 6.75 resp. 4.86. Algorithm A gives us results with highest uncertainty and residual variance. Statistical characteristics looks much better for Algorithm B and the first sample; however, fifth line of the Table 1 shows the failure of Algorithm B with incorrect estimator of CMC value, high sum of residual squares and low index of determination. Algorithm C combines advantages of Algorithms A and B together, it does not failed in the critical sample and gives us reasonably good results in statistical point of view.

4 Conclusions

In this paper, the problem of critical micelle concentration determination is solved. Although generally the solution of CMC problem is based on the intersection of two straight lines or two quadratic functions, we designed three different regression models grounded in explicitly defined approximation function of voltage potentials. Utilization of regression model into the CMC evaluation provides us CMC estimator augmented with uncertainty. Employment of these brand-new algorithms to chemical research will benefit whole chemistry community dealing with CMC evaluation problem.

The aim of future research is deeper investigation of Algorithm B failure. It is necessary to determine linearization domain for initial parameters and deal with convex-concave problem by using regression model with additional constraints for unknown parameters.

Acknowledgement

This research was supported at the University of Pardubice by MSMT Institutional funds.

References

- [1] MUKERJEE, P., MYSELS, K.J.: *Critical Micelle Concentrations of Aqueous Surfactant Systems*. Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. (US), Washington, 1971.

- [2] KUBACEK, L., KUBACKOVA, L., VOLAUFOVA, J.: *Statistical models with linear structures*. Veda, Publishing House of the Slovak Academy of Sciences, Bratislava 1995.
- [3] KUBACEK, L., KUBACKOVA, L.: *Statistics and Metrology* (in Czech). Publishing House of Palacký University, Olomouc 2000.
- [4] SEBER, G. A. F., WILD, C. J.: *Nonlinear Regression*, John Wiley & Sons, New Jersey, 2003.

Current address

Jana Heckenbergerová, Ph.D.

Faculty of Electical Engineering and Informatics, Department of Mathematics and Physics,
University of Pardubice, Studentska 95, 532 10 Pardubice, Czech Republic

&

Institute of Computer Science, Department of Nonlinear Modelling, Academy of Sciences
of the Czech Republic, Pod Vodarenskou vezi 271/2, 182 07 Prague 8, Czech Republic

tel: (+420) 466 037 096

email: jana.heckenbergerova@upce.cz

Jaroslav Marek, Ph.D.

Faculty of Electical Engineering and Informatics, Department of Mathematics and Physics,
University of Pardubice, Studentska 95, 532 10 Pardubice, Czech Republic

tel: (+420) 466 037 219

email: jaroslav.marek@upce.cz

Jitka Součková, Mgr.

Faculty of Science, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic,

tel: (+420) 585 634 442

email: jitka.souckova@centrum.cz

Pavel Tuček, Ph.D.

Faculty of Science, Palacký University, Tř. Svobody 26, 771 46 Olomouc, Czech Republic,

email: pavel.tucek@upol.cz

COMPARISON OF TWO BAYESIAN APPROACHES TO SPC

JAROŠOVÁ Eva, (CZ)

Abstract. Properties of two Bayesian control charts are examined. The m-chart is represented by the credibility interval derived from the posterior distribution of the process mean with the normal prior, the second approach consists in posterior odds determination based on two models involving a change point with a geometric prior. Performance of both types of charts is measured by means of the cumulative probability of false alarm and the median run length. Characteristics are estimated via the simulation study and compared with those of the classical Shewhart chart.

Key words. Current mean estimation, change point detection, cumulative probability of false alarm, median run length, posterior odds.

Mathematics Subject Classification: 62P10, 62-07

1 Introduction

In last twenty years many manufacturing organizations adopted new production strategies that are characterized by short runs during which only relatively small quantities of parts are made, or by discontinuous production due to frequent product changeover. Conventional Shewhart control charts cannot be applied anymore because in these short-run processes a sufficient amount of data needed to construct control limits is not available. Various frequentist methods have been introduced for analyzing data from these processes, see e.g. [2]. A Bayesian approach seems to be a good solution in situations when only few data are available and yet some decision must be made. Our uncertainty about a quantity, for example the process mean, before observations of the process are obtained, is expressed in form of a prior distribution of this quantity. The posterior distribution is derived making use of both observed data and the prior information. Decisions are then based on the posterior distribution.

In this work we confine to normal distribution of observations and to detection of a shift in the process mean. Sustained special causes are assumed that continue until they are identified and removed, what means that the shift is permanent. Two simple methods based on different approaches mentioned above are presented: the Bayes' estimation of the current mean (m-chart) and the Bayesian change point detection (posterior odds chart). For their use in statistical process

control (SPC), the performance in terms of the risk of false alarm or the speed of detection of a change in the process mean is of the main interest.

2 Bayesian m-chart

Suppose we have a random sample of r observations $\mathbf{x} = (x_1, x_2, \dots, x_r)^T$ from a normal distribution with mean μ and known variance σ^2 , where μ is regarded as a variable with the prior distribution $p(\mu)$. Using the Bayes' rule, the posterior density of μ is given by

$$p(\mu | \mathbf{x}) = \frac{p(\mu)p(\mathbf{x} | \mu)}{\int p(\mu)p(\mathbf{x} | \mu)d\mu}, \quad (1)$$

where the likelihood $p(\mathbf{x} | \mu)$ has the form

$$p(\mathbf{x} | \mu) = \frac{1}{(2\pi\sigma^2)^{\frac{r}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^r (x_i - \mu)^2\right). \quad (2)$$

Using the identity $\sum_{i=1}^r (x_i - \mu)^2 = r(\bar{x} - \mu)^2 + \sum_{i=1}^r (x_i - \bar{x})^2$, where $\bar{x} = \frac{1}{r} \sum_{i=1}^r x_i$ and ignoring the constant of proportionality in the resultant formula, the posterior density (1) can be expressed in the form

$$p(\mu | \mathbf{x}) = \frac{p(\mu) \exp\left(-\frac{r}{2\sigma^2} (\bar{x} - \mu)^2\right)}{\int p(\mu) \exp\left(-\frac{r}{2\sigma^2} (\bar{x} - \mu)^2\right) d\mu}. \quad (3)$$

When the conjugate normal prior

$$p(\mu) = \frac{1}{\sqrt{2\pi w_0^2}} \exp\left(-\frac{(\mu - m_0)^2}{2w_0^2}\right) \quad (4)$$

with hyperparameters m_0 and w_0^2 is chosen, we get

$$p(\mu | \bar{x}) \propto \exp\left\{-\frac{1}{2\sigma^2 w_0^2 / (\sigma^2 + r w_0^2)} \left(\mu - \frac{\sigma^2 m_0 + r w_0^2 \bar{x}}{\sigma^2 + r w_0^2}\right)^2\right\}. \quad (5)$$

It follows that the posterior mean is

$$m = \frac{\sigma^2 m_0 + r w_0^2 \bar{x}}{\sigma^2 + r w_0^2} = \frac{\sigma^2}{\sigma^2 + r w_0^2} m_0 + \frac{r w_0^2}{\sigma^2 + r w_0^2} \bar{x}. \quad (6)$$

and the posterior variance is

$$w^2 = \frac{\sigma^2 w_0^2}{\sigma^2 + r w_0^2}. \quad (7)$$

Suppose samples of fixed size and $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_r$ instead of individual observations are available. Then the posterior mean and variance are expressed by

$$m = \frac{\sigma^2}{\sigma^2 + r n w_0^2} m_0 + \frac{r n w_0^2}{\sigma^2 + r n w_0^2} \bar{\bar{x}} \quad (8)$$

$$w^2 = \frac{\sigma^2 w_0^2}{\sigma^2 + r n w_0^2}. \quad (9)$$

where $\bar{\bar{x}} = \sum_{i=1}^r \bar{x}_i / r$. It is convenient to use recursive formulas

$$m_i = \left(1 - \frac{n w_{i-1}^2}{\sigma^2 + n w_{i-1}^2} \right) m_{i-1} + \frac{n w_{i-1}^2}{\sigma^2 + n w_{i-1}^2} \bar{x}_i, \quad (10)$$

$$w_i^2 = \frac{\sigma^2 w_{i-1}^2}{\sigma^2 + n w_{i-1}^2}. \quad (11)$$

The hypothesis $H_0 : \mu = \mu_0$ against the two-sided hypothesis $H_1 : \mu \neq \mu_0$ is tested by means of the credibility interval

$$lcl_i, ucl_i = m_i \pm u_{1-\alpha/2} w_i, \quad (12)$$

where $u_{1-\alpha/2}$ is the upper $\alpha/2$ percentage point of the $N(0,1)$ distribution. Control limits lcl_i and ucl_i are computed sequentially for $i = 1, 2, \dots$. As soon as for some i μ_0 lies outside the interval, $H_0 : \mu = \mu_0$ is rejected and a signal is given that the process mean has changed.

The prior distribution $N(m_0, w_0^2)$ expresses our prior belief regarding the mean of the process. Naturally we expect that the mean of the process does not differ from the target value μ_0 at the beginning of SPC and we put $m_0 = \mu_0$. Small w_0^2 corresponds to strong belief about μ and it will require more samples than the posterior m_i moves so far from μ_0 that a shift could be detected. The magnitude of the prior parameter w_0^2 , which is the initial value of (11), affects the relative size of weights in (10) and thus the quickness of the reaction to a shift in the mean of the process. Based on the results of simulations, the choice $w_0^2 = \sigma^2 / n$ seems reasonable with regard to the chart's performance; estimated cumulative probability of false alarm (see further) is sufficiently low and the median run length is the same as for the higher value of w_0^2 .

3 Performance of Bayesian m-chart

In the classical Shewhart control chart (under the assumptions of normal distribution and independence), a type I error probability, called a risk of false alarm in SPC, equals to 0.0027. This small level of the risk is a key property in SPC because false alarms can have serious consequences. To enable comparison with other control charts where the probability of false alarm is varying, the cumulative probability that a false alarm will occur within a given number of plotted points can be used. The cumulative probability of false alarm (*CPFA*) of the Shewhart chart can serve as a benchmark when different control charts are examined. When r sequential samples are considered, the probability that at least one false signal will occur within r points is given by $CPFA(r) = 1 - 0.9973^r$.

The other important property is run length, i.e. a number of samples taken from a process till a signal occurs. Traditionally, the expected value of run lengths called the average run length (*ARL*) is used to measure sensitivity of control charts. Large in-control values of *ARL* and conversely small *ARL* when a shift in process parameter has occurred are desirable. The run length on the Shewhart chart follows a geometric distribution with the mean $1/p$, where p is the probability of the signal at any sampling time. Owing to the fact that the geometric distribution is quite skewed and its standard deviation is very large, using *ARL* has been criticized in recent years and the median or other percentiles or the run length distribution are recommended by some analysts. The median run length (*MRL*) of the Shewhart chart is $\lceil -\log(2)/\log(1-p) \rceil$, where $\lceil \cdot \rceil$ denotes the greatest integer function.

Properties of the *m*-chart were examined by means of simulations. The simulation study was realized according to the following scheme: The in-control state of the process was represented by the $N(10,1)$ distribution. When a shift δ at the sampling point c was considered, first $c-1$ samples were generated from $N(10,1)$ and next samples were from $N(10+\delta,1)$. As soon as the stopping condition was valid, the particular run was interrupted and the position of the break point recorded to provide the information about when a signal occurred. Under chosen conditions, each run of length 100 at maximum was repeated 10 000 times. *CPFA*() was estimated through the cumulative proportion of runs (out of 10 000) in which the first signal occurred within the interval $< i, i+1 >$. The run length was determined according to $i - c + 1$, where i corresponds to the breakpoint and c is the shift point. Consequently, the minimum run length is equal to 1. Because of false signal occurrences, the sample size was varying and less than 10 000 when the median was calculated. Values $m_0 = 10$ and $w_0^2 = \sigma^2 / n = 1/n$ of hyperparameters were chosen. The posterior characteristics and the 99% credibility limits lcl_i and ucl_i were computed sequentially. The size of samples ranged from $n = 1$ to $n = 5$, the position of the shift point varied from $c = 1$ to $c = 50$.

The cumulative proportions of false signals are displayed in Figure 1 together with the corresponding cumulative probabilities of the Shewhart chart given by $CPFA(i) = 1 - 0.9973^i$. The curves represent smoothed bar charts of the cumulative proportions over intervals of width equal to 5 and so they do not start at the origin. It is apparent that the cumulative proportion of false alarms increases slower than the cumulative probability of the Shewhart chart and so the risk of false alarm is sufficiently low for any $i = 1, 2, \dots$.

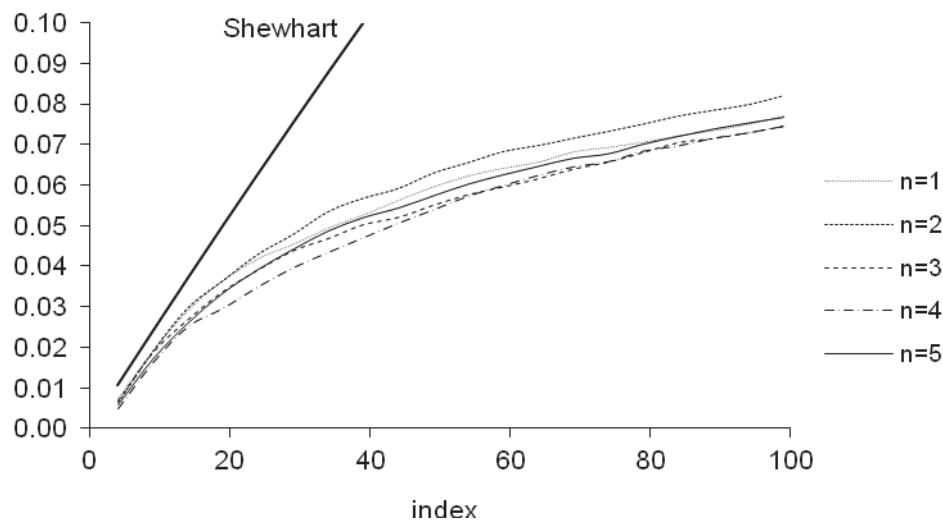


Figure 1. *m*-chart, estimated *CPFA* (in-control process)

The effect of sample size n on the distribution of run length is illustrated in Figure 2, where the estimated probability mass function of run lengths, i.e. the proportion of cases when given run length was observed, is displayed. Here the shift $\delta = 1.5\sigma$ at the sampling point $c = 10$ was considered. As n increases, the run lengths tend to be shorter and exhibit smaller spread.

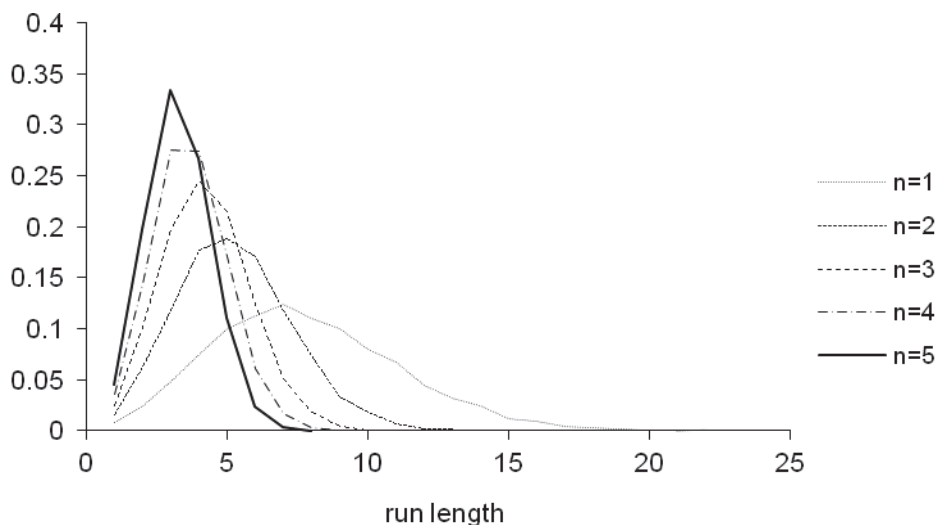


Figure 2. *m*-chart, run length distribution ($\delta = 1.5\sigma$, $c = 10$)

The run length distribution depends on the interval to the shift point c . Run lengths tend to increase when a shift occurs later. The increase is most apparent for small shifts and small sample sizes. Table 1 introduces median run lengths for some chosen values of n , δ , and c . The median run length of the Shewhart chart is given in the rightmost column.

Table 1. Bayesian m -chart, median run length

		$c = 1$	$c = 10$	$c = 20$	$c = 30$	$c = 40$	$c = 50$	Sh
$\delta = 1.5\sigma$	$n = 1$	4	7	10	11	13	14	11
	$n = 5$	2	3	4	5	6	6	1
$\delta = 3\sigma$	$n = 1$	2	4	5	6	6	7	1
	$n = 5$	1	2	2	3	3	3	1

4 Bayesian detection of the change point

Consider that sequential observations from the in-control process follow the distribution $f_0(x_i)$. Then a persistent shift in the mean occurs and another distribution $f_1(x_i)$ is valid. Suppose that r independent sequential observations x_1, \dots, x_r are available. Let $p(\mathbf{x}|\nu)$ denote the joint density function of $\mathbf{x} = (x_1, \dots, x_r)^T$ conditioned by the value of ν . Then

$$p(\mathbf{x}|\nu) = \begin{cases} \prod_{i=1}^r f_1(x_i) & \nu = 1 \\ \prod_{i=1}^{\nu-1} f_0(x_i) \prod_{i=\nu}^r f_1(x_i) & 2 \leq \nu \leq r \\ \prod_{i=1}^r f_0(x_i) & \nu > r \end{cases} \quad (13)$$

where $\nu > r$ corresponds to the case when the change point has not occurred until sampling time r . Using Bayes' rule, the posterior distribution of ν given $\mathbf{x} = (x_1, \dots, x_r)^T$ is

$$p(\nu|\mathbf{x}) = \frac{p(\nu)p(\mathbf{x}|\nu)}{\sum_{\nu=1}^{\infty} p(\nu)p(\mathbf{x}|\nu)}. \quad (14)$$

With the prior distribution of the change point $p(\nu) = p(1-p)^{\nu-1}$, $\nu = 1, 2, \dots$, the posterior probability that the change point has occurred by the r th sampling time is

$$\Pr(\nu \leq r | \mathbf{x}) = \frac{p \prod_{i=1}^r f_1(x_i) + \sum_{\nu=2}^r p(1-p)^{\nu-1} \prod_{i=1}^{\nu-1} f_0(x_i) \prod_{i=\nu}^r f_1(x_i)}{p \prod_{i=1}^r f_1(x_i) + \sum_{\nu=2}^r p(1-p)^{\nu-1} \prod_{i=1}^{\nu-1} f_0(x_i) \prod_{i=\nu}^r f_1(x_i) + (1-p)^r \prod_{i=1}^r f_0(x_i)}. \quad (15)$$

Cancelling out $(1-p)^r \prod_{i=1}^r f_0(x_i)$ and putting $R_i = \frac{f_1(x_i)}{f_0(x_i)}$, we have

$$\Pr(\nu \leq r | \mathbf{x}) = \frac{\frac{p}{(1-p)^r} \sum_{\nu=1}^r (1-p)^{\nu-1} \prod_{i=\nu}^r R_i}{\frac{p}{(1-p)^r} \sum_{\nu=1}^r (1-p)^{\nu-1} \prod_{i=\nu}^r R_i + 1} = \frac{p \sum_{\nu=1}^r \prod_{i=\nu}^r \frac{R_i}{(1-p)}}{p \sum_{\nu=1}^r \prod_{i=\nu}^r \frac{R_i}{(1-p)} + 1}. \quad (16)$$

Suppose that n observations are available at each sampling time and sample means \bar{x}_i are obtained. They are assumed to have a normal distribution with mean μ_0 and variance σ^2/n when the process is in control. Denoting $f_0(\bar{x}_i)$ and $f_1(\bar{x}_i)$ densities of normal distributions $N(\mu_0, \sigma^2/n)$ and $N(\mu_1, \sigma^2/n)$, respectively, and assuming $\mu_1 = \mu_0 + \delta^*$, we have

$$R_i = \frac{f_1(\bar{x}_i)}{f_0(\bar{x}_i)} = \exp\left(-\frac{n\delta^{*2}}{2\sigma^2} + \frac{n\delta^*}{\sigma^2}(\bar{x}_i - \mu_0)\right) \quad (17)$$

Equation (7.9) can be rewritten in the form

$$\Pr(\nu \leq r | \mathbf{x}_r) = \frac{pZ_{p,r}}{pZ_{p,r} + 1}, \quad (18)$$

where $Z_{p,r} = \sum_{\nu=1}^r \prod_{i=\nu}^r \frac{R_i}{(1-p)}$ can be calculated recursively according to

$$Z_{p,i} = \frac{R_i}{1-p} (Z_{p,i-1} + 1) \text{ for } 1 \leq i \leq r, \quad Z_0 = 0. \quad (19)$$

Decision is based on the posterior odds

$$\frac{\Pr(H_1 | \mathbf{x}_r)}{\Pr(H_0 | \mathbf{x}_r)} = pZ_{p,r}, \quad (20)$$

where $\Pr(H_1 | \mathbf{x}) = \Pr(\nu \leq r | \mathbf{x})$ and $\Pr(H_0 | \mathbf{x}_r) = 1 - \Pr(H_1 | \mathbf{x}_r)$. The posterior odds are calculated sequentially and as soon as $pZ_{p,r} > \frac{\pi^*}{1-\pi^*}$, where π^* is close to 1 ($0 < \pi^* < 1$), the hypothesis H_0 is rejected and the out-of-control signal is given. This condition is in agreement with the Shiryaev's stopping rule (see e.g. [4]).

The statistic $Z_{p,r}$ which is a part of the posterior odds involves the prior parameter p . Larger values of p imply a relatively high weight given to low values of ν and represent our belief that the process may move to the out-of-control state very soon. If there is no reason to expect a change soon after we started the process control, smaller value of p seems to be reasonable. On the other hand, it can be shown that a larger value of p contributes to a higher speed of detection. It is interesting that $pZ_{p,r}$ for various p close to 0 and $Z_r = \sum_{\nu=1}^r \prod_{i=\nu}^r R_i$ for $p=0$ proposed by Roberts [6]

are similar. Consequently, the Shiryaev-Roberts stopping rule can be formulated by means of the condition $Z_r \geq \frac{\pi^*}{1-\pi^*}$ [4].

So far the single value μ_1 representing the out-of-control state of a process has been considered. Usually both positive and negative larger shifts off a target are undesirable. Two statistics R and consequently Z_p (or Z) for two values of δ^* must be calculated, similarly as in the two-sided CUSUM chart. Under assumption that both directions are of equal importance following formulas can be used:

$$R_i^+ = \exp\left(-\frac{n\delta^{*2}}{2\sigma^2} + \frac{n\delta^*}{\sigma^2}(\bar{x}_i - \mu_0)\right), \quad R_i^- = \exp\left(-\frac{n\delta^{*2}}{2\sigma^2} - \frac{n\delta^*}{\sigma^2}(\bar{x}_i - \mu_0)\right) \quad (21)$$

and

$$Z_{p,r}^+ = \sum_{v=1}^r \prod_{i=v}^r \frac{R_i^+}{(1-p)}, \quad Z_{p,r}^- = \sum_{v=1}^r \prod_{i=v}^r \frac{R_i^-}{(1-p)}. \quad (22)$$

In case of the Shiryaev-Roberts statistic similarly Z_r^+ and Z_r^- are get for $p=0$. Value $Z_{p,r} = \max\{Z_{p,r}^+, Z_{p,r}^-\}$ is crucial to possible triggering a signal.

The stopping threshold is derived from some conventional high value of the probability π^* . Kenett and Zacks (1998) suggest the value $\pi^* = 0.95$ but it appears that the risk of false signal is too high. The value $\pi^* = 0.99$ and consequently the stopping threshold $\pi^*/(1-\pi^*) = 99$ perform well when the posterior odds statistic and known process variance are considered, but π^* has to be higher for the S-R statistic (e.g. $\pi^* = 0.9973$).

As was already noted, δ^* represents the minimum (positive or negative) shift of the process mean such that the process with the mean $\mu_1 = \mu_0 + \delta^*$ is considered to be out of control. Since changes of the mean up to $\pm 1.5\sigma$ are acceptable in Six Sigma methodology, the value $\delta^* = 1.5\sigma$ seems to be appropriate. It can be expected that true shifts δ for which $|\delta| > |\delta^*|$ tend to be detected earlier.

The prior parameter p affects both the run length and the risk of false alarm. The median run length decreases with the increase of p and it is shortest for the S-R statistic. The simulations indicated the similar run-length performance of the posterior odds with $p = 0.1$ and that of the S-R statistic.

5 Performance of the posterior odds chart

Both pZ_p and S-R statistics were studied together to make some comparisons possible, although the latter statistic is not the posterior odds. The basic scheme of the simulation study was the same as before but now the posterior odds pZ_p or the S-R statistic were calculated sequentially and the corresponding stopping criteria were applied. The in-control distribution was represented by $N(10,1)$. The shift $\delta^* = 1.5\sigma$ has been chosen to define the out-of-control state.

Curves of cumulative proportions of false signals for the posterior odds statistic in Figure 3 exhibit linear increase and indicate that the probability of false alarm at any sample time is constant similarly as in the Shewhart chart. But unlike the Shewhart chart the risk of false alarm decreases

with the sample size. The curves of cumulative proportions for the S-R statistic were similar and are not displayed. A character of false signal occurrences differs from the m -chart (cf. Figure 1).

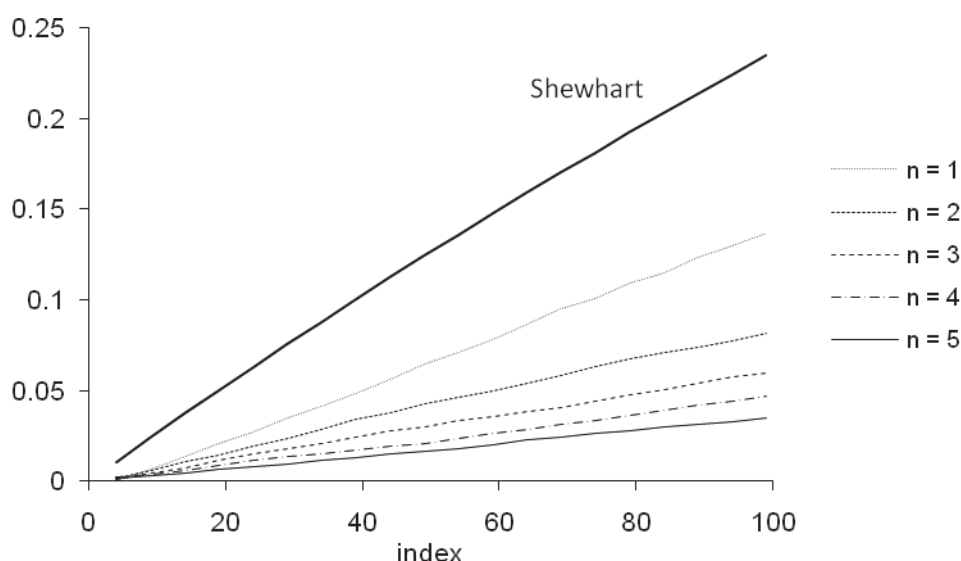


Figure 3. Posterior odds chart, estimated $CPFA$ (threshold 99)

The run length distribution is not influenced by the position of the shift point. The curves for different values of c overlap and so they are not displayed. Run lengths of course depend on the sample size n , see Figure 4 where $\delta = 1.5\sigma$ was considered.

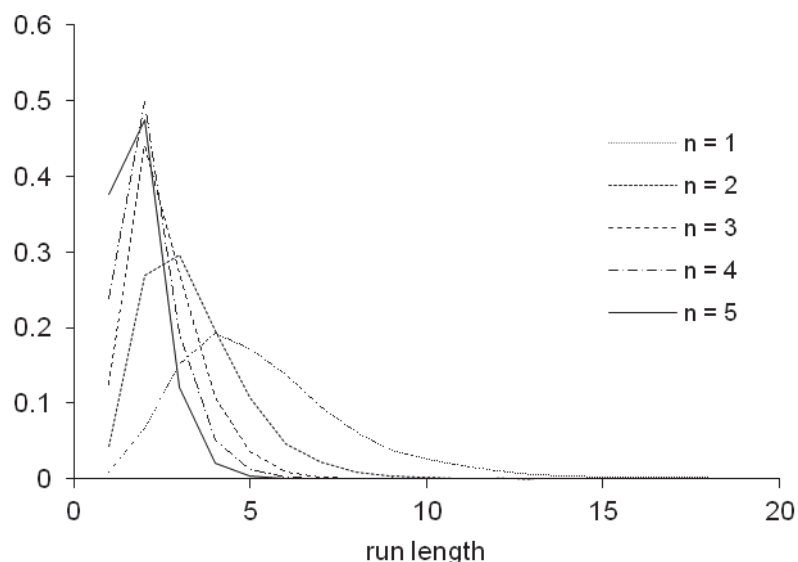


Figure 4. Posterior odds chart, run length distribution,

The median run lengths for both the posterior odds and the S-R statistics were mostly identical (see Table 2, where also MRL of the Shewhart chart is displayed).

Table 2. Posterior odds chart and median run length

		Post. odds	S-R	Sh
$\delta = 1.5\sigma$	$n = 1$	5	5	11
	$n = 5$	2	2	1
$\delta = 3\sigma$	$n = 1$	2	2	1
	$n = 5$	1	1	1

6 Conclusion

It is apparent that under the assumption of known process variance the run-length performance of both charts is worse than the performance of the Shewhart chart, with the exception of individual observations and small shifts about 1.5σ . Moreover, the m-chart can be usable only soon after the startup. The charts are intended to be used in situations when the classical Shewhart chart cannot be constructed because of the lack of data and then the assumption of known variance is spurious. Some possibilities of the process variance estimation were explored by means of simulations. The variance was estimated sequentially similarly as in the self-starting CUSUM chart [3] but results are not shown here. Obviously, some prior information about the variance is required otherwise the performance of these charts is quite poor.

References

- [1] BOLSTAD, W.M.: *Introduction to Bayesian Statistics*. John Wiley & Sons, Hoboken, 2007.
- [2] DEL Castillo, E.; GRAYSON, J.M.; MONTGOMERY, D.C.; and RUNGER, G.C.: *A Review of Statistical Process Control Techniques for Short Run Manufacturing Systems*. In *Communications in Statistics – Theory and methods*, Vol. 25, No. 11, pp. 2723-2737, 1996.
- [3] HAWKINS, D.M.: Self-Starting Cusums for Location and Scale. *The Statistician* 36, pp. 299-315, 1987.
- [4] KENETT, R.S. and ZACKS, S.: *Modern Industrial Statistics: Design and Control of Quality and Reliability*. Duxbury Press, Belmont CA, 1998.
- [5] MONTGOMERY, D.C.: *Statistical Quality Control: A Modern Introduction*, 6th ed. John Wiley & Sons, Inc., Hoboken, 2009.
- [6] ROBERTS, S.W.: *A comparison of some control chart procedures*. In *Technometrics*, Vol. 8, pp. 411-430, 1966.

Current address

doc. Ing. Eva Jarošová, CSc.

Skoda Auto University

Tř. Václava Klementa 864

293 60 Mladá Boleslav

Czech Republic

Phone Number 732469892

e-mail: jarosova@is.savs.cz

QUANTILE CHARACTERISTICS OF CONDITIONAL DISTRIBUTIONS OF FINITE MIXTURES

MALÁ Ivana (CZ)

Abstract. In the text conditional distributions of positive value continuous random variable are studied in the case of given information that a value of this variable is less or greater than a given quantity. The variable is supposed to have probability distribution with a density function given as a finite mixture of probability densities. The formulas for the conditional probability density and cumulative distribution function are derived in the text. Moreover, conditional quantile characteristics of location and variability are evaluated as a function of conditions. Per capita income of the households in the Czech Republic in 2008 is modelled with the mixture of three lognormal components and conditional distributions are studied. Conditional densities and conditional characteristics are shown in figures.

Key words finite mixture, conditional probability distribution, conditional characteristics, R

Mathematics Subject Classification: C13; C51

1 Introduction

In the text conditional distributions of a positive value random variable with continuous distribution given as a finite mixture are analysed. Conditions of interest are described by an additional information that the analysed random variable is less than a given value or it is greater than a given value. It requires only straightforward computations to derive conditional density and conditional distribution function of the variable under the condition that the variable is greater or less than the given quantity. Quantile characteristics of conditional distributions can be evaluated only with the use of numeric procedures. The results are used for the modelling of net annual income per capita (in Czech Crowns, CZK) for the Czech households in 2008. In the case of an analysis of incomes, information that the income is less or greater (the condition that the value is in a given finite interval is not treated in this text) that the given quantity is frequently known. The intervals in this text are of the form $(0; z)$ and $(z; \infty)$.

2 Methods

2.1 Finite mixture of probability distributions

Suppose that a probability distribution of a positive random variable Y is a mixture of K probability densities $f_j(x; \theta_j)$. These component densities depend on the unknown (p dimensional) vector parameter θ_j , $j=1, \dots, K$. Mixing probabilities (weights in the mixture) π_j should fulfil obvious constraints

$$\sum_{j=1}^K \pi_j = 1, \quad 0 \leq \pi_j \leq 1, \quad j = 1, \dots, K. \quad (1)$$

Under these assumptions the density of Y is given as a weighted average of densities f_j with weights π_j (mixing proportions) in the form

$$f(y; \psi) = \sum_{j=1}^K \pi_j f_j(y; \theta_j). \quad (2)$$

The mixture density (2) depends on the vector parameter ψ

$$\psi = (\pi_1, \dots, \pi_{K-1}, \theta_j, \quad j = 1, \dots, K), \quad (3)$$

with $(K-1)$ parameters π_j and Kp parameters θ_j . It follows from (2) that a cumulative distribution function of a mixture is defined as

$$F(y; \psi) = \sum_{j=1}^K \pi_j F_j(y; \theta_j), \quad (4)$$

where $F_j(x; \theta_j)$ is a distribution function of the j -th distribution. Generally, there is not a close formula for quantiles of a mixture and the $100P\%$ quantile y_P can be found from the definition as a (numeric) solution of equality

$$F(y_P; \psi) = \sum_{j=1}^K \pi_j F_j(y_P; \theta_j) = P, \quad 0 < P < 1. \quad (5)$$

For a numeric procedure we have a good first approximation, we can take weighted average of quantiles of components with weights equal to mixing proportions.

2.2 Conditional distributions

In this text we will analyse two possible conditions. For a given positive value of z the conditional distributions of Y given that $Y \leq z$ and $Y > z$ are of interest. The dependence of quantile characteristics on the value of z is analysed.

Suppose that Y has probability density (2) and we have additional information that $Y \leq z$. Straightforward computation gives for the cumulative distribution function $F(y|Y \leq z)$ the form (for given z)

$$\begin{aligned}
 F(y|Y \leq z) &= 0 & y \leq z, \\
 &= P(Y \leq y|Y \leq z) = \frac{P(Y \leq y)}{P(Y \leq z)} = \frac{F(y)}{F(z)} = \frac{\sum_{j=1}^K \pi_j F_j(y)}{\sum_{j=1}^K \pi_j F_j(z)}, & 0 < y < z, \\
 &= 1 & y \geq z.
 \end{aligned} \tag{6}$$

Derivating (6) we obtain conditional density of Y given that $Y \leq z$ by the formula

$$\begin{aligned}
 f(y|Y \leq z) &= \frac{f(y)}{F(z)} = \frac{\sum_{j=1}^K \pi_j f_j(y)}{\sum_{j=1}^K \pi_j F_j(z)} & 0 < y < z, \\
 &= 0 & otherwise.
 \end{aligned} \tag{7}$$

The conditional density (7) cannot be written as a mixture of component probability conditional densities f_j/F_j of mixture components or as a mixture of densities f_j with new weights.

For the $100P$ % quantile we obtain (from (6)) equation

$$P = F(y_{zP}|Y \leq z) = \frac{\sum_{j=1}^K \pi_j F_j(y_{zP})}{\sum_{j=1}^K \pi_j F_j(z)}.$$

The formula can be rearranged to the equality

$$P \sum_{j=1}^K \pi_j F_j(z) = \sum_{j=1}^K \pi_j F_j(y_{zP}), \tag{8}$$

or

$$PF(z) = F(y_{zP}). \tag{9}$$

It means, that $100P$ -per cent quantile of the conditional distribution is equal to the quantile of the original distribution, the percentage is given by the left part of the equality in (9). This formula corresponds with [3]. In the text the first and the ninth deciles ($P=0.1$ and 0.9), lower and upper quartiles ($P=0.25$ and 0.75) and median ($P=0.5$) were evaluated according to (9) as the functions of the condition z . From these values, characteristics of variability quartile range and deviation in the form $y_{z0.75} - y_{z0.25}$ and $0.5(y_{z0.75} - y_{z0.25})$, as well as these values for decile, can be evaluated.

For high values of z we obtain distribution similar to unconditional distribution of Y .

Suppose that Y has probability density (2) and we have additional information that $Y > z$. In this case formulas (6) and (7) turn to formulas

$$\begin{aligned}
 F(y|Y > z) &= P(Y \leq y|Y > z) = 0 & y \leq z \\
 &= \frac{P(z < Y \leq y)}{P(Y > z)} = \frac{F(y) - F(z)}{1 - F(z)} = \frac{\sum_{j=1}^K \pi_j (F_j(y) - F_j(z))}{1 - \sum_{j=1}^K \pi_j F_j(z)} & y > z,
 \end{aligned} \tag{10}$$

and

$$f(y|Y > z) = \frac{f(y)}{1 - F(y)} = \frac{\sum_{i=1}^K \pi_j f_j(y)}{1 - \sum_{i=1}^K \pi_j F_j(z)} \quad y > z, \quad (11)$$

$$= 0 \quad y \leq z.$$

Evaluation of quantile characteristics of this distribution is the same as above if we take conditional density and conditional distribution function form (10) and (11) instead of (6) and (7). Formulas (8) and (9) can be rewritten as

$$F(y_{zp}|Y > z) = \frac{\sum_{i=1}^K \pi_j (F_j(y_{zp}) - F_j(z))}{1 - \sum_{i=1}^K \pi_j F_j(z)} = P,$$

$$\sum_{i=1}^K \pi_j F_j(y) = P \left(1 - \sum_{i=1}^K \pi_j F_j(z) \right) + \sum_{i=1}^K \pi_j F_j(z). \quad (12)$$

It means, that 100P-per cent quantile of the conditional distribution is equal to the quantile of the original distribution, the percentage is given by the right part of the equality in (12). For small values of z the conditional distribution should be similar to the (unconditional) distribution of Y . For high values of z we obtain distribution of large values, in the case of incomes the distribution of high incomes.

All computations in this text were performed in R 2.13.1. Functions from the *stats*-package were used for searching for a root of a function and for a maximum of a function. Moreover, the *graphics*-package was used for three dimensional graphs (Figures 5-8). For the estimation of parameters in the mixture a package *flexmix* was used.

3 Data and Results

In the paper, data about Czech households from the Living Conditions Survey (a national module of the European Union Statistics on Income and Living Conditions (EU-SILC)) from 2009 are used, [6]. This dataset (includes 9,911 households) covers incomes of Czech households from 2008. From these data an annual net per capita income (in CZK) was evaluated as a ratio of total net income of a household and a number of members. In the problem of modelling of incomes we have frequently information that the income is less (or grater) than a given amount. The distribution of per capita income was modelled with the use of a mixture of three two-parametric lognormal distributions. In this model we have to estimate two parameters π_1 and π_2 ($\pi_3 = 1 - (\pi_1 + \pi_2)$) and six unknown parameters of the component densities $\mu_j, \sigma_j^2, i = 1, 2, 3$. Maximum likelihood estimates of these parameters were constructed by maximization of logarithm of likelihood function L

$$L(\psi) = \prod_{i=1}^n \sum_{j=1}^3 \pi_j \frac{1}{\sqrt{2\pi}\sigma_j y_i} \exp\left(-\frac{(\ln y_i - \mu_j)^2}{2\sigma_j^2}\right).$$

For this maximization usually EM algorithm is used ([5], [4]). As mentioned above, 8 unknown parameters were estimated with the use of a package *flexmix* [1]. According to the Akaike's criterion this model is comparable with more complicated mixture models with more components and it is capable to describe well distribution of per capita incomes.

In the Table 1 estimates of parameters and maximum likelihood estimates of characteristics of component lognormal distributions are given. Estimated components are ordered according to the expected values, the first component includes 27.5 per cent of households with low incomes, the second 61.3 per cent of households with medium income and the last one contains 11.2 per cent households with high income. The artificial subgroups are constructed and the procedure doesn't give group membership for households in the sample. The interpretation of components according to the location of incomes mentioned above is clear from the normal components of the mixture model for logarithms of incomes and is valid for lognormal mixture for estimated expected values and medians. If the mode is analysed, for lognormal distribution $LN(\mu, \sigma^2)$ this value is given by formula $e^{\mu - \sigma^2}$. This formula gives the smallest value of the mode for the third component with the highest values of both estimated parameters μ and σ^2 . This fact can be seen also in the Figure 1.

Table 1: Maximum likelihood estimates of parameters and estimated characteristics of the level and variability (in CZK)

component	$j=1$	$j=2$	$j=3$
$\hat{\pi}$	0.275	0.613	0.112
$\hat{\mu}$	11.681	11.822	11.850
$\hat{\sigma}$	0.144	0.382	0.830
expected value	119,535	146,527	197,689
median	118,302	136,216	140,084
mode	115,875	117,721	70,340
standard deviation	17,303	58,079	196,849
coefficient of variation	0.14	0.4	1

These values for components are completed in the Table 2 with the values of characteristics of the mixture. Only estimates of expected value of components can be weighted with component weights

Table 2: Sample values and estimates of characteristics of the level and variation for estimated mixture models

characteristics	$E(Y)$	$y_{0.5}$	mode	$\sqrt{D(Y)}$	quartile deviation
sample values	145,277	126,596	-	93,397	28,379
model	144,834	126,806	115,947	83,550	28,604

to the mixture expected value, remaining values were evaluated from the definition: close form was used for evaluating of standard error and numeric procedures for the evaluating of the mode and quartiles. In this case we can compare sample and estimated values. As expected, for income distribution medians (both sample and estimated) are less than the mean and estimated expected value.

In the Figure 1, component densities are shown together with the density of the mixture. In the figure the smallest value of the mode in the third component is visible as it was discussed above.

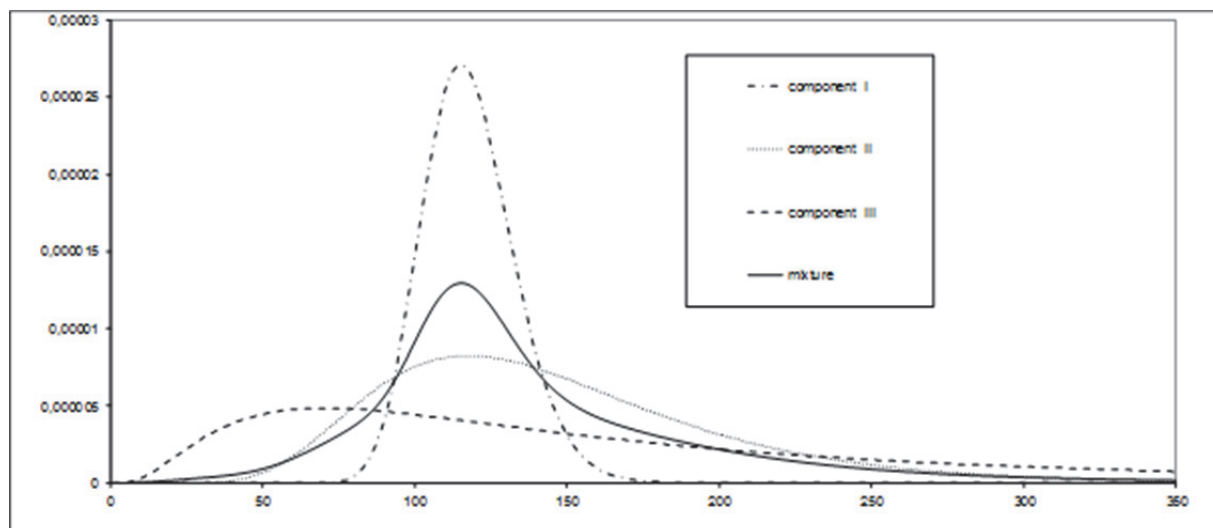


Figure 1: Estimated densities of components (dotted lines) and the mixture density (solid line), income in thousands of CZK is on the horizontal axis

In the Figure 2 maximum likelihood estimates of quantiles of conditional distributions are shown as a function of z for condition $Y \leq z$. For given z (on horizontal axis) the quartile range and the range between the ninth and the first decile are visible. The gap between these values increases with increasing z and it stabilizes for large values of the condition, when these values correspond with characteristics of unconditional model distribution.

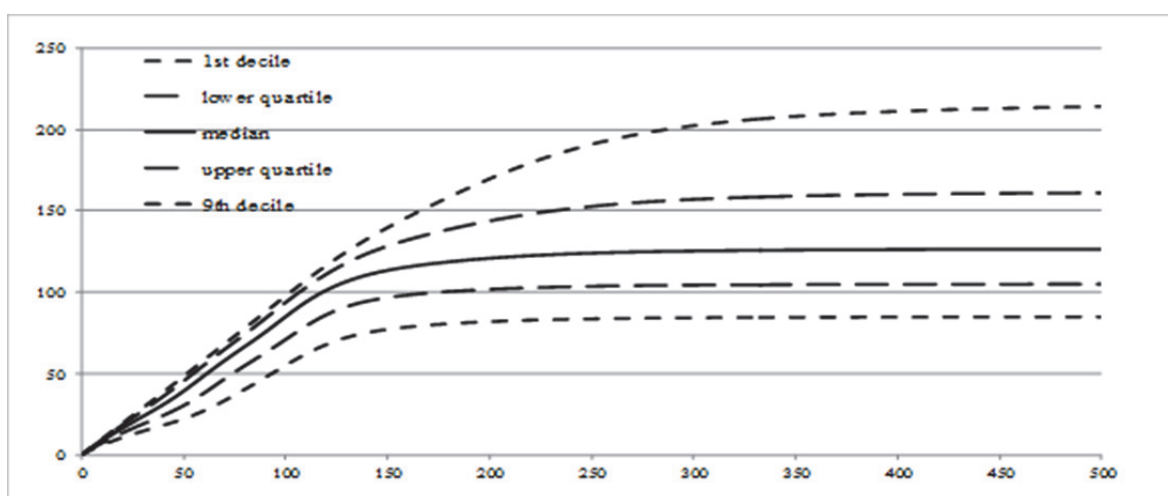


Figure 2: Estimated conditional quantiles (in 1,000CZK): the first and the ninth deciles, median, quartiles for the condition $Y \leq z$, z in 1,000CZK (horizontal axis)

In the Figure 3 estimated quantiles for the conditional distribution of Y given $Y > z$ are presented. For small values of z the conditional distribution is similar to the unconditional distribution, with increasing value of z all characteristics increase with z . In correspondance, differences between quartiles (and deciles) begin from values for unconditional distribution ($Y > 0$) and increase with increasing value of condition z .

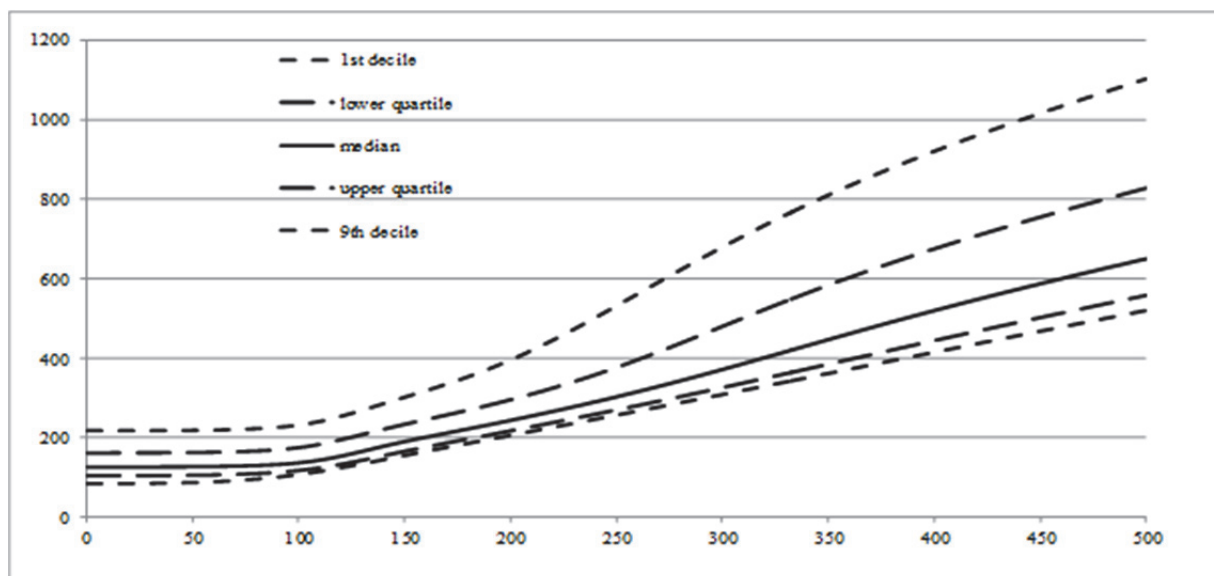


Figure 3: Estimated conditional quantiles (in 1,000CZK): the first and the ninth deciles, median, quartiles for the condition $Y > z$, z in 1,000CZK

In the Figure 4 values of conditional densities in modes are shown as function of conditions z for both types of conditions. The value of the unconditional mode is shown by the grey line. This value is important for the value of mode as will be shown later.

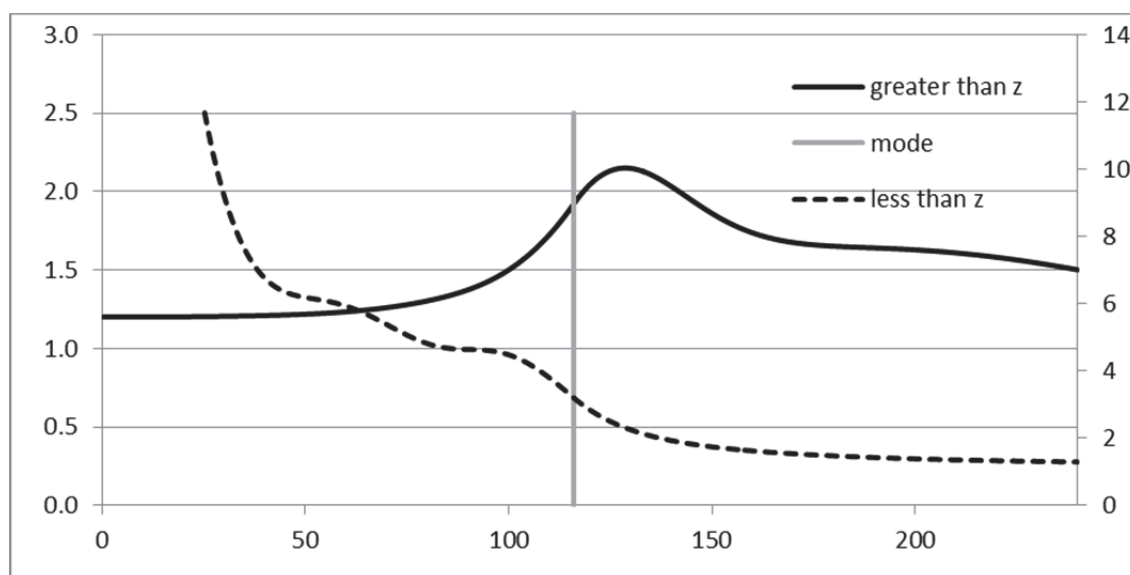


Figure 4: Maximum values of conditional densities (multiplied by 10^5) as a function of z (horizontal axis in 1,000CZK) for both conditional distributions (condition $Y \leq z$ left vertical axis, condition $Y > z$ right vertical axis)

The mixture density in the Figure 1 is unimodal. For such a distribution it is easy to find modes of conditional distributions, as these values depend on the mode of unconditional distribution (according to the Table 1 115,947 CZK for the analysed model). In the case of condition $Y > z$, the mode of conditional distribution is equal to the value of the mode of unconditional distribution for z less than mode (the unconditional mode is included in the interval $(z; \infty)$) and to z if the mode is not included in the $(z; \infty)$, it means for z greater than unconditional mode. For z less than unconditional mode the conditional densities are unimodal, for z greater than unconditional mode these densities

are decreasing functions of $y > z$ (and is equal to 0 for $y \leq z$). These densities are shown in Figures 5 ($z \leq 100,000$ CZK) and 6 ($z > 100,000$ CZK).

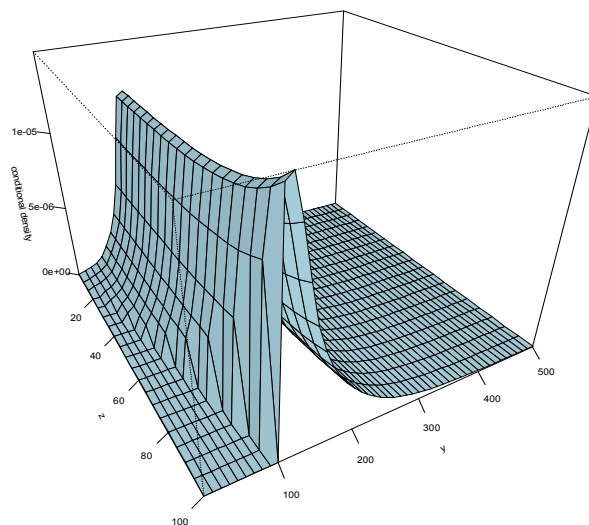


Figure 5: Estimated conditional densities for the condition $Y > z$, $z \leq 100,000$ CZK (values of y and z are given in 1,000CZK)

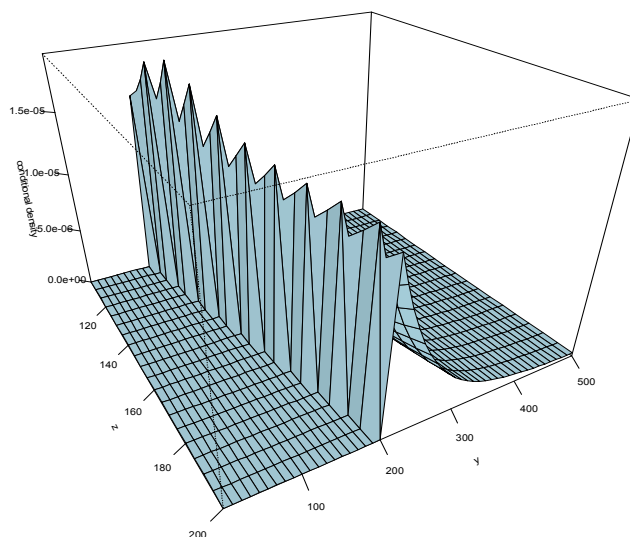


Figure 6: Estimated conditional densities for the condition $Y > z$, $z > 100,000$ CZK (values of y and z are given in 1,000CZK)

Suppose now the condition $Y \leq z$. If z is less than unconditional mode, the mode of conditional distribution is equal to z (as the value of the unconditional mode is not included in the interval $(0; z)$). For z greater than unconditional mode, the value of the mode of conditional distribution is equal to the unconditional mode. Conditional densities are increasing functions of y for $z \leq \text{mode}$ (Figure 7) and unimodal for $z > \text{mode}$ (Figure 8), and these densities are according to (7) equal to 0 for $y \geq z$.

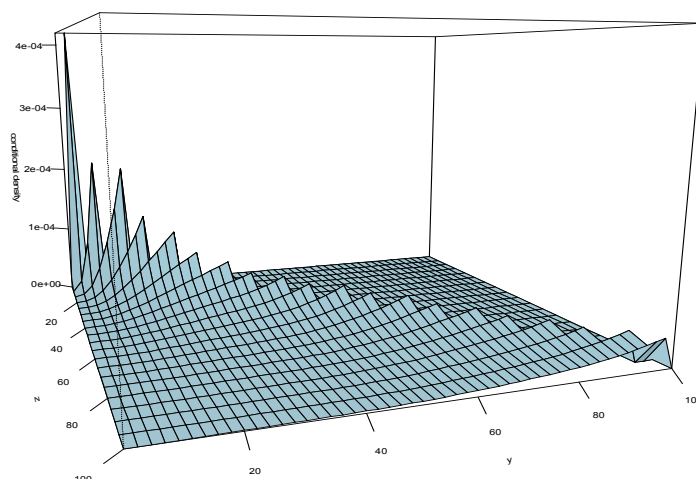


Figure 7: Estimated conditional densities for the condition $Y \leq z$, $z \leq 100,000$ CZK (values of y and z are given in 1,000CZK)

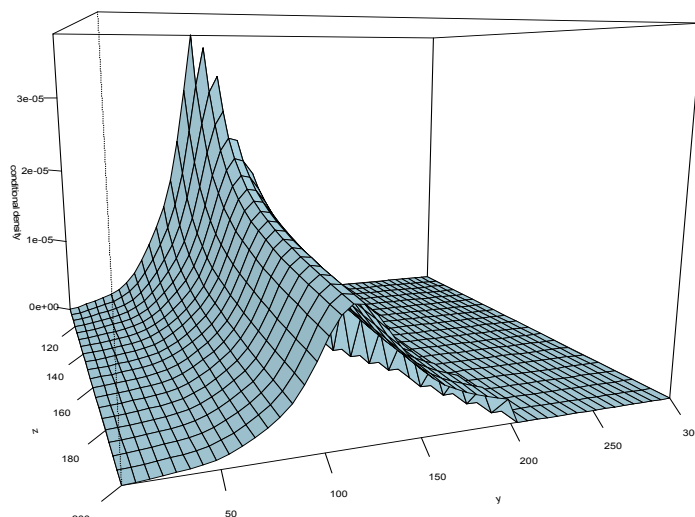


Figure 8: Estimated conditional densities for the condition $Y \leq z$, $z > 100,000$ CZK (values of y and z are given in 1000CZK)

Conclusions

In the text conditional distributions of a positive random variable with finite mixture probability distribution were constructed with the use of straightforward probability computations. For evaluation of conditional density or probability distribution close formulas are given based on component densities and distribution functions. For evaluation of quantile (or moment) characteristics numeric methods (solving nonlinear equations a numeric integration) are necessary to be used. If unknown parameters are estimated from a sample with the use of maximum likelihood method, numeric optimization is used for complete data problem and EM two-step numeric algorithm for incomplete data problem (as it is in this text).

The conditional modes depend strongly on the mode of original (unconditional) distribution. The analysed mixture distribution is unimodal and in this case, for both conditions $Y \leq z$ and $Y > z$, the conditional mode can be evaluated only with the use of the value of unconditional mode. It means that conditional mode is linear function of z for both conditions. For the condition $Y \leq z$ this relation

is linear (with the slope 1) for z less than the unconditional mode and constant for z greater than unconditional mode. For the condition $Y > z$ it is constant for z less and linear for z greater than the unconditional mode.

Additionally, the shape of the conditional density (unimodal, increasing, decreasing function) depends also on relation between the unconditional mode and the condition z .

References

- [1] GRÜN, B., LEISCH, F. FLEXMIX version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(4):1-35, 2008.
- [2] KLEIBER, C., KOTZ, S. *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley-Interscience, New York. ISBN 0-471-15069-9. 2003.
- [3] MALA. I.: Conditional Distributions of Incomes and their Characteristics, Dublin 21.08.2011 – 26.08.2011. In: *ISI 2011*. Dublin : ISI, 2011, s. 1–6
- [4] PAVELKA, R. Application of density mixture in the probability model construction of wage distributions, *Applications of Mathematics and Statistics in Economy: AMSE 2009*, Uherské hradiště, 2009, 341-350, 2009.
- [5] TITTERINGTON, D.M., SMITH, A.F., MAKOV, U.E. *Statistical analysis of finite mixture distributions*, Wiley, 1985.
- [6] www.czso.cz

Current address

Ivana Malá

University of Economics,
Faculty of Informatics and Statistics,
Department of Probability and Statistics, Prague,
W.Churchilla sq.4, Prague, 130 67,
Czech Republic,
+420224095486,
e-mail: malai@vse.cz

MULTISTATE LIFE TABLES: APPLICATION OF THE METHOD ON THE MARRIAGE CAREER

MISKOLCZI Martina, (CZ), LANGHAMROVÁ Jitka, (CZ)

Abstract. The article introduced application of multistate demographic methods onto the ‘marriage career’ based on real data. The analysis presented alternative approach to the analysis of marriages, divorces and behaviour of women toward this issue, and additional utilization of life tables’ methodology.

Objective of the article is to verify changes in the behaviour of women in the Czech Republic related to their marriage decision over last ten years: women more often decide not to marry and stay unmarried.

Calculation of multistate life table and modelling ‘marriage career’ of women in the Czech Republic 2001–2010 showed that women change their decision toward marriages. Younger women in the age 15–30 years in the Czech Republic changed substantially their behaviour related to their marriage over last 10 / 20 years. Tendency not to marry is stronger among young women, they stay unmarried and probably live in partnerships without official marriage. Women over 30 years changed their behaviour only little.

Key words. multistate life tables, mathematical demography, marital status, marriage, single, married, divorced, widowed

Mathematics Subject Classification: Primary 91D20; Secondary 91C99, 91D15

Introduction

Multistate demography is a part of demography that analyses states of demographic subjects and events that causes these states. For simplicity and mathematical modelling, usually only one type of demographic event is studied and the sequence of events is called ‘career’. For example, fertility of a woman can be analyzed – in such a case birth of woman, birth of first child, birth of second child, birth of third child etc., abortions and death of a woman would be in center of the interest. Another example is analysis of ‘educational career’, where each beginning of the study, successful completion (graduation) or unsuccessful end is observed according to the education stage.

Here, ‘marriage career’ or ‘marital status career’ of women is studied in the period of 2001–2010 in order to analyze changes in the trend in nuptiality (marriages) and divorcity in the Czech Republic after 2000. Method of multistate life tables enables to study “... occurrences of events and

transfers and on their association with the populations that are exposed to the risk of experiencing them.” (Rogers, 1980, p.497)

1.1 Definition of Terms

In the multistate demography, that had originated from multiregional demography (Rogers, 1975), following terms are used:

- *Events* U_i : birth (U_0), marriage (U_1), divorce (U_2), death of the partner (become a widow) (U_3), death (U_ω). ω is used in demography to denote first age that no one reaches in the population.
- *States* (Z_i): single (Z_0), married (Z_1), divorced (Z_2), widowed (Z_3) and dead (Z_ω). States could be absorbent or transient. Absorbent states cannot be left, subject, once he came here, cannot leave and remains in the state. Usually, it is represented by the state ‘dead’.
- *Randomness*: Occurrence of events is considered to be random. It is assumed that an individual with certain realization of his life cycle can be found in the population with some probability.
- *Probability distribution*: random event is characterized by probability distribution.
- *Multistate life tables* are the extension of standard (one state) life tables. They present additional dimension(s), in this case represented by the original marital status and studied marital status. Instead of one number in a column for each age there is a square matrix.

1.2 Scheme of Marital Status Life Cycle

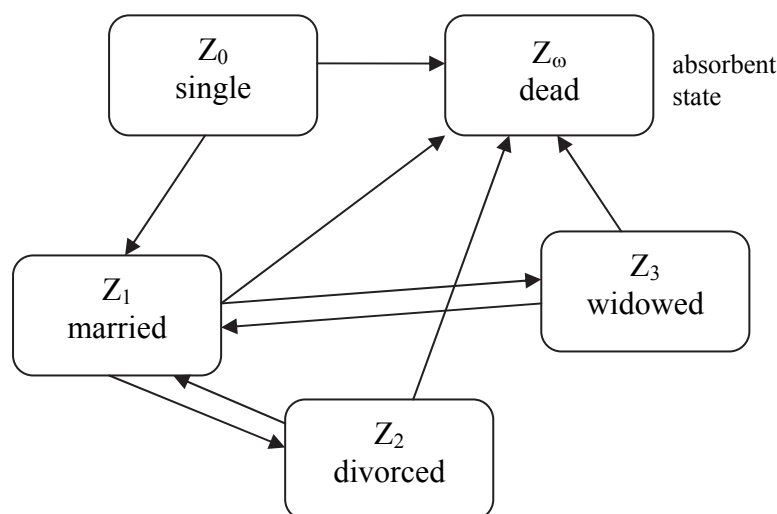


Figure 1: Scheme of marital status life cycle

1.3 Objective of the Research

Objective of this article is to verify changes in the behaviour of women in the Czech Republic related to their marriage decision over last ten years. In surveys it is usually concluded that people live together in cohabitation, in partnership without official marriage and increasing proportion of children is born outside of marriage. This hypothesis will be accepted or rejected based on real data from the Czech Republic 2001–2010.

2 Multistate Life Tables Calculation

2.1 Data

For the calculation, it is necessary to estimate unknown probability distribution $P(U, x, t | Z)$ or $P(U, x | Z)$ for each event U and state Z , where x_i is time between U_0 and U_i , t is time from the last transition into state. (Koschin, 1992)

Probability distribution can be characterized by probability density function, intensity probability (also called hazard rate or risk function) or distribution function. Briefly, using absolute frequencies (number of events, number of subjects exposed to a risk of event and length of the exposure) the intensity probability can be estimated. (Koschin, 1992) It was proved that such an estimate is the best unbiased estimate of the intensity probability, which is constant in given interval. (Rogers, 1975).

In demography, data are available in annual frequency, less often by month. One year is the most detailed distribution of data if official statistics should be used for entire population.

For intensity of mortality in the age x it can be used following estimate:

$$\frac{\text{number of deaths in the age } x \text{ during calendar year}}{\text{mid-period size for the group in the age } x \text{ in given calendar year} * 1},$$

where multiplication by 1 in the denominator means length of exposure. Each individual who is in the population during the whole year adds one year to the final sum. Each individual who came into the population during the year or left the population during the year represents weight of one half. This corresponds with the assumption of uniform distribution of demographic event during the year.

From demographic point of view, this estimate is usual indicator: specific mortality rate $m_x = \frac{M_x}{S_x}$,

$x = 0, 1, \dots, \omega-1$.

Analogously, intensity of fertility can be estimated by $\varphi_x = \frac{N_x}{\bar{S}_x - \frac{1}{2}N_x}$, where correction in the

denominator eliminates from the exposure for one half of the year those women who gave birth in the same calendar year. (Koschin, 1992)

Analogously, specific nuptiality and divorce rates will be used to estimate intensity of nuptiality and intensity of divorce.

2.2 Calculation

For each age $x = 15, 16, \dots, 59$ for women in the Czech Republic intensity of transition-probability matrix is prepared for states single, married, divorced and widowed:


$$H_x = \begin{pmatrix} \sigma_{Sx} + \mu_{Fx} & 0 & 0 & 0 \\ -\sigma_{Sx} & \rho_x + \mu_{Fx} + \mu_{Mx} & -\sigma_{Dx} & -\sigma_{Wx} \\ 0 & -\rho_x & \sigma_{Dx} + \mu_{Fx} & 0 \\ 0 & -\mu_{Mx} & 0 & \sigma_{Wx} + \mu_{Fx} \end{pmatrix},$$

Where σ denotes nuptiality (marriage rate) with the index according to marital status (S-single, D-divorced, W-widowed), ρ is divorce rate, μ denotes mortality of females (F) or males-husbands (M).

Then, transition-probability matrices p_x are calculated and, subsequently, other life tables indicators in the form of matrices: table number of survivors (matrices l_x), table number of person-years (matrices L_x), number of remaining years of life to be lived by the table generation (for entire group of individuals) in the age of x (matrices T_x) and matrices of expected length of stay (e_x ; this term is used rather than life expectancy). (Koschin, 1992; Land & Rogers, 1982; Rogers, 1975; Raymer & Willemkens, 2008)

2.3 Results

Example of results for 2010, women, the Czech Republic is published here:

	25,78	0,00	0,00	0,00
	14,24	32,66	18,54	15,27
	4,04	11,05	25,42	4,91
	0,34	0,68	0,43	24,21
...				
e_{25}	18,30	0,00	0,00	0,00
	12,53	25,78	14,05	7,30
	3,35	8,15	20,09	1,80
	0,31	0,56	0,35	25,39
...				
e_{59}	1,00	0,00	0,00	0,00
	0,00	0,99	0,00	0,00
	0,00	0,00	0,99	0,00
	0,00	0,01	0,00	1,00

This can be interpreted as following:

- Single woman in the age of 25 years may expect that she spends till the age of 59 years another 18.3 years as single, 12.5 years as married, 3.4 years as divorced and 0.3 years as widowed.
- Married woman in the age of 25 years may expect that she spends till the age of 59 years another 25.8 years as married, 8.2 years as divorced and 0.6 years as widowed.
- Divorced woman in the age of 25 years may expect that she spends till the age of 59 years another 14.1 years as married, 20.1 years as divorced and 0.4 years as widowed.
- Widowed woman in the age of 25 years may expect that she spends till the age of 59 years another 7.3 years as married, 1.8 years as divorced and 25.4 years as widowed.
- Once a woman is married and later divorced or widowed, she cannot return to the state ‘single’.
- State ‘dead’ is not part of matrices.

It table number of survivors (matrices l_x) it can be watched how many women survive and how do they change their status (see Figure 2). Following Figure 3 shows length of their stay in each state (calculated till the age of 59). Such a calculation was prepared for years 2002–2010 for ages $x = 15, 16, \dots, 59$ years. Further, abridged calculation of the year 2001 and 1990 is available for women in the Czech Republic (for 5-years long age intervals).

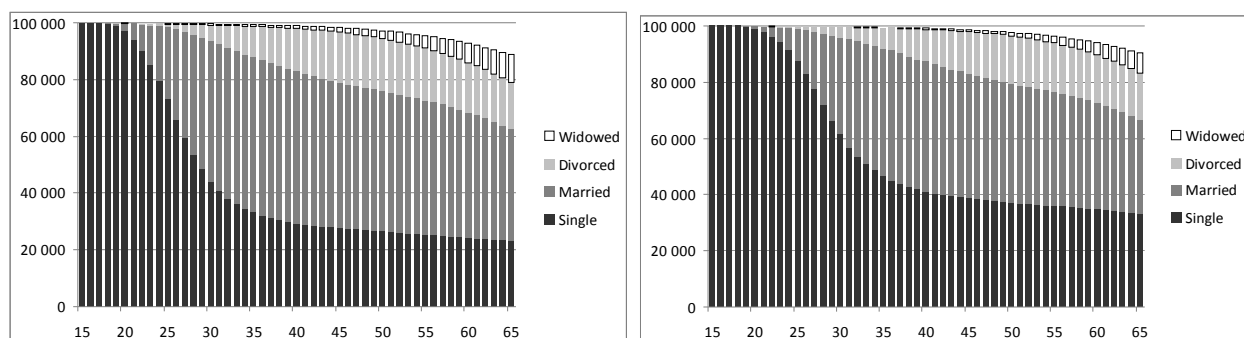


Figure 2: Table number of survivors, women, 2002 and 2010, Czech Republic

Majority of woman in young ages are single, their transition into the state ‘married’ is visible between the age 25 and 30 years. It can be seen that years 2002 and 2010 differs in the number and proportion of women that remain in the state ‘single’ after 35 years of age. In 2002 the proportion of those women formed approximately 30 % and slowly decreased, whereas in 2010 the proportion is 40 % of all women from the studied group (it is theoretical group of 100,000 women set up as the root of multistate life table for the age of 15 years).

Another trend can be seen here: proportion of woman single + married increased between years 2002 and 2010 by 5 percentage points. In the age of 40 years there were 82.9 % of single or married women, whereas in 2010 this proportion increased to 87.3 %. It shows that the state ‘single’ is in some cases preferred also by women who were divorced or widowed in 2002.

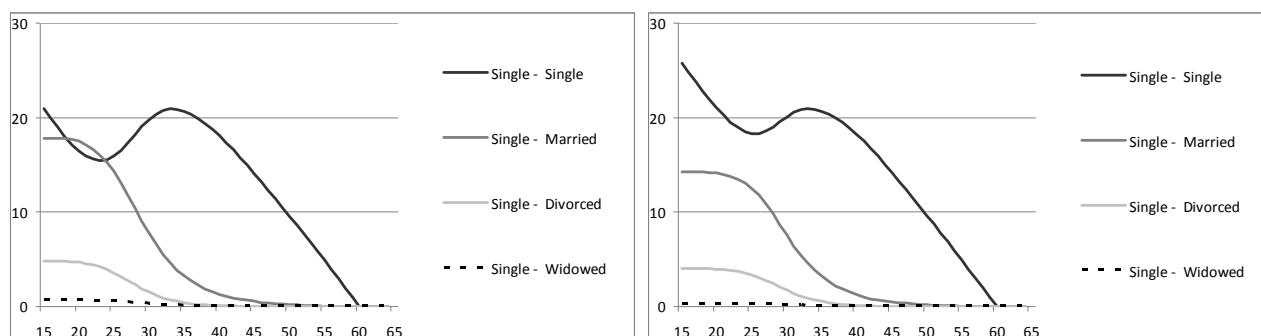


Figure 3: Expected length of stay in states for originally single women, women, 2002 and 2010, Czech Republic

It is interesting that young single women tend to marry and escape from the state ‘single’, but for women in ages 25 to 33 years number of expected years when they remain in the state ‘single’ even grows. For example in 2010, single woman in the age of 33 years has very low probability 0.044 to leave state ‘single’ and high probability 0.956 to remain in the state ‘single’. For older single women the probability of transition into other states (first state ‘married’, then possibly ‘divorced’ or ‘widowed’) is very low, expected length of stay in the state ‘single’ decreases proportionally with the age.

Expected number of years in the state ‘married’ is quite high for single women till 25–28 years of age, for higher ages it declines very rapidly. Comparison of years 2002 and 2010 shows that young single women in 2010 have to expect more years to live as a single individual, second choice ‘married’ is lower by almost 10 years. In 2002 both choices were comparable for young single woman in the age of 15–30 years and the difference became evident later.

2.4 Trends and Changes in Women's Behaviour

Comparing the same indicator from the matrices of expected length of stay over longer period of time, the trend can be commented and hypotheses assessed. Two cases were chosen as most interested: originally single women and their chance to remain single or transition to the state 'married'.

Note that expected length of stay is calculated till the age of 59 years.

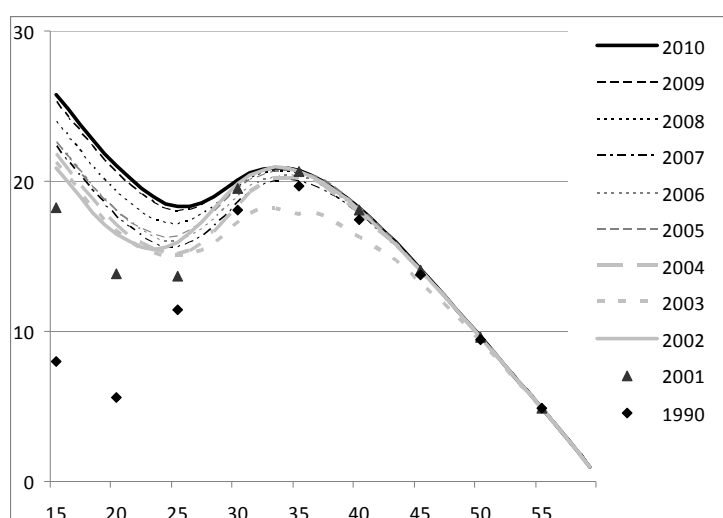


Figure 4: Expected length of stay single-single, women, 1990, 2001–2010, Czech Republic

Trend in the period of 2001–2010 shows that in the age of 15 to 28 years single women tend to postpone marriage and stay single. The expected length of stay in the state 'single' prolongs and difference between 2001 and 2010 is almost five years for the age of 25 years. The common characteristic is that all lines have the same shape, i.e. decrease of expected length of stay between 15 and 25 years and then approximately between 25 and 35 years increasing chance (measured both in probability and number of expected years) that woman remains single. The largest difference is visible in first 15 years of studied part of women's lives with one exception – year 2003 differs from others in the ages of 30–45 years. This could be explained by legislative impact (Joint taxation of married couples motivated many spouses to marry officially. This was massive economic benefit for families with children.)

Year 1990 (Koschin, 1992) shows the remarkable change that happened over last 20 years in the Czech Republic. Young single women till their 25 years could expect to remain single for less than 10 years whereas in 2010 women till 25 years might expect to remain single another 18 to 25 years of their lives. This represents more than double number of years. On the other hand, comparable results belong to ages 32 years and more. If a woman remains single till her 32 years then there is almost no difference over last 20 years in the indicator how long she might expect to remain such.

In this sense changes in behaviour of women in the Czech Republic related to their marriage decision over last 10 / 20 years are confirmed. Young single women more often stay legally unmarried and the expected number of years of living further in the state 'single' increases year-over-year. The change in behaviour happens mainly before the age 30 years. (This does not say anything about their real status and way of their partnership.)

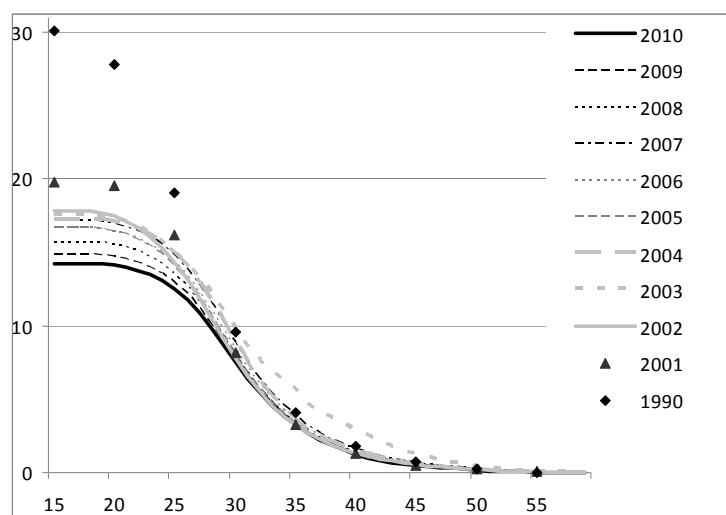


Figure 5: Expected length of stay for transition single-married, women, 1990, 2001–2010, Czech Republic

Transition of single woman to the state ‘married’ has evident trend. The expected length of stay in the state ‘married’ for originally single women declines since 1990 till 2010 each year. It correlates with the decreasing transition probability ‘single–married’. Single women may expect shorter and shorter length of period when is married, currently it is – for example – 12.5 years for a woman in the age of 25 years till her 59 years in comparison of 19.1 years in 1990.

For second half of studied part of women’s lives, after 30 years, the shape and level of the indicator is very similar. Thus, change has occurred and continues in the first half of women’s lives.

This again verifies fact that behaviour of women in the Czech Republic changes in respect to their decision regarding marriage. Women tend not to marry.

Conclusion

The article introduced application of multistate demographic methods onto the ‘marriage career’ based on real data. The analysis presented alternative approach to the analysis of marriages and divorces and additional utilization of life tables’ methodology.

Calculation of multistate life table and modelling ‘marriage career’ of women in the Czech Republic 2001–2010 showed that women change their decision toward marriages. The article concentrated on single women and their decision stay single or become married.

Objective of this article was verified for younger women: women 15–30 years old in the Czech Republic changed their behaviour related to their marriage over last 10 / 20 years. Tendency not to marry is stronger among young women, they stay unmarried and probably live in partnerships without official marriage. Women over 30 years changed their behaviour only little.

Acknowledgement

The paper was written with the support of Internal Grant Agency of the University of Economics in Prague F4/29/2011 “Analysis of Population Ageing and Impact on Labor Market and Economic Activity“.

References

- [1] KOSCHIN, F.: *Vicestavová demografie*. Prague, University of Economics in Prague, 1992. ISBN 80-7079-087-3.
- [2] LAND, K. C., ROGERS, A. (ed.): *Multidimensional Mathematical Demography*. New York, Academic Press, 1982.
- [3] RAYMER, J., WILLEKENS, F. (ed.): *International Migration in Europe. Data, Models and Estimates*. New York, John Wiley and Sons, 2008.
- [4] ROGERS, A. (ed.): *Essays in Multistate Mathematical Demography* (reprint from Environment and Planning A 12, 1980, s. 485–622). Laxenburg, IIASA, 1980.
- [5] ROGERS, A.: *Introduction to Multiregional Mathematical Demography*. New York, John Wiley and Sons, 1975.

Current address

Martina Miskolczi, Ing. Mgr. MBA

Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
Katedra demografie
nám. W. Churchilla 4
130 67 Praha 3, Česká republika
martina.miskolczi@svse.cz
martina.miskolczi@sseznam.cz

Jitka Langhamrová, Doc. Ing. CSc.

Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
Katedra demografie
nám. W. Churchilla 4
130 67 Praha 3, Česká republika
langhamj@vse.cz

TWO APPLICATIONS OF PROBABILITY IN THE THEORY OF RELIABILITY AND MAINTENANCE

MOŠNA František, (CZ)

Abstract. Contribution considers using of probability in the theory of reliability and maintenance. The first application deals with the characteristic OEE (Overall Equipment Effectiveness) of some production facility composed of two or more machines. In the second part, a formula for expected value of reliability of machine composed of regularly replacing components is derived.

Key words and phrases. probability, reliability, maintenance.

Mathematics Subject Classification. 60K10.

1 Introduction

Theory of probability is often used in considerations about the reliability and maintenance of machines. In this contribution, an illustration of two such very simple applications is presented. We met with these problems during my collaboration with the Department for Quality and Dependability of Machines at the Czech University of Life Sciences (see [1] or [2]).

2 Characteristic OEE of production facilities

The first illustration considers so called characteristic OEE Overall Equipment Effectiveness.

This characteristic is very often used in theory of reliability of machines or production facilities and it describes (roughly spoken) quantitative ratio between the actual and the ideal production of machines or production units.

Here are some basic concepts:

- a availability, i.e. use of working hours of the given machine in percentage,

$1 - a$ percentage of downtime (maintenance, repairs and so on),

- w actual current performance of machine, this performance is reduced compared to the declared rated performance due to wear-out etc.

- q percentage of quality products manufactured on the given machine,

$1 - q$ percentage of rejects,

- p number of products produced on the given machine during the time unit (for instance during the work shift) at full performance.

The characteristic OEE is defined as a ratio of actual products manufactured on the given machine and the number of products produced in ideal case, it means

$$\text{OEE} = \frac{awqp}{p} = awq .$$

For instance, a machine produces at full performance 400 products during working hours (8 hours), i.e. 50 pieces per hour.

Suppose it uses only 90% of working hours (due to maintenance or repairs etc.), $a = 0,9$, it works only 7.2 hours and it produces only 360 products at full performance during this period.

The machine operates only at 75% of its full performance (due to wear), $w = 0.75$, so it produces only 270 pieces.

Of this 20% are rejects and only 80% correspond to the required quality, $q = 0,8$, so only 216 products.

Compared to the ideal case (when the machine produces 400 quality pieces), in real case it produces only 216 products per shift. Thus Characteristic OEE can be calculated as a quotient

$$\text{OEE} = \frac{216}{400} = 0.54$$

or product

$$\text{OEE} = 0.9 \cdot 0.75 \cdot 0.8 = 0.54 .$$

In the theory of reliability and maintenance of machines we are often interested in getting characteristic OEE of all production facility composed of two or more machines.

2.1 Connected in parallel side by side

Let us consider at first production facility composed of two machines (with characteristics a_1, w_1, q_1, p_1 and OEE_1 or a_2, w_2, q_2, p_2 and OEE_2 , respectively) connected in parallel side by side.

The number of productions in ideal case is

$$p_1 + p_2 .$$

But the number of really produced pieces is equal to

$$a_1 \cdot w_1 \cdot q_1 \cdot p_1 + a_2 \cdot w_2 \cdot q_2 \cdot p_2 = \text{OEE}_1 \cdot p_1 + \text{OEE}_2 \cdot p_2 .$$

So we can very simply work out OEE for all product unit

$$\text{OEE} = \frac{\text{OEE}_1 \cdot p_1 + \text{OEE}_2 \cdot p_2}{p_1 + p_2} .$$

This formula can be simply generalized for more machines connected in parallel.

2.2 Connected in series in a row

Now let us focus our attention on the case when production process goes on the line consisting of two machines connected in series in a row. Suppose any product passes through both the machines and effective production takes place only if both machines are working simultaneously.

To calculate the overall characteristics OEE we need to know not only the individual data on availability, performance and product quality in both machines but also the information about synchronization of these two machines. For instance, we need to have the percentage of both machine downtime, let us denote it by a_{12} .

The number of products of the line in ideal case is $\min(p_1, p_2)$.

In order to derive percentage of using working hours we shall use formula for probability of union of two random events A and B and complementary event to A

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad P(\text{not}A) = 1 - P(A).$$

Downtime of the first machine is $1 - a_1$, downtime of the second machine $1 - a_2$, downtime of both machines at the same time is a_{12} . Hence percentage of working hours when at least one machine is not working (and the line does not work) is equal to the number

$$(1 - a_1) + (1 - a_2) - a_{12}.$$

Part of working hours when both machines are working (and line works) presents a complement of the previous, hence its probability is

$$a = 1 - [(1 - a_1) + (1 - a_2) - a_{12}] = a_1 + a_2 + a_{12} - 1,$$

more accurately the positive part of this number.

During this time $a \cdot w_1 \cdot p_1$ products can pass through the first machine and $a \cdot w_2 \cdot p_2$ products through the second machine. As production process can be successful only when both machines are working together, number of produced pieces is equal to the lesser number of both,

$$\min(a \cdot w_1 \cdot p_1, a \cdot w_2 \cdot p_2).$$

The whole quality of the line production can be got as a product $q_1 \cdot q_2$.

Number of products actually produced by this line is

$$(a_1 + a_2 + a_{12} - 1)^+ \cdot \min(w_1 \cdot p_1, w_2 \cdot p_2) \cdot q_1 \cdot q_2$$

and we can conclude that the characteristic OEE of line is

$$\text{OEE} = \frac{(a_1 + a_2 + a_{12} - 1)^+ \cdot \min(w_1 \cdot p_1, w_2 \cdot p_2) \cdot q_1 \cdot q_2}{\min(p_1, p_2)}.$$

The formula can be simply generalized for three (or more) machines by using the formula of probability of unit of three (or more) random events.

These calculations can serve in designing or planning of production lines and to synchronize its elements, too.

3 Mean life of age-related preventive maintained components

Let us monitor some production facility. Any failure of it can cause large damages, so the reliability of it is very important. This facility contains a certain so-called key component (for instance a bulb). In the case of its failure all the facility stops working (and causes damages mentioned above). So we desire to prolong its reliability. This can be achieved by regular replacing of the key component.

Now let us try to deduce formula for mean reliability provided by regular replacing such component after the period t_p using a new component with the same reliability properties.

Let us suppose that durability of any of k -th component is characterized by a random variable X with a continuous density function f and distribution function F . We suppose that random variables X_1, X_2, \dots are independent.

Let us denote by T a random variable which describes the life of age-related, preventively replaced components. Further, we derive the formula of the density function f_T and the distribution function F_T for the random variable T and we are particularly interested in the mean value ET .

Let us denote $p = P[X_k < t_p]$, $q = P[X \geq t_p] = 1 - p$ and $I = \int_0^{t_p} xf(x) dx$. We express the random variable T using X_k in the following way:

$$T = \begin{cases} X_1 & \text{for } X_1 < t_p \\ t_p + X_2 & \text{for } X_1 \geq t_p, X_2 < t_p \\ 2t_p + X_3 & \text{for } X_1 \geq t_p, X_2 \geq t_p, X_3 < t_p \\ \dots\dots\dots & \\ kt_p + X_{k+1} & \text{for } X_1 \geq t_p, X_2 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p. \end{cases}$$

With respect to independence X_1, X_2, \dots , from the total probability theorem

$$P(A) = \sum P(A/B_k)P(B_k) \text{ for } B_1, B_2, \dots \text{ mutually disjoint, } \cup B_k = \Omega, P(B_k) \neq 0$$

it holds for arbitrary $x \in \langle 0; \infty \rangle$:

$$\begin{aligned} F_T(x) &= P[T < x] = \\ &= P[T < x/X_1 < t_p] \cdot P[X_1 < t_p] + \\ &\quad + P[T < x/X_1 \geq t_p, X_2 < t_p] \cdot P[X_1 \geq t_p, X_2 < t_p] + \dots + \\ &\quad + P[T < x/X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p] \cdot \\ &\quad \cdot P[X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p] + \dots = \\ &= P[T < x/X_1 < t_p] \cdot p + P[T < x/X_1 \geq t_p, X_2 < t_p] \cdot pq + \dots + \\ &\quad + P[T < x/X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p] \cdot pq^k + \dots \end{aligned} \tag{1}$$

Further, we calculate according to the definition of conditional probability

$$P(A/B) = \frac{P(A \cup B)}{P(B)} \quad \text{for} \quad P(B) \neq 0$$

with respect to the independence of X_1, X_2, \dots

$$\begin{aligned}
 P[T < x/X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p] &= \\
 P[k t_p + X_{k+1} < x/X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p] &= \\
 = \frac{P[X_{k+1} < x - k t_p, X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p]}{P[X_1 \geq t_p, \dots, X_k \geq t_p, X_{k+1} < t_p]} &= \\
 = \frac{P[X_{k+1} < \min(x - k t_p, t_p)] \cdot q^k}{q^k \cdot p}.
 \end{aligned} \tag{2}$$

After substitution of (2) into equation (1) we obtain

$$\begin{aligned}
 F_T(x) &= P[T < x] = \sum_{k=0}^{\infty} P[X_{k+1} < \min(x - k t_p, t_p)] \cdot q^k \\
 &= \sum_{k=0}^{\infty} F(\min(x - k t_p, t_p)) \cdot q^k.
 \end{aligned}$$

It is possible to itemize the distribution function F_T around following intervals:

$$F_T(x) = \begin{cases} F(x) & \text{on } (0; t_p) \\ F(t_p) + qF(x - t_p) & \text{on } (t_p; 2t_p) \\ F(t_p) + qF(t_p) + q^2F(x - 2t_p) & \text{on } (2t_p; 3t_p) \\ \dots\dots\dots & \dots\dots\dots \end{cases}$$

We obtain density function f_T by differentiation of F_T

$$f_T(x) = \begin{cases} f(x) & \text{on } (0; t_p) \\ qf(x - t_p) & \text{on } (t_p; 2t_p) \\ q^2f(x - 2t_p) & \text{on } (2t_p; 3t_p) \\ \dots\dots\dots & \dots\dots\dots \end{cases}$$

Finally, we calculate the mean value of life (sum of particular operating time) for components that underwent age-related preventive replacement

$$\begin{aligned}
 ET &= \int_0^{\infty} x f_T(x) dx = \sum_{k=0}^{\infty} \int_{k t_p}^{(k+1)t_p} q^k x f(x - k t_p) dx = \\
 &= \sum_{k=0}^{\infty} q^k \int_0^{t_p} (x + k t_p) f(x) dx = \\
 &= \left(\int_0^{t_p} x f(x) dx \right) \sum_{k=0}^{\infty} q^k + q t_p \left(\int_0^{t_p} f(x) dx \right) \sum_{k=0}^{\infty} k q^{k-1} \\
 &= I \frac{1}{1-q} + p q t_p \frac{1}{(1-q)^2} = \frac{I + q t_p}{p},
 \end{aligned}$$

we used the formula for the sum of the geometrical series $1 + q + q^2 + \dots + q^k + \dots = \frac{1}{1-q}$ and from this formula through differentiation we obtained the derived formula $1 + 2q + 3q^2 + \dots + k q^{k-1} + \dots = \frac{1}{(1-q)^2}$.

Integral I can be modified using integration by parts

$$\begin{aligned} I &= \int_0^{t_p} x f(x) dx = t_p F(t_p) - \int_0^{t_p} F(x) dx \\ &= t_p p - t_p + \int_0^{t_p} R(x) dx = -q t_p + \int_0^{t_p} R(x) dx, \end{aligned}$$

where $R(x) = 1 - F(x)$ is so called survival function. For expected value ET of the life of the age-related preventively replaced components at time t_p we obtain the following formula

$$ET = \frac{1}{p} \int_0^{t_p} R(x) dx = \frac{\int_0^{t_p} R(x) dx}{1 - R(t_p)}.$$

For instance for random variable X_k governed by the Weibull distribution with parameters m and x_0 (distribution function is $F(x) = 1 - e^{-\frac{x^m}{x_0}}$) we can write

$$ET = \frac{\int_0^{t_p} e^{-\frac{x^m}{x_0}} dx}{1 - e^{-\frac{t_p^m}{x_0}}}.$$

4 Conclusion

It is possible to complete both examples by simple calculation spreadsheets in MS Excel (using for instance [3]). The first example can be generalized also for various combinations of parallel and serial connection of machines. Examples illustrate the importance and significance of probability calculus for technical computing.

References

- [1] LEGÁT, V., MOŠNA, F., ČERVENKA, V., JURČA, V.: *Preventive maintenance improve operating reliability*. In Acta Technologica Agriculturae, Vol. 8, No. 1, pp. 21-24, 2005.
- [2] LEGÁT, V., MOŠNA, F., ČERVENKA, V.: *Optimalizace preventivní údržby strojních prvků*. In Proceedings of conference Spolehlivost 2001 Brno, pp. 101-108, 2001.
- [3] JANČAŘÍK, A.: *Matematické funkce. [Mathematical functions]*. In Excel při ruce, 2007, No. 3, p. 6.

Current address

František Mošna, RNDr. Ph.D.

Czech Univ. of Life Sciences in Prague,
Kamýcká 129, 165 21 Praha 6,
tel. +420 22438 3276,
e-mail: mosna@tf.czu.cz

INFLATION MODELING AND COINTEGRATION

NEUBAUER Jiří, (CZ)

Abstract. The article deals with the question of modeling multidimensional non-stationary cointegrated processes. It is a modern method especially used for description of economic time series. Multidimensional non-stationary process is called cointegrated if there is a linear combination of its one-dimensional components, which is stationary or trend-stationary. For instance this property can be found in some series of economic indices which are predominantly non-stationary. Nevertheless, there are linear links which keep that whole system in so-called long-term equilibrium. The article is focused on a cointegration analysis of selected time series of the Czech Republic macro-economic indices.

Key words and phrases. cointegration, tests of cointegration, inflation

Mathematics Subject Classification. Primary 60A05, 08A72; Secondary 28E10.

1 Introduction

Empirical research in macroeconomics as well as in financial economics is largely based on time series. Non-stationarity, a property common to many macroeconomic and financial time series, means that a variable has no clear tendency to return to a constant value or a linear trend. Multidimensional non-stationary process is called cointegrated if there is a linear combination of its one-dimensional components, which is stationary or trend-stationary. For instance, this property can be found in some series of economic indices. The article deals with a question of modeling multidimensional non-stationary cointegrated processes. A detailed description of this method can be found for example in Hamilton (1994), Juselius (2006), Johansen (1995) or Lütkepohl (2007). One can find a range of useful applications of mentioned method, for example in the article Neubauer (2006) is presented the application of cointegration analysis on time series of selected exchange rates.

The model of inflation described in Kim (2001) is applied to the macroeconomic time series of the Czech Republic (quarterly data from 2002 to 2012). This model is tested using methods of a cointegration analysis.

1.1 Basic definitions

Definition 1.1 Let $\{\epsilon_t\}$ be independent identically distributed random variable with zero mean and variance matrix Ω . A stochastic process Y_t which satisfies that $Y_t - EY_t = \sum_{i=1}^{\infty} C_i \epsilon_{t-i}$ is called $I(0)$ process if $C = \sum_{i=0}^{\infty} C_i \neq 0$.

Definition 1.2 A stochastic process $\{Y_t\}$ is called integrated of order d , $I(d)$, $d = 1, 2, \dots$, if $\Delta^d(Y_t - EY_t)$ is $I(0)$ process.

Let in the following ϵ_t be a sequence of independent normally distributed n -dimensional random variables $\epsilon_t \sim N_n(0, \Omega)$.

Definition 1.3 A stochastic process $\{Y_t\}$ is called n -dimensional autoregressive process $VAR(p)$, if

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \Lambda D_t + \epsilon_t, t = 1, 2, \dots, T \quad (1)$$

for fixed values of Y_{-p+1}, \dots, Y_0 , where Φ_1, \dots, Φ_p are matrices of coefficients ($n \times n$), Λ is an ($n \times s$) matrix of coefficient of deterministic term D_t ($s \times 1$), which can contain a constant, a linear term, seasonal dummies, intervention dummies or other regressors that we consider non-stochastic.

The process defined by the equation (1) can be written in error correction form

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \Lambda D_t + \epsilon_t, t = 1, \dots, T, \quad (2)$$

where $\Pi = \sum_{i=1}^p \Phi_i - I$, $\Gamma_i = -\sum_{j=i+1}^p \Phi_j$. This error correction form of VAR process is used in the analysis of cointegration.

The basic idea of cointegration can be shown on 2 one-dimensional processes of order $I(1)$. We say that the processes X_t and Y_t are cointegrated if there exists any linear combination $aX_t + bY_t$ which is stationary.

Definition 1.4 Let Y_t be n -dimensional process integrated of order 1. We call this process cointegrated with a cointegrating vector β ($\beta \in R^n, \beta \neq 0$) if $\beta' Y_t$ can be made stationary by a suitable choice of its initial distribution.

If $n > 2$ then there may be two nonzero $n \times 1$ vectors β_1, β_2 such that $\beta_1' Y_t$ and $\beta_2' Y_t$ are both stationary, where β_1, β_2 are linearly independent. Indeed, there may be $r < n$ linearly independent cointegrating vectors. The *cointegrating rank* is the number of linearly independent cointegrating vectors and space spanned by these vectors is *cointegrating space*.

1.2 Maximum likelihood estimation of the cointegrating vector

We will deal in brief with the maximum the likelihood analysis of the cointegrated system in this part of the article.

Granger's theorem (see Johansen (1995)) gives necessary and sufficient conditions for $VAR(p)$ process to be $I(1)$ and cointegrated. According to the rank of the matrix Π in the error correction form we define $H(r)$ model of the process $I(1)$.

Definition 1.5 $H(r)$ model of the process $I(1)$ is defined as a model $VAR(p)$ such that

$$\Pi = \alpha\beta',$$

where α and β are $n \times r$ matrices. The reduced form error correction model is

$$\Delta Y_t = \alpha\beta'Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \Lambda D_t + \epsilon_t, \quad t = 1, \dots, T, \quad (3)$$

where the parameters $\alpha, \beta, \Gamma_1, \dots, \Gamma_{p-1}, \Lambda, \Omega$ vary freely.

Under hypothesis

$$H(r) : \Pi = \alpha\beta'$$

the maximum likelihood estimator of β is found by the following procedure (see Johansen (1995), Hamilton (1994)). First of all we solve the equation

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$$

for the eigenvalues $1 > \lambda_1 > \dots > \lambda_n > 0$ and eigenvectors $V = (V_1, \dots, V_n)$ which we normalize by $V'S_{11}V = I$ (S_{00}, S_{10}, S_{01} and S_{11} are matrices described in Johansen (1995)). The cointegrating relations are estimated by

$$\hat{\beta} = (V_1, \dots, V_r),$$

the maximalized likelihood function is

$$L_{\max}^{-2/T} = |S_{00}| \prod_{i=1}^r (1 - \lambda_i). \quad (4)$$

The likelihood ratio test $Q(H(r)|H(n))$ for testing $H(r)$ in $H(n)$ we obtain by comparing two expressions (4) for r and n , then

$$Q(H(r)|H(n))^{-\frac{2}{T}} = \frac{|S_{00}| \prod_{i=1}^r (1 - \lambda_i)}{|S_{00}| \prod_{i=1}^n (1 - \lambda_i)}.$$

The logarithm of this expression is called TRACE statistic

$$-2 \ln Q(H(r)|H(n)) = -T \sum_{i=r+1}^n \ln(1 - \lambda_i). \quad (5)$$

The test statistic for testing $H(r)$ in $H(r+1)$ – MAX statistic – is given by

$$-2 \ln Q(H(r)|H(r+1)) = -T \ln(1 - \lambda_{r+1}). \quad (6)$$

The asymptotic distribution of the statistics (5) and (6) depends on the deterministic terms present in the model (see Johansen (1995)).

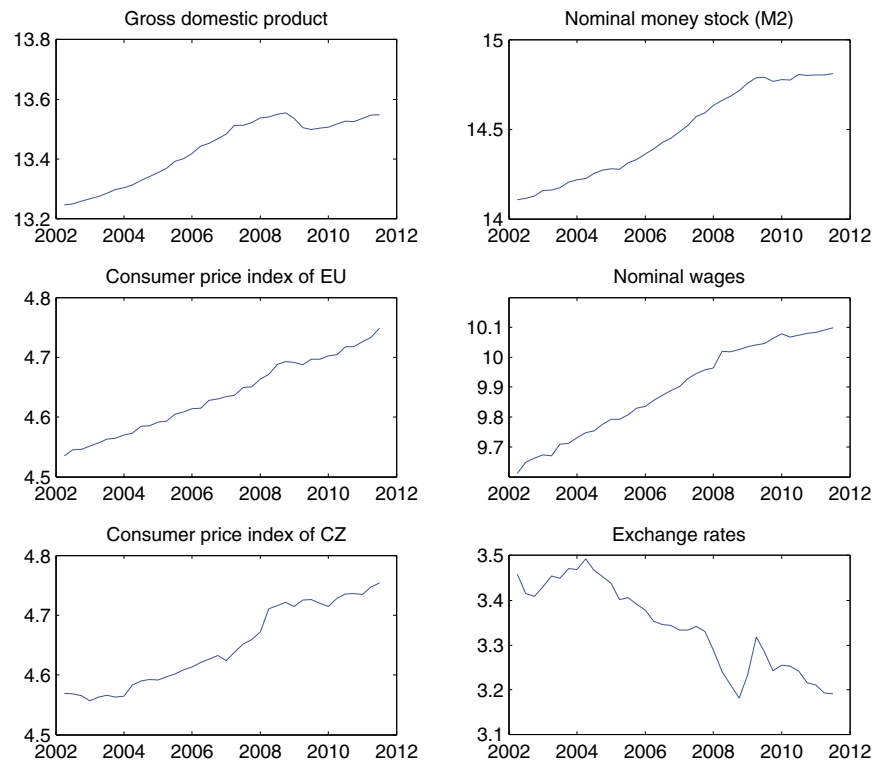


Figure 1: Natural logarithm of time series

2 Analysis of the inflation model of the Czech Republic

The article Kim (2001) analyses relative impacts of the monetary, labour and foreign sector on Polish inflation using cointegration and error-correction models. They use a structural system approach in which cointegration relationships are used to derive deviations from steady state levels. We apply the described model to the macroeconomic times series of the Czech Republic and try to verify validity of the model for given datasets. We use these time series in the model of inflation (quarterly data):

- y_t^R – natural logarithm of the real Czech gross domestic product (GDP)
- m_t – natural logarithm of the nominal money stock (M2) in domestic currency
- p_t^F – natural logarithm of the consumer price index of the European Union
- w_t – natural logarithm of nominal wages in the Czech republic
- p_t^D – natural logarithm of the Czech consumer price index
- e_t – natural logarithm of exchange rates between Czech crown and Euro

Data sources: EUROSTAT (epp.eurostat.ec.europa.eu) and the Czech national bank (www.cnb.cz).

r	Eigenvalue	TRACE	p -value	MAX	p -value
0	0.97542	258.99	0.0000	122.29	0.0000
1	0.85852	136.69	0.0000	64.535	0.0000
2	0.67917	72.160	0.0000	37.516	0.0000
3	0.59454	34.644	0.0000	29.790	0.0000
4	0.13677	4.8534	0.0276	4.8534	0.0276

Table 1: The test of cointegration in the model of inflation – TRACE statistic

We assume (according to Kim (2001)) that there are three sources of inflation in an open economy: wage inflation, monetary inflation and imported inflation. We can write

$$p_t^D = f(p_t^F, m_t - p_t^D, w_t - p_t^D, e_t, y_t^R),$$

where $m_t - p_t^D$ is real money supply, $w_t - p_t^D$ are real wages and y_t^R is the real output.

The first equation is based on the claim that the expansion of money supply in excess of real productive potential of economy leads to inflation. Assuming the homogeneity between real money balance and real output is possible to write

$$m_t - p_t^D = b + y_t^R. \quad (7)$$

If increases in real wages are greater than the levels warranted by productivity, inflation will occur. We can establish the following equation

$$w_t - p_t^D = a + (y_t - p_t^D). \quad (8)$$

An inequality between price levels of particular national economies results in imported inflation

$$p_t^D = e_t + p_t^F. \quad (9)$$

A detailed derivation of these equations is possible to find in Kim (2001).

We define multidimensional time series

$$Y_t = \begin{pmatrix} p_t^D \\ e_t + p_t^F \\ w_t - p_t^D \\ m_t - p_t^D \\ y_t^R \end{pmatrix}.$$

We used several unit root tests to analyze non-stationarity. It turned out that all components of Y_t are non-stationary $I(1)$ processes. The time series Y_t can be described as VAR(5) process with a constant. It should be noted that the choice of the parameter p in the VAR (p) is not quite unambiguous. It was chosen on the basis of several criterions (Akaike, Schwartz Bayesian and Hannan-Quinn criterion), values of cross-correlation functions and portmanteau tests.

In view of the theory developed in the previous part (Kim (2001)), we could expect 3 cointegrating vectors which are columns of the matrix

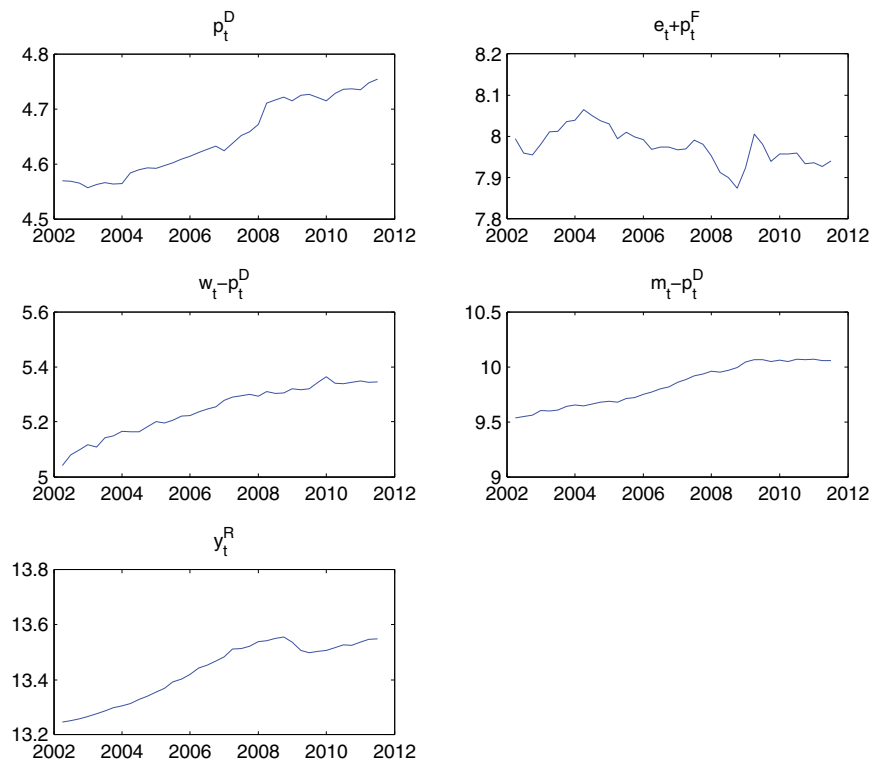


Figure 2: Multidimensional time series Y_t

$$\beta = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & 0 \end{pmatrix}.$$

Table 1 summarizes the results of the cointegration analysis. According to the TRACE and MAX test statistics, we can decide that there are 4 cointegrating relations in the time series Y_t . (with the risk 0.01) but our model should contain only 3 cointegrating vectors. The estimate of the matrix of cointegrating vectors is

$$\hat{\beta} = \begin{pmatrix} 1.0000 & 0.00000 & 0.00000 & 0.00000 \\ (0.00000) & (0.00000) & (0.00000) & (0.00000) \\ 0.00000 & 1.0000 & 0.00000 & 0.00000 \\ (0.00000) & (0.00000) & (0.00000) & (0.00000) \\ 0.00000 & 0.00000 & 1.0000 & 0.00000 \\ (0.00000) & (0.00000) & (0.00000) & (0.00000) \\ 0.00000 & 0.00000 & 0.00000 & 1.0000 \\ (0.00000) & (0.00000) & (0.00000) & (0.00000) \\ -0.74918 & 0.80640 & -0.85806 & -2.6412 \\ (0.049725) & (0.11633) & (0.019387) & (0.19438) \end{pmatrix}.$$

The numbers in parenthesis are standard errors. All calculations were done in software Gretl and Matlab.

We denote $\beta_1 = (0, 0, 0, 1, -1)'$, $\beta_2 = (0, 0, 1, 0, -1)'$ and $\beta_3 = (1, -1, 0, 0, 0)'$. We are able to test these linear restriction imposed on the estimated cointegrating vector. For the first restriction β_1 we get the test statistic 8.09638 (unrestricted loglikelihood = 745.49441, restricted loglikelihood = 741.44622) and p -value = 0.00444, for the second restriction β_2 we get the test statistic 1.25606 (restricted loglikelihood = 744.86638) and p -value 0.26240 and for the third restriction β_3 the test statistic 4.61338 (restricted loglikelihood = 743.18772) and p -value 0.03172. According to the results, we reject validity of the first equation describing monetary inflation (7). The second relation – wage inflation (8) – seems to be valid. The last restriction (imported inflation 9) can be rejected with the risk 0.05, but not with the risk 0.01. If we test two restrictions at once, all three combinations are rejected. The restriction consisting of all three vectors β_1 , β_2 and β_3 is rejected as well.

3 Conclusion

We described multidimensional time series as the model VAR(5) with a constant. We have found out that it is possible to identify four cointegrating vectors, nevertheless, the theoretical model (see Kim (2001)) should contain only three cointegrating vectors. Using some statistical tests we tried to verify that the given theoretical expectations of the inflation model are consistent with data sets describing trends in the Czech economy from 2002 to 2012. The assumptions that wage inflation can be described by the equation (7) and imported inflation is given by the equation (9) seem to be correct, but the remaining relation (monetary inflation) gave us different findings. These results are based on testing each restriction separately. If we want to verify more relations at once, such models are rejected.

Based on the previous cointegration analysis it is possible to say that the model of inflation described in the article Kim (2001) when applied to data of the Czech Republic is not fully applicable. The first indication of discrepancies can be seen in the description of the system at the beginning when it was difficult to unambiguously determine the parameter p in VAR process. This parameter affects the behavior of the majority of the tests as well as the behavior of the tests which were used. An inadequate model specification may result in other problems. This can be solved by adding new variables into the model. Another potential problem may be connected with the recent economic crisis and its impact on economic time series behavior. These questions require extensive consultation with experts in economic theory and will be done in future research.

Acknowledgement

The paper was supported by the grant GAČR P402/10/P209.

Reference

- [1] HAMILTON, J. D.: *Time Series Analysis*. Princeton: Princeton University Press, 1994. ISBN 0-691-04289-6.

- [2] JOHANSEN, S.: *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*. Oxford: Oxford University Press, 1995. ISBN 0-19-877450-7.
- [3] JUSELIUS, K.: *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press, 2006. ISBN 0-19-928566-7.
- [4] KIM, B. Y.: *Determinants of Inflation in Poland: A Structural Cointegration Approach*. Bank of Finland. Institute for Economies in Transition, BOTFIT, 2001.
- [5] LÜTKEPOHL, H.: *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag, 2007. ISBN 978-3-540-26239-8.
- [6] NEUBAUER, J.: Modelling of Economic Time Series and the Method of Cointegration. *Austrian Journal of Statistics*, Vol. 35, No. 2–3, p. 307–313, 2006. ISSN 1026-597X.

Current address

Mgr. Jiří Neubauer, Ph.D.

University of Defence,
Kounicova 65,
662 10 Brno,
Czech Republic,
tel. +420 973 442 029,
email: Jiri.Neubauer@unob.cz

APPLICATION OF RELEVANCE VECTOR MACHINE TO FORECASTING VOLATILITY IN CZECH FINANCIAL TIME SERIES

ŽIŽKA David, (CZ)

Abstract. The Relevance Vector Machine (RVM) is a probabilistic method which is very powerful for prediction problems. This paper introduces an application the RVM to predictions linear GARCH and nonlinear GJR-GARCH models. First, daily returns from Czech financial market were modeled by volatility models. Second, the results of estimating these models were used as input to the RVM. The best models obtained from training were used for forecasting. Final part was focused on comparison the predictive power of classical volatility models and models based on the RVM.

Keywords. Relevance vector machine, financial time series, volatility models

Mathematics Subject Classification: 62M10

1. Introduction

Forecasting volatility of financial time series is very important activity in the financial markets. Typically, parametric models of volatility are used for these predictions. This paper introduces an alternative probabilistic method called the Relevance Vector Machine (RVM) for obtain more accurate predictions. The main goal is to investigate the predictive power of the classical GARCH and the RVM models. This paper is divided into two parts; theoretical overview and experimental section.

The input data to the experimental section are close daily values of the PX index (Prague Market Index). The reference period is during 5.1.2000 – 29.6.2007. Logarithmic returns expressed as percentages will be used.

First part is focused on estimating the best volatility models during 5.1.2000-29.12.2006. Specifically, linear GARCH and nonlinear GJR-GARCH models are used for the in-sample estimation. Subsequently, these results are used as input to train the RVM. This training set is used to learn a model of the dependency of the targets on the inputs with the objective of making accurate predictions.

Final part is focused on comparing the predictive ability of the RVM with parametric volatility models. For this purpose is used out-of-sample time period during 2.1.2007-29.6.2007, i.e. it covers 126 daily values.

2. Overview of Methods

Forecasts of Volatility Models

GARCH model

GARCH means Generalized Autoregressive Conditional Heteroscedasticity model. Bollerslev (1986) proposes a useful extension of ARCH model known as the generalized ARCH model. Bollerslev extended ARCH model of delayed conditional variance. Detailed description of this model can be found in [1].

One step ahead forecast of GARCH (1,1)

$$\sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 . \quad (2.1)$$

GJR GARCH model

GJR GARCH model by Glosten, Jagannathan and Runkle (1993) is similar to Threshold GARCH model by Zakoian [8]. Though the GJR model is designed to capture the leverage effect between asset return and volatility, the way is not the same as Exponential GARCH model (EGARCH). The leverage coefficients of the EGARCH model are directly applied to the actual innovations while the leverage coefficients of the GJR model can connect to the model through an indicator variable. For this case if the asymmetric effect occurs the leverage coefficients should be negative for the EGARCH model and positive for the GJR model. Detailed description of this model can be found in [2].

One step ahead forecast of GJR (1,1)

$$\begin{aligned} \sigma_{t+1}^2 &= \omega + \beta_1 \sigma_t^2 + \alpha_1 \varepsilon_t^2 \text{ for } \varepsilon_t > 0 \text{ and} \\ \sigma_{t+1}^2 &= \omega + \beta_1 \sigma_t^2 + \alpha_1 \varepsilon_t^2 + \gamma_1 \varepsilon_t^2 \text{ for } \varepsilon_t < 0 . \end{aligned} \quad (2.2)$$

Relevance Vector Machine

The Relevance Vector Machine (RVM) introduced by Tipping (2000) [6] has become a powerful tool for prediction problems as it uses a Bayesian approach with a functional form identical to the Support Vector Machine (SVM) introduced by Vapnik (1995) [7]. It enjoys the benefit of the SVM based techniques, including generalization and sparsity. Noticeably, the RVM does not have the limitations of SVM. Hence it has more advantages over SVM in the sense that the RVM generates probability based prediction. This approach relaxes Mercer's condition on Kernel basis functions used for training and the RVM does not need to estimate the trade-off parameter C.

Tipping (2001) [5] illustrated the RVM's predictive ability on some popular benchmarks by comparing it with the SVM. The empirical analysis also proved that the RVM outperformed the SVM. Hence are investigated the volatility models based on the RVM.

In supervised learning we are given a set of examples of input vectors $\{x_n\}_{n=1}^N$ along with corresponding targets $\{t_n\}_{n=1}^N$. From this learning set we wish to learn a model of the dependency of the targets on the inputs with the objective of making accurate predictions of t for previously unseen values of x .

The predictions are based on some function $y(x)$ defined over the input space and learning is the process of inferring this function. A flexible and popular formula for $y(x)$ is:

$$y(x; w) = \sum_{j=1}^M \omega_j \phi_j(x) = w^T \phi(x), \quad (2.3)$$

where basic function $\phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_m(x))^T$ is nonlinear, $w = (\omega_1, \omega_2, \dots, \omega_m)^T$ is weight vector and x is input vector.

3. Experimental Results

The Prague Market Index (PX) is investigated in this section. The reference period is during 5.1.2000 – 29.6.2007 (it contains 1882 values). This period is divided into two parts: 5.1.2000–29.12.2006 (1756 values) and 2.1.2007–29.6.2007 (126 values). First part is used for in-sample estimating the volatility models and training the RVM. Second part is taken for out-of-sample forecasting. Logarithmic returns expressed as percentages are used for this analysis.

Subsequently, Maximum likelihood estimates of parametric models are computed in GiveWin software. Further, these models are trained in R 2.13.2 with Kernlab package by Karatzoglou (2004). Detailed description of this package can be found in [3].

Volatility Models

Estimations of parameters are statistically significant for all models (see Tables 1-2). The Portmanteau test does not confirm an autocorrelation for all models. Further, the asymptotic test does not confirm a normal distribution and the ARCH test does not confirm a conditional heteroscedasticity for all models.

The GJR GARCH (1,1) is more adequate than the GARCH (1,1) model in accordance with AIC.T and log-likelihood (LL) criteria.

Table 1. GARCH (1,1) Model Parameters

GARCH	Coefficient	Std.Error	t-value
ω	0.0480	0.0133	3.36*
α_1	0.0959	0.0152	5.22*
β_1	0.8755	0.0192	41.1*

Note: Significant at the: 1% level *, 5% level **

Table 2. GJR GARCH (1,1) Model Parameters

GJR	Coefficient	Std.Error	t-value
ω	0.0703	0.0189	3.44*
α_1	0.0399	0.0159	2.78*
β_1	0.8588	0.0218	41.4*
Threshold	0.1220	0.0371	2.89*

Note: Significant at the: 1% level *, 5% level **

Relevance Vector Machine for PX

The training set (2000-2006) is used to learn the RVM model. Results are in Table 3 and Fig.1-4. In this training has GARCH-RVM model the lowest values of relevance vectors, variance and training error. On the other hand, it is important to note that these values are strongly affected by the choice of sigma parameter (estimated or chosen).

Table 3. Training Results for the RVM

	GARCH	GJR-GARCH
Hyperparameter sigma	102	162
No. of Relevance Vectors	80	97
Variance	1.3196	1.6397
Training Error	1.2707	1.5658

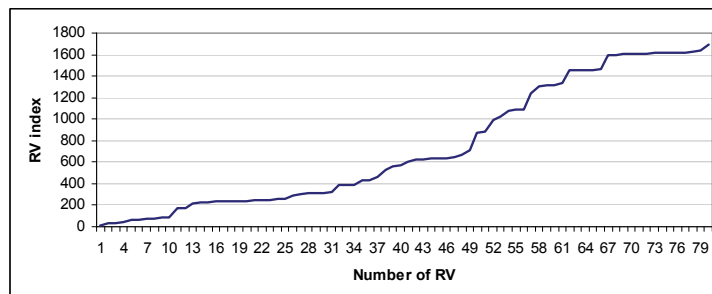
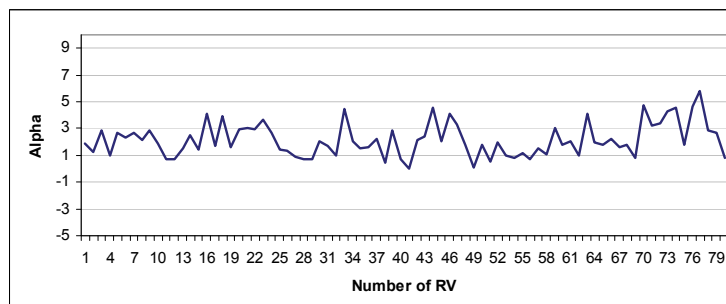
Fig. 1. GARCH RVM: Relevance Vector Index

Fig. 2. GARCH RVM: Alpha values


Fig. 3. GJR RVM: Relevance Vector Index

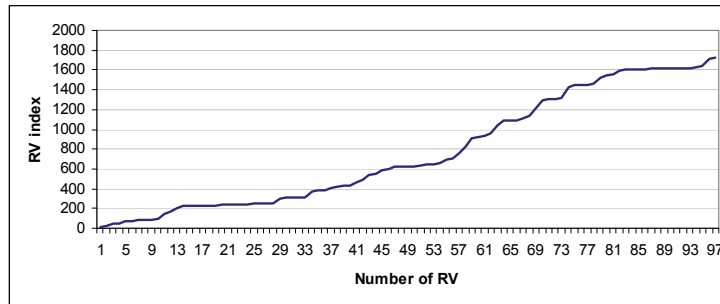
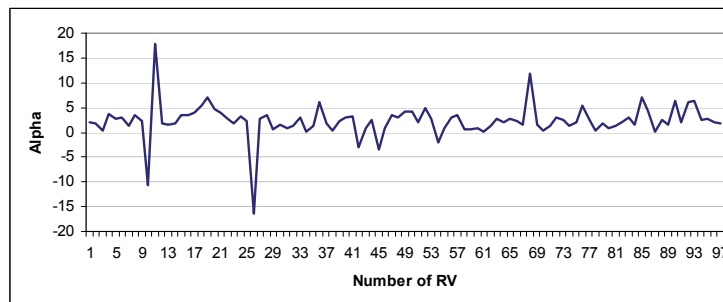


Fig. 4. GJR RVM: Alpha values



Predictive Power of Forecasts

The predictive power of models is measured by Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE):

$$\begin{aligned}
 MSE &= \frac{1}{n_2} \sum_{t=1}^{n_2} (y_t^2 - \hat{\sigma}_t^2)^2, \\
 RMSE &= \sqrt{\frac{1}{n_2} \sum_{t=1}^{n_2} (y_t^2 - \hat{\sigma}_t^2)^2}, \\
 MAE &= \frac{1}{n_2} \sum_{t=1}^{n_2} |y_t^2 - \hat{\sigma}_t^2|,
 \end{aligned} \tag{2.4}$$

where y_t^2 = actual values, $\hat{\sigma}_t^2$ = forecasted volatility and n_2 = predicted sample size.

Table 4 includes complete results. The best predictive power of parametric models is achieved in the case of the GARCH model but values of the GJR are very close. The best predictive power of hybrid models is achieved in the case of the GJR-RVM model. The RVM models have better forecasts than classical volatility models for this case.

Table 4. Goodness of Fit Measures

	MSE	RMSE	MAE
GARCH	3.3308	1.8250	0.9384
GARCH RVM	3.2236	1.7954	0.8752
GJR GARCH	3.3521	1.8309	0.9395
GJR GARCH RVM	3.1575	1.7769	0.6844

4. Conclusion

First, the best linear and nonlinear models of volatility were fitted for returns of the PX index. The different effects of positive and negative returns on the conditional variance were confirmed. The GJR GARCH model was more adequate than the GARCH model based on chosen criteria.

Subsequently, these results were used as input to train the RVM. The GARCH-RVM model had lower values of relevance vectors, variance and training error.

Second, the predictive power was measured by MSE, RMSE and MAE. The best predictive power of parametric models was achieved in the case of the GARCH model. The best predictive power of hybrid models was achieved in the case of the GJR-RVM model. Generally, the RVM models had better forecasts than classical volatility models, at least for the PX index. In future, it is planned to apply time series of exchange rates as well.

References

- [1] BOLLERSLEV, T., CHOU, R., KRONER, K.: ARCH modeling in finance, *Journal of Econometrics*, Vol 52 (1992).
- [2] GLOSTEN L., JAGANNATHAN R., RUNKLE D.: On the relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, Vol 46 (1993), 779-801.
- [3] KARATZOGLU, A., SMOLA, A., HORNIK, A., ZEILEIS, A.: An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 2004.
- [4] NELSON, D.B.: Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, Vol 59 (1991), 347-370.
- [5] TIPPING, M.E.: Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, Vol 1 (2001), 211-244.
- [6] TIPPING, M.E.: Relevance Vector Machine. Microsoft research, Cambridge, UK 2000.
- [7] VAPNIK, V.N.: The nature of statistical learning theory. Springer-Verlag, New York 1995.
- [8] ZAKOIAN, M.: Threshold Heteroscedastic Models, *Journal of Economic Dynamics and Control*, Vol 18 (1994) 931-955.

Current address

Žižka David, Ing.

University of Economics Prague,

W.Churchill Square 4, 130 67 Prague, Czech Republic, xzizd02@vse.cz