Mathematica Balkanica

New Series, Vol. 24, 2010, Fasc. 3-4

Preface

Muharem Avdispahic [1]

Coordinator of the TEMPUS Project
SEE Doctoral Studies in Mathematical Sciences
(144703-TEMPUS-2008-BA-TEMPUS-JPCR)

The main goals of the TEMPUS Project "SEE Doctoral Studies in Mathematical Sciences", funded by European Commission under the TEMPUS IV first call, consist of the development of a model of structured doctoral studies in Mathematical Sciences involving the network of Western Balkans universities, the curricula design based on the existing strenghts and tendencies in the areas of Pure Mathematics, Applied Mathematics and Theoretical Computer Science and the first phase of implementation of the agreed model during the SEE Doctoral Year in Mathematical Sciences 2011.

A decisive step in this direction was "SEE Young Researchers Workshop" held in Ohrid, FYR Macedonia, September 16-20, 2009, as a part of the Mathematical Society of South-Eastern Europe (MASSEE) International Congress on Mathematics - MICOM 2009. MICOM 2009 continued the tradition of two previous highly successful MASSEE congresses that took place in Bulgaria in 2003 and in Cyprus in 2006.

This volume of the journal Mathematika Balkanica contains the talks delivered at Ohrid Workshop by South-Eastern European PhD students in various stage of their research towards a doctoral degree in mathematics or informatics. Facilitating publication efforts of young researchers from the universities of Sarajevo, Tuzla, Belgrade, Skopje, Stip, Graz, and Sofia fully coincides with MASSEE goals to promote, organize and support scientific, research and educational activities in South-Eastern Europe.

The consent of the Editorial Board of Mathematica Balkanica to publish "SEE Young Researchers Workshop" contributions aptly meets intentions of European reform processes aimed at creating the European Higher Education Area and European Research Area. It is an encouragement to these young researchers in the first place and at the same time an encouragement to their institutions in overcoming fragmentation and enhancing their capacities through fostering reciprocal development of human resources.

[1]Department of Mathematics, University of Sarajevo, Zmaja od Bosne 35, 71000 Sarajevo, BOSNIA and HERZEGOVINA, E-mail: mavdispa@pmf.unsa.ba

Mathematical Society of South-Eastern Europe

Mathematica Balkanica

A quarterly published
by the Bulgarian Academy of Sciences—
National Committee for Mathematics

Honorary Editor — ***Blagovest Sendov***
Editor-in-Chief — ***Petar Kenderov***

Mathematica Balkanica - Editorial Office
Acad. G. Bonchev str. Bl.25A, 1113 Sofia, Bulgaria
Phone: +359-2-979 6311, Fax: +359-2-870 7273
E-mail: balmat@math.bas.bg

WEB PAGE: http://www.mathbalkanica.info

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Classification of Smoking Cessation Status Using Various Data Mining Methods

Aleksandar Kartelj

This study examines different approaches of binary classification applied to the problem of making distinction between former and current smokers. Prediction is based on data collected in national survey performed by the National center for health statistics of America in 2000. The process consists of two essential parts. The first one determines which attributes are relevant to smokers status, by using methods like basic genetic algorithm and different evaluation functions [1]. The second part is a classification itself, performed by using methods like logistic regression, neural networks and others [2]. Solving these types of problems has its real contributions in decision support systems used by some health institutions.

AMS Subj. Classification: 62P10, 62H30, 68T01

Key Words: data mining, classification, induction learning

1. Introduction

Today data mining is one of the most popular and most exciting disciplines of applied informatics. It enables us to discover complex and hidden patterns in data, which can potentially bring to totally new conclusions in different disciplines, where sometimes even those disciplines experts cannot do better. One of especially interesting areas today in which data mining is often applied is certainly medicine. Decision support systems in health are developing more than ever, and their backbones are often theoretically founded on proved mathematical and physical principles.

1.1. **Some applications.** Classification problems are recognized in group of data mining methods. Beside, there are also methods of association, clustering techniques, regression etc. Classification can have exact mathematical models, heuristic models or random based models.

Some of the most important applications in real life are tumor classification, spam filters, decision making about good candidates for bank credits etc. Practically every real problem where the output is yes or no, can be solved using this technique. Naturally, the problem context must also be included in the process.

The smoking cessation status problem is a classification problem. It is highly dimensional, which means that it is dependent on large number of factors. Those factors are important in understanding of goal problem. One more potential gain of solving this problem is noticing new, until now not recognized correlated factors (in medical terminology dimension can be translated to symptom).

1.2. **Data.** Data collected by the National center for health statistics are presented in structured way (tabular form). In this study the main interest is put on only one of these surveys, the one oriented to questions related to adult persons. There were about 30000 respondents in it, and all answers are presented in one file (2000 NHIS sample adult file), which is publicly available at the site of this institution. Every person in this survey answered a set of 1429 questions, and some of these questions were related to smoking behaviour. Smoking status was cached in attribute SMKSTAT1. There are five possible statuses: `current´, `former´, `never´, `smoker, but currently unknown´and `unknown´. In this study we are solving the problem of binary classification, so we are interested only in instances which have values `current´or `former´. The idea is at first to filter only the relevant set of questions [1], and then based on their answers only produce binary classificator which will perform prediction of smoking status: former or current smoker.

2. Preprocessing

2.1. **Manual attribute elimination.** From more than 30000 instances, in the first step total of 14416 was selected (7421 current and 6995 former smoker). The ratio is well balanced, which contributes to the algorithm efficiency, as it will be seen later. The major of 1429 attributes is of nominal type, every with two or more possible values. A large part of the proposed set is irrelevant, so the first technique is a manual reduction of the attribute set, and then the automatic subset algorithm using data mining tool Weka v3.6 was used.

The manual attribute reduction is performed logically. So it is obvious on the first sight, that some attributes are redundant. For example, the prediction will not depend on rase of respondent, well not enough to make some important contribution to decision.

One other important heuristic, during the manual procedure is the value frequency on some question. Sometimes, there exists obviously correlated dependency, but still we should eliminate some feature (question). The best example are the questions about pregnancy. It is clear that most pregnant women will stop smoking, but only one small percent of women who were answering this inquiry were pregnant, so the relevance of these questions is not important in forming of general predictor [2].

An attribute can be relevant to the prediction, but sometimes we remove it because of redundancy. For example FRUITNO, FRUITTP, FRUITY and FRUITW all represent information whether and how much somebody eats fruits, with having in mind that FRUITY and FRUITW are generated using FRUITNO (mass of fruit) and FRUITTP (frequency of taking fruit). From this fact obvious is the redundancy of keeping attributes FRUITY and FRUITW, because they are already implicitly expressed through others.

3. **Selecting best attribute subset**

In learning algorithms often there is a problem with large input dimension. In data mining there are generally two techniques for solving this kind of problem [4]:

(1) Selecting subset of instances
(2) Selecting subset of features (attributes)

Selecting subset of instances is sometimes called sampling, and it represents one of the basic statistical tecniques. The goal is to select a representative sample, which is the sample that will contain the same or almost the same distribution, mathematical expectance, and dispersion as the original set. The easy way is certainly to select a random sample, but sometimes it is imperative to have a good distribution (the random sample can fail), so one possible approach is doing cluster analysis prior selection, and after that a proportional random selection from each cluster.

Selecting subset of features is a technique where we try to decrease the problem dimension. As it is described in 1.2 our problem is highly dimensional. This algorithm in the most general case tries to find the best subset of features from possible 2^N subsets, accordingly to evaluation function. The brute force algorithm is obviously time consuming even for sets with relativly small input dimension (N). So, there are different heuristical and random based principles, which can gain some performance. According to [1] there are 4 basic steps in the typical feature subset selection algorithm:

(1) generation procedure
(2) search procedure
(3) stopping criterion (eg. evaluation function)
(4) validation procedure

3.1. **Evaluation function.** The evaluation function is practically the condition of optimality in our case, but generally it can be statistical, heuristical or some other metric. The most common categorization of evaluation functions is on:

- Filter methods
- Wrapper methods

Evaluation function	Generality	Time complexity	Accuracy
Distance metric	Yes	Small	-
Info gain	Yes	Small	-
Dependency degree	Yes	Small	-
Consistency metric	Yes	Medium	-
Classifier error	No	Large	Very large'

TABLE 1. Evaluation functions

Method	Number of features	Correctly classified	Kappa statistic
Genetic algorithm (CfsSubsetEval)	24	71.52%	0.4289
Best first (CfsSubsetEval)	25	73.05%	0.4569
Ranker method (GainRatioAttributeEval)	25 (fixed)	71.31%	0.4244
Greedy algorithm (CfsSubsetEval)	25 (fixed)	73.00%	0.4587

TABLE 2. Selection results

Filter methods evaluate the subset quality accordingly to some prior criterion convention: distance measure, info gain, degree of dependence, consistency etc. **Wrapper methods** are not predefined, so they form criterion in dependence with learning algorithm. In most cases the criterion is the classification error itself. In table1 a comparison of evaluation functions using some key features is shown.

3.2. **Results comparison.** In this study there are few techniques used for feature subset selection algorithm 3, and they are all performed in Weka 3.6 data mining tool.

3.2.1. *Selected features.* Results 2 show that complete search (Best first) in combination with distance measure based evaluation function gave best results. Training was performed on random selected sample consisted of 66,6% of all instances. Afterward, the testing was performed using the rest of 33,3% instances. Selected attributes are shown in Table 3

4. **Classification**

Mathematically, the classification problem is defined as:
Let $\alpha = \{(x_1, y_1), \ldots, (x_n, y_n) | x_i \in R^n, y_i \in \{-1, 1\}\}$ be the training set. The function f, sometimes called predictor or classifier is formed using some learning algorithm and training set. After performing the learning algorithm, f maps

Feature	Description	Feature	Description
AGEP	Age	RATCAT	Income
HYPEV	Blood pressure	HEARAID	Hearing problems
RESTLESS	Restless	WRKLYR2	Had job in last 12m
ALCAMT	Alcohol	BMI	Body mass index
AUSUALPL	Where to go when sick	AHCAFYR1	Drug availability
ADNLONGR	Dental hygiene	FOBHAD	Blood analysis
AHCSYR2	Went to stomatologist	AHCSYR8	Went to doctor
SHTFLUYR	Flu vaccine	SHTPNUYR	Pneumonia vaccine
STD	Had infective illness	MILKKND	Milk kind
FRUITY	Fruit	VITEM	Vitamins in past 12m
CALC	Calcium use	MDTOB1	Asked about tobacco
SMHARM	Asked about tobacco risks	INCR150 INCR150	Opinion about tobacco increase
SKNX	Done complete head to toe check		

TABLE 3. Selected features

arbitrary test instance t from R^n in ˝an appropriate˝class c from $\{-1, 1\}$. Under ˝appropriate ˝we mean the one that minimizes empirical risk

$$I_{emp}[f; n] = \frac{1}{n} \sum_{i=1}^{n} V(y_i - f(x_i)) \,,$$

where V represents some task-specific loss function and y_i is the correct value for given x_i [3].

Classification methods. There are many techniques, for implementing above given preposition [6]. This study analyzed four different techniques:

(1) Logistic regression
(2) Multilayer perceptron
(3) SVM
(4) Decision tree C4.5

4.1. **Logistic regression.** This method tries to fit training samples under curve of sigmoid (logistic) function. Similarly, there is also linear regression, which tries to fit data under the linear function. Logistic regression represents referent politics in classification methods.

Method	Correctly classified	Kappa statistics	Running time (seconds)
Logistic regression	73.05%	0.4596	8
Multilayer perceptron	69.09%	0.3843	3021
SVM	55.07%	0.0822	603
J48 (C4.5)	70.90%	0.4166	6

TABLE 4. Classification results

4.2. **Multilayer perceptron.** Neural network fits highly dimensional data very well. It has good degree of generalization, which makes it possible to perform well on unseen (test) data.

4.3. **SVM - Support vector machine.** One of today most prominent technique. It performs extremely good in work with high dimensional data. This method tries to put margin between positive and negative instances [5].

4.4. **Decision tree C4.5.** In every step of the algorithm one attribute is chosen, and that attribute then represents new node in tree. By that node the tree branches on each possible value for that attribute. Then the algorithm is recursively called for each of newly created nodes. In Weka 3.6 this algorithm is called j48.

4.5. **Results comparison.** All algorithms were performed using Weka 3.6. Table 4 shows the results. Logistic regression made best result. Multilayer perceptron is still the best potential candidate, and the reason why it did not perform best here is probably because networks incorporated in Weka 3.6 have very general character, so they are not meant for working with this specific problem. Decision tree C4.5 is generally applicable on shallow data, and the proposed result is probably very close to the best possible one that can be gained using this technique. Even SVM matters today as the best and most popular technique, in this situation it failed. The potential reason is the large diversity of values on some attributes. Since SVM reduces the multiclass problem (more than 2 values per attribute) to binary, this probably comes out as a performance issue, and reduces the algorithm accuracy in some way.

5. Conclusion and further work

The selection of best classification method is almost always dependent on data itself. In future, different architectures of neural networks should be investigated, considering the fact that in this study only the simplest version of the multilayer perceptron was used.

References

[1] M. Dash, H. Liu, Feature Selection for Classification, *Jour. Intelligent Data Analysis,* **1**, 1997, 131–156.

[2] Mollie R. Poynton, Anna M. McDaniel, Classification of smoking cessation status with a backpropagation neural network, *Journal of Biomedical Informatics*, **39**(6), 2006, 680–686.

[3] Theodoros Evgeniou, Massimiliano Pontil and Tomaso Poggio, Statistical Learning Theory: A Primer, *International Journal of Computer Vision*, **38**, 1, 2000, 9–13.

[4] Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, *Addison-Wesley*, 2005.

[5] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge 2000.

[6] S.B. Kotsiantis, *Supervised Machine Learning: A Review of Classification Techniques*, Informatica, **31**, 2007, 249–268.

Department of Informatics
Faculty of Mathematics
University of Belgrade
Studentski Trg 16
11000 Belgrade, Serbia
E-mail: kartelj@matf.bg.ac.rs

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Cryptographic Primitives with Quasigroup Transformations

Aleksandra Mileva

The intention of this research is to justify deployment of quasigroups in cryptography, especially with new quasigroup based cryptographic hash function NaSHA as a runner in the First round of the ongoing NIST SHA-3 competition. We present new method for fast generation of huge quasigroup operations, based on the so-called extended Feistel networks and modification of the Sade's diagonal method. We give new design of quasigroup based family of cryptographic hash functions - NaSHA, which deploy the new method and with a novel approach - different quasigroups for every application of component quasigroup transformations in every iteration of the compression function and, much more, the used quasigroups are functions of the processed message block.

AMS Subj. Classification: Primary 20*N*05, Secondary 94*A*60

Key Words: cryptographic hash function, NaSHA, quasigroup transformations

1. Introduction

This thesis is the final result of four years of research done in the Institute of Informatics at "Ss Cyril and Methodius" University in Republic of Macedonia. In the following sequel we give the motivation and our research goals.

The most known constructions of the cryptographic primitives, error detecting and error correcting codes use structures from the associative algebra as groups, rings and fields. Two eminent specialists on quasigroups, J. Dénes and A. D. Keedwell [2], once proclaimed the advent of a new era in cryptology, consisting in the application of non-associative algebraic systems as quasigroups and neo-fields. The quasigroups and their combinatorial equivalent Latin squares are very suitable for this aim, because of their structure, their features, their big number and because they lead to particular simple and yet efficient primitives. Nevertheless, at present, very few researchers use these tools and cryptographic community still hesitate about them.

On October 9, 2007 NIST announced the request for candidate algorithm nominations for a new cryptographic SHA-3 hash algorithm family. The reason were Wang′s differential attacks [4, 3] on SHA-1 from 2005. The last standard -

SHA-2 hash functions are in the same general family of hash functions as SHA-1. They could potentially be attacked with similar techniques, but they are much stronger than SHA-1.

With this thesis we wanted to justify deployment of quasigroups in cryptography, especially with new quasigroup based cryptographic hash function as a runner in the NIST SHA-3 competition. Several questions were raised, as: (1) What kind of quasigroups are suitable for cryptographic purposes? (2) How to generate and how to compute fast operation of huge quasigroups? (3) What kind of features have huge quasigroups obtained by new construction method? (4) Do some old or new quasigroup transformations exist that can use quasigroups obtained by new method? (5) Design of cryptographic primitives with quasigroup transformations.

In the following sections we will present, without proofs, some of the main results of this thesis.

2. How to choose a quasigroup?

In a quasigroup based cryptography you can find that different authors are seeking quasigroups with different properties. One needs $CI-$quasigroups, the other needs multivariate quadratic quasigroups, the third needs quasigroups with less possible structure, the fourth needs exponential quasigroups, the fifth needs orthogonal quasigroups etc. Some cryptographic primitives need special kind of quasigroups. There are special cryptosystems build on some particular subsets of quasigroups. Our interest is to find what properties should have a quasigroup in order to be used as a non-linear building block in cryptographic primitives and to be able to contribute to the defence against linear and differential attacks. When we try to find quasigroups suitable for cryptography in this sense, we started from shapeless quasigroups, defined by Gligoroski et al. [6].

Definition 1. [6] *A quasigroup $(Q,*)$ of order r is said to be* **shapeless** *iff it is non-idempotent, non-commutative, non-associative, it does not have neither left nor right unit, it does not contain proper sub-quasigroups, and there is no $k < 2r$ for which identities of the kinds are satisfied:*

$$\underbrace{x(...*(x}_{k}*y)) = y,\ y = ((y*\underbrace{x)*...)*x}_{k} \tag{2.1}$$

Shapeless quasigroups are a good choice, but sometimes even a quasigroup with some structure is preferable (when the structure does not affect the security). In other cases quasigroups with additional restriction to the structure may be needed, for example, not to be either semisymmetric or Stein quasigroup or Schroeder quasigroup, etc. Most often quasigroups are used for creating quasigroup transformations, and for them, usually it is enough a quasigroup

to be shapeless. Some quasigroup transformations, like $\mathcal{A}$ and $\mathcal{RA}$, even defined by linear quasigroups [5], can produce non-linear Boolean functions [11]. Some quasigroup transformations, like E transformation, preserve linearity of the used quasigroup [11]. At the end, it is important the quasigroup string transformations to be non-linear Boolean functions without any linear component Boolean function, without nontrivial difference propagations with prop ratio 1 and restriction weight of 0 and with every nonzero output selection vector correlated to more than one input selection vector.

3. **Fast generation of huge quasigroup operations**

We introduced the so called extended Feistel networks (which are Feistel networks with additional properties) as orthomorphisms to define huge quasigroups [8]. A Feistel network [15] takes any function and transforms it into a bijection, so it is a commonly used technique for creating a non-linear cryptographic function. Using a Feistel network for creating a huge quasigroup is not a novel approach. Kristen [14] presents several different constructions using one or two Feistel networks and isotopies of quasigroups. Complete mappings, introduced by Mann [13] (the equivalent concept of orthomorphism was introduced explicitly in [12]), are also useful for creation of huge quasigroups. In [14] complete mappings with non-affine functions represented by Cayley tables or with affine functions represented by binary transformations, are used for that aim. The main disadvantages of the previously mentioned constructions are the lack of efficiency in one case and the lack of security in the other case. Namely, the Cayley table representations need a lot of memory, and also the affine functions do not have good cryptographic properties.

Our approach uses the extended Feistel networks as orthomorphisms, to generate huge quasigroups of order $R = 2^{s2^t}$. We only need to store small permutations of order 2^s, $s = 4, 8, 16$. We use the generalization of Sade's diagonal method [10] to the complete mappings and the orthomorphisms, given by the following Theorem. For the Abelian group $(\mathbb{Z}_2^n, \oplus_n)$ they are equivalent with Sade's diagonal method.

Theorem 1. *Let ϕ be a complete mapping of the admissible group $(G, +)$ and let θ be an orthomorphism associated to ϕ. Define an operations $\circ$ and $\bullet$ on G by:*

$$x \circ y = \phi(y - x) + y \tag{3.1}$$

$$x \bullet y = \theta(x - y) + y \tag{3.2}$$

where $x, y \in G$. Then $(G, \circ)$ and $(G, \bullet)$ are quasigroups.

Definition 2. *Let $(G,+)$ be an Abelian group, let $f : G \to G$ be a mapping and let $a, b, c \in G$ are constants. The* **extended Feistel network** *$F_{a,b,c} : G^2 \to G^2$ created by f is defined for every $l, r \in G$ by*

$$F_{a,b,c}(l,r) = (r + a, l + b + f(r + c)).$$

The extended Feistel network $F_{a,b,c}$ is a bijection with inverse

$$F^{-1}_{a,b,c}(l,r) = (r - b - f(l + c - a), l - a).$$

One of the main results of this thesis, that we will frequently use, is the following one.

Theorem 2. *Let $(G,+)$ be an Abelian group and $a, b, c \in G$. If $F_{a,b,c} : G^2 \to G^2$ is an extended Feistel network created by a bijection $f : G \to G$, then $F_{a,b,c}$ is an orthomorphism of the group $(G^2,+)$.*

In the sequel we will consider only extended Feistel networks of the Abelian groups $(\mathbb{Z}_2^n, \oplus_n)$.

Proposition 3.1. *Let $a, b, c \in \mathbb{Z}_2^k$ and let $F_{a,b,c} : \mathbb{Z}_2^{2k} \to \mathbb{Z}_2^{2k}$ be an extended Feistel network of the group $(\mathbb{Z}_2^{2k}, \oplus_{2k})$ created by a mapping $f : \mathbb{Z}_2^k \to \mathbb{Z}_2^k$. Then $F_{a,b,c}$ is affine iff f is affine.*

Proposition 3.2. *Let $f, g : \mathbb{Z}_2^k \to \mathbb{Z}_2^k$ be bijections, $a, b, c, a', b', c' \in \mathbb{Z}_2^k$ and let $F_{a,b,c}, F_{a',b',c'} : \mathbb{Z}_2^{2k} \to \mathbb{Z}_2^{2k}$ be extended Feistel networks of the group $(\mathbb{Z}_2^{2k}, \oplus_{2k})$, created by f and g respectfully. Then the composite function $F_{a,b,c} \circ F_{a',b',c'}$ is a complete mapping and orthomorphism on $\mathbb{Z}_2^{2k}$ too.*

Corollary 1. *If $F_{a,b,c}$ is an extended Feistel network of the group $(\mathbb{Z}_2^{2k}, \oplus_{2k})$ created by bijection f, then $F^2_{a,b,c}$ is a complete mapping and orthomorphism too.*

The following Theorem shows us that extended Feistel network $F_{a,b,c}$ has the same algebraic degree as its starting bijection f.

Theorem 3. *Let $f : \mathbb{Z}_2^k \to \mathbb{Z}_2^k$ be a bijection of algebraic degree $deg(f) \geq 1$ and let $F_{a,b,c} : \mathbb{Z}_2^{2k} \to \mathbb{Z}_2^{2k}$ be an extended Feistel network of the group $(\mathbb{Z}_2^{2k}, \oplus_{2k})$, created by f. Then $deg(F_{a,b,c}) = deg(f)$.*

Here is a method for definition of huge quasigroups from small bijection.

Algorithm (Creation of extended Feistel network of order 2^{2^r})
Step 1 Take a suitable non-affine bijection of desired algebraic degree $f : \mathbb{Z}_2^{2^t} \to \mathbb{Z}_2^{2^t}$ where $t < r$ is a small positive integer ($t = 2, 3, 4$). Let $f' = f$ and $k = t$.
Step 2 Create Extended Feistel network $F_{a,b,c} : \mathbb{Z}_2^{2^{k+1}} \to \mathbb{Z}_2^{2^{k+1}}$ for some $a, b, c \in \mathbb{Z}_2^{2^k}$ using f' as starting bijection. Let $f' = F_{a,b,c}$ and $k = k + 1$.

Step 3 If $k < r$, go to step 2 else output the f'.

In applications one needs effectively constructed quasigroups of order 2^{256}, 2^{512},2^{1024},.... A huge quasigroup of order 2^{2^r} can now be designed as it follows. Take a suitable non-affine bijection of desired algebraic degree $f : \mathbb{Z}_2{}^{2^t} \to \mathbb{Z}_2{}^{2^t}$, where $t < r$ is a small positive integer ($t = 2, 3, 4$). We use the previous algorithm and we obtain F as the output extended Feistel network of order 2^{2^r}. Define a quasigroup operation $\circ$ on the set $\mathbb{Z}_2{}^{2^r}$ by 3.2, i.e.,

$$x \circ y = F(x \oplus y) \oplus y, \text{ for every } x, y \in \mathbb{Z}_2{}^{2^r}.$$

Note that we need only $r - t$ iterations for getting F and a small amount of memory for storing the bijection f. Hence, the complexity of our algorithm for construction of quasigroups of order 2^{2^r} is $\mathcal{O}(\texttt{log}(\texttt{log}\ r))$.

Example 1. We use a starting bijection $f : \mathbb{Z}_2^4 \to \mathbb{Z}_2^4$. So, $t = 2$. We choose constants $(a^{(i)}, b^{(i)}, c^{(i)}) = (i, 0, 0) \in \mathbb{Z}_2{}^{2^{t+i}}$, $i = 1, 2, \ldots, 7$. Now we can construct the following orthomorphisms, where $l_i, r_i \in \mathbb{Z}_2^i$, $i = 4, 8, 16, \ldots$:

$F^{(1)}_{1,0,0} : \mathbb{Z}_2^8 \to \mathbb{Z}_2^8$ as $F^{(1)}_{1,0,0}(l_4, r_4) = ((r_4 \oplus_4 1), (l_4 \oplus_4 f(r_4)))$,

$F^{(2)}_{2,0,0} : \mathbb{Z}_2^{16} \to \mathbb{Z}_2^{16}$ as $F^{(2)}_{2,0,0}(l_8, r_8) = ((r_8 \oplus_8 2), (l_8 \oplus_8 F^{(1)}_{1,0,0}(r_8)))$,

$F^{(3)}_{3,0,0} : \mathbb{Z}_2^{32} \to \mathbb{Z}_2^{32}$ as $F^{(3)}_{3,0,0}(l_{16}, r_{16}) = ((r_{16} \oplus_{16} 3), (l_{16} \oplus_{16} F^{(2)}_{2,0,0}(r_{16})))$,

$F^{(4)}_{4,0,0} : \mathbb{Z}_2^{64} \to \mathbb{Z}_2^{64}$ as $F^{(4)}_{4,0,0}(l_{32}, r_{32}) = ((r_{32} \oplus_{32} 4), (l_{32} \oplus_{32} F^{(3)}_{3,0,0}(r_{32})))$,

$F^{(5)}_{5,0,0} : \mathbb{Z}_2^{128} \to \mathbb{Z}_2^{128}$ as $F^{(5)}_{5,0,0}(l_{64}, r_{64}) = ((r_{64} \oplus_{64} 5), (l_{64} \oplus_{64} F^{(4)}_{4,0,0}(r_{64})))$,

$F^{(6)}_{6,0,0} : \mathbb{Z}_2^{256} \to \mathbb{Z}_2^{256}$ as $F^{(6)}_{6,0,0}(l_{128}, r_{128}) = ((r_{128} \oplus_{128} 6), (l_{128} \oplus_{128} F^{(5)}_{5,0,0}(r_{128})))$,

$F^{(7)}_{7,0,0} : \mathbb{Z}_2^{512} \to \mathbb{Z}_2^{512}$ as $F^{(7)}_{7,0,0}(l_{256}, r_{256}) = ((r_{256} \oplus_{256} 7), (l_{256} \oplus_{256} F^{(6)}_{6,0,0}(r_{256})))$.

So we need $7 = 9 - 2$ iterations for getting $F^{(7)}_{7,0,0} : \mathbb{Z}_2^{512} \to \mathbb{Z}_2^{512}$.

Further on in this section we consider the algebraic properties of the quasigroups obtained by the above mentioned algorithm. For that aim we take a somewhat simplified situation when $f : \mathbb{Z}_2^k \to \mathbb{Z}_2^k$ is a bijection and $F_{a,b,c} : \mathbb{Z}_2^{2k} \to \mathbb{Z}_2^{2k}$ is an extended Feistel network created by f. We denote by $(Q, \circ)$ the quasigroup on the set $Q = \mathbb{Z}_2^{2k}$ derived by the orthomorphism $F_{a,b,c}$.

Proposition 3.3. *The quasigroup $(Q, \circ)$ is non-idempotent iff $f(c) \neq b$ or $a \neq 0$.*

Proposition 3.4. *The equality*

$$(x \circ y) \circ (y \circ x) = x \tag{3.3}$$

is an identity in $(Q, \circ)$, i.e. $(Q, \circ)$ is a Schroeder quasigroup.

Corollary 2. *The quasigroup $(Q, \circ)$ is non-commutative and, much more, no different elements of Q commutes.*

Proposition 3.5. *The quasigroup $(Q, \circ)$ has neither left nor right unit.*

Proposition 3.6. *If $a \neq 0$, or $f(c) \neq b$, or $\phi \circ F_{a,b,c}(x) \neq F_{a,b,c} \circ \phi(x)$ for some $x \neq 0 \in Q$, then the quasigroup $(Q, \circ)$ is non-associative.*

Proposition 3.7. *a) The identity*

$$y = (\underbrace{(y \circ x) \circ \dots) \circ x}_{l}$$

holds true in $(Q, \circ)$ iff $F^l_{a,b,c} = I$.

b) The identity

$$\underbrace{x \circ (\cdots \circ (x}_{l} \circ y)) = y$$

holds true in $(Q, \circ)$ iff $\phi^l = I$, where $\phi = I \oplus_{2k} F_{a,b,c}$.

Regarding the subquasigroups of the quasigroup $(Q, \circ)$, we notice the following property, where $< A >$ denotes the subquasigroup generated by the subset A of Q.

Proposition 3.8. $< 0 >=< \{\theta^i(0) |\ i = 1, 2, \dots\} >$.

Proposition 3.9. *The quasigroup $(Q, \bullet)$, created by an affine orthomorphism θ of a group $(\mathbb{Z}_2^n, \oplus_n)$ is totally anti-symmetric (TA-quasigroup).*

So, even affine extended Feistel network can find some application also, for example, for creating TA-quasigroups that can be used for the definition of the check digit systems, where the early typing errors have to be recognized.

4. Cryptographic hash function NaSHA

We use the quasigroup transformation $\mathcal{MT}$ for definition of a new family of hash functions NaSHA-(m, k, r) [7]. The parameters m, k and r denote the length of the output hash result (the message digest), the complexity of $\mathcal{MT}$ and the order 2^{2^r} of used quasigroup respectively, so k is a positive even integer and m and r are positive integers. We showed that, the transformation $\mathcal{MT} : Q^t \to Q^t$ can be considered as a one-way function when $Q = \mathbb{Z}_{2^n}$ is enough big.

NaSHA-(m,k,r) hash algorithm

Input: A positive even integer k and positive integers m and r such that $m > 2^r$, and an input message M.

Output: A hash value NaSHA-$(m,k,r)(M)$ of m bits.

1. Denote by n the smallest integer such that $m \leq 2^n$.
   (For example, n=8 for m=224 and n=9 for m=384.)
2. Pad the message M, so that the length of the padded message M' is a multiple of 2^{n+1}, $|M'| = 2^{n+1}N$ for some N.
   Separate M' in N 2^{n+1}-bit blocks, $M' = M_1||M_2||\ldots||M_N,\ |M_i| = 2^{n+1}$.
3. Initialize the initial value H_0, which is a 2^{n+1}-bit word.
4. The first message block M_1 and the initial value H_0 separate to $q = 2^{n-r+1}$ 2^r-bits words:
   $M_1 = S_1||S_3||S_5||\ldots||S_{2q-3}||S_{2q-1}$,
   $H_0 = S_2||S_4||S_6||\ldots||S_{2q-2}||S_{2q}$, ($|S_i| = 2^r$) and form the word
   $S^{(0)} = S_1||S_2||S_3||S_4||\ldots||S_{2q-3}||S_{2q-2}||S_{2q-1}||S_{2q}$.
5. Choose leaders l_i as functions that depend on $S_1, S_2, S_3, \ldots, S_{2q}$ and a suitable linear transformation $LinTr_{2^{n+2}}$.
6. Choose two quasigroups $(\{0,1\}^{2^r}, *_1)$ and $(\{0,1\}^{2^r}, *_2)$ (one for $\mathcal{A}$ and one for $\mathcal{RA}$ transformation) and compute the string of bits $S^{(N-1)}$ as follows:
   for $i = 1$ to $N-1$ do
   $A_1||A_2||A_3||\ldots||A_{2q} \leftarrow \mathcal{MT}(LinTr^{2q}_{2^{n+2}}(S^{(i-1)}))$
   $B_1||B_2||B_3||\ldots||B_{q-1}||B_q \leftarrow M_{i+1}$,
   $S^{(i)} := B_1||A_2||B_2||A_4||\ldots||B_{q-1}||A_{2q-2}||B_q||A_{2q}$,
   end
7. Choose two quasigroups $(\{0,1\}^{2^r}, *_1)$ and $(\{0,1\}^{2^r}, *_2)$ and compute $\mathcal{MT}(LinTr^{2q}_{2^{n+2}}(S^{(N-1)})) := A_1||A_2||A_3||\ldots||A_{2q}$. Then
   NaSHA-$(m,k,r)(M) = A_4||A_8||\ldots||A_{2q-4}||A_{2q}$ (`mod` 2^m).

We give a complete implementation of NaSHA-$(m,2,6)$ algorithm where $m \in \{224, 256, 384, 512\}$ in [7]. It supports internal state sizes of 1024 and 2048 bits, and arbitrary output sizes between 125 and 512 bits. The used quasigroups of order $2^{2^6} = 2^{64}$ are constructed by extended Feistel networks, because they allow to insert tunable parameters in their definition. We used that feature to obtain the novel design: different quasigroups for every application of component quasigroup transformations in every iteration of the compression function and, much more, the used quasigroups are functions of the processed message block. This implementation has been accepted as a 1^{st} Round candidate in the SHA-3 competition of The American National Institute of Standards and Technology, NIST, but did not pass to the 2^{nd} Round. Some improvements are given in [9]. We obtain performance of up to 23.06 cycles per byte on an Intel Core 2 Duo in 64-bit mode.

In the implementation, as a starting bijection $f : \mathbb{Z}_2^8 \to \mathbb{Z}_2^8$ we use an improved AES S-box with the APA structure from Cui and Cao [1]. From the starting bijection f we define three extended Feistel networks F_{a_1,b_1,c_1}, F_{a_2,b_2,c_2},

$F_{a_3,b_3,c_3} : \mathbb{Z}_2^{16} \to \mathbb{Z}_2^{16}$ by

$$F_{a_i,b_i,c_i}(l_8||r_8) = (r_8 \oplus a_i)||(l_8 \oplus b_i \oplus f(r_8 \oplus c_i)),$$

where l_8 and r_8 are 8-bit variables, and $a_i,\ b_i,\ c_i$ are 8-bit words that are defined before each application of $\mathcal{MT}$. Denote by f' the bijection $F_{a_1,b_1,c_1} \circ F_{a_2,b_2,c_2} \circ F_{a_3,b_3,c_3} : \mathbb{Z}_2^{16} \to \mathbb{Z}_2^{16}$.

By using the bijection f' we define a quasigroup operation on $\mathbb{Z}_2^{64}$ which is going to be used for the additive string transformation $\mathcal{A}$ as follows. Create the Feistel networks $F_{\alpha_1,\beta_1,\gamma_1} : \mathbb{Z}_2^{32} \to \mathbb{Z}_2^{32}$ and $F_{A_1,B_1,C_1} : \mathbb{Z}_2^{64} \to \mathbb{Z}_2^{64}$ by

$$F_{\alpha_1,\beta_1,\gamma_1}(l_{16}||r_{16}) = (r_{16} \oplus \alpha_1)||(l_{16} \oplus \beta_1 \oplus f'(r_{16} \oplus \gamma_1)),$$

$$F_{A_1,B_1,C_1}(l_{32}||r_{32}) = (r_{32} \oplus A_1)||(l_{32} \oplus B_1 \oplus F_{\alpha_1,\beta_1,\gamma_1}(r_{32} \oplus C_1)),$$

where l_{16}, r_{16} are 16-bit variables, $\alpha_1, \beta_1, \gamma_1$ are 16-bit words, l_{32}, r_{32} are 32-bit variables and A_1, B_1, C_1 are 32-bit words. The constant words will be defined latter. The function F_{A_1,B_1,C_1} is an orthomorphism (complete mapping) in the group $(\mathbb{Z}_2^{64}, \oplus)$, and then the operation defined by

$$x *_{a_1,b_1,c_1,a_2,b_2,c_2,a_3,b_3,c_3,\alpha_1,\beta_1,\gamma_1,A_1,B_1,C_1} y = F_{A_1,B_1,C_1}(x \oplus y) \oplus y$$

is a quasigroup operation in $\mathbb{Z}_2^{64}$.

By using the bijection f' we define also a quasigroup operation in $\mathbb{Z}_2^{64}$ which is going to be used for the reverse additive string transformation $\mathcal{RA}$ as follows. Create the Feistel networks $F_{\alpha_2,\beta_2,\gamma_2} : \mathbb{Z}_2^{32} \to \mathbb{Z}_2^{32}$ and $F_{A_2,B_2,C_2} : \mathbb{Z}_2^{64} \to \mathbb{Z}_2^{64}$ by

$$F_{\alpha_2,\beta_2,\gamma_2}(l_{16}||r_{16}) = (r_{16} \oplus \alpha_2)||(l_{16} \oplus \beta_2 \oplus f'(r_{16} \oplus \gamma_2)),$$

$$F_{A_2,B_2,C_2}(l_{32}||r_{32}) = (r_{32} \oplus A_2)||(l_{32} \oplus B_2 \oplus F_{\alpha_2,\beta_2,\gamma_2}(r_{32} \oplus C_2)),$$

where l_{16}, r_{16} are 16-bit variables, $\alpha_2, \beta_2, \gamma_2$ are 16-bit words, l_{32}, r_{32} are 32-bit variables and A_2, B_2, C_2 are 32-bit words. The constant words will be defined later. The function F_{A_2,B_2,C_2} is an orthomorphism (complete mapping) in the group $(\mathbb{Z}_2^{64}, \oplus)$, and then the operation defined by

$$x *_{a_1,b_1,c_1,a_2,b_2,c_2,a_3,b_3,c_3,\alpha_2,\beta_2,\gamma_2,A_2,B_2,C_2} y = F_{A_2,B_2,C_2}(x \oplus y) \oplus y$$

is a quasigroup operation in $\mathbb{Z}_2^{64}$.

Before every computation $\mathcal{MT}(S_1||S_2||S_3||\dots||S_{2q-1}||S_{2q})$, where S_i are 64-bit words, we define the 64-bit leaders l_1 of $\mathcal{RA}$ and l_2 of $\mathcal{A}$, the 8-bit words $a_1, b_1, c_1,\ a_2, b_2, c_2,\ a_3, b_3, c_3$, the 16-bit words $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2,\ \gamma_2$ and the 32-bit words $A_1, B_1,\ C_1,\ A_2, B_2, C_2$.

For $m = 224$ and 256, necessary definitions are:

$$l_1 = S_1 + S_2, \quad l_2 = S_3 + S_4,$$

$$a_1||b_1||c_1||a_2||b_2||c_2||a_3||b_3 = S_5 + S_6, \quad c_3 = a_1$$

$$\alpha_1||\beta_1||\gamma_1||\alpha_2 = S_7 + S_8, \quad \beta_2||\gamma_2 = (S_9 + S_{10})(\mathtt{mod}\ 2^{32}),$$

$$A_1||B_1 = S_{11} + S_{12}, \quad C_1||A_2 = S_{13} + S_{14}, \quad B_2||C_2 = S_{15} + S_{16}.$$

For $m = 384$ and 512, necessary definitions are (this is improved version, suggested in [9]):

$$l_1 = S_1 + S_2 + S_{28} + S_{30}, \quad l_2 = S_3 + S_4 + S_{29} + S_{31},$$
$$a_1||b_1||c_1||a_2||b_2||c_2||a_3||b_3 = S_5 + S_6 + S_{17} + S_{18}, \quad c_3 = a_1$$
$$\alpha_1||\beta_1||\gamma_1||\alpha_2 = S_7 + S_8 + S_{19} + S_{20},$$
$$\beta_2||\gamma_2 = (S_9 + S_{10} + S_{21} + S_{22})(\mathtt{mod}\ 2^{32}),$$
$$A_1||B_1 = S_{11} + S_{12} + S_{23} + S_{27}, \quad C_1||A_2 = S_{13} + S_{14} + S_{24} + S_{26},$$
$$B_2||C_2 = S_{15} + S_{16} + S_{25} + S_{32}.$$

Here, the addition + is modulo 2^{64}.

The linear transformations are given as follows. Denote by $LinTr_{512}$ and by $LinTr_{256}$ the transformations of the sets $\{0,1\}^{2028}$ and $\{0,1\}^{1024}$ respectively, defined by

$$LinTr_{512}(S_1||S_2||\dots||S_{31}||S_{32}) = (S_7 \oplus S_{15} \oplus S_{25} \oplus S_{32})||S_1||S_2||\dots||S_{31},$$
$$LinTr_{256}(S_1||S_2||\dots||S_{15}||S_{16}) = (S_4 \oplus S_7 \oplus S_{10} \oplus S_{16})||S_1||S_2||\dots||S_{15},$$

where S_i are 64-bits words, $\oplus$ denotes the operation XOR on 64-bits words, and the operation $||$ denotes the concatenation of words.

For more information about NaSHA, see [7].

5. Future work

In the future we plan an examination of quasigroups produced by Extended Feistel networks of other Abelian groups. Also we plan to use them in the design of quasigroup based block cipher.

References

[1] L. Cui, Y. Cao, A new S-box structure named Affine-Power-Affine, *International Journal of Innovative Computing, Information and Control,* **3**(3), 2007, 751–759

[2] J. Dénes, D. Keedwell, Some applications of non-associative algebraic systems in cryptology, *Pure Mathematics and Applications*, **12**(2), 2001, 147–195

[3] X. Wang, Y. L. Yin, H. Yu, Finding Collisions in the Full SHA-1, *Advances in Cryptology - CRYPTO 2005, LNCS 3621*, 2005, 17–36

[4] X. Wang, H. Yu, Y. L. Yin, Efficient Collision Search Attacks on SHA-0, *Advances in Cryptology - CRYPTO 2005, LNCS 3621*, 2005, 1–16

[5] D. Gligoroski, V. Dimitrova, S. Markovski, *Quasigroups as Boolean Functions, Their Equation Systems and Gröbner Bases, Gröbner Bases, Coding, and Cryptography*, Springer 2009, 415–420

[6] D. Gligoroski, S. Markovski, L. Kocarev, Edon-$\mathcal{R}$, an Infinite Family of Cryptographic Hash Functions, *The Second NIST Cryptographic Hash Workshop, UCSB, Santa Barbara, CA,* 2006, 275–285
[7] S. Markovski, A. Mileva, *NaSHA*, Submission to NIST, 2008
[8] S. Markovski, A. Mileva, Generating huge quasigroups from small non-linear bijections via extended Feistel function, *Quasigroups and Related Systems*, **17**, 2009, 91–106
[9] S. Markovski, A. Mileva, NaSHA - cryptographic hash functions, *NIST The First SHA-3 Candidate Conference, 25-28 February 2009, Leuven, Belgium*
[10] A. Sade, Groupoides automorphes par le groupe cyclique, *Canadian Journal of Mathematics*, **9**(3), 1957, 321–335
[11] A. Mileva, Analysis of some Quasigroup transformations as Boolean Functions, *MASSEE International Congress on Mathematics MICOM 2009, 16-20 September, Ohrid*
[12] D. M. Johnson, A. L. Dulmage, N. S. Mendelsohn, Orthomorphisms of groups and orthogonal latin squares I, *Canad. J. Math.*, **13**, 1961, 356–372
[13] H. B. Mann, The construction of orthogonal Latin squares, *The Annals of Mathematical Statistics*,**13**, 1942, 418–423
[14] K. A. Meyer, *A new message authentication code based on the non-associativity of quasigroups*, PhD thesis, Iowa State University, 2006
[15] H. Feistel, Cryptography and computer privacy, *Scientific American*, **228** (No. 5), 1973, 15–23

Faculty of Computer Science
and Information Technology,
University "Goce Delcev",
Štip, Republic of Macedonia

E-Mail: aleksandra.mileva@ugd.edu.mk

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

On Li's Coefficients for Some Classes of L-Functions

Almasa Odžak

We study the generalized Li coefficients associated with the class $\mathcal{S}^{\sharp\flat}$ of functions containing the Selberg class and (unconditionally) the class of all automorphic L-functions attached to irreducible unitary cuspidal representations of $GL_N(\mathbb{Q})$ and the class of L-functions attached to the Rankin-Selberg convolution of two unitary cuspidal automorphic representations π and π' of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$. We deduce a full asymptotic expansion of the Archimedean contribution to these coefficients and investigate the contribution of the non-archimedean term. Obtained results are applied to automorphic L-functions. Also, a bound towards a generalized Ramanujan conjecture for the Archimedean Langlands parameters $\mu_\pi(v, j)$ of π is derived.

AMS Subj. Classification: 11M41, 11M26, 11S40

Key Words: Li's coefficients, Selberg class, Rankin-Selberg L-functions, Generalized Ramanujan conjecture, Generalized Riemann hypothesis

1. Introduction

The Riemann hypothesis (RH), formulated by B. Riemannn in 1859 is one of the most important conjectures in mathematics. It states that all non-trivial zeros of the Riemann zeta function are on the critical line Re $s = 1/2$ and has been resisting all attempts to prove it.

Various arithmetical, geometrical and algebraic objects can be described by the so-called global L-functions, which are formally similar to the Riemann zeta-function. They can be associated to elliptic curves, number fields, Maass forms, Dirichlet characters, ... The hypothesis that all non-trivial zeros of global L-functions are on the line Re $s = 1/2$ is nowadays called Generalized Riemann hypothesis (GRH). While the global L-functions are seemingly independent of each other, they have similar analytic properties and they are also assumed to satisfy GRH. This was a motivation for mathematicians to try to understand, or at least classify the class of all objects for which GRH holds true.

The Langlands program is an attempt to understand all global L-functions and to relate them to automorphic forms. Common properties of functions which conjecturally satisfy GRH are appointed. A. Selberg [16] has given a set of precise axioms which are believed to characterize the L-functions for which GRH

holds. Elements of the Selberg class are Dirichlet series with an Euler product representation, meromorphic continuation and a functional equation of the right shape. Although the exact nature of the class is conjectural, it is assumed that it is possible to classify its elements and give us an insight into their relationship to automorphic forms and the GRH.

Consequences of RH or GRH are various and important. They include many propositions which are known to be true under these hypotheses and some statements equivalent to the RH or GRH. Statements equivalent to RH give us an opportunity to restate the Riemann hypothesis in a different language, even in an entirely different disciplines, so we gain more possible tools for proving it. There are three main categories of statements equivalent to RH: purely number-theoretical statements, statements closely related to the analytic properties of the zeta function and cross-disciplinary statements.

One of the statements equivalent to RH, closely related to analytic properties of zeta function is the Li positivity criterion, proved by X.-J. Li in 1997 [12], stating that RH is equivalent to the non-negativity of the set of coefficients

$$\lambda_n = \sum_{\rho}^{*} \left(1 - \left(1 - \frac{1}{\rho}\right)^n\right). \tag{1}$$

Here, the sum runs over the non-trivial zeros of the Riemann zeta function, counted with multiplicities, and * indicates that the sum is taken in the sense of the limit as $|\mathrm{Im}\rho| \to \infty$.

Recently it was proved that it is actually sufficient to probe the Li coefficients for their large n behavior. Namely, A. Voros [22] has proved that the Riemann hypothesis true is equivalent to the tempered growth of λ_n (as $\frac{1}{2}n\log n$), determined by its archimedean part, while the Riemann hypothesis false is equivalent to the oscillations of λ_n with exponentially growing amplitude.

The Li criterion motivated many numerical calculations concerning RH. The generalized Li coefficients attached to global L functions and the generalized Li criterion as well may serve as a motivation for numerical calculations concerning GRH, possibly producing some new ideas for proving it.

E. Bombieri and J. C. Lagarias [3] have noticed that the Li criterion can be formulated for a general multiset of complex numbers and obtained an arithmetic expression of the Li coefficient λ_n in the form

$$\lambda_n = S_\infty(n) - S_{NA}(n) + 1, \tag{2}$$

where S_∞ denotes a contribution from archimedean places (Γ-factors) and S_{NA} is a contribution of finite (non-archimedean) places.

J. C. Lagarias [11] has defined the generalized Li coefficient attached to an irreducible cuspidal unitary automorphic representation π of $GL_m(\mathbb{Q})$ and proved the generalized Li criterion in this case. He obtained an arithmetic expression for these coefficients completely analogous to (2) and determined the

asymptotic behavior of both the Archimedean and the finite part of generalized Li coefficients in this case.

F. C. S. Brown [4] has determined zero-free regions of Dirichlet and Artin L-functions (under the Artin hypothesis) in terms of sizes of the corresponding generalized Li coefficients.

Our main objective is to formulate and prove generalization of the Li criterion for some classes of L functions, to find the arithmetic expression and investigate the asymptotic behavior of generalized Li coefficients. Also, we shall analyze consequences of the obtained results under GRH. Classes which will be treated are the Selberg class and the class of L-functions attached to the Rankin-Selberg convolution of two unitary cuspidal automorphic representations π and π' of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$.

2. **Preliminaries**

2.1. **The Selberg class of functions.** The Selberg class of functions $\mathcal{S}$, introduced by A. Selberg in [16], is a general class of Dirichlet series F satisfying the following properties:

(i) (Dirichlet series) F posses a Dirichlet series representation

$$F(s) = \sum_{n=1}^{\infty} \frac{a_F(n)}{n^s},$$

that converges absolutely for $\mathrm{Re}s > 1$.

(ii) (Analytic continuation) There exists an integer $m \geq 0$ such that $(s-1)^m F(s)$ is an entire function of finite order. The smallest such number is denoted by m_F and called a polar order of F.

(iii) (Functional equation) The function F satisfies the functional equation $\Phi_F(s) = w\overline{\Phi_F(1-\bar{s})}$, where $\Phi_F(s) = F(s)Q_F^s \prod_{j=1}^{r} \Gamma(\lambda_j s+\mu_j)$, with $Q_F > 0$, $r \geq 0$, $\lambda_j > 0$, $|w| = 1$, $\mathrm{Re}\mu_j \geq 0$, $j = 1, \ldots, r$. Though the numbers $\lambda_1, \ldots, \lambda_r$ are not unique, it can be shown (see, e.g. [15]) that the number $d_F = 2\sum_{j=1}^{r} \lambda_j$ is an invariant, called the degree of F. Furthermore, the number $\xi_F = 2\sum_{j=1}^{r}(\mu_j - 1/2)$ is also an invariant (see [15] , p. 43]) called the ξ invariant.

(iv) (Ramanujan conjecture) For every $\epsilon > 0$ $a_F(n) \ll n^{\epsilon}$.

(v) (Euler product)

$$\log F(s) = \sum_{n=1}^{\infty} \frac{b_F(n)}{n^s},$$

where $b_F(n) = 0$, for all $n \neq p^m$ with $m \geq 1$ and p prime, and $b_F(n) \ll n^\theta$, for some $\theta < \frac{1}{2}$.

An extended Selberg class $\mathcal{S}^\sharp$ is a class of functions satisfying conditions (i), (ii) and (iii).

It is conjectured that the Selberg class coincides with the class of all automorphic L-functions. Some properties such as Ramanujan conjecture, boundedness of coefficients in the Dirichlet series representation of $\log L(s,\pi)$ and the bound $\mathrm{Re}\mu_j \geq 0$ on the archimedean Langlands parameters have not yet been proved.

In order to apply our results unconditionally to automorphic L-functions attached to irreducible unitary automorphic representations of $GL_N(\mathbb{Q})$, we shall focus on the class $\mathcal{S}^{\sharp\flat}$ of functions satisfying axioms (i), (ii) and the following two axioms:

(iii') (Functional equation) The function F satisfies the functional equation $\Phi_F(s) = w\overline{\Phi_F(1-\bar{s})}$, where $\Phi_F(s) = F(s)Q_F^s \prod_{j=1}^{r} \Gamma(\lambda_j s+\mu_j)$, with $Q_F > 0$, $r \geq 0$, $\lambda_j > 0$, $|w| = 1$, $\mathrm{Re}\mu_j > -\frac{1}{4}$, $\mathrm{Re}(\lambda_j + 2\mu_j) > 0$, $j = 1, \ldots, r$.
Let us note that (iii') implies that $\mathrm{Re}(\lambda_j + \mu_j) > 0$.

(v') The logarithmic derivative of the function F possesses a Dirichlet series representation

$$\frac{F'}{F}(s) = -\sum_{n=1}^{\infty} \frac{c_F(n)}{n^s},$$

converging absolutely for $\mathrm{Re}s > 1$.

It can be shown that introduced class $\mathcal{S}^{\sharp\flat}$ contains Selberg class.

2.2. **Rankin-Selberg L-functions.** The Rankin-Selberg L-function attached to the product $\pi \times \widetilde{\pi}'$ of two unitary cuspidal automorphic representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ is given, for $\mathrm{Re}s > 1$, by an absolutely convergent Euler product of local factors

$$L\left(s, \pi_f \times \widetilde{\pi}'_f\right) = \prod_{\mathfrak{p}<\infty} L(s, \pi_\mathfrak{p} \times \widetilde{\pi}'_\mathfrak{p}), \tag{3}$$

as proved in [8, Th. 5.3.]. Here, $\widetilde{\pi}$ denotes the contragradient representation of π. For any place v of F, $\widetilde{\pi}_v$ is equivalent to the complex conjugate $\overline{\pi}_v$ [6], hence $L\left(s, \pi_f \times \widetilde{\pi}'_f\right) = \overline{L\left(\overline{s}, \widetilde{\pi}_f \times \pi'_f\right)}$.

Let us put

$$L\left(s, \pi_\infty \times \widetilde{\pi}'_\infty\right) = \prod_{v \in S_\infty} L(s, \pi_v \times \widetilde{\pi}'_v),$$

where S_∞ denotes the set of all infinite places of F. Then, as proved in [7, 8, 9, 17, 18, 19, 20] (see also [5, Th. 9.1. and Th. 9.2.], the complete Rankin-Selberg L-function

$$\Lambda(s, \pi \times \widetilde{\pi}') = L(s, \pi_f \times \widetilde{\pi}'_f) L(s, \pi_\infty \times \widetilde{\pi}'_\infty)$$

extends to a meromorphic function of order 1 on the whole complex plane, bounded (away from its possible poles) in vertical strips. It has simple poles at $s = 1 + it_0$ and $s = it_0$, arising form $L(s, \pi_f \times \widetilde{\pi}'_f)$ if and only if $m = m'$ and $\pi' \cong \pi \otimes |\det|^{it_0}$, for some $t_0 \in \mathbb{R}$. Otherwise, it is a holomorphic function. Finally, $\Lambda(s, \pi \times \widetilde{\pi}')$ satisfies the functional equation

$$\Lambda(s, \pi \times \widetilde{\pi}') = \epsilon\left(\pi \times \widetilde{\pi}'\right) Q\left(\pi \times \widetilde{\pi}'\right)^{\frac{1}{2}-s} \Lambda(1-s, \widetilde{\pi} \times \pi'), \tag{4}$$

where $Q(\pi \times \widetilde{\pi}') > 0$ is the arithmetic conductor and $\epsilon(\pi \times \widetilde{\pi}')$ is a complex number of modulus 1.

3. **Results**

The arithmetic formulas for generalized Li coefficients are obtained using explicit formulas, proved in [21] for the class $\mathcal{S}^{\sharp\flat}$ and in [14] for the Rankin-Selberg L-function. Proofs are based on results of Jorgenson and Lang [10] on explicit formulas in the fundamental class of functions and results on expanding their class of test functions to which the explicit formula applies, obtained in papers [1] and [2].

3.1. **Generalized Li coefficients and Li criterion.** The generalized Li coefficient attached to $F \in \mathcal{S}^{\sharp\flat}$ can be defined analogously as (1), by

$$\lambda_F(n) = {\sum_{\rho \in Z(F)}}^{*} \left(1 - \left(1 - \frac{1}{\rho}\right)^n\right), \tag{5}$$

where $Z(F)$ denotes set of all nontrivial zeros of the function F.

The *-convergence of the series (5) was proved in [21] using the explicit formula with suitably chosen test function.

Proposition 3.1 ([21] Li coefficient for the class $\mathcal{S}^{\sharp\flat}$)**.** *Let $F \in \mathcal{S}^{\sharp\flat}$ such that $0 \notin Z(F)$. Then, the series (5) is *-convergent for every integer n. Moreover, the series $\mathrm{Re}\lambda_F(n) = {\sum_{\rho \in Z(F)}}^{*} \mathrm{Re}\left(1 - \left(1 - \frac{1}{\rho}\right)^n\right)$ converges absolutely for all integers n.*

The Generalized Li criterion for the class $\mathcal{S}^{\sharp\flat}$ was also proved in [21].

Proposition 3.2 ([21] Li criterion for the class $\mathcal{S}^{\sharp\flat}$)**.** *Let $F \in \mathcal{S}^{\sharp\flat}$ such that $0 \notin Z(F)$. Then, all non-trivial zeros of F lie on the line $\mathrm{Re}s = \frac{1}{2}$ if and only if $\mathrm{Re}\lambda_F(n) \geq 0$, for all $n \in \mathbb{N}$.*

The definition of the generalized Li coefficient attached to the product $\pi \times \widetilde{\pi}'$ is given by

$$\lambda_{\pi,\pi'}(n) = \sum_{\rho \in Z(L)}{}^{*} \left(1 - \left(1 - \frac{1}{\rho}\right)^n\right). \tag{6}$$

The existence of coefficients (6) and the generalized Li criterion in this setting are proved in the following propositions.

Proposition 3.3 ([14] Li coefficient for the Rankin-Selberg L-function)**.** *The generalized Li coefficients $\lambda_{\pi,\pi'}(n)$ are well defined for every integer n.*

Proposition 3.4 ([14] Li criterion for the Rankin-Selberg L-function)**.** *All non-trivial zeros of $L(s, \pi_f \times \widetilde{\pi}'_f)$ lie on the line* $\mathrm{Re}s = \frac{1}{2}$ *if and only if* $\mathrm{Re}\lambda_{\pi,\pi'}(n) \geq 0$, *for all $n \in \mathbb{N}$.*

3.2. **An arithmetic formula for the Li coefficients.** The arithmetic formulas for the generalized Li coefficients are given in the next two theorems.

Theorem 3.5 ([21] Arithmetic formula for the class $\mathcal{S}^{\sharp\flat}$)**.** *Let $F \in \mathcal{S}^{\sharp\flat}$ be a function such that $0 \notin Z(F)$. Then, for all $n \in \mathbb{N}$*

$$\lambda_F(-n) = m_F + n \log Q_F + \sum_{l=1}^{n} \binom{n}{l} \gamma_F(l-1) + \sum_{l=1}^{n} \binom{n}{l} \eta_F(l-1)$$

where

$$\eta_F(0) = \sum_{j=1}^{r} \lambda_j \frac{\Gamma'}{\Gamma}(\lambda_j + \mu_j) \quad \textit{and} \quad \eta_F(l-1) = \sum_{j=1}^{r} (-\lambda_j)^l \sum_{k=0}^{\infty} \frac{1}{(\lambda_j + \mu_j + k)^l},$$

for $l \geq 2$.

Theorem 3.6 ([14] Arithmetic formula for the Rankin-Selberg L-function)**.** *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$, respectively. Then, for all $n \in \mathbb{N}$ and $t_0 \in \mathbb{R}\backslash\{0\}$*

$$\begin{aligned} \lambda_{\pi,\pi'}(-n) &= \sum_{j=1}^{n} \binom{n}{j} \gamma_{\pi,\pi'}(j-1) + \delta_{\pi,\pi'}(0) \\ &\quad + \frac{n}{2}(\log Q_{\pi \times \widetilde{\pi}'} - dmm' \log \pi) \\ &\quad + \delta_{\pi,\pi'}(t_0) \left(1 - \left(1 + \frac{1}{it_0}\right)^n + 1 - \left(1 - \frac{1}{1 - it_0}\right)^n\right) \\ &\quad - dmm' + \sum_{l=1}^{dmm'} \left(\frac{\mu_{\pi \times \widetilde{\pi}'}(l)}{1 + \mu_{\pi \times \widetilde{\pi}'}(l)}\right)^n + \sum_{j=1}^{n} \binom{n}{j} \eta_{\pi,\pi'}(j-1) \end{aligned} \tag{7}$$

where

$$\eta_{\pi,\pi'}(0) = \frac{1}{2}\sum_{l=1}^{dmm'} \frac{\Gamma'}{\Gamma}\left(\frac{3+\mu_{\pi\times\widetilde{\pi}'}(l)}{2}\right),$$

$$\eta_{\pi,\pi'}(j-1) = \frac{(-1)^j}{2^j}\sum_{l=1}^{dmm'}\sum_{t=0}^{\infty}\frac{1}{\left(t+\frac{3+\mu_{\pi\times\widetilde{\pi}'}(l)}{2}\right)^j}, \quad \text{for } j \geq 2$$

and $\gamma_{\pi,\pi'}(k)$ are the coefficients in the Laurent (Taylor) series expansion of $\frac{L'}{L}$ at $s = 1$.

3.3. **Asymptotic behavior of the Li coefficients.** In order to investigate the asymptotic behavior of the Li coefficients we will treat the archimedean and non-Archimedean (finite) contribution to the nth Li coefficient separately. First, we shall write the Li coefficient (5) as

$$\lambda_F(-n) = S_\infty(n,F) + S_{NA}(n,F),$$

where

$$S_\infty(n,F) = m_F + n\log Q_F + \sum_{l=1}^{n}\binom{n}{l}\eta_F(l-1)$$

is the Archimedean contribution, while

$$S_{NA}(n,F) = \sum_{j=1}^{n}\binom{n}{l}\gamma_F(l-1)$$

is the finite (non-Archimedean) term.

The next theorem gives us the full asymptotic expansion of the Archimedean contribution to the nth Li coefficient in terms of $n\log n$, n, n^0 and odd negative powers of n.

Theorem 3.7 ([13] Archimedean contribution to the Li coefficient for $\mathcal{S}^\sharp$)**.** *Let $F \in \mathcal{S}^\sharp$ be a function non-vanishing at zero. Then, for an arbitrary $K \in \mathbb{N}$*

$$\begin{aligned} S_\infty(n,F) &= \frac{d_F}{2}n\log n + nC_F + \frac{\xi_F}{2} \\ &+ m_F + \frac{d_F}{4} - \frac{d_F}{2}\sum_{k=1}^{K}\frac{B_{2k}}{2k}n^{1-2k} + O_{F,K}(n^{-2K}), \end{aligned}$$

as $n \to \infty$, where $C_F = \log Q_F + \frac{d_F}{2}(\gamma-1) + \sum_{j=1}^{r}\lambda_j\log\lambda_j$ and B_{2k} are Bernoulli numbers.

Let

$$\lambda_F(n,T) = \sum_{\substack{\rho \\ |\mathrm{Im}\rho|<T}} \left(1-\left(1-\frac{1}{\rho}\right)^n\right).$$

denote an incomplete nth Li coefficient to the height T. The following result gives us the representation of the finite contribution to the nth Li coefficient in terms of the incomplete nth Li coefficient to the height $\sqrt{n}$, up to the error term $O(\sqrt{n}\log n)$.

Theorem 3.8 ([13] Non-Archimedean contribution to Li coefficient for the class $\mathcal{S}^{\sharp\flat}$)**.** *Let $F \in \mathcal{S}^{\sharp\flat}$ be a function non-vanishing at zero. Then,*

$$S_{NA}(n,F) = -\lambda_F(-n,\sqrt{n}) + O(\sqrt{n}\log n).$$

Combining the last two theorems we obtain an asymptotic behavior of the nth Li coefficient attached to the function $F \in \mathcal{S}^{\sharp\flat}$, as $n \to \infty$.

Corollary 3.9 ([13])**.** *Let $F \in \mathcal{S}^{\sharp\flat}$ be a function non-vanishing at zero. Then, for all $n \in \mathbb{N}$*

$$\lambda_F(-n) = \frac{d_F}{2} n\log n + nC_F - \lambda_F(-n,\sqrt{n}) + O(\sqrt{n}\log n)$$

where $C_F = \log Q_F + \frac{d_F}{2}(\gamma-1) + \sum_{j=1}^{r} \lambda_j \log \lambda_j$ and γ is the Euler constant.

Since $\mathcal{S}^{\sharp\flat}$ contains all automorphic L-functions, attached to irreducible unitary automorphic representations of $GL_N(\mathbb{Q})$, we immediately obtain the following corollary.

Corollary 3.10 ([13])**.**

$$\begin{aligned} S_\infty(n,\pi) &= \frac{N}{2} n\log n + n\left(\frac{1}{2}\log Q(\pi) + \frac{N}{2}(\gamma - 1 - \log(2\pi))\right) \\ &+ \frac{1}{2}\sum_{j=1}^{N} \kappa_j(\pi) + \delta(\pi) - \frac{N}{4} - \frac{N}{2}\sum_{k=1}^{K} \frac{B_{2k}}{2k} n^{1-2k} + O_{\pi,K}(n^{-2K}), \end{aligned}$$

as $n \to \infty$, for an arbitrary, fixed $K \in \mathbb{N}$.

The above result improves the result of Lagarias [11], showing that $S_\infty(n,\pi)$ has a full asymptotic expansion in terms of n^{-k} (k-odd). Furthermore, it shows that only the constant term depends on the sum of Archimedean Langlands parameters κ_j and the terms with negative degrees depend singly on N.

An analogous decomposition of the generalized Li coefficient attached to the Rankin-Selberg L-function is somewhat more complicated as well as the results for the asymptotic behavior of the Archimedean term. Coefficient (6) can

be written as
$\lambda_{\pi,\pi'}(-n) = S_\infty(n,\pi,\pi') + S_{NA}(n,\pi,\pi')$,where

$$S_\infty(n,\pi,\pi') = \delta_{\pi,\pi'}(0) + \frac{n}{2}\left(\log Q_{\pi\times\pi'} - dmm'\log\pi\right)$$
$$-dmm' + \sum_{l=1}^{dmm'}\left(\frac{\mu_{\pi\times\widetilde{\pi}'}(l)}{1+\mu_{\pi\times\widetilde{\pi}'}(l)}\right)^n + \sum_{j=1}^{n}\binom{n}{j}\eta_{\pi,\pi'}(j-1)$$

is an Archimedean contribution and

$$S_{NA}(n,\pi,\pi') = \sum_{j=1}^{n}\binom{n}{j}\gamma_{\pi,\pi'}(j-1)$$
$$+\delta_{\pi,\pi'}(t_0)\left(1-\left(1+\frac{1}{it_0}\right)^n + 1 - \left(1-\frac{1}{1-it_0}\right)^n\right)$$

is a finite term. An incomplete nth Li coefficient to the height T is defined as

$$\lambda_{\pi,\pi'}(n,T) = \sum_{\substack{\rho \\ |\mathrm{Im}\rho|<T}}\left(1-\left(1-\frac{1}{\rho}\right)^n\right).$$

The main results in this case are the following two theorems and a corollary.

Theorem 3.11 ([14] Archimedean contribution - Rankin-Selberg L-function)**.** *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ respectively. Then, for an arbitrary $K \in \mathbb{N}$*

$$S_\infty(n,\pi,\pi') = \delta_{\pi,\pi'}(0) + \frac{n}{2}dmm'\log n + \frac{1}{2}\sum_{l=1}^{dmm'}\mu_{\pi\times\widetilde{\pi}'}(l)$$
$$+\frac{n}{2}\left(\log Q_{\pi\times\pi'} - dmm'\left(\log 2\pi - \gamma + 1\right)\right) + \sum_{l=1}^{dmm'}\left(\frac{\mu_{\pi\times\widetilde{\pi}'}(l)}{1+\mu_{\pi\times\widetilde{\pi}'}(l)}\right)^n$$
$$-\frac{1}{4}dmm' - \frac{dmm'}{2}\sum_{k=1}^{K}\frac{B_{2k}}{2k}n^{1-2k} + O_{\pi,\pi',K}(n^{-2K}),$$

as $n\to\infty$, where B_{2k} are the Bernoulli numbers.

Theorem 3.12 ([14] Non-Archimedean contribution - Rankin-Selberg L-function)**.** *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ respectively. Then,*

$$S_{NA}(n,\pi,\pi') = -\lambda_{\pi,\pi'}(-n,\sqrt{n}) + O(\sqrt{n}\log n).$$

Corollary 3.13 ([14]). *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ respectively. Then, for all $n \in \mathbb{N}$*

$$\begin{aligned}\lambda_{\pi,\pi'}(-n) &= \frac{dmm'n}{2}\log n + nC_{\pi,\pi'} - \lambda_{\pi,\pi'}(-n,\sqrt{n}) \\ &+ O(\sqrt{n}\log n) + \sum_{l=1}^{dmm'}\left(\frac{\mu_{\pi\times\widetilde{\pi}'}(l)}{1+\mu_{\pi\times\widetilde{\pi}'}(l)}\right)^n,\end{aligned}$$

where $C_{\pi,\pi'} = \frac{1}{2}\left(\log Q_{\pi\times\pi'} + dmm'\left(\gamma - 1 - \log 2\pi\right)\right)$.

3.4. **Consequences of the Generalized Riemann hypothesis.** The simplification of the corollaries 3.9 and 3.13 can be made in the case when GRH holds true. The results are as follows.

Corollary 3.14. *Let $F \in \mathcal{S}^{\sharp\flat}$ be a function non-vanishing at zero. If the generalized Riemann hypothesis holds for $F(s)$ then*

$$\lambda_F(n,\sqrt{n}) = O(\sqrt{n}\log n)$$

and

$$\lambda_F(-n) = \frac{d_F}{2}n\log n + nC_F + O(\sqrt{n}\log n).$$

Corollary 3.15. *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ respectively. Assume the GRH for $L\left(s,\pi_f\times\widetilde{\pi}'_f\right)$. Then,*

$$\begin{aligned}\lambda_{\pi,\pi'}(-n) = \frac{dmm'}{2}n\log n + \sum_{l=1}^{dmm'}\left(\frac{\mu_{\pi\times\widetilde{\pi}'}(l)}{1+\mu_{\pi\times\widetilde{\pi}'}(l)}\right)^n \\ + \frac{n}{2}\left(\log Q_{\pi\times\widetilde{\pi}'} - dmm'(1+\log 2\pi-\gamma)\right) + O_{\pi,\pi'}\left(\sqrt{n}\log n\right).\end{aligned}$$

An interesting consequence of the GRH for $L\left(s,\pi_f\times\widetilde{\pi}'_f\right)$ is a bound towards the Ramanujan conjecture for the Archimedean Langlands parameters. This result is obtained by comparison of the expressions for the generalized Li coefficients under GRH obtained in two different ways in [14].

Theorem 3.16 ([14]). *Let π and π' be two automorphic unitary cuspidal representations of $GL_m(\mathbb{A}_F)$ and $GL_{m'}(\mathbb{A}_F)$ respectively. Under the GRH for $L\left(s,\pi_f\times\widetilde{\pi}'_f\right)$ one has*

$$\mathrm{Re}\mu_{\pi\times\widetilde{\pi}'}(l) \geq -\frac{1}{2}, \text{ for all } l = 1,\ldots,dmm'.$$

Corollary 3.17 ([14]). *Let π be an automorphic unitary cuspidal representation of $GL_m(\mathbb{A}_F)$, unramified at the archimedean place $v \in S_\infty$. Then GRH for the function $L(s, \pi_f \times \widetilde{\pi}_f)$ implies the bound*

$$|\mathrm{Re}\mu_\pi(v, j)| \leq \frac{1}{4}$$

for all $j = 1, \ldots, m$.

References

[1] M. Avdispahić, L. Smajlović, Explicit formula for a fundamental class of functions. *Bull. Belg. Math. Soc. Simon Stevin* , **12**, 2005, 569-587.

[2] M. Avdispahić, L. Smajlović, A note on Weil's explicit formula. In: Khrennikov, A.Yu., Rakić, Z., Volovich, I.V. (eds.), *p-adic Mathematical Physics: 2^{nd} International Conference on p-adic Mathematical Physics*, American Institute of Physics, 2006, 312-319.

[3] E. Bombieri, J. C. Lagarias, Complements to Li's criterion for the Riemann hypothesis. *J. Number Theory*, **77**, 1999, 274-287 .

[4] F. C. S. Brown, Li's criterion and zero-free regions of L-functions. *J. Number Theory*, **111**, 2005, 1-32.

[5] J. W. Cogdell, *Lectures on L-Functions, Converse Theorems and Functoriality for GL_n,* Fields Institute Lectures, Spring, 2003, available at http://www.math.ohio-state.edu/˜cogdell.

[6] I. M. Gelfand, D. Kazhdan, Representation of the group $GL(n, K)$, where K is a local field. In: Gelfand, I.M. (ed.)*Lie groups and their representations*. Wiley, New York, 1974, 95-118.

[7] S. S. Gelbart, F. Shahidi, Boundedness of automorphic $L-$functions in vertical strips, *J. Amer. Math. Soc.*, **14**, 2001, 79–107.

[8] H. Jacquet, J. A. Shalika, On Euler products and the classification of automorphic representations, *I. Amer. J. Math.*, **103**, 1981, 499-558.

[9] H. Jacquet, J. A. Shalika, On Euler products and the classification of automorphic representations II, *Amer. J. Math.*, **103**, 1981, 777–815.

[10] J. Jorgenson, S. Lang, *Explicit formulas for regularized products and series.* Lecture Notes in Mathematics 1593, Springer-Verlag, Berlin-Heidelberg, 1994, 1-134.

[11] J. C. Lagarias, Li's coefficients for automorphic L-functions. *Ann. Inst. Fourier*, **57**, 2007, 1689-1740.

[12] X.-J. Li, The positivity of a sequence of numbers and the Riemann hypothesis. *J. Number Theory*, **65**, 1997, 325-333.

[13] A. Odžak, L. Smajlović, On Li's coefficients for the Selberg class, (submitted)

[14] A. Odžak, L. Smajlović, On Li's coeffcients for the Rankin-Selberg L-functions, *The Ramanujan J.* , **21**(3), 2010, 303-334.

[15] A. Perelli, A survey of the Selberg class of L-functions, part I, *Milan. J. Math*, **73**, 2005, 19–52.

[16] A. Selberg, Old and new conjectures and results about a class of Dirichlet series. In: *Proc. Amalfi Conf. Analytic Number Theory,* ed. by E. Bombieri et al., Universitia di Salerno, 1992, 367–385.

[17] F. Shahidi, On certain L-functions.*Amer. J. Math.*, **103**, 1981, 297-355.

[18] F. Shahidi, Fourier transforms of intertwinting operators and Plancherel measures for $GL(n)$. *Amer. J. Math.*, **106**, 1984, 67-111.

[19] F. Shahidi, Local coefficients as Artin factors for real grups. *Duke Math. J.*, **52**, 1985, 973-1007.

[20] F. Shahidi, A proof of Langlands' conjecture on Plancherel measures, Complementary series for p-adic groups. *Ann. Math.*, **132**, 1990, 273-330.

[21] L. Smajlović, On Li's criterion for the Riemann hypothesis for the Selberg class, *J. Number Theory*, **130**(4), 2010, 828-851.

[22] A. Voros, Sharpenings of Li's criterion for the Riemann hypothesis. *Math. Phys. Anal. and Geom.*, **9**, 2006, 53-63.

Department of Mathematics,
University of Sarajevo
Zmaja od Bosne 33-35,
Sarajevo, Bosnia and Herzegovina

E-Mail: almasa@pmf.unsa.ba

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Modeling Pulsatility in the Human Cardiovascular System

Aurelio de los Reyes V and Franz Kappel

In this study we investigated, modified and combined two existing mathematical cardiovascular models, a non-pulsatile global model and a simplified pulsatile left heart model. The first goal of the study was to integrate these models. The main objective is to have a global lumped parameter pulsatile model that predicts the pressures in the systemic and pulmonary circulation, and specifically the pulsatile pressures in the the finger arteries where real-time measurements can be obtained. Modifications were made in the ventricular elastance to model the stiffness of heart muscles under stress or exercise state. The systemic aorta compartment is added to the combined model. A finger artery compartment is included to reflect measurements of pulsatile pressures. Parameters were estimated to simulate an average normal blood pressures. Preliminary simulation results are presented.

AMS Subj. Classification: 92C30

Key Words: mathematical cardiovascular model, pulsatility, contractility, left ventricular elastance

Introduction

Cardiovascular modeling has a longstanding desire to understand the behavior of the blood pressures in the peripheral and systemic compartments, cardiac outputs, ventricular elastance and contractility in the human circulatory system. In Kappel and Peer (1993) [4] and Timischl (1998) [12], efforts have been done to model non-pulsatile blood flow simulating values of quantities taken over one heart beat respectively over one breath. These models are considered to be global in the sense of considering all essential subsystems such as systemic and pulmonary circulation, left and right ventricles, baroreceptor loop, etc. These are used to describe the overall reaction of the cardiovascular-respiratory system under a constant ergometric workload imposed on a test person on a bicycle-ergometer. The basic control autoregulatory mechanisms were constructed assuming that the regulation is optimal with respect to a cost criterion. The model was extended and used to describe the response of the

cardiovascular-respiratory system under orthostatic stress condition, see for example Fink et al. (2004) [5] and Kappel et al. (2007) [3]. In the study done by Olufsen et. al (2009) [9], a simple lumped parameter cardiovascular model was developed to analyze cerebral blood flow velocity and finger blood pressure measurements during orthostatic stress (sit-to-stand).

1. **Cardiovascular Modeling**

The cardiovascular model presented here is the combination of two existing cardiovascular models: a non-pulsatile global model adapted from the earlier work of Kappel and Peer (1993) [4] and a simplified pulsatile left heart model by Olufsen et al. (2009) [9]. The non-pulsatile global model is based on Grodin's mechanical part of the cardiovascular system. It incorporates all the essential subsystems such as systemic and pulmonary circulation, left and right ventricles, baroreceptor loop, etc. This model considered the mean values over one heart cycle instead of the instantaneous values. On the other hand, the pulsatile left heart model utilizes a minimal cardiovascular structure to close the circulatory loop. The model consists of two arterial compartments and two venous compartments combining vessels in the body and the brain, and a heart compartment representing the left ventricle.

The combined cardiovascular model includes arterial and venous pulmonary, left and right ventricles, systemic aorta, finger arteries, and arterial and venous systemic compartments as shown in Figure 1. The pressures and the compliances in the compartments are denoted by P and c, respectively, while resistances are denoted by R. In the right ventricle, Q is the cardiac output and S is the contractility. The subscripts mainly stand for the name of the compartments. That is, $ap, vp, lv, sa, fa, as, vs$, and rv correspond to arterial pulmonary, venous pulmonary, left ventricle, systemic aorta, finger arteries, arterial systemic, venous systemic and right ventricle compartments, respectively. In addition, subscripts mv and av denote the mitral valve, respectively aortic valve. Also, sa_1 and sa_2 as subscripts for R (i.e., R_{sa_1} and R_{sa_2}) correspond to two different resistances connecting the systemic aorta to finger arteries and systemic aorta to arterial systemic compartment, respectively.

The current model is mathematically formulated in terms of an electric circuit analog. The blood pressure difference plays the role of voltage, the blood flow plays the role of current, the stressed volume plays the role of an electric charge, the compliances of the blood vessels play the role of capacitors, and the resistors are the same in both analogies. The stressed volume in a compartment is the difference between total and unstressed volume (i.e., the volume in a compartment at zero transmural pressure). Thus, stressed volume is

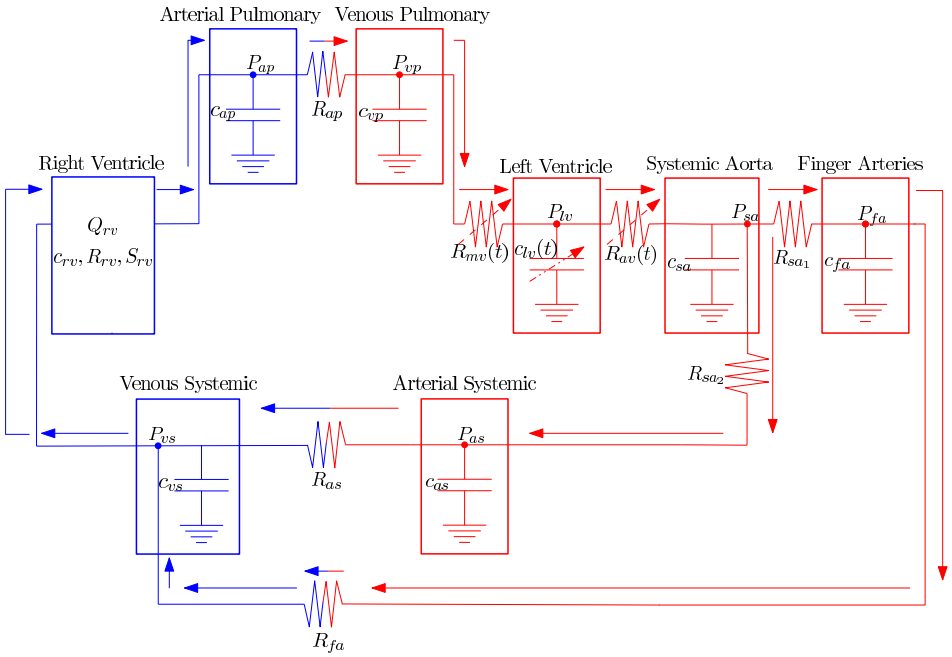


FIGURE 1. The electric analog of the global pulsatile model depicting the blood flow in the pulmonary and systemic circuit including the systemic aorta and finger arteries compartment.

the additional volume added to the unstressed volume when positive transmural pressure causes a stretching of the vascular walls.

The following are the basic assumptions of the modeling process:

- The vessels in the arterial and venous parts of the systemic or pulmonary circuits are lumped together as a single compartment for each of these parts. Each compartment is considered as a vessel with compliant walls in which its volume is characterized by the pressure in the vessel. Thus, these vessels are called *compliance vessels*.
- The systemic peripheral or pulmonary peripheral region is composed of capillaries, arterioles, and venules which are lumped together into a single vessel. These vessels are considered to be pure resistances to blood flow and characterized only by flow through the vessel. Therefore these vessels are called *resistance vessels*.
- The atria are not represented in the model. It is assumed that the right atrium is part of the venous systemic compartment and the left atrium is part of the venous pulmonary compartment.

1.1. **Blood Volume in the Compartment.** For each compartment, we associate the pressure $P(t)$ and the volume $V(t)$ of the blood. Assuming linear relationship between the transmural pressure and the total volume, we have

$$V(t) = cP(t), \tag{1}$$

where c represents the compliance of the compartment which is assumed to be constant. In this case, the unstressed volume is zero and the stressed volume equals the total volume in the compartment. Generally, the total volume in the compartment can be expressed as

$$V(t) = cP(t) + V_u, \tag{2}$$

where V_u denotes the unstressed volume. A more physiologically realistic approach is to consider that the relation between pressure and total volume is $V = f(P)$, which is nonlinear. In this case, the unstressed volume is given by $V_u = f(0)$ and the compliance, $c(P)$ at pressure P is $f'(P)$ assuming smoothness on f.

For simplicity, we used (2) assuming $V_u = 0$ in most of the compartments except in the left ventricle. This is mainly to avoid introduction of additional parameters which cannot be observed directly. This however introduces a modeling error that needs to be considered for further investigations.

1.2. **Blood Flow and Mass Balance Equations.** The blood flow is described in terms of the mass balance equations, that is, the rate of change for the blood volume $V(t)$ in a compartment is the difference between the flow into and out of the compartment. For a generic compartment, we have

$$\frac{d}{dt}(cP(t)) \;=\; F_{in} - F_{out}, \tag{3}$$

where c denotes the compliance, $P(t)$ the blood pressure in the compartment and F_{in} and F_{out} are the blood flows into and out of the compartment, respectively. The loss term in the compartment is the gain term in the adjacent compartment. Also, the flow F between two compartments can be described by Ohm's law. That is, it depends on the pressure difference between adjacent compartments and on the resistances R against blood flow. Thus we have the relation

$$F \;=\; \frac{1}{R}\left(P_1 - P_2\right), \tag{4}$$

where P_1 and P_2 are pressures from adjacent generic compartments 1 and 2, respectively.

The blood flow out of the venous systemic compartment is the cardiac output $Q_{rv}(t)$ which is the blood flow into the arterial pulmonary. The cardiac

output generated by the right ventricle is

$$Q_{rv}(t) \;=\; HV_{str}(t), \tag{5}$$

where H is the heart rate and $V_{str}(t)$ is the stroke volume, that is the blood volume ejected by one beat of the ventricle. Following the discussions given in Batzel et. al (2007) [1], the cardiac output of the right ventricle can be expressed as

$$Q_{rv}(t) \;=\; H\frac{c_{rv}P_{vp}(t)a_{rv}(H)f(S_{rv}(t),P_{ap}(t))}{a_{rv}(H)P_{ap}(t)+k_{rv}(H)f(S_{rv}(t),P_{ap}(t))}\;, \tag{6}$$

where we have chosen that the function $f(S_{rv}(t), P_{ap}(t))$, according to Timischl (1998) [12] is given by

$$f(S_{rv}(t),P_{ap}(t)) \;=\; 0.5\,(S_r(t)+P_{ap}(t))-0.5\,((P_{ap}(t)-S_r(t))+0.01)^{1/2}\;. \tag{7}$$

This function chooses the minimum value between S_{rv} and P_{ap} at a specific time point t. Also,

$$k_{rv}(H) \;=\; e^{-(c_{rv}R_{rv})^{-1}t_d(H)} \quad \text{and} \quad a_{rv}(H) \;=\; 1-k_{rv}(H)\;, \tag{8}$$

and

$$t_d(H) \;=\; \frac{1}{H^{1/2}}\left(\frac{1}{H^{1/2}}-\kappa\right)\;, \tag{9}$$

where κ is in the range of $0.3-0.4$ when time is measured in seconds and in the range of $0.0387-0.0516$ when time is measured in minutes.

Moreover, the change in the volume in the left ventricle $dV_{\ell v}(t)/dt$ as modeled in [9] is

$$\frac{dV_{\ell v}(t)}{dt} \;=\; \frac{P_{vp}(t)-P_{\ell v}(t)}{R_{mv}(t)}-\frac{P_{\ell v}(t)-P_{sa}(t)}{R_{av}(t)} \tag{10}$$

where $P_{vp}(t), P_{\ell v}(t)$ and $P_{sa}(t)$ are respectively, the blood pressures in the venous pulmonary, left ventricle and systemic aorta compartments and the time-varying elastances $R_{mv}(t)$ and $R_{av}(t)$ in the mitral valve and aortic valve, respectively.

1.3. **Opening and Closing of the Heart Valves.** In order to model the left ventricle as a pump, the opening and closing of the mitral and aortic valves must be included. During the diastole, the mitral valve opens allowing the blood to flow to the ventricle while the aortic valve is closed. Then the heart muscles start to contract, increasing the pressure in the ventricle. When the left ventricular pressure exceeds the aortic pressure, the aortic valve opens, propelling the pulse wave through the vascular system [6].

Rideout (1991) [10] originally proposed a model of the succession of opening and closing of these heart valves. A piecewise continuous function was later

developed by Olufsen et al., see for example [6] and [9]. This function represents the vessel resistance which characterized the *open* valve state using a small baseline resistance and the *closed* state using a value of larger magnitudes. The time-varying resistance is given as

$$
\begin{aligned}
R_{mv}(t) &= \min\left(R_{mv,open} + e^{(-2(P_{vp}(t)-P_{lv}(t)))}, 10\right) \ , \\
R_{av}(t) &= \min\left(R_{av,open} + e^{(-2(P_{lv}(t)-P_{sa}(t)))}, 10\right) \ ,
\end{aligned} \tag{11}
$$

where $R_{mv}(t)$ and $R_{av}(t)$ are the time varying mitral valve and aortic valve resistances, respectively. The first equation suggests that when $P_{lv}(t) < P_{vp}(t)$, the mitral valve opens and the blood enters the left ventricle. As $P_{lv}(t)$ increases and becomes greater than $P_{vp}(t)$, the resistance exponentially grows to a large value. A similar remark can be deduced from the second equation. The value 10 is chosen to ensure that there is no flow when the valve is closed and remains there for the duration of the closed valve phase. The open and closed transition is not discrete. An exponential function is used for the partially opened valve, with the amount of *openness* [9].

1.4. **Time-Varying Elastance Function.** The slope of a pressure-volume curve which has pressure on the y-axis and volume on the x-axis is called the *ventricular elastance* or simply the *elastance*. It is a measure of stiffness of the ventricles. Elastance and compliance are inverse of each other.

According to Ottesen et al. (2004) [7], the relationship between the left ventricular pressure $P_{\ell v}$ and the stressed left ventricular volume $V_{\ell v}(t)$ is described by

$$
P_{\ell v}(t) = E_{\ell v}(t)\left(V_{\ell v}(t) - V_d\right) \ , \tag{12}
$$

where $E_{\ell v}(t)$ is the time-varying ventricular elastance and V_d (constant) is the ventricular volume at zero diastolic pressure.

In [9], the time-varying elastance function $E_{\ell v}(t)$ is given by

$$
E_{\ell v}(t) = \begin{cases}
E_m + \frac{E_M - E_m}{2}\left[1 - \cos\left(\frac{\pi t}{T_M}\right)\right], & 0 \le t \le T_M \\
E_m + \frac{E_M - E_m}{2}\left[\cos\left(\frac{\pi}{T_r}\left(t - T_M\right)\right) + 1\right], & T_M \le t \le T_M + T_r \\
E_m, & T_M + T_r \le t < T \ .
\end{cases} \tag{13}
$$

This is a modification of a model developed by Heldt et al. (2002) [2]. Here, T_M denotes the time of peak elastance, and T_r is the time for the start of diastolic relaxation. These are both functions of the length of the cardiac cycle T. These parameters are set up as fractions where $T_{M,frac} = T_M/T$ and $T_{r,frac} = T_r/T$.

Moreover, E_m and E_M are the minimum and maximum elastance values, respectively. The above elastance function (13) is sufficiently smooth. Its derivative can be easily computed as follows

$$\frac{dE_{\ell v}(t)}{dt} = \begin{cases} \frac{E_M - E_m}{2}\left[\frac{\pi}{T_M}\sin\left(\frac{\pi t}{T_M}\right)\right], & 0 \le t \le T_M \\ \frac{E_M - E_m}{2}\left[-\frac{\pi}{T_r}\sin\left(\frac{\pi}{T_r}\left(t - T_M\right)\right)\right], & T_M \le t \le T_M + T_r \\ 0, & T_M + T_r \le t < T\ . \end{cases} \tag{14}$$

In our model, further modifications of the elastance function in (13) has been done. The maximum elastance E_M can be interpreted as a measure of the contractile state of the ventricle, see [8] and [11]. For normal resting heart, E_M can be a parameter constant. However, during exercise state, the contractility of the heart muscles may vary and could depend on the heart rate. That is, an increase in heart rate may result to an increase ventricular elastance. Thus we considered E_M as a function dependent on the heart rate H. Such function must be positive-valued, bounded and continuous. We chose the Gompertz function for $E_M(H)$, a sigmoidal function given by

$$E_M(H) = a\exp(-b\exp(-cH))\ , \tag{15}$$

where a, b, c are positive constants. The constant a determines the upper bound of the function, b shifts the graph horizontally and c is the measure of the steepness of the curve.

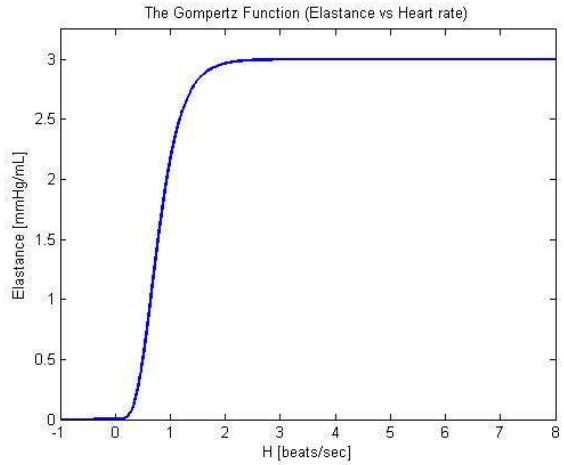
In Ottesen (2004) [7] and Olufsen et al. (2009) [9], E_M=2.49 [mmHg/mL]. Figure 2(a) depicts the maximum elastance curve where constants a, b and c were estimated obtaining $E_M = 2.4906$ [mmHg/mL] at $H = 70/60$ beats per second.

Since, E_M is now H-dependent, T_M which is the time of peak elastance should be H-dependent as well. We considered T_M as the time for systolic duration which is defined by the Bazett's formula given by

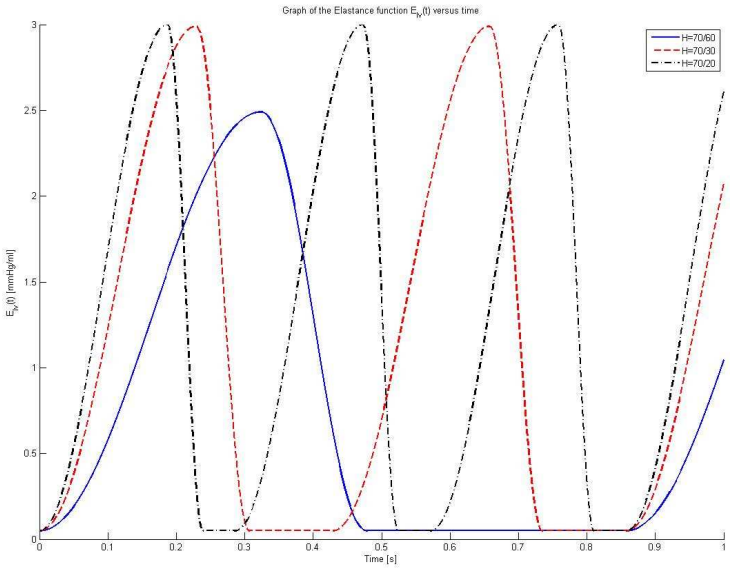
$$T_M = \frac{\kappa}{H^{1/2}}\ . \tag{16}$$

Figure 2(b) depicts the elastance function with varying heart rates. As the heart rate increases, the maximum elastance value increases as well. Notice also the decrease in the time for peak elastance and the smaller support of the elastance curve.

1.5. **Filling Process in the Right Heart.** The filling process in the right ventricle depends on the pressure difference between the filling pressure and the pressure in the right ventricle when the inflow valve (tricuspid valve) is open. Following Batzel et al. (2007) [1], the blood inflow into the right ventricle is



(a)



(b)

FIGURE 2. (a) The maximum elastance E_M expressed as a sigmoidal function dependent on the heart rate H. (b) The elastance function with varying heart rates.

given by

$$\frac{dV_{rv}(t)}{dt} = \frac{1}{R_{rv}} \left(P_v(t) - P_{rv}(t)\right), \tag{17}$$

where $V_{rv}(t)$ is the volume in the right ventricle at time t after the filling process has started, $P_v(t)$ is the venous filling pressure, $P_{rv}(t)$ is the pressure in the right ventricle, and R_{rv} is the total resistance to the inflow into the right ventricle.

As in [1], it is assumed that $P_v(t)$ is constant during the diastole, $P_v(t) \equiv P_v$, the end-systolic volume at the end of a heart beat equals the end-systolic volume of the previous heart beat and the compliance c_{rv} of the relaxed ventricle remains constant during the diastole.

1.6. **The Contractility of the Right Ventricle.** There is a heart mechanism called the *Bowditch effect*. It roughly states that changing the heart rate causes a concordant change in the ventricular contractilities. In this study, we adapted the model presented in Batzel et al. (2007) [1] (see also [4]), where sympathetic and parasympathetic activities were not considered directly. Thus, the variations of the contractilities can be described by the following second order differential equation

$$\ddot{S}_r + \gamma_r \dot{S}_r + \alpha_r S_r = \beta_r H, \tag{18}$$

where α_r, β_r and γ_r are constants. This set-up guarantees that the contractility S_r varies in the same direction as the heart rate H. Introducing the state variable $\sigma_r = \dot{S}_r$ and transforming (18) into systems of first order differential equations, we have

$$\begin{aligned} \dot{S}_r &= \sigma_r, \\ \dot{\sigma}_r &= -\alpha_r S_r - \gamma_r \sigma_r + \beta_r H. \end{aligned} \tag{19}$$

1.7. **The Combined Model Equations.** The model can be described as a system of coupled first order of ordinary differential equations with state variables $x(t) = (P_{sa}, P_{fa}, P_{as}, P_{vs}, P_{ap}, P_{vp}, P_{\ell v}, S_r, \sigma_r)^T \in \mathbb{R}^9$, representing pressures in the systemic aorta, finger arteries, arterial systemic, venous systemic, arterial pulmonary and left ventricle compartments, right ventricular contractility and

its derivative, respectively. These are given by

$$
\begin{aligned}
\frac{dP_{sa}}{dt} &= \frac{1}{c_{sa}}\left(\frac{P_{lv}-P_{sa}}{R_{av}(t)}-\frac{P_{sa}-P_{fa}}{R_{sa_1}}-\frac{P_{sa}-P_{as}}{R_{sa_2}}\right),\\
\frac{dP_{fa}}{dt} &= \frac{1}{c_{fa}}\left(\frac{P_{sa}-P_{fa}}{R_{sa_1}}-\frac{P_{fa}-P_{vs}}{R_{fa}}\right),\\
\frac{dP_{as}}{dt} &= \frac{1}{c_{as}}\left(\frac{P_{sa}-P_{as}}{R_{sa_2}}-\frac{P_{as}-P_{vs}}{R_{as}}\right),\\
\frac{dP_{vs}}{dt} &= \frac{1}{c_{vs}}\left(\frac{P_{as}-P_{vs}}{R_{as}}+\frac{P_{fa}-P_{vs}}{R_{fa}}-Q_r\right),\\
\frac{dP_{ap}}{dt} &= \frac{1}{c_{ap}}\left(Q_r-\frac{P_{ap}-P_{vp}}{R_{ap}}\right),\\
\frac{dP_{vp}}{dt} &= \frac{1}{c_{vp}}\left(\frac{P_{ap}-P_{vp}}{R_{ap}}-\frac{P_{vp}-P_{lv}}{R_{mv}(t)}\right),\\
\frac{dP_{lv}}{dt} &= E_{lv}\left(\frac{\frac{dE_{lv}}{dt}P_{lv}}{{E_{lv}}^2}+\frac{P_{vp}-P_{lv}}{R_{mv}(t)}-\frac{P_{lv}-P_{sa}}{R_{av}(t)}\right),\\
\frac{dS_r}{dt} &= \sigma_r,\\
\frac{d\sigma_r}{dt} &= -\alpha_r S_r-\gamma_r\sigma_r+\beta_r H,
\end{aligned}
\tag{20}
$$

where the time-varying resistances $R_{av}(t)$ and $R_{mv}(t)$ are given in equation (11), the cardiac output of the right ventricle Q_{rv} is given in equation (6) and other auxiliary equations such as for k_{rv} and a_{rv} are given in (8), the duration for diastole t_d is given in (9) and the right ventricular contractility S_r is given in (19).

2. **Simulation Results and Discussions**

Figure 3 shows the preliminary simulation results of the combined cardiovascular model (1.7) using the values of the parameters given in Table 1. The parameters are estimated to produce an average normal pulsatile pressures in the finger arteries which is 120/90 mmHg. Less pulsatility is observed in the arterial systemic compartment. In the venous pulmonary and arterial pulmonary compartments, the pulsatility is very small which is observed physiologically. Here, the contractility of the right ventricle is assumed to be constant considering a rest normal condition. Furthermore, it is also observed numerically that when the heart rate is increased, pulsatility is increased. The blood pressures in most of the compartments except the venous systemic compartment increased.

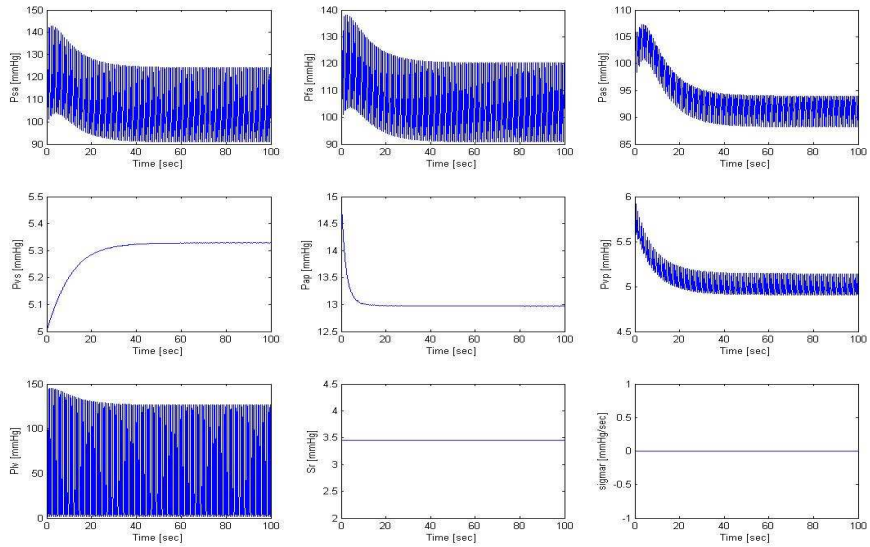
This is due to the Frank- Starling mechanism assumed in the filling process of the right ventricle as assumed in [1]. On the other hand, decreasing the heart rate produces the opposite result.

Parameter	Meaning	Value	Units
c_{sa}	Compliance of the systemic aorta compartment	1.15	mL/mmHg
c_{fa}	Compliance of the finger arteries compartment	0.105	mL/mmHg
c_{as}	Compliance of the arterial systemic compartment	3.75	mL/mmHg
c_{vs}	Compliance of the venous systemic compartment	750.95	mL/mmHg
c_{rv}	Compliance of the relaxed right ventricle	44.131	mL/mmHg
c_{ap}	Compliance of the arterial pulmonary compartment	25.15	mL/mmHg
c_{vp}	Compliance of the venous pulmonary compartment	175.75	mL/mmHg
$R_{mv,open}$	Resistance when the mitral valve is open	0.001	mmHg s/mL
$R_{av,open}$	Resistance when the aortic valve is open	0.001	mmHg s/mL
R_{sa_1}	Resistance between systemic aorta and finger arteries	0.4745	mmHg s/mL
R_{sa_2}	Resistance between systemic aorta and arterial systemic	0.1834	mmHg s/mL
R_{fa}	Resistance between finger and venous systemic compartment	44.9980	mmHg s/mL
R_{as}	Resistance in the peripheral region of the systemic circuit	1.2229	mmHg s/mL
R_{rv}	Inflow resistance of the right ventricle	0.002502	mmHg s/mL
R_{ap}	Resistance in the peripheral region of the pulmonary circuit	0.1097	mmHg s/mL
E_m	Minimum elastance value of the left heart	0.029	mmHg/mL
V_d	Unstressed left heart volume	10	mL
κ	Constant in the Bazett's formula	0.35	s
α_r	Coefficient of S_r in the differential equation for S_r	0.00797	s^{-2}
β_r	Coefficient of H in the differential equation for S_r	0.02355	mmHg/s
γ_r	Coefficient of $\dot{S}_r$ in the differential equation for S_r	0.03102	s^{-1}
a	Constant in the Gompertz function	3	mmHg/mL
b	Constant in the Gompertz function	10	
c	Constant in the Gompertz function	3.415	s^{-1}

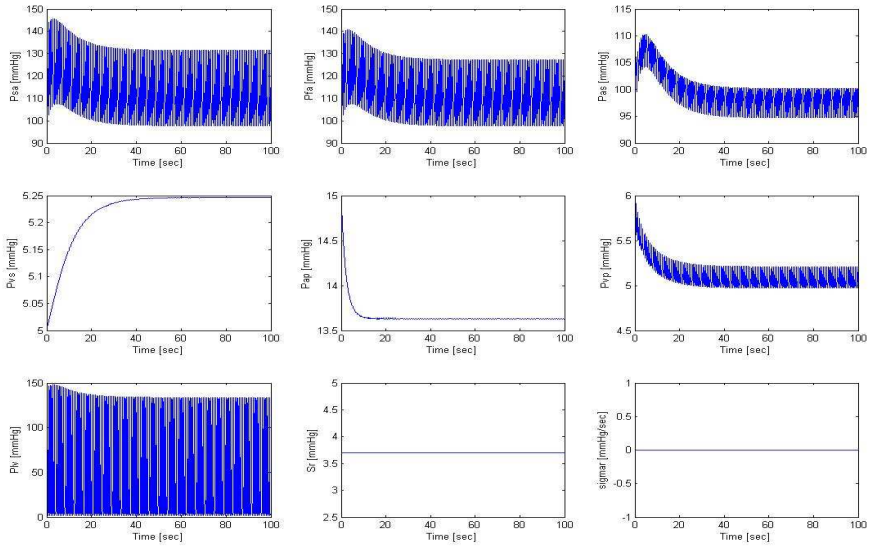
TABLE 1. The table of parameters.

3. **Ongoing and Future Work**

This modeling effort is a work in progress. A lot of considerations can be accounted to have a more holistic view of the overall behavior of the human cardiovascular system under specific conditions. The following are ongoing and future work on this area:



(a)



(b)

FIGURE 3. Simulations depicting the plots of the state variables at the heart rates (a) $H = 70/60$ beats per second and (b) $H = 75/60$ beats per second.

- to look closer into the relationship between the right ventricle contractility S_r and the left ventricle elastance E_M with regard to its physiological relevance,
- to investigate the behavior of the pulsatile model under a constant workload,
- to design a feedback law mechanism which controls the heart rate,
- to estimate and identify sensitive parameters,
- to investigate further the role of the unstressed volumes in the modeling process, and
- to include the respiratory system in the global pulsatile model.

Moreover, the model can be extended to describe the response of the cardiovascular system under a constant workload as in [4] and [12], its behavior under orthostatic stress as in [5] and [3], and to study blood loss due to haemorrhage.

Acknowledgment. The authors are grateful to the Austrian Academic Exchange Service (ÖAD) for granting scholarship to Mr. de los Reyes under Technologiestipendien Südostasien (Doktorat) program to work on this project.

References

[1] J.J. Batzel, F. Kappel, D. Schneditz, and H.T. Tran, *Cardiovascular and respiratory systems: Modeling, analysis and control*, SIAM, Philadelphia, PA, 2007.

[2] T. Heldt, E.B. Shim, R.D. Kamm, and R.G. Mark, Computational modeling of cardiovascular response to orthostatic stress, *Journal of Applied Physiology* **92**, 2002, 1239–1254.

[3] F. Kappel, M. Fink, and J.J. Batzel, Aspects of control of the cardiovascular-respiratory sytem during orthostatic stress induced by lower body negative pressure, *Mathematical Biosciences*, **206**(2), 2007, 273–308.

[4] F. Kappel and R.O. Peer, A mathematical model for fundamental regulation processes in the cardiovascular system, *Journal of Mathematical Biology*, **31**(6), 1993, 611–631.

[5] M.Fink, J.J. Batzel, and F. Kappel, An optimal control approach to modeling the cardiovascular-respiratory system: An application to orthostatic stress, *Cardiovascular Engineering*, **4**(1), 2004, 27–38.

[6] M.S. Olufsen, J.T. Ottesen, H.T. Tran, L.M. Ellwein, L.A. Lipsitz, and V. Novak, Blood pressure and blood flow variation during postural change from sitting to standing: Model development and validation, *Journal of Applied Physiology*, **99**, 2005, 1523–1537.

[7] J.T. Ottesen, M.S. Olufsen, and J.K. Larsen, *Applied mathematical models in human physiology*, SIAM, Philadelphia, PA, 2004.

[8] J.L. Palladino and A. Noordergraaf, A paradigm for quantifying ventricular contraction, *Cellular and Molecular Biology Letters*, **7** (2), 2002, 331–335.

[9] S.R. Pope, L. Ellwein, C. Zapata, V. Novak, C.T. Kelley, and M.S Olufsen, Estimation and identification of parameters in a lumped cerebrovascular model, *Mathematical Biosciences and Engineering*, **6**(1), 2009, 93–115.

[10] V. Rideout, *Mathematical and computer modeling of physiological systems*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[11] K. Sunagawa and K. Sagawa, Models of ventricular contraction based on time-varying elastance, *Critical Revies in Biomedical Engineering*, **7**(3), 1982, 193–228.

[12] S. Timischl, *A global model of the cardiovascular and respiratory system*, Ph.D. thesis, University of Graz, Institute for Mathematics and Scientific Computing, 1998.

Institute for Mathematics
and Scientific Computing,
University of Graz,
Heinrichstraße 36,
8010 Graz, Austria

E-Mail: 06delosr@edu.uni-graz.at

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Integral Representations of the Logarithmic Derivative of the Selberg Zeta Function

Dženan Gušić

We point out the importance of the integral representations of the logarithmic derivative of the Selberg zeta function valid up to the critical line, i.e. in the region that includes the right half of the critical strip, where the Euler product definition of the Selberg zeta function does not hold. Most recent applications to the behavior of the Selberg zeta functions associated to a degenerating sequence of finite volume, hyperbolic manifolds of dimension 2 and 3 are surveyed. The research problem consists in extending this kind of integral representations to the setting of the locally symmetric spaces of rank 1.

AMS Subj. Classification: MSC2010: 11F72, 11M36, 58J37

Key Words: Selberg zeta function, Selberg trace formula, degenerating hyperbolic manifolds

1. Introduction

The Selberg trace formula, introduced by A. Selberg in 1956, describes the spectrum of the hyperbolic Laplacian in terms of geometric data involving the lengths of geodesics on the Riemann surface. It is formally similar to the explicit formulas relating the zeros of the Riemann zeta function to prime numbers. The Selberg zeta zeros correspond to eigenvalues of the Laplacian, and the primes corresponding to primitive geodesics. Motivated by this analogy, Selberg [14] introduced the Selberg zeta function of a Riemann surface, whose analytic properties are encoded by the Selberg trace formula. The Selberg zeta function has properties similar to the Riemann zeta function. It is defined by the infinite product similar to the classical Euler product, but taken over closed geodesics rather than primes. It also satisfies a functional equation relating values at s with values at $1-s$. Selberg proved that the Selberg zeta functions satisfy the analogue of the Riemann hypothesis, i.e. all zeros of the Selberg zeta function in the case of the compact Riemann surface lie on the critical line $\mathrm{Re}(s) = 1/2$.

2. The Selberg zeta function

Let $H = \{z = x + iy : y > 0\}$ denote the upper half-plane equipped with the hyperbolic metric $ds^2 = (dx^2 + dy^2)/y^2$. Möbius transformations $z \to (az + b)/(cz + d)$ where $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$ form the group PSL(2,$\mathbb{R}$) that acts on H by homeomorphisms which preserve the hyperbolic distance. Discrete subgroups of PSL(2,$\mathbb{R}$) are called Fuchsian groups.

Let X be a non-compact, hyperbolic Riemann surface of a finite volume with $n_1 \geq 1$ cusps. Then, X can be identified with $\Gamma\backslash H$, where $\Gamma \subseteq \mathrm{PSL}(2, \mathbb{R})$ is a Fuchsian group of the first kind containing n_1 inequivalent parabolic elements. Let $\Im$ denote the fundamental domain of X and $|\Im|$ its volume. We put

$$\bar{\Gamma} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \right\}$$

and denote by v the multiplier system of a weight $m \in \mathbb{R}$ on $\bar{\Gamma}$. For an r dimensional unitary representation ψ of Γ, the function W defined by $W(T) = \psi(t)v(t)$, $T \in \Gamma$ is a unitary $r \times r$ multiplier system of the weight m. This kind of a system has been introduced in order to develop L^2 spectral theory of operators

$$\Delta_m = y^2(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}) - imy\frac{\partial}{\partial x}$$

on X, since $-\Delta_m$ is essentially a self-adjoint operator on the space D_m of all twice continuously differentiable functions $f : H \to V$ (V is an r dimensional vector space over $\mathbb{C}$) such that f and $\Delta_m f$ are square integrable on $\Im$ and satisfy the equality

$$f(Sz) = \frac{(cz + d)^m}{|cz + d|^m} W(S)f(z)$$

for all $z \in H$ and $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \bar{\Gamma}$. The operator$-\Delta_m$ has the unique self-adjoint extension $-\tilde{\Delta}_m$ to the space $\tilde{D}_m$. Let $\{T_j\}$, $j = 1, 2, ..., n_1$ denote all parabolic classes of the group Γ. By m_j we shall denote the multiplicity of 1 as an eigenvalue of the matrix $W(T_j)$, and $n_1^* = \sum_{j=1}^{n_1} m_j$ will be the degree of singularity of W. The operator $-\tilde{\Delta}_m$ has both a discrete and a continuous spectrum in the case $n_1^* \geq 1$ and only a discrete spectrum in the case $n_1^* = 0$. The discrete spectrum will be denoted as $\{\lambda_n\}_{n \geq 0}$ ($0 \leq \lambda_0 < \lambda_1 < ..., \lambda_n \to \infty$). One of the properties of the continuous spectrum is that, for a fixed $j \in \{1, 2, ..., n_1\}$ and $1 \leq h \leq r$ it is possible to choose the orthonormal vector columns f_{h_j} so that $W(T_j^{-1})f_{h_j} = e^{2\pi i \alpha_{h_j}} f_{h_j}$, where $0 \leq \alpha_{h_j} < 1$ and $\alpha_{h_j} = 0$ iff $1 \leq h \leq m_j$ (see [9]).

The group Γ contains inequivalent hyperbolic, elliptic and parabolic classes. Following [9], we denote the set of inequivalent hyperbolic resp. elliptic classes by $\{P\}$ resp. $\{R\}$ whereas the set of inequivalent, primitive hyperbolic classes is denoted by $\{P_0\}$. The Selberg zeta function associated to the pair (Γ,W) is defined as the Euler product

$$Z_{\Gamma,W}(s)=\prod_{\{P_0\}_\Gamma}\prod_{k=0}^{\infty}\det(I_r-W(P_0)N(P_0)^{-s-k}) \tag{1}$$

converging absolutely for $\mathrm{Re}(s)>1$, where $N(P)$ denotes the norm of the class $\{P\}$. All elements of an elliptic class are conjugate in $\mathrm{SL}(2,\mathbb{R})$ to a rotation for some angle $\theta\in(0,\pi)$. The order of the primitive element R_0 associated to R is denoted by $M_R/2$.

3. The Selberg trace formula

The most important tool in analysis of the spectrum of the operator $-\tilde{\Delta}_m$ is the Selberg trace formula [[9], p. 412, Th. 6.3.]. M. Avdispahić and L. Smajlović [5] have proved that the Selberg trace formula (written in a different form) holds true for a larger class of test functions. They have proved the following theorem.

Theorem 3.1. *Let $n_1^*\geq 1$ and suppose that $h(r)$ satisfies the conditions*

S1. $h(r)=h(-r)$;

S2. $h(r)$ *is analytic in the strip* $|\mathrm{Im}(r)|<\frac{1}{2}+\delta$, *for some* $\delta>0$;

S3'. *For* $j=0,1,2$, $h^{(j)}(r)=O\left(\left(1+|r|^2\right)^{-1}\right)$ *in the strip* $|\mathrm{Im}(r)|<\frac{1}{2}+\delta$, with $0<\delta<1/4$. *Then, the formula*

$$\sum_{n=0}^{K}h(r_n)+\int_0^\infty h(t)dR(t)-\frac{|\mathfrak{F}|}{2\pi}\int_0^\infty th(t)r(t)\,dt+\frac{n_1^*}{2\pi}\int_0^\infty h(t)H(t)\,dt=$$

$$\frac{2}{\sqrt{2\pi}}\sum_{\substack{\{P_0\}_\Gamma\\ TrP_0>2}}\sum_{m=1}^{\infty}\frac{Tr\left(W(P_0)^m\right)\log N(P_0)}{N(P_0)^{m/2}-N(P_0)^{-m/2}}\widehat{h}\left(m\log N(P_0)\right)+ \tag{2}$$

$$\sum_{\substack{\{R\}_\Gamma\\ 0<\theta(R)<\pi}}\frac{Tr\left(W(R)\right)ie^{i(m-1)\theta}}{2\sqrt{2\pi}M_R\sin\theta}\int_{-\infty}^{\infty}\frac{e^u-e^{2i\theta}}{\cosh u-\cos 2\theta}\widehat{h}(u)\,e^{\frac{m-1}{2}u}du+$$

$$\frac{1}{\sqrt{2\pi}}\sum_{\alpha_{h_j}\neq 0}\left(\frac{1}{2}-\alpha_{h_j}\right)\int\limits_{-\infty}^{\infty}\frac{e^u-1}{\cosh u-1}\widehat{h}\left(u\right)e^{\frac{m-1}{2}u}du+$$

$$\frac{1}{2}h\left(0\right)Tr\left(I_r-\Phi\left(\frac{1}{2}\right)\right)+\frac{2n_1^*}{\sqrt{2\pi}}\int\limits_0^{\infty}\frac{2\widehat{h}\left(u\right)}{\sinh\frac{u}{2}}\left(1-\cosh\frac{mu}{2}\right)du$$

holds true

Here, we put

$$r\left(t\right)=\frac{\sinh 2\pi t}{\cosh 2\pi t+\cos\pi m}-1,$$

$$H(t)=\frac{\Gamma'}{\Gamma}\left(1+it\right)+\frac{\Gamma'}{\Gamma}\left(1-it\right)-2\log t,$$

$$R\left(t\right)=N\left[0\leq r_n\leq t\right]-\frac{1}{4\pi}\int\limits_{-t}^{t}\frac{\phi'}{\phi}\left(\frac{1}{2}+iu\right)du-r\frac{|\mathfrak{F}|}{4\pi}t^2+\frac{n_1^*}{\pi}t\log t$$

$$-\frac{t}{\pi}\left(n_1^*-n_1^*\log 2-\sum_{\alpha_{h_j}\neq 0}\log\left|1-e^{2\pi i\alpha_{h_j}}\right|\right),$$

where $N\left[0\leq r_n\leq t\right]$ counts the non-negative solutions of the equations $\frac{1}{4}+r_n^2=\lambda_n$. For $\lambda_n<\frac{1}{4}$, $n=0,...,K$, we put $r_n=-i\sqrt{\frac{1}{4}-\lambda_n}$, Φ denotes the hyperbolic scattering matrix and $\phi=\det\Phi$. (More about the hyperbolic scattering determinant can be found in [3]. Finally,

$$\hat{h}(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}h|_{\mathbb{R}}(x)\cdot e^{-itx}dx$$

denotes the Fourier transform of the function h considered as a function on $\mathbb{R}$.

4. Integral expressions for the logarithmic derivative of the Selberg zeta function

We shall denote by

$$D_{\Gamma,W}(s)=\frac{Z'_{\Gamma,W}(s)}{Z_{\Gamma,W}(s)}$$

the logarithmic derivative of the Selberg zeta function $Z_{\Gamma,W}(s)$. From (1) it is easily seen that, for $\mathrm{Re}(s)>1$ one has

$$D_{\Gamma,W}(s)=\sum_{\{P\},TrP>2}\frac{\Lambda(P)}{N(P)^s}Tr(W(P)),$$

where $\Lambda(P) = \dfrac{\log N(P_0)}{1 - N(P)^{-1}}$, for a primitive element P_0 such that $P = P_0^n$ for some $n \in \mathbb{N}$. Another representation of the logarithmic derivative $D_{\Gamma,W}(s)$ (with $m = 0$, hence $W = \psi$) given by absolutely and uniformly convergent series in any half plane $\mathrm{Re}(s) > 1 + \varepsilon$ ($\varepsilon > 0$) was obtained by Wakayama in [15].

In many applications it is important to have a representation of the logarithmic derivative $D_{\Gamma,W}(s)$ inside the critical strip, i.e. in the half-plane Re $(s) > 1/2$. M. Avdispahić and L. Smajlović have used Theorem 3.1. in order to obtain new integral representations of the logarithmic derivative of the Selberg zeta function valid up to the critical line. Namely, the functions

$$h_s(r) = \frac{1}{(s - \frac{1}{2})^2 + r^2} \quad \text{and} \quad h_{\alpha,y}(t) = \frac{2\alpha \exp y\alpha}{\alpha^2 + t^2} \cos yt$$

where $\mathrm{Re}(s) > \frac{1}{2}$ and $y > 0$ satisfy assumptions of Theorem 3.1. and do not satisfy assumptions of the classical trace formula [[9], p. 412, Th. 6.3.]. Inserting those functions into the display (2) they have proved the following theorem.

Theorem 4.1. [5] a) *For* $\mathrm{Re}(s) > 1/2$

$$\frac{D_{\Gamma,W}(s)}{2s-1} = \sum_{n=0}^{K} \frac{1}{(s-1/2)^2 + r_n^2} + \int_0^\infty \frac{dR(t)}{\left(s - \frac{1}{2}\right)^2 + t^2} - \frac{|\mathfrak{F}|}{2\pi} \int_0^\infty \frac{t \cdot r(t)\, dt}{\left(s - \frac{1}{2}\right)^2 + t^2} +$$

$$\frac{n_1^*}{2\pi} \int_0^\infty \frac{H(t)\, dt}{\left(s - \frac{1}{2}\right)^2 + t^2} - \frac{1}{2} \sum_{\alpha_{h_j} \neq 0} \left(\frac{1}{2} - \alpha_{h_j}\right) \int_0^\infty \frac{\sin \pi m \cdot dt}{\left(\left(s - \frac{1}{2}\right)^2 + t^2\right)(\cosh 2\pi t + \cos \pi m)}$$

$$(3) \quad - \sum_{\substack{\{R\}_\Gamma \\ 0 < \theta(R) < \pi}} \frac{Tr(W(R))}{2M_R \sin\theta} \int_0^\infty \frac{1}{\left(s - \frac{1}{2}\right)^2 + t^2} \frac{\cosh 2(\pi - \theta)t + e^{i\pi m} \cosh 2\theta t}{\cosh 2\pi t + \cos \pi m} dt -$$

$$\frac{n_1^*}{2(2s-1)} \left(\frac{\Gamma'}{\Gamma}\left(s + \frac{m}{2}\right) + \frac{\Gamma'}{\Gamma}\left(s - \frac{m}{2}\right) - 2\frac{\Gamma'}{\Gamma}(s)\right) - \frac{1}{(2s-1)^2} Tr\left(I - \Phi\left(\frac{1}{2}\right)\right)$$

b) *For* $\mathrm{Re}\alpha > 0$, $y > 0$ *and* $x = e^y$ *one has*

$$(4) \quad D_{\Gamma,W}\left(\frac{1}{2} + \alpha\right) = \frac{1}{1 + x^{2\alpha}} \sum_{N(P) < x} \frac{TrW(P)\Lambda(P)}{N(P)^{\alpha + \frac{1}{2}}} \left(x^{2\alpha} - N(P)^{2\alpha}\right)$$

$$+ \frac{4\alpha x^\alpha}{1 + x^{2\alpha}} \left(\sum_{n=0}^{K} \frac{\cos y r_n}{\alpha^2 + r_n^2} + \int_0^\infty \frac{\cos yt\, dR(t)}{\alpha^2 + t^2} - \frac{|\mathfrak{F}|}{2\pi} \int_0^\infty \frac{t \cdot r(t) \cos yt}{\alpha^2 + t^2} dt \right.$$

$$+\frac{n_1^*}{2\pi}\int_0^\infty \frac{\cos ytH\left(t\right)dt}{\alpha^2+t^2}-\sum_{\alpha_{h_j}\neq 0}\left(\frac{1}{2}-\alpha_{h_j}\right)\int_0^\infty \frac{\cos yt\sin\pi m dt}{\left(\alpha^2+t^2\right)\left(\cosh 2\pi t+\cos\pi m\right)}dt$$

$$-\sum_{\substack{\{R\}_\Gamma\\ 0<\theta(R)<\pi}}\frac{Tr\left(W\left(R\right)\right)}{2M_R\sin\theta}\int_0^\infty \frac{\cos yt}{\alpha^2+t^2}\frac{\cosh 2\left(\pi-\theta\right)t+e^{i\pi m}\cosh 2\theta t}{\cosh 2\pi t+\cos\pi m}dt\Bigg)$$

$$+n_1^*\int_0^\infty \frac{e^{-\alpha t}\left(\cosh\frac{mt}{2}-1\right)}{2\sinh\frac{t}{2}}dt+\frac{n_1^*}{1+x^{2\alpha}}\int_0^y \frac{e^{\alpha t}-e^{(2y-t)\alpha}}{2\sinh\frac{t}{2}}\left(\cosh\frac{mt}{2}-1\right)dt.$$

Theorem 4.1. may also be derived through interpretation of the Selberg trace formula as an explicit formula in the Jorgenson-Lang fundamental class of functions, as in [2].

5. **Some applications**

In order to illustrate the importance of the representation of the logarithmic derivative of the Selberg zeta function that is valid in the right half of the critical strip, we shall state here the main result of the paper [1] and indicate how the representation (3) of the logarithmic derivative of the Selberg zeta function and its three dimensional analogue are used in the proof of this result.

The authors have considered a sequence $\{M_k\}$ of low dimensional, finite volume hyperbolic manifolds whose limiting manifold is M_∞. By $Z_{M_k}(s)$ we shall denote the Selberg zeta function associated to the manifold M_k, while the $\mathbf{Z}_k(s)$ denotes the product of local factors in the Euler product expansion of $Z_{M_k}(s)$ corresponding to pinching geodesics of M_k. The problem of the behavior of the function $Z_{M_k}(s)/\mathbf{Z}_k(s)$ through degeneration (i.e. as $k\to\infty$) was firstly posed by Wolpert in [16]. In the case of dimension $d=2$, the convergence of $Z_{M_k}(s)/\mathbf{Z}_k(s)$ to $Z_{M_\infty}(s)$ in the half plane $\mathrm{Re}(s)>1/2$ was proved by M. Schulze [13] by means of analysis of certain operators. In [1], the unified approach in both dimensions $d=2$ or $d=3$ is provided and the following theorem is proved

Theorem 5.1. *Let $\{M_k\}$ be a degenerating sequence of low dimensional, finite volume hyperbolic manifolds whose limiting manifold is M_∞. Let $\{Z_{M_k}(s)\}$ be the associated sequence of Selberg zeta functions, normalized to have convergent Euler product in the half-plane Re $s>(d-1)$ and functional equations which*

relate values at s to values at $(d-1)-s$. Then, for all Re $s > (d-1)/2$ we have

$$Z_{M_k}(s)/\mathbf{Z}_k(s) = Z_{M_\infty}(s) + o(1), \quad as \quad k \to \infty.$$

The statement of the above theorem was proved in [12] and [6] (in dimensions $d = 2$ and $d = 3$ respectively) but in the region that does not include the right half of the critical strip. (Actually, the right half of the critical strip is, due to the functional equation, the maximal region where the statement of Theorem 5.1. could hold true).

It was Theorem 4.1. a) that enabled the authors of [1], to use investigations conducted in [12], [10] and [6] about the behavior of heat kernels through degeneration, and prove that the convergence through degeneration of the Selberg zeta functions holds in the half-plane $\mathrm{Re}(s) > (d-1)/2$.

The representation (4) is itself of a number-theoretical interest, since it allows one to consider a limiting process as $y \to \infty$ on the right-hand side of (4), the left hand side being independent of y. This observation was used in [4] in order to evaluate analogues of Euler-Stieltjes constants for the Selberg zeta functions and determine the upper and lower bounds for some special values of $D_{\Gamma,Id}$ that improve the bounds obtained in [11].

6. **A possible generalization**

Let G be a connected semisimple Lie group with finite center, K its maximal compact subgroup. By H we shall denote the symmetric space G/K, assumed to be of a rank one. Let Γ be a discrete, torsion-free subgroup of G such that $\Gamma\backslash G$ is compact, T finite dimensional unitary representation of Γ and χ its character. By the work of R. Gangolli [7], it is possible to define zeta function $Z_\Gamma(s,\chi)$ attached to the data (G, K, Γ, χ), as an Euler product similar to (1). Actually, Gangolli has first proved the trace formula in this setting and derived the representation of logarithmic derivative of the function $Z_\Gamma(s,\chi)$ and the functional equation relating values of $Z_\Gamma(s,\chi)$ at s with the values at $2\rho_0 - s$, for a constant $\rho_0 > 0$, whose dependance on the data (G, K, Γ, χ) is fully explained in [7]. This yields to a natural notion of the critical strip $0 \leq \mathrm{Re}(s) \leq 2\rho_0$ and the critical line $\mathrm{Re}(s) = \rho_0$. A similar zeta function can also be defined in the case of a non-compact space $\Gamma\backslash G$.

The main goal of our future research is to use the trace formula proved in [7] and [8] and try to obtain the analogue of Theorem 3.1. in the setting of rank one symmetric spaces. A further goal is to obtain the representation of the logarithmic derivative of the function $Z_\Gamma(s,\chi)$ (analogous to Theorem 4.1.)

valid in the half plane $\mathrm{Re}(s) > \rho_0$ (i.e. up to the critical line) in both compact and non-compact setting.

References

[1] M. Avdispahić, J. Jorgenson and L. Smajlović, Asymptotic behavior of the Selberg zeta functions for degenerating families of hyperbolic manifolds (submitted).

[2] M. Avdispahić, L. Smajlović, An explicit formula and its application to the Selberg trace formula. *Monatsh. Math.*, **147**, no. 3, 2006, 183–198.

[3] M. Avdispahić, L. Smajlović, Explicit formula for the hyperbolic scattering determinant,*Acta Math. Sinica, English Series.*, **23**, no. 5, 2007, 889–894.

[4] M. Avdispahić, L. Smajlović, Euler constants for a Fuchsian group of the first kind. *Acta Arith.*, **131**, 2008, 125-143.

[5] M. Avdispahić, L. Smajlović, On the logarithmic derivative of the Selberg zeta function, submitted.

[6] J. Dodziuk, J. Jorgenson, Spectral asymptotics on degenerating hyperbolic 3-manifolds. *Memoirs of AMS*, **643**, 1998.

[7] R. Gangolli, Zeta functions of Selberg's type for compact space forms of symmetric spaces of rank one, *Illinois J. Math.*, **21**, 1977, 1-41.

[8] R. Gangolli, G. Warner, Zeta functions of Selberg's type for some noncompact quotients of symmetric spaces of rank one, *Nagoya Math. J.*, **78**, 1980, 1–44.

[9] D. A. Hejhal, The Selberg trace formula for PSL(2,$\mathbb{R}$) *Vol. II. Lecture Notes in Mathematics 1001.* Springer-Verlag, Berlin-Heidelberg (1983).

[10] J. Huntley, J. Jorgenson and R. Lundelius, On the asymptotic behavior of counting functions associated to degenerating hyperbolic Riemann surfaces. *J. Funct. Analysis*, **149**, 1997, 58-82.

[11] J. Jorgenson, J. Kramer,Bounds for special values of Selberg zeta functions of Riemann surfaces, *J. Reine Angew. Math.*, **541**, 2001, 1-28.

[12] J. Jorgenson, R. Lundelius, A regularized heat trace for hyperbolic Riemann surfaces of finite volume, *Comment. Math. Helv.*, **72**, 1997, 636-659.

[13] M. Schulze, On the resolvent of the Laplacian on functions for degenerating surfaces of finite geometry, *J. Funct. Anal.*, **236**, 2006, 120-160.

[14] A. Selberg Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, *J. Indian Math. Soc.*, **20**, 1956, 47-87.

[15] M. Wakayama, A formula for the logarithmic derivative of Selberg's zeta function, *J. Math. Soc. Japan* . **41**, 1989, 463-471.

[16] S. Wolpert, Asymptotics of the spectrum and the Selberg zeta function on the space of Riemann surfaces. *Comm. Math. Phys.*, **112**, 1987, 283-315.

Department of Mathematics,
University of Sarajevo,
Zmaja od Bosne 33-35,
71000 Sarajevo,
Bosnia and Herzegovina

E-Mail: dzenang@yahoo.com

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Semi-Smooth Newton Methods for the Time Optimal Control of Nonautonomous Ordinary Differential Equations [1]

Jelena Rubeša and Karl Kunisch

The control problem of minimal time transition between two stationary points are formulated in a framework of an indirect numerical method. The problem is regularized and the monotone behavior of the regularisation procedure is investigated. Semi-smooth Newton method applied on the regularized problems converge superlinearly and usually produce a very accurate solution. Differently from other methods, this one does not need a-priory knowledge of the control switching structure. A code was developed in the C++ language and the NVIDIA CUDA technology made it even faster.

AMS Subj. Classification: 49J15, 49M15

Key Words: time optimal control, switching structure, regularization, semi-smooth Newton method, CUDA technology

Introduction

This paper addresses time optimal control problems for a class of linear non-autonomous multi-input controls systems for ordinary differential equations. Due to structural difficulties, time optimal control has been receiving a considerable amount of attention for decades. Much of the literature up to the late sixties is covered in [5]. Many recent results can be found or are referenced in [2, 8, 10]. Time optimal control for infinite dimensional systems is considered in [3], for example.

The optimality system associated to time optimal control problems with pointwise constraints on the controls is complicated due to lack of smoothness of the optimal controls. In fact, the first order optimality system for time optimal control problems contains a multi-valued operation which complicates the

[1]This work was completed with the support of project NAWI Graz.

use of fast numerical methods. For this reason we introduce a regularization to the time optimal problem. In section 1 the behavior of the solutions of the regularized problems as the regularization parameter ε tends to zero is investigated. In particular monotonic structure of the solutions with respect to ε is shown. An optimality system for the regularized problems is derived under a normality condition. The optimal controls of the regularized problems are $W^{1,\infty}$ regular and converge to a minimum norm solution of the original problem as the regularization parameter tends to zero.

The optimality system of the regularized problems is still not C^1 but the optimal controls are now Lipschitz functions. In finite dimensions, locally Lipschitz continuous functions are almost everywhere differentiable, by Rademacher's theorem. This concept is not available in infinite dimensions so second order methods with local quadratic convergence order are not directly applicable. However, sufficient conditions will be obtained in section 2 which imply that semi-smooth Newton methods [6] are well-posed and locally superlinearly convergent.

Section 3 contains a brief description of numerical results. We compare the chosen regularization to an alternative one, which has stronger regularization properties. Since the optimal controls of the original time optimal problems are typically not continuous, it appears that our choice of regularization which leads to $W^{1,\infty}$ regularized controls is preferable over others which provide smoother ones.

Let us note that the approach that we propose for solving time optimal problems deviates from traditional approaches, which are frequently grouped into direct and indirect methods. The basic idea of direct methods is to discretize the control problem in order to obtain nonlinear programming (NLP) problem, which may be solved by NLP techniques such as Initial Value Solver (IVS) and Sequential Quadratic Programming (SQP). These methods use only control and state variables. In this sense direct methods are easier to implement but less accurate than the indirect methods; cf., e.g. Reference [1], [8].

The second group consists of *indirect optimization methods*, which are based on the solution of the first order necessary conditions of optimality, the Pontryagin minimum principle (PMP). We shall use this approach. The two point boundary value problem (TPBVP) arising from the PMP is solved by the *Semi-smooth Newton's method*, and quite contrary to the multiple shooting method, does not need a priori knowledge of switching structure. In practice, it is usually rather difficult to determinate the optimal switching structure in

advance, especially for the multi switch problems. The Newton's method will converge superlinearly and usually produces a very accurate solution.

1. The Time-optimal problem and its Regularization

Consider the time-optimal control problem for the non-autonomous linear multi-input system

$$\text{(P)} \qquad \begin{cases} \min \int_{t_0}^{t_1} dt \\ \text{subject to} \\ \qquad \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t),\ |u(t)|_{\ell^\infty} \le 1, \\ \qquad x(t_0) = x_0,\ x(t_1) = x_1, \end{cases}$$

where t denotes a scalar time variable on some nonempty closed interval $J \doteq [t_0, t_1]$ of the positive real line. Here $A \in L^\infty(t_0, t_1; \mathbb{R}^{n\times n})$ and $B \in L^\infty(t_0, t_1; \mathbb{R}^{n\times m})$ are two given coefficient matrices. The columns of $B(t)$ are denoted by $b_i(t)$ for $i = 1, 2, \ldots, m$. The vector function $x(\cdot)$ satisfying problem (P) is a system trajectory and a vector x is a state of the system. We assume that the initial state x_0 and terminal state x_1 are two given n-dimensional column vectors. The vector $u(\cdot)$ is control used to modify the system. We will assume that the control u is a bounded measurable m-vector, and $|\cdot|_{\ell^\infty}$ denotes the infinity-norm on $\mathbb{R}^m$. Also we assume that x_1 can be reached in finite time by an admissible control. Then (P) admits a solution with optimal (τ^*, x^*, u^*), where τ^* is the optimal time, x^* is the optimal state and u^* is the optimal control. For the proof we refers to [3].

The first order optimality system for (P) can be expressed in terms of the adjoint p and the Hamiltonian

$$H(x, u, p_0, p) = p_0 + p^T(Ax + Bu),$$

as

$$\text{(1.1)} \qquad \begin{cases} \dot{x} = Ax + Bu,\ x(t_0) = x_0,\ x(t_1) = x_1, \\ -\dot{p} = A^T p, \\ u = argmin_{|u|_{\ell^\infty}\le 1} H(x, u, p_0, p), \\ p_0 + p(t_1)^T(A(t_1)x(t_1) + B(t_1)u(t_1)) = 0,\ p_0 \ge 0, \end{cases}$$

where the superscript T denotes transposition. For details, see page 27 of [4]. Due to the special structure of H the optimal control can be expressed as

$$u_i = -\sigma(s_i), \tag{1.2}$$

where $s_i = p^T b_i$ denotes the switching function for the ith control variable and σ denotes the coordinate-wise operation

$$\sigma(s_i) \;=\; \begin{cases} -1 & \text{if } s_i < 0 \\ [-1,\,1] & \text{if } s_i = 0 \\ 1 & \text{if } s_i > 0. \end{cases} \tag{1.3}$$

Introducing the transformation $\hat{t} = \frac{t}{\tau}$ and setting $t_0 = 0$, $t_1 = \tau$,

$$\hat{x}(\hat{t}) \doteq x(\tau\hat{t}) = x(t),\ \hat{p}(\hat{t}) \doteq p(\tau\hat{t}) = p(t),\ \hat{u}(\hat{t}) \doteq u(\tau\hat{t}) = u(t),$$
$$\hat{A}(\hat{t}) \doteq A(\tau\hat{t}) = A(t),\ \hat{B}(\hat{t}) \doteq B(\tau\hat{t}) = B(t),$$

we obtain the following equivalent system to (1.1):

$$\begin{cases} \dot{\hat{x}} = \tau\hat{A}\hat{x} + \tau\hat{B}\hat{u},\ \hat{x}(0) = x_0,\ \hat{x}(1) = x_1 \\ -\dot{\hat{p}} = \tau\hat{A}^T\hat{p} \\ \hat{u} = argmin_{|\hat{u}|_{\ell^\infty}\le 1}\, H(\hat{x}, \hat{u}, \hat{p}) \\ 1 + \hat{p}(1)^T(\hat{A}(1)\hat{x}(1) + \hat{B}(1)\hat{u}(1)) = 0 \end{cases} \tag{1.4}$$

where $\hat{t} \in [0, 1]$. The non-differentiable operation involved in characterizing the optimal control,

$$u = -\sigma(B^T p),$$

compare (1.2), prohibits the use of Newton-type methods for solving (1.4) numerically.

Therefore a family of regularized problems

$$(P_\varepsilon) \qquad \begin{cases} \min_{\tau\ge 0} \int_0^\tau (1 + \frac{\varepsilon}{2}\,|u(s)|^2)\,ds \\ \text{subject to} \\ \qquad \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t),\ |u(t)|_{\ell^\infty} \le 1, \\ \qquad x(0) = x_0,\ x(\tau) = x_1, \end{cases}$$

with $\varepsilon > 0$ is considered. The norm $|\cdot|$ used in the cost-functional denotes the Euclidean norm. It is straightforward to argue the existence of a solution $(u_\varepsilon, x_\varepsilon, \tau_\varepsilon)$.

The convergence of the solutions $(x_\varepsilon, p_\varepsilon, u_\varepsilon, \tau_\varepsilon)$ of (P_ε) to a solution (x^*, p^*, u^*, τ^*) of (P) is considered next. Note that (x^*, p^*, u^*, τ^*) is unique if the normality condition is assumed.

Theorem 1.1. *For $\varepsilon \to 0^+$ we have $\tau_\varepsilon \to \tau^*$ and every convergent subsequence of solutions $\{(u_\varepsilon, x_\varepsilon)\}_{\varepsilon>0}$ to (P_ε) converges in $L^2(0, \tau_\epsilon; \mathbb{R}^m) \times W^{1,2}(0, \tau_\epsilon; \mathbb{R}^n)$ to a solution (u^*, x^*) of (P), where u^* is a minimum norm solution.*

Corollary 1.2. *If normality holds, then the solution u^* to (P) is unique, it is bang-bang, and $u_\varepsilon \to u^*$ in L^2 as $\varepsilon \to 0^+$.*

We turn to the optimality condition for (P_ε). Let

$$\sigma_\varepsilon(s) \;=\; \begin{cases} -1 & \text{if } s \le -\varepsilon \\ \frac{s}{\varepsilon} & \text{if } |s| < \varepsilon \\ 1 & \text{if } s \ge \varepsilon. \end{cases} \tag{1.5}$$

If σ_ε is applied to a vector, then it acts coordinate-wise.

Theorem 1.3. *Assume that normality of $(\hat{A}, \hat{B})$ holds on $[\alpha, 1]$ for every $\alpha \ge 0$, that $\hat{A} \in W^{1,\infty}(0,1; \mathbb{R}^{n\times n})$, $\hat{B} \in W^{1,\infty}(0,1; \mathbb{R}^{n\times m})$ and let $(x_\varepsilon, u_\varepsilon, \tau_\varepsilon)$ be a solution of (P_ε). If there exist an interval $J_1 \doteq (\alpha, \alpha+\delta)$, $\delta > 0$, $\eta > 0$ such that*

$$|\,\hat{u}_\varepsilon(t)|_{\ell^\infty} \le 1 - \eta \;\; \textit{for a.e. } t \in J_1, \tag{1.6}$$

then there exists an adjoint state p_ε such that

$$\begin{cases} \dot{x}_\varepsilon = A\,x_\varepsilon + B\,u_\varepsilon,\; x_\varepsilon(0) = x_0,\; x_\varepsilon(\tau_\varepsilon) = x_1 \\ -\dot{p}_\varepsilon = A^T\,p_\varepsilon \\ u_\varepsilon = -\sigma_\varepsilon(B^T\,p_\varepsilon) \\ 1 + \frac{\epsilon}{2}|u_\varepsilon(\tau_\varepsilon)|^2_{\mathbb{R}^m} + p_\varepsilon(\tau_\varepsilon)^T\big(A(\tau_\varepsilon)\,x_\varepsilon(\tau_\varepsilon) + B(\tau_\varepsilon)\,u_\varepsilon(\tau_\varepsilon)\big) = 0. \end{cases} \tag{1.7}$$

The third claim in (1.7) reveals the extra regularity of u_ε:

Corollary 1.4. *Under the assumptions of Theorem 1.3 we have*

$$u_\varepsilon \in W^{1,\infty}(0, \tau_\varepsilon; \mathbb{R}^m)\,.$$

2. Semi-smooth Newton methods

In this section the semi-smooth Newton method for solving the regularized optimality system (1.7) is described and analyzed. It will allow that (1.7) can be solved efficiently in spite of the fact that σ_ε is not Fréchet differentiable.

Throughout we fix $\varepsilon > 0$ and denote by $(x_\varepsilon, u_\varepsilon, \tau_\varepsilon) \in W^{1,2}(0,1) \times L^2(0,1) \times \mathbb{R}$ a solution to (P_ε) with associated adjoint $p_\varepsilon \in W^{1,2}(0, 1)$. It is assumed that normality of (A, B) holds on $[\alpha, 1]$ for every $\alpha \geq 0$, $A \in W^{1,\infty}(0,1; \mathbb{R}^{n\times n})$ and $B \in W^{1,\infty}(0,1; \mathbb{R}^{n\times m})$, that

(H1) there exists $\bar{s} \in (0,1)$ such that $|\frac{1}{\varepsilon} B(\bar{s})^T p_\varepsilon(\bar{s})|_{\ell^\infty} = |u_\varepsilon(\bar{s})|_{\ell^\infty} < 1$

and

(H2) $$|b_i(1)^T p_\varepsilon(1)| \neq \varepsilon, \text{ for all columns } i = 1, \dots, m \text{ of } B.$$

We now use the fact that u_ε is continuous and assumption (H1) implies (1.6) in some neighborhood $(\alpha, \alpha + \delta)$ containing $\bar{s}$. The existence of $p_\varepsilon \in W^{1,2}(0, 1)$ follows, by Theorem 1.3, such that (1.7) holds. With (H1) and (H2) holding there exists a neighborhood $\mathcal{U}_{p_\varepsilon}$ of p_ε in $W^{1,2}(0,1;\mathbb{R}^n)$, $\bar{t} \in (0,1)$ such that for the interval $(\alpha, \alpha + \delta) \subset (0,1)$ we have for every $p \in \mathcal{U}_{p_\varepsilon}$

$$|B(t)^T p(t)|_{\ell^\infty} < \varepsilon \text{ for } t \in (\alpha, \alpha + \delta) \tag{2.1}$$

and

$$|b_i(t)^T p(t)| \neq \varepsilon \text{ for all } t \in [\bar{t}, 1], \text{ and } i = 1, \dots, m.$$

We set $U = \{u \in L^2(0, 1; \mathbb{R}^m) : u|_{[\bar{t}, 1]} \in W^{1,2}(\bar{t}, 1; \mathbb{R}^m)\}$ endowed with the norm

$$|u|_U = (|u|^2_{L^2(0,1)} + |\dot{u}|^2_{L^2(\bar{t},1)})^{\frac{1}{2}},$$

and introduce

$$F : D_F \subset X \to L^2(0,1;\mathbb{R}^n) \times L^2(0,1;\mathbb{R}^n) \times U \times \mathbb{R}^n \times \mathbb{R},$$

where

$$D_F = W^{1,2}(0,1) \times \mathcal{U}_{p_\varepsilon} \times U \times \mathbb{R},$$

$$X = W^{1,2}(0,1;\mathbb{R}^n) \times W^{1,2}(0,1;\mathbb{R}^n) \times U \times \mathbb{R},$$

and

$$F(x,\,p,\,u,\,\tau) = \begin{pmatrix} \dot{x} - \tau A x - \tau B u \\ -\dot{p} - \tau A^T p \\ u + \sigma_\varepsilon(B^T p) \\ x(1) - x_1 \\ 1 + \frac{\varepsilon}{2}|u(1)|^2 + p(1)^T\,(A(1)x(1) + B(1)u(1)) \end{pmatrix}. \tag{2.2}$$

Note that $F = (F_1, \ldots, F_5)$ is well-defined. This is obvious for F_1, F_2 and F_3. For F_4, F_5 it follows from the fact that $W^{1,2}(0,1)$ embeds continuously into $C(0,1)$. Moreover $F(x_\varepsilon,\, p_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon) = 0$. We shall keep $x(0) = x_0$ as an explicit constraint.

Applying Newton's method to $F = 0$ is impeded by the non-differentiability of σ_ε. We use

$$G\sigma_\varepsilon(s) := \begin{cases} \frac{1}{\varepsilon} & \text{if } |s| < \varepsilon \\ 0 & \text{if } |s| \geq \varepsilon \end{cases} \tag{2.3}$$

as a generalized derivative. The Newton iteration step is given by

$$DF(x,\,p,\,u,\,\tau)(\delta x,\,\delta p,\,\delta u,\,\delta\tau) = -F(x,\,p,\,u,\,\tau) \tag{2.4}$$

where $\delta x(0) = 0$ and DF denotes the Fréchet derivative in all terms of F except for $p \to \sigma_\varepsilon(B^T p)$, for which the generalized derivative is taken according to (2.3). For further reference we give the detailed form of (2.4):

$$\begin{cases} \frac{d}{dt}\delta x - \tau A\,\delta x - \tau B\,\delta u - \delta\tau\,(A\,x + B\,u) = -F_1 \quad \delta x(0) = 0 \\ -\frac{d}{dt}\delta p - \tau A^T\,\delta p - \delta\tau\,A^T\,p = -F_2 \\ \delta u + G\sigma_\varepsilon(B^T\,p)\,B^T\,\delta p = -F_3 \\ \delta x(1) = -F_4 \\ p(1)^T\,\big(-A(1)\,F_4 + B(1)\,\delta u(1)\big) + \varepsilon u(1)^T\,\delta u(1) \\ \qquad\qquad +\delta p(1)^T\,\big(A(1)\,x(1) + B(1)\,u(1)\big) = -F_5, \end{cases} \tag{2.5}$$

where the coordinates of $G\sigma_\varepsilon(B^T p)B^T\delta p$ are given by $G\sigma_\varepsilon((B^T p)_i))(B^T\delta p)_i$.

A possible initialization may consist of choosing $((u)_0,\, \tau_0)$, setting $(x)_0$ as the linear interpolation between x_0 and x_1, and determining $(p)_0$ such that the transversality condition and the adjoint equation are satisfied.

We now briefly summarize those facts from semi-smooth Newton methods which are relevant for this paper. Proofs can be found in [6, 11, 9, 7]. Let X and Z be Banach spaces and let $F : D_F \subset X \to Z$ be a nonlinear mapping with an open domain D_F.

Definition 2.1. The mapping $F : D_F \subset X \to Z$ is called Newton-differentiable at $\mathbf{x}$, if there exists an open neighborhood $N(\mathbf{x}) \in D_F$ and mappings $DF : N(\mathbf{x}) \to \mathcal{L}(X, Z)$ such that

$$\lim_{h \to 0} \frac{1}{|h|_X} |F(\mathbf{x} + h) - F(\mathbf{x}) - DF(\mathbf{x} + h)h|_Z = 0. \tag{2.6}$$

The family $\{DF(s) : s \in N(\mathbf{x})\}$ is called a Newton-derivative of F at $\mathbf{x}$.

Note that F does not need to be Fréchet-differentiable in order to have the property (2.6). In general, there exists a set of Newton-derivatives at $\mathbf{x}$ which becomes a singleton whenever F is Fréchet-differentiable. If the mapping F is Newton-differentiable for each $\mathbf{x}$ in an open subset $\mathcal{U} \subset D_F$, then we say that F is Newton-differentiable on $\mathcal{U}$.

Lemma 2.2. *[Chain rule] Suppose that $H : D_H \subset X \to Y$ is Newton-differentiable at $\mathbf{x} \in D_H$ and $F : Y \to Z$ is Newton-differentiable at $H(\mathbf{x})$. Then the mapping $g = F(H)$ is Newton-differentiable at $\mathbf{x}$.*

Theorem 2.3. *Suppose that $\mathbf{x}^* \in \mathcal{U}$ is a solution to $F(\mathbf{x}) = 0$ and that F is Newton-differentiable in an open set $\mathcal{U}$ containing $\mathbf{x}^*$ with Newton-derivative DF. If further $\{\|DF(\mathbf{x})^{-1}\| : \mathbf{x} \in \mathcal{U}\}$ is bounded, then the Newton-iteration*

$$\mathbf{x}_{k+1} = \mathbf{x}_k - DF(\mathbf{x}_k)^{-1} F(\mathbf{x}_k)$$

converges q-superlinearly to $\mathbf{x}^*$, provided that $|\mathbf{x}_0 - \mathbf{x}^*|_X$ is sufficiently small.

For the statement and the proof of the superlinear convergence of the time-optimal control problem, some further notation is required. For $(x, p, u, \tau) \in D_F$ we define $\mathcal{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ by

$$\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix},$$

where

$$A_{11} = \varepsilon^{-1}\tau \int_0^1 \Phi(1,t)\, B(t)\, \chi_I\, B(t)^T\, \Phi(1,t)^T\, dt \in \mathbb{R}^{n\times n} \tag{2.7}$$

$$
\begin{aligned}
A_{12} = \varepsilon^{-1}\tau \textstyle\int_0^1 \Phi(1,t)\, B(t)\, \chi_I\, B(t)^T \int_t^1 \Phi(t,s)^{-T}\, A(s)^T\, p(s)\, ds\, dt \\
- \textstyle\int_0^1 \Phi(1,t)\big(A(t)\, x(t) + B(t)\, u(t)\big)\, dt \in \mathbb{R}^n
\end{aligned} \tag{2.8}
$$

$$
\begin{aligned}
A_{21} = \big(A(1)\, x(1) + B(1)\, u(1)\big)^T \\
-\big(p(1)^T\, B(1) + \varepsilon\, u(1)^T\big) G\sigma_\varepsilon\big(B(1)^T\, p(1)\big)\, B(1)^T \in (\mathbb{R}^n)^T,
\end{aligned} \tag{2.9}
$$

where $\chi_I = diag(\chi_{I_1}, \ldots, \chi_{I_m})$ and χ_{I_i} is the characteristic function of the set

$$
I_i = I_i(p) = \{t : |b_i^T p| < \varepsilon\}, \quad i = 1, \ldots, m
$$

which is nonempty for every $p \in \mathcal{U}_{p_\varepsilon}$, i and $t \in (\alpha,\, \alpha + \delta)$. The normality assumption together with (H1) implies that the symmetric matrix A_{11} is invertible with uniformly bounded inverse with respect to $p \in \mathcal{U}_{p_\varepsilon}$ and τ in compact subsets of $(0, \infty)$. In fact, since $I_i(p) \supset (\alpha,\, \alpha + \delta)$ holds for every i and $p \in \mathcal{U}_{p_\varepsilon}$, the matrix χ_I is an identity matrix on $(\alpha,\, \alpha + \delta)$ for every $p \in \mathcal{U}_{p_\varepsilon}$ and hence

$$
\begin{aligned}
A_{11} \quad &\geq \varepsilon^{-1}\, \tau \textstyle\int_\alpha^{\alpha+\delta} \Phi(1,t)\, B(t)\, B(t)^T\, \Phi(1,t)^T\, dt \\
&= \varepsilon^{-1}\, \tau \Phi(1) \textstyle\int_\alpha^{\alpha+\delta} \Phi(t)^{-1}\, B(t)\, B(t)^T\, \Phi(t)^{-T}\, dt\, \Phi(1)^T
\end{aligned} \tag{2.10}
$$

where τ belongs to the family of closed bounded neighborhoods of τ_ε in $\mathbb{R}^+$.

We assume that

$$
\text{(H3)} \qquad \begin{cases} \text{there exists a bounded neighborhood} \\ \mathcal{U} \subset D_F \subset X \text{ of } (x_\varepsilon,\, p_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon) \text{ and } c > 0 \text{ such that} \\ |A_{21}\, A_{11}^{-1}\, A_{12}| \geq c \text{ for all } (x,\, p,\, u,\, \tau) \in \mathcal{U}. \end{cases}
$$

Theorem 2.4. *If normality,* (H1)–(H3) *hold and* $(x_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon)$ *denotes a solution to* (P_ε) *with associated adjoint* p_ε*, then the semi-smooth Newton algorithm converges superlinearly, provided that the initialization is sufficiently close to* $(x_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon)$.

3. A numerical example

We consider

$$\begin{cases} \min_{\tau\geq 0} \int_0^\tau dt \\ \text{subject to} \\ \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t) \\ |u(t)| \leq 1,\ x(0) = x_0,\ x(\tau) = x_1, \end{cases} \tag{3.1}$$

where

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad x_0 = \begin{pmatrix} 1/2 \\ 1/6 \end{pmatrix}, \qquad x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 25t & 0 \\ 0 & 2t \end{pmatrix} \chi_{\left[0,\frac{1}{3}\right)} + \begin{pmatrix} -t & 0 \\ 0 & 2t \end{pmatrix} \chi_{\left[\frac{1}{3},\frac{1}{2}\right)} + \begin{pmatrix} -t & 0 \\ 0 & -t \end{pmatrix} \chi_{\left[\frac{1}{2},\frac{2}{3}\right]} + \begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix} \chi_{\left(\frac{2}{3},1\right]}.$$

To solve (3.1) numerically a time discretization based on the Crank-Nicolson method on the equidistant mesh with the mesh size $h = 1/(N+1)$ is applied to (2.5). The initialization for the state was chosen as a semicircle connecting x_0 and x_1. Then $u(1)$ and $u(1-h)$ were chosen to be active, τ such that the state equation held and p was chosen so that the transversality condition and the adjoint equation were satisfied. The Newton system (2.5) was solved iteratively and the iteration was stopped when the equation residual was smaller than 10^{-6} in the L^2-norm. The optimal time of steering the state form x_0 to the origin is $\tau^* = 0.272752$. This reference measure was obtained by discretization for $N = 8192$ and $c = 500$, since the exact solution is not available.
In this paper we choose to regularize σ by the ramp functions

$$\sigma_\varepsilon(s_i) \ = \ \begin{cases} -1 & \text{if } s_i \leq -\varepsilon \\ \frac{s_i}{\varepsilon} & \text{if } |s_i| < \varepsilon \\ 1 & \text{if } s_i \geq \varepsilon. \end{cases} \tag{3.2}$$

In Table 1 we show the number of iterates of the Newton iteration (outer loop) that was required for this procedure with respect to the different $c = \frac{1}{\varepsilon}$. Also in Table 1 we depict the optimal minimal times $\tau^*(c)$. These results are obtained for $N = 512$.
Certainly other alternatives are possible for approximation of the function σ, as

for instance

$$\sigma_{atan}(s_i) = \frac{2}{\pi}\, atan\big(\varepsilon^{-1}\, s_i\big). \tag{3.3}$$

This family of $C^\infty-$ functions also has the property that it converges to σ as $\varepsilon \to 0^+$, but it appears to be less appropriate for the purpose of approximating the discontinuous switching structure of the optimal controls. We refer to Table 2, which corresponds to the results in Table 1, only with the approximation procedure (3.2) changed to (3.3).

c	1	10	20
# It	34	37	29
$\tau^*(c)$	0.3115	0.2731	0.2731

TABLE 1. Approximation σ_ε.

c	1	10	50	100	500
# It	19	140	50	84	168
$\tau^*(c)$	0.4765	0.3012	0.277	0.2746	0.2733

TABLE 2. Approximation σ_{atan}.

For the purpose of comparison, the same values for c in the first two computational processes were taken. Clearly c has to be taken significantly larger in the case of (3.3) than for (3.2) to obtain comparable results. The final time for the σ_ε approximation with $c = 10$ is quite accurate when compared to the optimal τ^* and further computations do not considerably improve the accuracy. This is not the case for the σ_{atan} approximation where the result improves as c is increased. With $c = 500$ we obtain the same four-digit accuracy as for $c = 10$ in the σ_ε approximation.

The small discrepancy between this converged value and τ^* is considered to be the discretization error. In some cases, typically at the beginning of the iterations and for the lowest values of c the full Newton step was too large. Therefore we used a one-dimensional line search based on a quadratic polynomial interpolation for the L^2- norm of the residual combined with an Armijo rule.

The graphs for the corresponding controls for σ_ε approximation and $N = 512$ are given in Figure 1. The plots were obtain for $c = 10$. Since the example

(3.1) has two coordinate for the control value, $m = 2$, the upper plots present the first coordinate of the control and the lower ones the second coordinate of u. The first coordinate jumps two times while the second one only one time.

It. no	25	26	27	28	29	30	31	32
c_k	0.94	0.8069	0.5947	0.3814	0.3772	0.3023	0.0307	0.0009

TABLE 3. Approximation σ_ε for $c = 10$ and $N = 512$.

Table 3 shows the quotients $c_k = \dfrac{|u^{k+1} - u^*(c)|_{L^2}}{|u^k - u^*(c)|_{L^2}}$, where $u^*(c)$ is the solution to the discretize version of (3.1) using the ramp function (3.2) for $c = 10$ and $N = 512$. It shows that the algorithm is in fact superlinearly convergent.

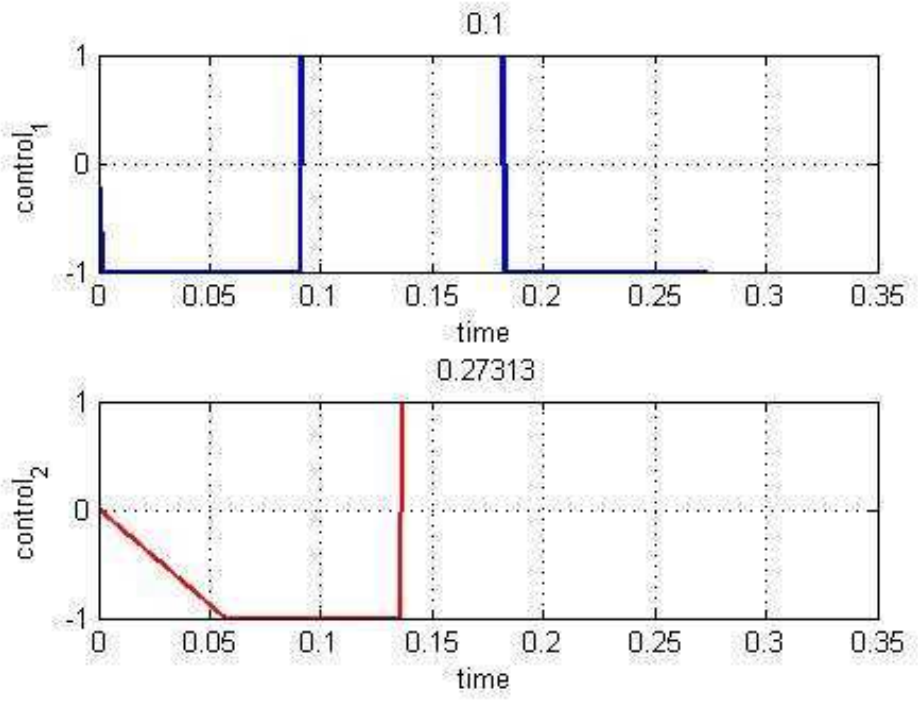


FIGURE 1. Approximation σ_ε for $N = 512$ and $c = 10$.

Precision	*Hardware*	*It.no*	*Time*
Single	Intel Pentium4 2.8GHz	3871	50614.2 ms
Single	Nvidia 8800 GT	3825	996.941 ms
Single	AMD Phemon 9950 2.6GHz	3401	20162.9 ms
Single	Nvidia GTX 280	3859	1073.21 ms
Double	AMD Phemon 9950 2.6GHz	1342	9347.58 ms
Double	Nvidia GTX 280	1342	405.809 ms

TABLE 4. Difference between a CPU and GPU real-time computing at each iteration of Newton's method.

3.1. **CUDA Architecture.** The algorithm for solving the time optimal control problem (3.1) has the outer loop which present the Newton iteration of the semi-smooth Newton method. Each Newton loop has the inner loop which solves the sparse linear system using a CGNR method.

In the Table 1 for $c = 10$ we needed 37 Newton's loop but each Newton loop needs a huge number (3871) of iterations in inner loop to find an appropriate Newton step, see Table 4. Firstly a code was written in the C/C++ language and additionally a GPU enabled version of the code was developed using Nvidia's CUDA technology. The key component for the GPU implementation was an efficient sparse matrix-vector multiplication kernel for the CGNR iteration. Two workstations with two different graphics boards were used for the numerical tests. Overall the single precision benchmark gives a speedup of 51x and 19x. While the double precision benchmark gives a speedup of 23x.

References

[1] C. Büskens and H. Maurer, Sqp-methods for solving optimal control problems with control and state sonstraints:adjoint variables, sensitivity analysis and real-time, *Journal of Computational and Applied Mathematics*, **120**, 2000, 85–108.

[2] H. J. Pesch Ch. Büskens and S. Winderl, Real-time solutions for bang-bang and singular optimal control problems, *Online Optimization of Large Scale Systems*, 2001, 129–142.

[3] H. O. Fattorini, *Infinite dimensional linear control systems: The time optimal and norm optimal problems*, Elsevier, Amsterdam, 2005.

[4] Wendell H. Fleming and Raymond W. Rishel, Deterministic and stochastic optimal control, *Bulletin of the American Mathematical Society*, **82**, 1976, 869–870.

[5] H. Hermes and J. LaSalle, *Functional analysis and time optimal control*, Academic Presse, 1969.

[6] K. Ito and K. Kunisch, *On the Lagrange multiplier approach to variational problems and applications*.

[7] K. Ito and K. Kunisch, The primal-dual active set method for nonlinear optimal control problems with bilateral constraints, *SIAM Journal on Control and Optimization*, **43**, 2004, 357–376.

[8] G. L. Lippi, J.-H. R. Kim and H. Maurer, Minimzing the transition time in lasers by optimal control methods. single-mode semiconductor lasers with homogenous transverse profile, *Physica. D* **191**, 2004, 238–260.

[9] K. Ito, M. Hintermüller and K. Kunisch, The primal–dual active set strategy as a semi–smooth newton method, *SIAM Journal on Optimization*, **13**, 2002, 865–888.

[10] H. Maurer and N. P. Osmolovskii, Second order sufficient optimality conditions for time-optimal bang-bang control, *SIAM J. Control and Optimization*, **42**, 2004, 2239–2263.

[11] M. Hintermüller and K. Kunisch, *Pde-constrained optimization subject to pointwise control and zero- and first-order state constraints*, *Siam Journal on Optimization*, **20**, 2009, 1133–1156.

Institute for Mathematics
and Scientific Computing,
University of Graz,
Heinrichstrasse 36
A-8010 Graz, Austria

E-Mail:jelena.rubesa@uni-graz.at

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

A Metaheuristic Approach to Solving the Generalized Vertex Cover Problem

Marija Milanović

The topic is related to solving the generalized vertex cover problem (GVCP) by genetic algorithm. The problem is NP-hard as a generalization of well-known vertex cover problem which was one of the first problems shown to be NP-hard. The definition of the GVCP and basics of genetic algorithms are described. Details of genetic algorithm and numerical results are presented in [8]. Genetic algorithm obtained high quality solutions in a short period of time.

AMS Subj. Classification: 90C27, 05C85, 90C59

Key Words: vertex cover, genetic algorithm, evolutionary approach, combinatorial optimization, graph algorithms

1. The vertex cover problem

In 1972 Karp [5] showed that 24 diverse problems from graph theory and combinatorics are NP-complete. The vertex cover problem (VCP) was one of them. The VCP is defined over an undirected graph $G = (V, E)$ and searches for a set of vertices S such that for each edge $e \in E$ at least one of its endpoints belongs to S and $|S|$ is as small as possible.

The vertex cover problem has many real-world applications. Examples of the areas where this problem occurs are communications, civil and electrical engineering, and bioinformatics (the vertex cover problem finds applications in the construction of phylogenetic trees, in phenotype identification, and in analysis of microarray data).

Up to now, numerous researchers have studied this problem, mostly from the aspect of approximation. Nevertheless, there is significantly smaller number of researchers who have given experimental results.

In [2] Gilmour and Dras have presented a framework that allows the exploitation of existing techniques and resources to integrate structural knowledge of the problem into the ant colony system metaheuristic, where the structure

was determined through the notion of kernelization from the field of parameterized complexity. They have given experimental results using vertex cover as the problem instance.

In [6] Kotecha and Gambhava have presented a hybrid genetic algorithm for solving the vertex cover problem. They have added local optimization technique to a genetic algorithm, and also developed a new heuristic vertex crossover operator especially for the vertex cover problem.

Pelikan et. al. [11] have analyzed the hierarchical Bayesian optimization algorithm (hBOA) on vertex cover for standard classes of random graphs and transformed SAT instances. The performance of hBOA was compared with that of the branch-and-bound problem solver, the simple genetic algorithm, and the parallel simulated annealing.

Richter [12] has introduced a novel stochastic local search algorithm for the vertex cover problem and tested its performance on a suite of problems drawn from the field of biology, and also on the commonly used DIMACS benchmarks for the related clique problem.

2. **Generalizations**

Several papers dealing with various generalizations of the vertex cover problem are [1, 3, 4, 10].

Broersma et. al. [1] have introduced the minimum weight processor assignment problem (MWPAP), showed that the MWPAP is a generalization of several problems known in the literature, including minimum multiway cut, graph k-colorability, and minimum (generalized) vertex covering. They have analyzed the complexity of the MWPAP, and showed that it is NP-hard, even when restricted to very specific classes of instances. For a number of classes of instances they have shown that the MWPAP is a polynomial.

In [3] Guo et. al. have investigated parameterized complexity of the following problems which all represent different generalizations of the vertex cover problem: connected vertex cover, capacitated vertex cover, and maximum partial vertex cover. They have shown that, with the size of the desired vertex cover as a parameter, the connected vertex cover and the capacitated vertex cover are both fixed-parameter tractable while the maximum partial vertex cover is W[1]-hard.

In his thesis [10], Moser has developed exact algorithms for the same three generalizations of the vertex cover problem.

3. **The generalized vertex cover problem**

In this paper, the formulation from [4] is chosen.

Let $G = (V, E)$ be an undirected graph, with three numbers $d_0(e) \geq d_1(e) \geq d_2(e) \geq 0$ for each edge $e \in E$. The solution is a subset $S \subseteq V$ and $d_i(e)$ represents the cost contributed to the solution by the edge e if exactly i of its endpoints are in the solution. The cost of including a vertex v in the solution is $c(v)$. The solution has a cost that is equal to the sum of the vertex costs and the edge costs. The generalized vertex cover problem (GVCP) is to compute a minimum cost set of vertices.

One of the problems that were a motivation for the generalized vertex cover problem is presented in [7]: given a budget that can be used to upgrade vertices, the goal is to upgrade a vertex set such that in the resulting network the minimum cost spanning tree is minimized.

In [4] Hassin and Levin have studied the complexity of GVCP with the costs $d_0(e) = 1$, $d_1(e) = \alpha$, $d_2(e) = 0$ for every $e \in E$ and $c(v) = \beta$ for every $v \in V$ for all possible values of α and β. They have also provided 2-approximation algorithms for the general case.

In the special case when $d_0(e) = 1$, $d_1(e) = d_2(e) = 0$ for every $e \in E$ and $c(v) = 1$ for every $v \in V$, the GVCP is reduced to the VCP. Thus, the generalized vertex cover problem is NP-hard as a generalization of the vertex cover problem which is proved to be an NP-hard problem. Hassin and Levin have also proved that there are some cases when the GVCP can be solved in polynomial time. Those cases are:

- $\frac{1}{2} \leq \alpha \leq 1$
- $\alpha < \frac{1}{2}$
- $\alpha < \frac{1}{2}$ and there exists an integer $d \geq 3$ such that $d(1-\alpha) \leq \beta \leq (d+1)\alpha$

4. **Mathematical formulation**

Let $G = (V, E)$ be an undirected graph. For every edge $e \in E$ three numbers $d_0(e) \geq d_1(e) \geq d_2(e) \geq 0$ are given and for every vertex $v \in V$ a number $c(v) \geq 0$ is given.

For a subset $S \subseteq V$ denote $\overline{S} = V \backslash S$, $E(S)$ is the set of edges whose both end-vertices are in S, $E(S, \overline{S})$ is the set of edges that connect a vertex from S with a vertex from $\overline{S}$, $c(S) = \sum_{v \in S} c(v)$, and for $i = 0, 1, 2$ $d_i(S) = \sum_{e \in E(S)} d_i(e)$ and $d_i(S, \overline{S}) = \sum_{e \in E(S, \overline{S})} d_i(e)$.

The generalized vertex cover problem is to find a vertex set $S \subseteq V$ that minimizes the cost $c(S) + d_2(S) + d_1(S, \overline{S}) + d_0(\overline{S})$. Thus, the value $d_i(e)$ represents the cost of the edge e if exactly i of its endpoints are included in the solution, and the cost of including a vertex v in the solution is $c(v)$.

v	1	2	3	4
c(v)	1	2	3	4

TABLE 1. c(v) costs

An integer programming formulation of the generalized vertex cover problem, introduced in [4], is shown below.

(1)
$$min \qquad \sum_{i=1}^{n} c(i)x_i \;+\; \sum_{(i,j)\in E} \Big(d_2(i,j)z_{ij} + d_1(i,j)(y_{ij} - z_{ij}) + d_0(i,j)(1 - y_{ij}) \Big)$$

subject to:

(2) $$y_{ij} \leq x_i + x_j \qquad for\ every\ (i,j) \in E$$

(3) $$z_{ij} \leq x_i \qquad for\ every\ i \in V, (i,j) \in E$$

(4) $$z_{ij} \leq x_j \qquad for\ every\ j \in V, (i,j) \in E$$

(5) $$x_i, y_{ij}, z_{ij} \in \{0,1\}$$

x_i is an indicator variable that is equal to 1 if vertex i is included in solution; y_{ij} is an indicator variable that is equal to 1 if at least one of the vertices i and j is included in the solution, z_{ij} is an indicator variable that is equal to 1 if both i and j are included in the solution.

Example 1 Let $|V| = 4$ and $|E| = 5$. The costs $c(v)$ of the including vertices in the solution are given in Table 1. For every edge its end-points and d_0, d_1, and d_2 costs are given in Table 2. The graph is shown in Figure 1.

The optimal objective value in this example is 15 and the generalized vertex cover consists of only one vertex (vertex 1). The corresponding vertex cost $c(S) = c(1) = 1$ and the edge costs are $d_1(S, \overline{S}) = d_1(1,2) + d_1(1,3) + d_1(1,4) = 3 + 4 + 2 = 9$, $d_0(\overline{S}) = d_0(2,3) + d_0(3,4) = 3 + 2 = 5$ and $d_2(S) = 0$. The optimal solution is obtained by CPLEX solver using the integer programming formulation (1)-(5). ♢

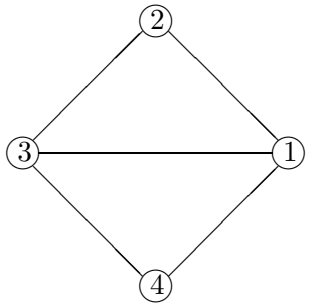


FIGURE 1. The graph

start	end	d_0	d_1	d_2
1	2	5	3	2
1	3	4	4	3
1	4	5	2	2
2	3	3	2	1
3	4	2	2	2

TABLE 2. Edges and their d_0, d_1, d_2 costs

5. **Genetic algorithm**

The genetic algorithms (GAs) are a class of optimization algorithms. The GAs solve problems by creating a population consisting of individuals. An individual represents an encoded solution of the problem. Individuals from the current population are evaluated by using a fitness function to determine their qualities. By applying genetic operators of selection, crossover, and mutation on the current generation, the next generation of individuals is produced. The process is repeated until stopping criterion is satisfied.

Individuals in the initial population are generated randomly or heuristically. The appropriate representation (encoding scheme) of the problem is the most important factor for a successful application of GA.

the selection operator favors better individuals to survive through the generations. The crossover operator provides a recombination of the genetic material by exchanging portions between the parents' genetic codes with the chance that good solutions can generate even better ones. The mutation causes sporadic and random changes by modifying individual's genetic material with some small probability.

The outline of genetic algorithm is shown below:

```
Input_data();
Population_init();
while not Stopping_criterion() do
    Objective_function();
    Fitness_function();
    Selection();
    Crossover();
    Mutation();
endwhile
Output_data();
```

In [9], a detailed description of GA can be found.

6. **Numerical experiments**

In [8], the very first numerical results for the GVCP were presented. The problem was solved by using an evolutionary based approach. A binary representation and standard genetic operators were used along with the appropriate objective function. The experiments were carried out on randomly generated instances with up to 500 vertices and 10000 edges. The performance of the genetic algorithm was compared with the CPLEX solver and 2-approximation algorithm based on LP relaxation.

The integer programming formulation (1)-(5) was implemented and tested by the CPLEX solver in order to obtain optimal solutions. Also, 2-approximation algorithm introduced in [4] was implemented and tested. The 2-approximation algorithm is based on LP relaxation of (1)-(5) integer program, fixing all relaxed binary variables with values greater or equal then $\frac{1}{2}$ to 1. Other variables (with relaxed value less than $\frac{1}{2}$) were fixed to 0. For solving this LP relaxation, the CPLEX solver was also used.

The generation of instances was performed in such way that the cases solvable in polynomial time were omitted. A detailed description of this process can be found in [8].

Numerical experiments have shown that genetic algorithm outperformed both the CPLEX solver and the 2-approximation heuristic.

7. **Conclusions**

The experimental results from [8] indicate that the genetic algorithm approach seems to be a good candidate for solving the GVCP.

Future research will be directed to parallelization of the GA, incorporation in exact methods, and application for solving similar problems.

References

[1] H.J.Broersma, D.Paulusma, G.J.M.Smit, F.Vlaardingerbroek, G.J.Woeginger, The computational complexity of the minimum weight processor assignment problem. *Lecture Notes in Computer Science,* **3353**, 2004, 189–200.

[2] S.Gilmour, M.Dras, Kernelization as heuristic structure for the vertex cover problem. *Lecture Notes in Computer Science*, **4150**, 2006, 452–459.

[3] J. Guo, R. Niedermeier, S. Wernicke, Parameterized complexity of generalized vertex cover problems. *Lecture Notes In Computer Science*, **3608**, 2005, 36–48.

[4] R. Hassin, A. Levin , The minimum generalized vertex cover problem. *ACM Transactions on Algorithms*, **2**, 2006, 66–78.

[5] R. M. Karp, Reducibility among combinatorial problems. *Complexity of Computer Computations, Plenum Press*, 1972, 85-103.

[6] K. Kotecha, N. Gambhava , A hybrid genetic algorithm for minimum vertex cover problem. *Proceedings of the First Indian International Conference on Artificial Intelligence*, 2003, 904–913.

[7] S.O. Krumke, M.V. Marathe, H. Noltemeier, R. Ravi, S.S. Ravi, R. Sundaram, H.C. Wirth, Improving minimum cost spanning trees by upgrading nodes. *Journal of Algorithms*, **3**, 1999, 92-111.

[8] M. Milanović, Solving the generalized vertex cover problem by genetic algorithm. *Computing and Informatics*, in press.

[9] M. Mitchell, *Introduction to genetic algorithms.* MIT Press, Cambridge, Massachusetts, 1999.

[10] H. Moser, *Exact algorithms for generalizations of vertex cover.* M.Sc. thesis, Friedrich-Schiller-University Jena, Faculty of Mathematics and Informatics, 2005.

[11] M.Pelikan, R.Kalapala, A.K.Hartmann, Hybrid evolutionary algorithms on minimum vertex cover for random graphs.*Proceedings of the Genetic and Evolutionary Computation Conference-GECCO(2007)*,547–554.

[12] S. Richter, M. Helmert, C. Gretton, A stochastic local search approach to vertex cover. *Lecture Notes in Computer Science*, **4667**, 2007, 412–426.

Faculty of Mathematics,
University of Belgrade,
Studentski trg 16/IV
11 000 Belgrade, Serbia
E-Mail: marija.milanovic@gmail.com

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Multiplicative Systems on Ultra-Metric Spaces

Nacima Memić

We perform analysis of certain aspects of approximation in multiplicative systems that appear as duals of ultrametric structures, e.g. in cases of local fields, totally disconnected Abelian groups satisfying the second axiom of countability or more general ultrametric spaces that do not necessarily possess a group structure. Using the fact that the unit sphere of a local field is a Vilenkin group, we introduce a new concept of differentiation in the field of p-adic numbers. Some well known convergence tests are generalized to unbounded Vilenkin groups, i.e. to the setting where the standard boundedness assumption related to the sequence of subgroups generating the underlying topology is absent. A new Fourier multiplier theorem for Hardy spaces on such locally compact groups is obtained. The strong L_q, $q > 1$, and weak L_1 boundedness of Fourier partial sums operators in the system constructed on more general ultrametric spaces is proved.

AMS Subj. Classification: MSC2010: 42C10, 43A50, 43A75

Key Words: p-adic derivative, Vilenkin group, Fourier multiplier, multiplicative system, ultrametric space

1. Introduction

The field of reals $\mathbb{R}$, obtained as a completion of rationals $\mathbb{Q}$ with respect to the metric generated by the ordinary absolute value as the Archimedean norm on $\mathbb{Q}$, is just one among infinitely many completions of the rationals. According to the Ostrowski theorem, any nontrivial norm on $\mathbb{Q}$ is either the ordinary absolute value or a p-adic norm for some prime number p. The completion of the field $\mathbb{Q}$ with respect to the p-adic norm leads to the field $\mathbb{Q}_p$ of p-adic numbers. Any p-adic norm is non-Archimedean. Twenty years ago, when Volovich [16] explicitely stated the hypothesis of non-Archimedean structure of space-time at ultra-small distances, the already rich areas of applications of p-adic analysis to number theory or algebraic geometry were extended also into direction of the mathematical physics (see. e.g. [15]) .

Ultrametric analysis we are concerned with concentrates on complex valued functions of an argument belonging to an ultrametric space. (Tate's thesis [14] is the most remarkable representative in this interpretation of p-adic analysis.)

p-Adic differential calculus differs from the real case since piece-wise constant functions depending on a finite number of digits have vanishing derivative. All approaches to a p-adic derivative of complex valued functions have been based on the idea that additive characters should be eigenfunctions of the differentiation operator ([4], [11]). Our attempt is to use the multiplicative system of multiplicative characters on the unit sphere and then extend the p-adic derivative to the field $\mathbb{Q}_p$. The first of three parts of the dissertation is devoted to this goal.

The additive group of the ring of integers of a local field as well as the multiplicative group of units in such a field are particular examples of so-called Vilenkin groups ([1]). In the second part, we are concerned with certain aspects of harmonic analysis on Vilenkin groups that are present in the general setting, i.e. without the boundedness assumption related to the sequence of subgroups that determines the topology of a group under consideration.

Further generalizations to multiplicative systems constructed on compact ultra-metric spaces that need not have a group structure form the content of the concluding part of the dissertation.

2. Differentiation on local fields

Differentiation on totally disconnected local fields is based on the classical relation that the differentiation operator should be diagonalized by some given orthogonal system. We use this relation to define a derivative on the p-adic field and the dyadic field.

2.1. **Differentiation on the p-adic field.** Let $\mathbb{Q}_p$ be the field of p-adic numbers endowed with the p-adic norm $\|.\|_p$. For every $\gamma \in \mathbb{Z}$, the sphere of radius p^γ is given by

$S_\gamma = \{x \in \mathbb{Q}_p : \|x\|_p = p^\gamma\}$.

Denote by $(\theta_n)_n$ the system of multiplicative characters on S_0.

We extend the characters θ_n to $\mathbb{Q}_p^*$ by the relation $\theta_n(x) = \frac{1}{\|x\|^{\frac{1}{2}}}\theta_n(\|x\|x)$.

We introduce a new definition of derivative on the field of p-adic numbers.

Definition 2.1. Let φ be a locally constant function. If the series $\sum_{n=0}^{\infty} n^{\alpha}\hat{\varphi}_{\gamma}(n)\theta_n(x)$ converges at a point $x \in S_{\gamma}$ for some $\alpha > 0$, then the function φ is said to be α-differentiable at x and $\varphi^{(\alpha)}(x) = \sum_{n=0}^{\infty} n^{\alpha}\hat{\varphi}_{\gamma}(n)\theta_n(x)$ is called its α-derivative at x.

The following properties of the α-derivative are proved.

Lemma 2.2. *Locally constant functions are infinitely differentiable.*

Lemma 2.3. *The α-derivative of any radial function is equal to 0. Every locally constant function whose derivatives of integer orders vanish is radial.*

Theorem 2.4. *Let φ be a locally constant function. For $\alpha > 0$, the α-derivative of φ has the form $\varphi^{(\alpha)}(x) = k^{\alpha}\varphi(x)$ if and only if $\varphi(x) = M(x)\theta_k(x)$ for some radial function $M(x)$.*

Proposition 2.5. *If a function φ is α-differentiable at some $x_0 \neq 0$, then the function $\varphi_z(x) = \varphi(\frac{x}{z})$ is α-differentiable at $x_0 z$ for every $z \neq 0$, and $\varphi_z^{(m)}(x_0 z) = (\varphi^{(m)})_z(x_0 z)$.*

Definition 2.6. f is a regular distribution if it is defined by

$$\langle f, \varphi\rangle = \int \psi(x)\overline{\varphi}(x)dx\,,$$

for any test function φ, where ψ is a fixed locally constant function.

Proposition 2.7. *The α-derivative of a regular distribution f defined by a locally constant function ψ is given by $\langle f^{(\alpha)}, \varphi\rangle = \langle \psi, \varphi^{(\alpha)}\rangle = \langle \psi^{(\alpha)}, \varphi\rangle$.*

Theorem 2.8. *Let f be a distribution on $\mathbb{Q}_p^*$ whose α-derivative is equal to $k^{\alpha}f$. Then $\langle f, \varphi\rangle = \int \theta_k(x)N(x)\overline{\varphi}(x)dx$ for every test function φ, where $N(x)$ is a fixed radial function .*

Definition 2.9. The multiplicative convolution of functions φ and ψ is given by the formula $(\varphi * \psi)(x) = \int_{S_0} \varphi(\frac{x}{t})\psi(t)dt, x \in \mathbb{Q}_p^*$, if the integral converges.

Notice that $\varphi * \psi \neq \psi * \varphi$.

The next proposition illustrates some properties of the multiplicative convolution.

Proposition 2.10.

(1) *On S_0, the relation $(\varphi * \psi)_\gamma = \varphi_\gamma * \psi$ holds.*
(2) *There is no identity element of the multiplicative convolution in the space of locally integrable functions.*

2.2. **Differentiation on the dyadic field.** On $\mathbb{R}_+$ endowed with the dyadic norm, the pseudo-differential operator with symbol x^α is defined as follows.

Definition 2.11. Let $\alpha \in \mathbb{R} \setminus \{-1\}$. We define a distribution $\Lambda^{\{\alpha\}}$ on $\mathbb{R}_+$ by

$$\langle \Lambda^{\{\alpha\}}, \varphi \rangle := \langle t^\alpha, F\varphi \rangle$$

where $\langle x^\alpha, \varphi \rangle := \int_0^{+\infty} x^\alpha \varphi(x) dx$, if $\alpha > -1$, and $\langle x^\alpha, \varphi \rangle := \int_0^{+\infty} x^\alpha (\varphi(x) - \varphi(0)) dx$, when $\alpha < -1$.

For $f \in D'(\mathbb{R}_+)$, we put $D^\alpha f := f * \Lambda^{\{\alpha\}}$ if the convolution exists.

For all $x \in \mathbb{R}_+$, $n \in \mathbb{N}$, let $x_n = [2^n x](mod 2)$,and $x_{-n} = [2^{1-n} x](mod 2)$.
As $x_{-n} = 0$ for n sufficiently large, the functions
$t(x,y) = \sum_{n=1}^{\infty} (x_n y_{-n} + x_{-n} y_n)$ and $\psi(x,y) = (-1)^{t(x,y)}$
are well defined on $\mathbb{R}_+ \times \mathbb{R}_+$.

Proposition 2.12. *Let $\alpha \in \mathbb{R}\setminus\{-1\}$, then we have $D^\alpha \psi(x,y)(y) = x^\alpha \psi(x,y)(y)$.*

3. Vilenkin groups

In this chapter we generalize Salem's and Lebesgue's tests for convergence of Fourier Vilenkin series to unbounded Vilenkin groups.

3.1. **A local Salem test on unbounded Vilenkin groups.** Theorem 1 obtained in [2] can be generalized to unbounded Vilenkin groups as follows.

Theorem 3.1. *Let G be a Vilenkin group, and f a continuous function at some point x satisfying:*

(1) $p_{k+1} \sup_{t \in U_k} |f(x+t) - f(x)| = o(1), \quad k \to \infty,$

(2) $$\lim_{l \to \infty} \lim_{k \to \infty} \sup_{t \in U_{k+1}} (C_k + 1) \sum_{\alpha=1}^{\frac{m_k}{m_l} - 1} \frac{1}{\alpha} \left| \sum_{j=0}^{p_{k+1}-1} f(x - z_\alpha^{(k)} - j x_k - t) \zeta_k^{j a_k} \right| = 0,$$

uniformly in $1 \le a_k < p_{k+1}$, *where* $C_k = \sup\limits_{i \le k} \dfrac{p_i}{p_{k+1}}$.

Then, $\lim\limits_{n\to\infty} S_n(x,f) = f(x)$.

3.2. **Lebesgue test on unbounded Vilenkin groups.** We generalize the results of [7] to unbounded Vilenkin groups.

Definition 3.2. We introduce the function

$$f^{\diamond}(x) = lim_{k\to\infty}\frac{1}{m(G_k)}\int_{G_k} f(x-t)\sum_{r=0}^{a_k}\chi_{m_k}^r(t)dt\,,$$

if the limit exists uniformly with respect to $a_k \in \{0,1,2,\dots,p_{k+1}-1\}$, at the point $x \in G$.

Definition 3.3. An element $x \in G$ is said to be a Lebesgue point of an integrable function f if

$$\frac{1}{m(G_k)}\int_{G_k}|f(x+t)-f(x)|dt = o(1)(k\to\infty).$$

Theorem 3.4. *If* G *is bounded and* x *is a Lebesgue point of* f, *then* $f^{\diamond}(x)$ *exists and is equal to* $f(x)$. *However, the existence of* $f^{\diamond}(x)$ *does not imply that* x *is a Lebesgue point. Moreover, if* G *is unbounded then* $f^{\diamond}(x)$ *need not exist even at Lebesgue points.*

Theorem 3.5. *Let* G *be any Vilenkin group, and* $f \in L^1(G)$. *Let* $n = a_k m_k + r$, *where* $1 \le a_k < p_k$ *and* $r < m_k$. *Suppose that*

$$f^{\diamond}(x) = lim_{k\to\infty}\frac{1}{m(G_k)}\int_{G_k} f(x-t)\sum_{r=0}^{a_k}\chi_{m_k}^r(t)dt$$

exists uniformly in $a_k \in \{0,1,2,\dots,p_{k+1}-1\}$ *at a point* $x \in G$. *Then*

$$S_n(f;x) - f^{\diamond}(x) = o(1) + \sum_{\alpha=1}^{m_k-1}\chi_{m_k}^{a_k}(z_\alpha^{(k)})D_r(z_\alpha^{(k)})\int_{G_k} f(x-z_\alpha^{(k)}-t)\chi_{m_k}^{a_k}(t)dt$$

as $n\to\infty$. *Thus the necessary and sufficient condition that the Fourier series of* f *converges at* x *is that*

$$(*)\ \sum_{\alpha=1}^{m_k-1}\chi_{m_k}^{a_k}(z_\alpha^{(k)})D_r(z_\alpha^{(k)})\int_{G_k} f(x-z_\alpha^{(k)}-t)\chi_{m_k}^{a_k}(t)dt = o(1)$$

uniformly in a_k *and* r *as* $k\to\infty$.

3.3. **Fourier multipliers on Hardy spaces.**

Definition 3.6. A complex function a is called an atom on G if

(1) $supp(a) \subset y + G_n$,
(2) $\|a\|_\infty \leq \frac{1}{\mu(G_n)}$,
(3) $\int_G a(x)dx = 0$.

If the group G is compact, then the function $a \equiv 1$ is also considered as an atom.

The atomic Hardy space H^1 consists of integrable functions f which can be represented as $f = \sum_{i=1}^{\infty} \lambda_i a_i$, where each a_i is an atom and $\sum_{i=1}^{\infty} |\lambda_i| < +\infty$.

The norm in H^1 is given by $\|f\|_{H^1} = \inf \sum_{i=1}^{\infty} |\lambda_i|$, where the infimum is taken over all such decompositions of f.

Definition 3.7. For any distribution f, let $Mf(x) = \sup_n |f * (\mu(G_n))^{-1} 1_{G_n}(x)|$.

The space H consists of all distributions f such that $Mf \in L^1$. The norm is given by $\|f\|_H = \|Mf\|_1$.

Onneweer and Quek [12] proved that $H = H^1$ on bounded locally compact Vilenkin groups.

G. Gat [8] established the strict inclusion $H^1 \subsetneq H$ on some compact unbounded Vilenkin groups.

There exists a maximal function that defines the space H^1 on compact unbounded Vilenkin groups constructed by P.Simon [13].

Our results obtained in [3] show that the situation remains the same on locally compact unbounded Vilenkin groups. Namely, the maximal function

$\tilde{M}f(x) = sup_{n,I_n} |f * (\mu(I_n))^{-1} 1_{I_n}(x)|$, generates the space H^1 on both bounded and unbounded locally compact Vilenkin groups, where I_n is an interval of the form $I_n = \biguplus_{i=\alpha}^{\beta} ix_n + G_{n+1}$ $0 \leq \alpha \leq \beta < p_{n+1}$.

Definition 3.8. Let $\phi \in L^\infty(\Gamma)$. ϕ is a multiplier in H^1 if the operator $Tf = (\phi f^\wedge)^\vee$ is bounded on H^1.

ϕ is a multiplier if and only if $Tf = (\phi f^\wedge)^\vee$ is bounded on the set of atoms.

Theorem 3.9.

In the next theorem, we generalize previous results of Kitada [10], Daly-Phillips [6] and Theorem 2(i) of [5] to the case of unbounded locally compact Vilenkin groups.

Theorem 3.10. *([3]) Let G be any Vilenkin group, $\phi \in L^\infty(\Gamma)$ and*
$\sup_N \int_{G_N^c} |(\phi - \phi_{N+1})^\vee(y)|dy = O(1)$,
where $\phi_{N+1} = \phi 1_{\Gamma_{N+1}}$ and $\wedge, \vee$ denote respectively the Fourier transform and the inverse Fourier transform. Then ϕ is a multiplier on H^1.

An example has been constructed to show that our estimate is sharper in comparison with Daly and Phillips' results.

The following result is an extension of the Marcinkiewicz multiplier theorem for Hardy spaces in the bounded case.

Corollary 3.11. *([3]) If $\phi \in L^\infty(\Gamma)$ on a bounded compact Vilenkin group G fulfills the requirement*

$$m_N^{p-1} \sum_{k=m_{N+1}}^{m_{N+2}-1} |\triangle\phi(k)|^p = O(1) \text{ for some } p \in (1,2]\,,$$

where $\triangle\phi(k) = \phi(k) - \phi(k+1)$, then $\phi \in m(H^1)$.

4. Multiplicative systems on ultra-metric spaces

Our setting in this chapter are ultrametric spaces that need not possess a group structure. We construct a multiplicative system $(\chi_n)_n$ on a given space X and deduce some of its basic properties.

Definition 4.1. Let G be a compact, 0-dimensional metric space. Suppose that $(C_n)_n$ is a sequence of covers of G with the following properties:

(1) Elements of a given C_n are disjoint and clopen.
(2) Each element of C_n is properly contained in some element of C_{n-1}.
(3) $C_0 = G$.
(4) $\bigcup_n C_n$ is a base for the topology of G.

Proposition 4.2. *If for every n, all elements of C_n are given the same measure, then X is homeomorphic to some additive Vilenkin group G determined by the sequence $(p_n)_n$, where p_n is the number of elements from C_{n+1} contained in one element from C_n.*

In the general case, when elements from C_n need not have the same measure, except for those contained in the same element of C_{n-1}, a multiplicative system is constructed with the following properties.

Proposition 4.3.

(1) $\int \chi_n = 0$ *for every* $n \neq 0$.
(2) *The family* $(\chi_n)_n$ *is orthonormal.*

If $U_n(x)$ is the unique element from C_n that contains the point x, then $m_n(x)$ will denote the measure of $U_n(x)$. The analogue of the property $D_{m_n}(t) = m_n 1_{G_n}(t)$ of the Dirichlet kernel is given in the following proposition.

Proposition 4.4. *Let* $x, t \in X$, *and* $n \geq 0$. *Then we have*

$$\sum_{k=0}^{m_n(x)-1} \chi_k(x)\overline{\chi}_k(t) = m_n(x) 1_{U_n(x)}(t)\,.$$

An analogue of the Calderon-Zygmund decomposition is also proved.

Lemma 4.5. *Let* $f \in L^1(X)$, $y > 0$ *and* $(\alpha_n)_n$ *be a sequence of integers. Suppose* $\|f\|_1 \leq y$. *Then, there exist functions* g, b *and a sequence* $B = \{\omega_j\}$ *of disjoint intervals of* X *such that:*

(1) $f = g + b$.
(2) $|g| \leq Cy$, *a.e.*
(3) $\|g\|_1 \leq \|f\|_1$
(4) $B = \bigcup_{n=0}^{\infty} B_n$, *where every* $\omega_j \in B_n$ *is strictly contained in some element from* C_n, *and forms a union of elements of* C_{n+1}.
(5) b *is supported in* $\bigcup_j \omega_j$.
(6) $\int_{\omega_j} b = 0$ *for every* $\omega_j \in B$ *and* $\int_{\omega_j} b\theta_n^{\alpha_n} = 0$, *if* $\omega_j \in B_n$.
(7) $\int_{\omega_j} |b| \leq C \int_{\omega_j} |f|$ *for every* $\omega_j \in B$.
(8) $\sum_j \mu(\omega_j) \leq y^{-1}\|f\|_1$.

If S_n is the n-th partial sum with respect to the system $(\chi_k)_k$, then using the previous lemma we obtain the following result on boundedness of the respective operators.

Theorem 4.6. *There exist constants C_p, $p \geq 1$, such that*

(1) $\|S_n f\|_p \leq C_p \|f\|_p$, *and*

(2) $\mu(\{|S_n f| > y\}) \leq C_1 y^{-1} \|f\|_1$.

Finally, on a class of locally compact ultra-metric spaces, we construct a wavelet system of a Khrennikov-Kozyrev type ([9]) and prove that it is a basis of eigenfunctions of an ultra-metric diffusion operator.

The operator has the following form:

$$Tf(x) = \int T(x,y)(f(x) - f(y))dy\,,$$

where the kernel $T(x,y)$ is symmetric, positive, locally constant and only depends on the distance $\|x-y\|_p$.

For the system of functions $\psi_{n,k,j}(x) = \dfrac{1_{U_n^k}(x) e^{\frac{2\pi i l j}{p_n^k}}}{\sqrt{\mu(U_n^k)}}$, where U_n^k is a basis for the topology of the space, and the numbers p_n^k, l depend on U_n^k and x respectively, we prove the following:

Theorem 4.7. *The system $\psi_{n,k,j}$ is an orthonormal total basis for the operator T.*

References

[1] G. N. Agaev, N. Ya. Vilenkin, G. M. Dzhafarli and A. I. Rubinshtein, *Mul´tiplikativnye sistemy funkcii i garmonicheskii analiz na nul´mernykh gruppakh*, Elm, Baku 1981.

[2] M. Avdispahić, Concepts of generalized variation on Vilenkin groups and convergence of Fourier-Vilenkin series, *Colloquia Mathematica Societies Janos Bolyai* **49.** *Alfred Haar Memorial Conference, Budapest (1985)*, North Holland, Amsterdam 1987, 145-163.

[3] M. Avdispahić and N. Memić, Fourier multiplier theorem for atomic Hardy spaces on unbounded Vilenkin groups, *J. Math. Anal. Appl.* **363**, no. 2, 2010, 588-595.

[4] P. L. Butzer and H. J. Wagner, Walsh Fourier series and the concept of a derivative, *Applicable Anal.* **3**, 1973, 29-46.

[5] J. Daly and S. Fridli, Translation invariant operators on Hardy spaces over Vilenkin groups, *Acta Math. Acad. Paedagog. Nyhazi* **20**, no. 2, 2004, 131-140.

[6] J. Daly and K. Phillips, A note on H^1 multipliers for locally compact Vilenkin groups, *Canad. Math. Bull.* **41**, 1998, 392-397.

[7] David H. Dezern, *Fourier series on Vilenkin groups*, Ph.D. Dissertation, Syracuse University, New York 1988.

[8] G. Gat, Investigations of certain operators with respect to the Vilenkin system, *Acta. Math. Hungar.* **61**, no. 1-2, 1993, 131-149.

[9] A.Yu.Khrennikov, S. V. Kozyrev, Wavelets on ultra metric spaces, *Applied and Computational Harmonic Analysis*. 2005. V. 198. N.1.P.103-126.

[10] T. Kitada, H^p Multiplier theorems on certain totally disconnected groups. *Sci. Rep. Hirosaki Univ.* **34** (1987), 1-7.

[11] C. W. Onneweer, On the definition of dyadic differentiation, *Applicable Anal.***9**, 1979, 267-278.

[12] C. W. Onneweer and T. S. Quek, Multipliers on weighted Hardy spaces over locally compact Vilenkin groups, *J. Austral. Math. Soc.* **48**, 1990, 472-496.

[13] P. Simon, Investigations with respect to the Vilenkin system, *Ann. Univ. Sci. Budapest. Sect. Math.* **27**, 1982, 87-101.

[14] J. T. Tate, *Fourier analysis in number fields and Hecke's zeta functions*, Ph.D. Thesis, Princeton University, Princeton, N.J. 1950; reproduced in: *Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965)*, 305-347, Thompson, Washington, D.C. 1967

[15] V.S.Vladimirov, I. V. Volovich, Ye.I.Zelenov, *p-Adic analysis and mathematical physics*, World Scientific, Singapore 1994.

[16] I. V. Volovich, Number theory as the ultimate physical theory, CERN-TH 87, 1987, 4781-4786.

Department of Mathematics,
University of Sarajevo,
Zmaja od Bosne 33-35,
71000 Sarajevo,
Bosnia and Herzegovina

E-Mail: nacima.o@gmail.com

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Organization and Security of the Audio and Video Archive for Unique Bulgarian Bells

Nikolay G. Noev

The purpose of this investigation is to study and to identify the most valuable unique bells as well as to develop a digital archive with the help of advanced technologies. The main tasks are: development of the audio archive of information gathered from artifacts, analysis, optimization and addition of metadata for indexing of digital data, compression and data protection, prevention of data loss, design, organization and maintenance of archive. We investigate the methods of protection with watermarking which can be used against illegal use of data. We create and protect samples for additional applications and web sites.

AMS Subj. Classification: H.3.7 Digital Libraries, K.6.5 Security and Protection

Key Words: digital archives, indexing of digital data, data protection

1. Introduction

The key to prosperity in today's world is access to digital content and skills to create new content. Digitization of analog materials and the creation of digital resources in the field of cultural heritage is a major contributor to e-Europe. The aim of this article is to study and identify several dozens of the most valuable bells in Bulgarian churches, monasteries and museums as well as to develop an audio, photo and video digital archive (with the help of advanced technologies) for analysis, preservation and protection of the data. Main tasks in the research are:

- Development of audio, photo and video archive with information collected by artifacts;
- Analysis and indexing the digital data;
- Design and maintenance of digital archive;
- Compression and optimization of the archive;
- Preventing data loss;
- Development of software for adding watermark against illegal use of data;
- Development of functions for creating, formatting and protection of samples for additional applications and web-sites.

In the second chapter we consider modern methods for digitization and data collection. In the third chapter we provide an analysis of the objects in order to determine the metadata of selected artifacts from selected collections and problem areas. Fourth chapter contains research of advanced technologies and methods for protecting intellectual property. In the fourth chapter has examined the organization of the digital archive with unique Bulgarian bells.

2. **Digitization and data collection**

Modern technologies have changed the way the information was presented in the archives and have made possible new services, unthinkable a few years ago.

Digitization is creation of an object, image, audio, document, or a signal (usually an analog signal) from a discrete set of its points or samples. The result is called "digital image", or more specifically "digital images" of the object and "digital form of the signal". The tasks of digitalization can be synthesized in certain key areas:

- Retention of funds and records - many of digitalized objects are fragile or brittle structure is influenced by weather conditions and over time their digitization is the only hope for preservation;
- Simultaneous access to materials - most objects are subject to the digitization of rare and unique items of historical past and have a priceless value, the process of digitization will allow more users involved to touch them;
- Conservation funds in digital formats - archives, websites, digital libraries. Strengthening international exchange and promotion;
- Providing access via computer networks - easy access to digital archives, access to records of persons with disabilities;
- Providing new opportunities to work with digitalized materials funds - all the functionality is available to users of web space to be copied, multiplied, forwarded, etc., without jeopardizing the its integrity and strength;
- Full text search - digital archives organization contributes to easier detection of the searched object among all the crowd, advanced search, unification of search results;
- Classification of digital funds via metadata - entire photo meta data wealth funds may be accompanied by important information about copyright, creation date, identification number, etc.

Development of digital technology hit the storage and processing of information. Today, almost every unit of information is created digitally: digital photography, digital sound recording, digital communications, text, saved in a file, videos, movies, multimedia presentations stored on digital media, etc.

Storage of such digital multimedia data in digital archives became subject to the same challenges, such as archives, we know before the invention of digital computing devices.

3. **Analysis, optimization and indexing the digital data**

3.1. **Essence of metadata.** Metadata are text fields, built-in media files or additional text files (XML, XMP) for recording information on the nature of the digital resource. In other words, meta data is "data about data" describing any electronic or non-electronic source of information on pre-established standard. Meta data contribute to finding and sharing information. In other words, this is the last stage of information management. Once digitized (implemented in electronic form) and structured (arranged in a specific sequence and line), information is visualized in an appropriate form. To be fully justified the efforts of both phases scheduled for managing the information it needs to be found and used by as many users. This is possible thanks to the meta data [3]. The presence of meta data with correctly placed points of connection ensures speed and accuracy of the application, and interactive user interaction. The shared experience of developers and users for the metabase warrants defining them as a new generation of data, integrating new technologies and therefore requires a new legal and economic regulation.

3.2. **Analysis of the objects in order to determine the metadata of selected artifacts collections and problem areas.** Following the studies and consultations with specialists, it was concluded the following organization of metadata:

- Title (name of subject);
- Creator (name of digitalizer);
- Description (additional data);
- Date (date of creation);
- Type (type of media);
- Format (file format, codec and parameters);
- Identifier (geographic coordinates);
- Rights (owner of property rights).

3.3. **Adding text metadata fields.** The process of creating the meta data to digitize files could be done by several basic approaches:

- Using the software that serves as a system for collecting and managing the meta data for individual sites in the process of digitization. The system can be either web-oriented and desktop application depending on team needs for digitization. Data can be stored in a database and easily accessible for further processing;

- Using the application to scan individual files and embed and extract from them the necessary meta data. For this purpose, one may use different standards for meta data and software tools for working with them. An example is the standard Adobe XMP and means of work with it Adobe XMP ToolKit;
- Presentation of metadata in a text editor, subject to certain rules and the standard XMP templates and languages used to define meta fields. The name of the file should completely coincide with the name of the digital resource, but with a different extension;
- Using embedded resources to work with meta data used to support the digitization software or operating systems. They keep the information that is specified in the study as necessary meta data in this digitization project.

There are three ways to view the meta data:

- Main menu interface: For example, from MS Word document a user can choose File from the menu option Properties, then on General, Summary, Statistics, Contents, Custom to find the main categories of meta data embedded document;
- Computer mouse: By positioning the mouse pointer on the version in MS Word document, which has enabled feature Track Changes, a dialog box informs who made and when the change was made;
- Software for viewing metadata. Windows Vista and Windows 7 operating system has built in tools for meta data. Thus facilitate the process of organizing files with tags "metadata" which indicate that the information belongs to a project or category.

FotoWare FotoStation [4] application language uses XML technology to define META text boxes. Definitions can be saved by RDF language for ontologies. META text fields are added to the digital archive unit as follows:

- Indicate the object(s) for META data adding;
- The Metadata button is selected from the toolbar or right-click the selected object is selected, then the menu is selected metadata edit, and which is made by pressing keys (Ctrl + T);
- From the window to edit the META fields cause textual data.

4. **Modern techniques for intellectual property protection**

With the development of digital technologies increasing part of the audio, video and any other information is available for fast, easy and high quality copying. This fact entails the problem of protecting information from unauthorized distribution. Research in this area is considered in several aspects. One of the most important of these is steganography [5], [6]. Steganography deals

with the concealment of information, steganography hides the message which should remain hidden. Like steganography, watermark protection aims carrying hidden information. However, there are significant differences between the two techniques. Digital watermark is visible or preferably invisible to the identification code that is permanently embedded into digital data and maintained a presence in them after extracting it [2].

4.1. **Methods for image watermarking in the spatial region.** In these methods data are incorporated directly into the original image. The main advantage is that the key is not necessary to do any preconditions transformations. The watermark is embedded by changing the illumination or color components. The main disadvantage is the low resistance. An example of this method is the method of Kutter [7]. To derive the value of the embedded bit is calculated assumed difference between value and actual value of the pixels. The sign of the difference determines the value of the embedded bit. Extracting bits is done without the knowledge of the original message. The method is robust to filtering, JPEG compression and geometric transformations.

4.2. **Methods for audio watermarking using low-bit coding.** By replacing the least significant bit of each sampling point by a coded binary string, we can encode a large amount of data in an audio signal The major disadvantage of this method is its poor immunity to manipulation. Encoded information can be destroyed by channel noise, resembling etc. We improve robustness using by error-correcting codes.

4.3. **Visual watermarking.** Visible mark added in digital picture and video records.

5. Approaches and tools for building digital archive of unique Bulgarian bells

In our research we also consider some previous experience for creation of digital archives [1], [8].

We use the fooling software tools in order to digitize and secure objects:

- Photo image processing - Adobe Photoshop, Fotoware Fotostation;
- Sound processing - specialized acoustic software from Bruel and Kjer;
- Text include - MS word;
- Video - Adobe Premiere, Virtual Dub;
- Watermarking - our own software;
- Media file organizer - Fotoware Fotostation, Fotoware Camelion.

We also use specialized hardware tools as follows:

- Photo-camera - Sony Alpha DSLR-A100;
- Video-camera - Sony HDR-SR8E HD AVCHD Camcorder;

FIGURE 1. Visual watermarking example.

- Audio-system - PULSE11 from company Bruel and Kjer.

Archive creation and organization is performed using Fotoware Fotostation, Camelion. This software package has following main characteristics:

- Media file organizer;
- Automating repetitive tasks - actions;
- Share files via network or upload to internet;
- Create web pages from workflows;
- User-friendly interface that is available in 10 different languages;
- Sophisticated full text search;
- Import from digital cameras, scanners, DVD, CD or any other source;
- Storage capacity of up to 200 000 files.

Organization of the archive can be separated in following steps:

- Creating an archive;
- Creating own functions - actions, macros;
- Adding objects to a digital archive;
- Converting file format, codec, size;
- Adding META data to objects in a digital media collection;
- Adding watermarking - visible and invisible.

Acknowledgement

For this work I want to thank Galina Bogdanova and Todor Todorov for collaboration and colleagues connected with the project "BELL - Research and Identification of Valuable Bells of the Historic and Culture Heritage of Bulgaria and Development of Audio and Video".

References

[1] G. Bogdanova, R. Pavlov, G. Todorov, V. Mateeva,, Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage. *Advances in Bulgarian Science Knowledge, National Centre for Information and Documentation*, **3**, 2006,7–15.

[2] I.Cox, J. Kilian, T. Leighton, G. Shamoon, Secure spread spectrum watermarking for multimedia. In *Proceedings of the IEEE International Conference on Image Processing*, **6**, 1997.

[3] Dublin Core Metadata Initiative - *http://dublincore.org/groups/education.*

[4] Fotoware FotoStation - *http://www.fotoware.com/en/Products/FotoStation/.*

[5] G.Gribunin, I., Okov, I. Turincev, *Cifrovaia steganographia.* Solon-Press, 2002.

[6] A. Karasev, Komputernaia tainopis, grafika i zvuk priobretaut podtekst. *Mir PK.* **1/97**, 1997, 132-134.

[7] M. Kutter, Digital Signature of Color Images using Amplitude Modulation. *Journal of Electronic Imaging*, 1998, 326–332.

[8] D. Paneva, K. Rangochev, D. Luchev, Knowledge Technologies for Description of the Semantics of the Bulgarian Folklore Heritage. In *Proceedings of the Fifth International Conference Information Research and Applications* i.Tech, Varna, Bulgaria, **1**, 2007, 19–25.

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev str., block 8,
1113 Sofia, Bulgaria

E-Mail:nickey.noev@gmail.com

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

File Format for Storage of Multimedia Information

Peter Armyanov

This article studies problems referring to storage, editing and streaming vector graphics and animation. The advantages and disadvantages of existing formats are reviewed and some suggestions to solve the problems are given. An approach for generating vector stream in a format, that allows embedding of additional media such audio and descriptive meta data is introduced. In parallel the problem of conversion multimedia files from one container format to another is described and solution for virtual conversion, based on file system filtering is mentioned. The ways for future development are pointed and some improvements of work already done are suggested.

AMS Subj. Classification: 68U05, 68P30

Key Words: vector graphics, stream transfer, computer animation, raster graphics, format-containers

1. Introduction

In the recent years vector graphics is gaining more and more popularity. It is in the basis of most of the engineering systems, geographic information systems and in the grounds of 3D modeling systems and has also well gained application in the profesional sphere. Vector graphic images are composed of multiple geometric elements, primarily points, lines and curves, analytically described. The basic advantages compared to the raster graphics is the ability for accurate and easy application of various transformations on the images, such as scaling or rotation [3]. One of the most well-known applications of the vector graphics is representation of 3D objects. With standard geometric transformations of a 3D object, described as a set of primitive elements, it may be represented as a two-dimensional image. In addition to that, without changing the description of the object itself, one may change the way a 3D object is constructed in two-dimensions by changing projection center or the projection plane for example. By applying various geometric transformations upon some components of the vector image, motion may be represented and 3D animation created. Representing the 3D animation as a description of a scene, composed of multiple independent objects and describing the motion with multiple geometric

transformations, applied to a given subset of geometric primitives composing the scene, is used frequently when creating and editing a clip, but is inconvenient when reproducing on a computer screen, as it is a raster device [1]. Given this facts, to make representation easier usually the final clips are saved in a raster format, as a sequence of continuous images (frames). An inconvenience in this representation is, that the advantages of the vector graphics are lost for example, when zooming a given part of the image one cannot achieve the same accuracy and preciseness, as in vector representation. It turns out to be more convenient a vector clip to be saved in its original representation and before the representation of each frame to be rasterized. For a movement to be represented there are two approaches every frame to be saved as a vector scene or information about the transformations to be transmitted between two frames.

In the current moment there are few and undeveloped formats for representation of vector animation, especially when concerned streaming transmittance. The contemporary characteristics of the processors allow a fast and convenient processing of vector images the problems with this type of animation gain more attraction. This paper examines the problems evolving from the description of animation in a vector format and the various approaches for overcoming these problems [5].

2. **Characteristics of vector graphics and animation**

Every vector image is a set of geometric primitives points, lines, curves, often in more complex groups polygons, shapes, letters. For the description of these components their analytical mathematical description is used. This type of representation gives one of the main characteristics of vector graphics namely the means various affine transformations to be applied on the image, so that it does not loose its image quality. Such common transformation is vector scaling to practically enlarge the image infinitely with no quality loss as in raster graphics. The most well-known example are the symbols in computer systems for text editing. On the other hand uniting the primitive components lets a given group to be treated in a transformation as single and independent from the other primitives in the image. This enables easy description of effects simulating movement animation. The union of multiple primitive components enables easy changing the details of the image when being viewed to show or hide given details. Thus, it is possible the visualization to be adjusted according to the characteristics of the particular machine which performs the image processing [1, 3].

In succession with the advantages of vector graphics there are some drawbacks. The most significant of them is the complex method of representation on the screen. Because the screens are raster devices, in order to represent vector

format graphics, it must be first transferred into a matrix of pixels. Taking into consideration the advantages of vector representation, as well as their gaining popularity, modern computer science gives special attention to the development of optimized algorithms for effective rasterization as well as developing special processors for this types of calculations. Another disadvantage is the comparatively huge amount of information, which has to be stored when describing complex and detailed images. While in raster graphics we have a rather constant amount of information needed to describe each image, when working with vector graphics it depends strongly on the type and detail quantity in the image. When transmitting an animation clip, represented in raster format, even with the usage of time compression and motion vectors, the size of each image of the clip, called frame, is relatively equal. In the vector approach for description, when the transformations between two frames are described analytically, the quantity of transmitted information decreases significantly. This is at the expence of the more complex process of representation of the clip. Thus when describing the clips in vector formats there are two basic approaches: describing each frame independently from the others, and describing the frame with multiple transformations, which differ from the previous frame. The advantages of the first approaches is the easier process of representation, as well as the opportunity to transfer the clip as a continuous stream. In the second approach the quantity of transferred information is much less, but the process of depicting is more complex first, the scene is generated, next the previous scene is made with the necessary transformations, after which its being rasterized on the screen. Another disadvantage is the impossibility the clip to be transmitted as a stream in order a scene to be represented, information is needed to be present for the current scene and the full list of all transformations frame by frame. This problem may be solved with the integration of the so-called key scenes, similar to the key frames in the MPEG raster compression. Thus, a stable balance between the amount of the transmitted information and the time, needed to collect information from the carrier stream can be achieved, before the clip itself is ready for representation [5].

There are various formats for representation and transfer of static images and video clips when it concerns raster graphics. When storing a video, two types of compression are mainly used. The first compresses each frame independently. In the other method part of the frames are compressed independently from the rest, for which only the part different from the previous or the following frame is described. The advantage in the first case is that the clips are easily modified they may be cut off at any given point. The disadvantage is the comparatively low quality of the image and the big size of data. In the second case it is possible to preserve the quality and decrease the amount of

generated data. The disadvantage though is the inability to cut the clip at a random frame.

The multimedia data itself is merged and preserved in files containers. Up to this moment the multimedia containers can be divided into two basic categories depending on their purpose: formats facilitating the transfer and preservation of multimedia information and formats facilitating the editing (and mostly the assembly) of multimedia. Additionally, we can divide the formats into professional and user-oriented. The transferring formats may also be divided into stream and index.

The purpose of the streaming formats is to allow transmittance of continuous stream of information, so that users begin to read it at a random place without loss of multimedia. Because of this requirement, the describing information in this type of formats is repeated at a certain period in the stream. Such formats are mostly used in the digital television transmissions. They do not support a random access to the data. A typical representative is the MPEG format family for example mpeg2 TS which transmits satellite television. The indexed formats for preservation aim to permit an easy access to a relatively random time in the file. This is usually achieved by keeping the indexed table with records for each frame from the media. These formats are used most often as final storage formats and are convenient for representation or editing. Because the describing information is stored primarily at the beginning or at the end of the file, these files are difficult to represent unless a part of them is not accessed and thus cannot be used in stream transfer. Often these formats allow preservation of text media for example subtitles. Typical representatives are AVI, VOB, as well as the usage in the professional video representation format MXF.

The formats facilitating the editing are mostly created especially for a given professional system for video or audio assembly and thus contain multiple specific descriptive data. The most important between them is the information concerning the average time needed to record each frame, called timecode. These formats allow an easy editing and cutting the media in a random frame. They may also contain multiple different media of one type, for example several sections with video information in different compressions or formats. Many of the formats in this category may be considered as index transferring format. The most well-known formats of these types are MXF, Apple's MOV and its open version MP4.

A comparison of some well-known file formats based on the type of information they might contain and additional features is given in Table 1.

TABLE 1. Comparison of some multimedia formats

Container format	Supported video compression types	Supported audio compression types	VBR video support	VBR audio support	Inplace editing	User defined information	Sub-titles	Stream-ing	Random access
3GP	MPEG-4, H.263, H.264	AMR, AAC	yes	yes	no	no	no	yes	no
ASF	no limitations	problems with Vorbis	yes	yes	no	no	yes	yes	limited
AVI	no limitations	problems with Vorbis	yes	yes	no	no	yes	no	limited
Flash Video	Sorenson, VP6	Raw, ADPOM, LPCM, MP3	no	no	no	no	no	yes	no
Matroska	no limitations	no limitations	yes	yes	no	yes	yes	yes	yes
MPEG-2 PS	MPEG-1, MPEG-2	MPEG-1, MPEG-2 AC3	yes	yes	no	no	yes	no	yes
MPEG-2 TS	MPEG-1, MPEG-2 H.264, MPEG-4	MPEG-1, MPEG-2 LPCM, AAC	yes	yes	no	no	yes	yes	no
MOV	no limitations	no limitations	yes	yes	yes	yes	yes	yes	yes
MP4	MPEG-1, MPEG-2 H.263, MPEG-4	MPEG-1, MPEG-2 Vorbis, AAC	yes	yes	yes	yes	yes	yes	yes
MXF	DV,raw,MPEG SD, MPEG HD,XdCAM,IMX	LPCM, AES3, MPEG-1, MPEG-2	no	yes	no	yes	no	no	yes
OGG	no limitations	no limitations	no	yes	no	yes	yes	yes	yes
VOB	MPEG-2 part 2	AC-3 LPCM MPEG-1, MPEG-2	yes	yes	no	no	yes	no	limited

Because of the different characteristics of the given formats it is acceptable for a given clip to lose part of its information when being transformed from one format to the other and this conversion to be impossible without trans coding this clip to other formats.

One possible solution to this problem is examining the creation of virtual multimedia files, representing the video or audio data in a format different from the one they have initially been saved in, without being represented in this virtual format in a physical medium. A copy may be placed on the media in one format, and the others to be derived while processing virtually with file system level filters. This strategy benefits from the fact that modern computers have powerful processors and huge operating memory. For speeding up the creation of the virtual files it is necessary some preliminarily processed information to be kept for the original file in order to make the creation of the virtual files and the work with them more efficient and quicker. In [6] a detailed information is given on the necessary preliminarily processed information for multimedia files. It also describes the storing format so that it can be easily created physical or virtual copies of a given multimedia file in various formats.

3. Basic formats for storing and transferring vector animation

In the current moment there are several formats for description of vector animation. The most frequently used one is the open format SVG (Scalable Vector Graphics). It serves as a basis for the more wide-spread amongst the users Macromedia's SWF (Small Web Format), later bought by Adobe. In this format the description of motion is implemented by affine transformation description. In this current form it does not allow the stream transfer of clips. Because of the huge user interest in the transmittance of the clips in Internet an expanded format for stream transmittance is developed - Flash Video (flv). This extension transmits the data in a raster form. SVG is a text format when data are transmitted as a XML document. This representation is chosen because of the simple formal description and interpretation, and the possibility for integration, as well as its portability in spite of the differences in the physical representation of the various hardware platforms. To decrease the size of the information transmitted the format allows compression of data. One disadvantage of this format is that it does not support the transition of non-graphic information, for example sound [3].

Another well-known format is Apple Animation. It describes the clips either as a set of transformations, or as a sequence of independent scenes. There is no approach for merging both descriptions. It supports RLE compression. The data are transmitted in a special binary representation. The format allows parallel transmittance of sound or raster described clips. A serious disadvantage

is that its description is closed and it is bind with a series of formats, patents of Apple Inc, that is why it is not much popular in Europe. It does not support a stream except an annotated MOV container [2].

In the end of the 90's of the last century, the company Silicon Graphics started developing the standard for vector graphics OpenGL (Open Graphic Library), which in the moment is the most wide-spread interface for programming of vector graphic, preliminarily 3D models and CAD systems. Approximately at the same time GLS (Graphic Library Stream) was developed, in order to transmit the tracing commands in a stream, so that OpenGL could be used as a format for storing and transmittance of graphics and animations. It has both text and binary described version. A huge advantage is the fact that the commands are directed straight into the GPU (Graphical Processing Unit) of the video-card. Because of the little interest of programmers and users at the time OpenGL was invented and also the slightly-developed graphical processors at that time the development on this format were ceased soon after its creation [4].

Summarizing the characteristics of the well-known formats at the moment for representation and transmittance of vector graphics animation we notice the lack of stream formats for transmittance as well as the inefficiency and restrictiveness of the approaches, chosen for representation of information in comparison with the stream formats, used to transmit raster clips.

4. **Approaches for effective description of streaming vector animation**

The main fields in optimizing the description of the stream vector graphic are the size of the information transferred, the efficiency of the representation process, the opportunity for continuous stream transfer and the opportunity for transfer of additional multimedia information.

The simplest and yet most convenient approach for finding the size of information is by using compression. Modern processors have enough power, so that the decompression of the data stream does not reflect in slowing down the representation process. On the other hand, when using a compression algorithm without dependency usage it is possible to transfer data in a continuous stream. When using compression dictionaries it is possible that they are transmitted periodically in the stream together with the rest of the descriptive information [3].

Another approach is increasing the number of graphical primitives by using the so-called macros. These are means for parametric description of more complex graphic objects, mostly in three-dimensional graphics. Thus a cube may be represented as a set of 12 lines, describing its edges, each defined with its coordinates in both ends and as coordinates of a single vertex, orientation

of a given edge and length of the edge. The same may be related to a set of more complex, but often used in 3D modeling and component animation. The increasing of the graphical primitives with such macros should not lead to less efficiency in representation, as well as not be affected by the transmission of data as a continuous stream [1].

A summary of the mentioned approach is the idea for introducing of specific for the given clip catalogs of the parametric objects a whole element from a scene may be described with such a catalog, parametrized at least by its position, orientation and size. When we want to include such an object in a given scene we simply quote the catalog record with a unique identifier (for example number) and we chose its parameters. This approach may decrease multiple times the size of the transmitted information, again with minimal delay of the visualization process. One disadvantage is the necessity of transmitting periodically these catalogs when making a continuous stream, which even when using compression may lead to a drastic rise in the size of transmitted information. Another, more serious disadvantage is the algorithmic complexity of the attachment of the primitives in a scene into an entity parametrized objects. When the scene is being drawn by a specific software, it stores data for these objects, but if the logic for the specific catalog is not compatible with the specific application, for example a codec processing a stream of vector graphic, finding and summarizing and parameterizing of objects pose serious algorithmic problems [1].

Another approach for decreasing the size of information, often used in computer animation is transformation macros. This approach is used when transmitting the animation as a set of continuous transformations which are applied to the objects in the main scene. Macros here are a brief parameterized description of a frequently used set of continuous elementary transformations, as well as repeating for example it is given that the last few transformations are repeated again and again a certain amount of time. This approach is particularly suitable for describing smooth movements in animation each frame differs from the previous by the position of an object moved with a fixed vector. Again, a summary may be made of this approach for more complex transformations, specific for the given clip, described in catalogs, but here the filtering and extracting of the same information for a record in the catalog after the clip is finished is practically impossible.

The application of some or all of the listed approaches when describing a vector animation leads to multiple decreasing of the size of information, describing a given clip by its main description with basic primitives without a serious loss of productivity in representation. This is a wanted effect, mostly when transferring the clips in a network with limited transfer speed and at

the current moment several big companies have aimed their attention in this direction.

5. Characteristics of continuous video streams

With continuous video streams the client starts receiving information when attached to the stream at any random moment. Is it essential that when he starts collecting data after a given period of time to be in condition to begin reproduction of the clip in the stream. These streams are most often used in digital television, where, on the contrary, the information is transmitted only in raster representation. In the current moment continuous streams of vector graphics have broadest usage in development of rich Internet applications [5].

A basic requirement to the continuous streams is their descriptive information to be transmitted in a certain amount of time, especially chosen so that it does not affect the stream at one hand and on the other, so that the user will not wait too long for the information to arrive on the stream because without it, the reproduction is not possible. When this information is small in size it is acceptable to be repeated in short intervals of time(under 300 ms). If the information has a bigger size, it is divided into smaller parts with different priorities. Thus the information with highest priority, without which the representation is impossible, is repeated quite often, usually in every frame, the less priority information rarely and the user data those which have no direct relation with the clip representation are repeated in largest time intervals once in 3-4 seconds [5].

Often to the main carrier stream of video information other streams are added, carrying different media types sound, for example. It is possible these different streams to be carried in a parallel manner into different physical channels, or they might be mixed into one channel. If we use a parallel transfer, the descriptive information is usually placed in a separate stream, even if it is specific for each of the media streams. The parallel carriage allows the video clip to be transmitted at the same time in several streams, differing by the detail of the image and the user may choose which of them to represent depending on the characteristics of the carrier environment used or depending on the computer configuration. It is also accepted that dynamic change in the representation of the channel in a given moment may occur without its representation itself. Such a technology is realized in Apple's annotated multimedia streams [2, 5].

6. Summary

In the contemporary applications for vector graphics procession more and more features for creation of animated clips are invented. The presence of a small number of standard formats for description of vector animation imposes

the clips to be stored in raster formats. The current paper examined the most-famous formats for storage of vector animation, as well as some multimedia containers. The characteristics of vector representation originate some specific problems concerning the developing format inefficiency in the size of needed data for storing or transmitting clips, as well as little or no options for streaming their representation. There are some well examined approaches for solving the basic problems, which arise when representing animation in vector format. The aim of the author is to develop a decision which gives streaming to vector graphics, which copes with the problems of the previously listed solutions.

References

[1] J. Vince , *Vector Analysis for Computer Graphics*, Springer, 2007

[2] *QuickTime File Format Specification* , **7** , Apple inc., 2009

[3] A. Adam , *SVG. Scalable Vector Graphics* , Franzis Verlag, 2002

[4] C. DunWoody, *The OpenGL Stream Codec* - Silicon Graphics - 1996

[5] B. Gilmer, *File Interchange Handbook for images, audio and metadata*, Focal Press, 2004

[6] P. Armyanov Minimal set of metadata for virtualization of multimedia files - In *Third International Conference on Information Systems and Grid Technologies, 28-29 V 2009, Sofia, Bulgaria,* 46-53

Sofia University St. Kliment Ohridski
Faculty of Mathematics and Invormatics
5, "James Bourchier" str,
1164 Sofia, Bulgaria

E-Mail:parmyanov@fmi.uni-sofia.bg

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Mathematical Optimization for the Train Timetabling Problem

Predrag Stanojević [1], *Miroslav Marić* [1], *Jozef Kratica* [2], *Nebojša Bojović* [3], *Miloš Milenković* [3]

Rail transportation is very rich in terms of problems that can be modelled and solved using mathematical optimization techniques. The train scheduling problem as the most important part of a rail operating policy has a very significant impact on a rail company profit considering the fact that from the quality of a train timetable depends a flow of three most important resources on rail network: cars, locomotives and crews. The train timetabling problem aims at determining a periodic timetable for a set of trains that does not violate track capacities and satisfies some operational constraints. In this paper, we developed an integer programming approach for determining an optimal train schedule for a single, one-way track linking two major stations, with a number of intermediate stations between. The application has been tested on a realistic example suggested by the PE "Serbian Railways". Obtained results show a potential for a practical application of proposed approach.

AMS Subj. Classification: 90C57; 90C10;

Key Words: rail transportation, scheduling, timetabling, integer programming

1. Introduction

In most countries the railway traffic system represents a very important part of the backbone transport system. Traffic and transport policies are striving towards decreasing road traffic pollution by increasing railway usage when appropriate. At the same time, the available railway systems are partly oversaturated creating bottlenecks on major links. An important issue is thus how to best use the existing capacity while ensuring sustainability and attractiveness. The train scheduling problem arises in several contexts [1],[2]. In real-time scheduling [3], it helps the dispatchers who are managing the traffic to make optimal decision about which trains to stop and where, as updated data about train positions becomes available [6],[8]. In tactical planning, it consists of finding the best master schedule or timetable on a regular basis (weekly, monthly, yearly)

[5], [7]. In strategic planning, it relates to investment decisions in rail infrastructure such as building new stations, extending current station track lengths, and upgrading single-line segments [9], [4] to double lines. The timetable planning problem aims at determining a timetable for a set of trains which does not violate track capacities and satisfies some operational constraints. The timetable is a schedule of trains on a given rail infrastructure. It contains the arrival and departure times of the trains at each intermediate point of their route. Due to its complex nature, the construction of the actual timetable as it is operated in practice, is still mainly a human planning process. Traditionally, planners use two types of graphs as the main tools for constructing a timetable. The so-called time-space diagram graphically represents the train movements that take place in between stations, on the tracks. One axis displays time, and the other axis depicts space. Each line in the time-space diagram depicts a train, which is indicated by the number next to the line. Lines with a positive inclination correspond to trains from Station E to Station D, and lines with a negative inclination to trains in opposite direction. Flat lines indicate fast trains, because those trains cover a large distance in a short time, and steep lines indicate slow trains. When a train dwells for some time at a station, this gives a vertical line at that station, because time moves on while the location of the train remains unchanged. Intersecting lines indicate that the two corresponding trains meet at the point of intersection. This is clearly only allowed at stations, or if the trains are using different tracks. The advantage of this diagram is that it makes it much more simple and intuitive to read the timetable and to detect conflicts. In Fig.1 is an example of a train diagram with 21 trains and 19 stations.

The X axis represents time of day, the Y axis the sequence of stations in distance scale. Lines indicate the movement of trains, with the slope indicating direction and speed, horizontal meaning stand-still. Usually, the outbound direction is defined upwards. Fig 2. displays a possible movement of four trains on a single track line connecting the stations A-E. Intermediate meet-points are located at all stations from B to D. The horizontal and vertical axes represent time and space, respectively.

Even in this simple example there are many possible combinations of stations and times for trains to be pulled over to allow meets and passes. Therefore, the train meet-pass problem is a very large-scale combinatorial optimization problem. When several trains in both directions are scheduled, many conflicts may arise. Each conflict involves trains moving either in opposite directions or in the same direction. Depending on the chosen solution for a conflict involving two trains, the location and the time of later conflicts may change, new conflicts at different locations and times may arise, and existing conflicts may

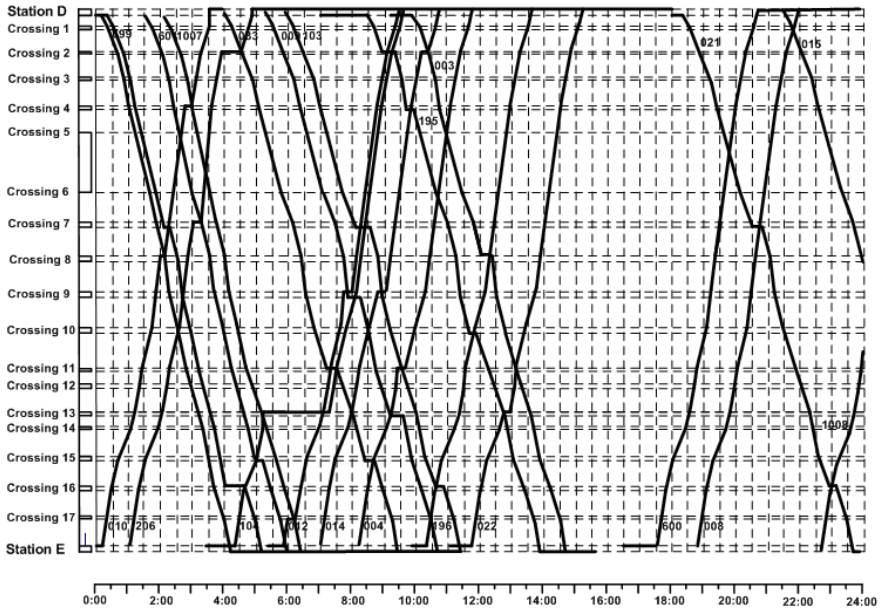


FIGURE 1. Train diagram example

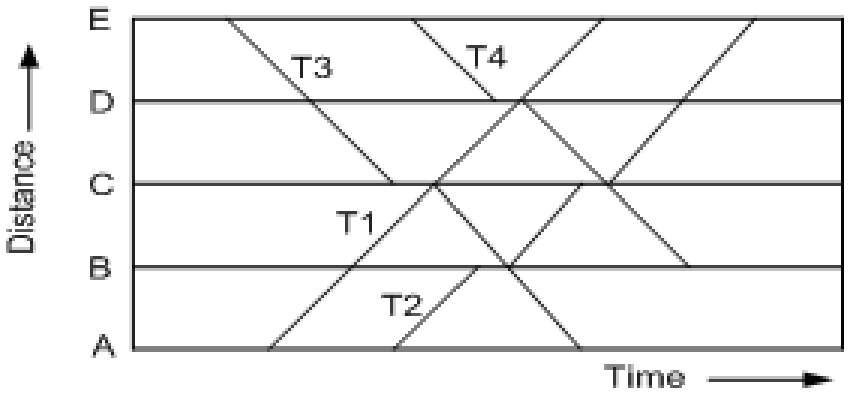


FIGURE 2. Graph of train movements

disappeared. Thus, the number of feasible solutions to train conflicts can be very large. Also, it is very important to notice that train scheduling decision may have different economic effects. Sequences that reduce travel times may decrease investments in cars and engines as well as crew and fuel costs. Regularity and transportation velocity, which also depend on the schedule, are important factors to the satisfaction of customers needs and thus are related to railway company revenues.

In practice, a timetable is constructed by specifying, a time-space path for each train through a railway network, which is drawn in the time-space diagrams, and by specifying the platform tracks that the train occupies, which are drawn in the platform-occupation charts. Generally, a timetable is not constructed from scratch. Rather, adjustments are made to an existing timetable, typically, to the timetable of the previous year. Trains are added to the existing timetable, deleted from it, or the schedule of an already existing train is adjusted. During the process of adding, deleting, and adjusting the schedules of trains, one usually runs into problems at a certain moment in time. It may not be possible to schedule a new train as desired, because there is no capacity on the track or in the station, or because capacity is only available when a connection can not be realized. In such a case, some of the already scheduled trains have to be rescheduled. This can be achieved by shifting an already scheduled train to some earlier or later point in time, by adjusting the planned trip time of a train, or by relocating a train to a different platform. Cycling through this process of scheduling trains, and backtracking on previously made choices in case of a 'dead end', one may eventually arrive at a complete timetable. The timetable construction process is quite complex, and it may take a team of planners several months to create a complete timetable. Modification of the last year timetable implies that the new timetable inherits properties that may be unnecessary and costly. The construction process is very time consuming. Thus, for time reasons, there is no possibility to optimize the timetable, and the planner is often satisfied to find a feasible timetable.

In this paper we present an optimization model for the problem of timetable optimization. This algorithm treats some kind of trade-off between departure times for each of a given set of trains and the total waiting time needed for train conflict resolution. Considering the fact that an optimal timetable gives a good base for making the optimal decision for routing of locomotives, cars and crews, this strategy of making timetables based on a flexible train departures can be a very efficient strategy for improving the timetable construction. We test this algorithm on a realistic example. The computation times are moderate considering the sizes of the optimization problems. Of course, there are many

aspects we do not consider, but these may be added later. We hope that the suggested procedure will be a first step towards intelligent computer assisted profit maximizing allocation of track capacity in timetabling.

2. **Model formulation**

We consider a single, one-way track linking two major stations with a number of intermediate stations. This case is particularly interesting because railway networks are mostly contained from single track sections. On these sections made of a track carrying traffic in opposite directions, track resource is limited by great traffic densities.

The following data represent problem parameters:

T - set of trains

S - set of stations

Q - set of discrete time intervals

d^t - departure station of train t

a^t - arrival station of train t

$Path(t)$ - set of stations between d^t and a^t

ed^t - estimated departure of train t

md^t - maximum allowed delay of train t

$ld^t = ed^t + md^t$

$f(t, s)$ - travel time of train t alongside the track segment s

$Prev(t, s)$ - previous station in the path of train t

$Next(t, s)$ - next station in the path of train t

$ea_{t,s}$ - earliest arrival of train t to station s (calculated from ed_t and $f(t, l)$)

$$ea_{t,s} = ed^t + \sum_{l=d^t}^{s} f(t, l)$$

mw_s - maximum number of trains waiting simultaneously at station s

mi_s - maximum number of trains simultaneously waiting at, or exiting from station s

$x^{dep}_{t,d^t,k}$ - an arc which denotes train t departing at time k

$x^{arr}_{t,a^t,k}$ - an arc which denotes train t arriving at time k

$x^{wait}_{t,s,k}$ - an arc denoting train t waiting in station s at time k

$x^{out}_{t,s,k}$ - an arc denoting that train t is leaving station s at time k

Arcs - set of all arcs $x^{dep}_{t,d^t,k}, x^{arr}_{t,a^t,k}, x^{wait}_{t,s,k} and x^{out}_{t,s,k}, \forall t, \forall k, \forall s$

$$Arcs = \{x^{dep}_{t,d^t,k} | \forall t, \forall k\} \cup \{x^{arr}_{t,a^t,k} | \forall t, \forall k\} \cup \{x^{wait}_{t,s,k} | \forall t, \forall k, \forall s\} \cup \{x^{out}_{t,s,k} | \forall t, \forall k, \forall s\}$$

$Exit(t, s, k)$ - set of arcs denoting train t leaving node (s,k) except for departure arcs, because there is no enter arc for them

$$Exit(t, s, k) = \left\{x^{out}_{t,s,k}, x^{wait}_{t,s,k,}\right\} \cup \left\{x^{arr}_{t,a^t,k} | s = a^t\right\}$$

$Enter(t,s,k)$ - set of arcs denoting train t entering node (s,k) except for arrival arcs, because there is no exit arc for them

$$Enter(t,s,k) = \left\{x^{out}_{t,\Pr ev(s),k-f(t,s)}, x^{wait}_{t,s,k-1,}\right\} \cup \left\{x^{dep}_{t,a^t,k} | s = d^t\right\}$$

The problem is formulated as an integer programming model with decision variables $x \in Arcs$ representing the flow of a train on an arc. The formulation is as follows:

$$minimize\ z = \sum_{t\in T}\Big(\sum_{k=ed_t}^{ld_t}(k-ed_t)x^{dep}_{t,d^t,k} + \sum_{s\in Path(t)}\sum_{k=ea_{t,s}}^{ea_{t,s}+md_t} x^{wait}_{t,s,k}\Big)\ (1)$$

Subject to:

$$\sum_{k=ed_t}^{ld_t} x^{dep}_{t,d^t,k} = 1, \forall t \in T\ (2)$$

$$\sum_{k=sa_t}^{la_t} x^{arr}_{t,a^t,k} = 1, \forall t \in T\ (3)$$

$$\sum_{x\in Enter(t,s,k)} x - \sum_{x\in Exit(t,s,k)} x = 0,$$

$$\forall t \in T, \forall s \in Path(t), \forall k \in Q, ea(t,s) \le k \le ea(t,s) + md(t)\ (4)$$

$$\sum_{t\in T} x^{wait}_{t,s,k} \le mw_s, \forall s \in S, \forall k \in Q\ (5)$$

$$\sum_{t\in T}(x^{wait}_{t,s,k} + x^{out}_{t,s,k}) \le mi_s, \forall s \in S, \forall k \in Q\ (6)$$

$$\sum_{t\in T^{out}}\sum_{\substack{l\le k-1,\\ l+f(t,s)\ge k}} x^{out}_{t,s,l} + \sum_{t\in T^{in}}\sum_{\substack{l\le k-1,\\ l+f(t,s+1)\ge k}} x^{out}_{t,s+1,l} \le 1, \forall s \in S, \forall k \in Q\ (7)$$

$$x \in \{0,1\}, \forall x \in Arcs$$

The objective function (1) minimizes the sum of train delays since the arc costs represent the delays in discretized time units. Equation (2) ensures that for each train there is a unit outflow from its departure node to one of the nodes that is the copy of its origin station at a time within its possible departure window. The constraint of equation (3) ensures that for each train there is a unit flow from one of the nodes that is the copy of its destination node to its arrival node. Constraint set (4) provides the flow conservation constraints. Constraint set (5) ensures that the number of trains waiting simultaneously in station be equal or less than the maximum number of tracks serving for train operations. Constraint set (6) ensures that the total number of trains staying in station and trains which at the same time pass through the station be less than the maximum number of available tracks. Constraint (7) is dedicated to capacity of sections between the stations. Therefore, the total number of trains can not be greater than one, if we are considering inter-station system of traffic regulation.

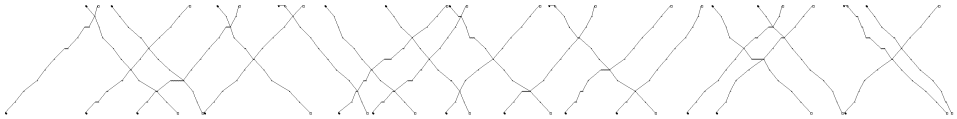


FIGURE 3. Optimal solution for a train timetabling problem

3. **Computational results**

In this section, we present computational results for a real world example of the train dispatching problem. The example is based on actual data for a 24-hour planning horizon on a part of the major line of Serbian railways. Within this time interval we have 25 trains of opposite directions and 12 potential meet/pass points for each of them. According to data from the graphical train timetable, for solving all conflicts raising on this set of trains a total waiting time of 42 minutes is planned. Next table contains all necessary data.

Of course, we also considered the possibility that is not for any train allowed to wait in any station due to incompatibility of its length and the length of station tracks or for some other reasons. Fig. 3 contains an optimal schedule with minimal train delays for given example. Table 2. presents the computational results for a given set of trains.

The minimum total waiting time, as a sum of the time spent in departure stations and the time needed for meeting/overtaking operations for this set of trains on a part of Serbian rail network is 185 minutes. We implemented our algorithm in C programming language and CPLEX 10.0 solver. All computational tests were conducted on a 3.0-GHz Intel Pentium 4 processor with 2Gb RAM. For this example CPLEX loading time was 8 minutes long and optimal solution has been found for 25 seconds of CPU time.

4. **Conclusions**

Due to the complexity of rail operations, the expected growth of traffic and the limited possibilities of enhancing the infrastructure, effective timetable design strategy play a key role in improving the level of railway service. In this paper we presented a model for tactical train scheduling on a single line track. Unlike most models, we tried here to make a trade-off between the time of train departing and the train schedule. For each train, depending on its importance and future train connections we defined an interval of departure as a rolling horizon within which we can depart our train. Obtained results are very promising. Considering the fact that time for deriving the optimal solution is the crucial parameter if we want to use this model as a part of decision

Train No.	Departure station	Arrival station	Departure time	Arrival time
1	Novi Sad	Subotica	22:52	24:50
2	Subotica	Novi Sad	02:40	03:37
3	Novi Sad	Subotica	09:04	11:00
4	Subotica	Novi Sad	17:00	18:54
5	Novi Sad	Subotica	05:09	07:16
6	Subotica	Novi Sad	18:40	20:45
7	Novi Sad	Subotica	11:35	13:32
8	Subotica	Novi Sad	05:45	07:42
9	Novi Sad	Subotica	20:19	22:18
10	Subotica	Novi Sad	14:22	16:19
11	Novi Sad	Subotica	10:10	12:16
12	Subotica	Novi Sad	11:17	13:17
13	Novi Sad	Subotica	13:05	15:17
14	Subotica	Novi Sad	07:20	09:30
15	Novi Sad	Subotica	15:25	17:55
16	Subotica	Novi Sad	04:33	06:54
17	Novi Sad	Subotica	04:33	06:42
18	Subotica	Novi Sad	10:29	12:53
19	Novi Sad	Subotica	07:38	09:49
20	Subotica	Novi Sad	15:48	18:04
21	Novi Sad	Subotica	19:15	21:26
22	Subotica	Novi Sad	13:00	15:18
23	Novi Sad	Subotica	22:20	24:25
24	Subotica	Novi Sad	19:21	21:38
25	Novi Sad	Subotica	22:22	24:34

TABLE 1. Timetable for a given set of trains

support system for on-line train dispatching, this model can be used as a tool for real-time train dispatching.

Train No.	Optimal departure time	Time deviation in minutes respect to earliest departure time
1	22:54	2
2	02:49	9
3	09:05	1
4	17:01	1
5	05:09	0
6	18:50	10
7	11:43	9
8	06:12	27
9	20:30	11
10	14:34	12
11	10:10	0
12	11:27	10
13	13:15	10
14	07:27	7
15	15:25	0
16	04:43	10
17	04:33	0
18	10:31	2
19	07:39	1
20	15:48	0
21	19:32	17
22	13:04	4
23	22:39	19
24	19:21	0
25	22:23	1

TABLE 2. Optimal solution for a given train timetabling problem

References

[1] C a p r a r a , M . F i s c h e t t i , P . T o t h , Modeling and Solving the Train Timetabling Problem.*Operations Research* **50**, 2002, 851-861.

[2] J . C o r d e a u , P . T o t h , D . V i g o , A Survey of Optimization Models for Train Routing and Scheduling.*Transportation Science* , **32**, 1998, 380-404.

[3] A . D ' A r i a n o , D . P a c c i a r e l l i , M . P r a n z o , Effects of Flexible Timetables in Real-Time Scheduling of Rail Operations.*Preprints of the 6th*

Triennial Symposium on Transportation Analysis, TRISTAN 2007, 1-7.

[4] A. Higgins, E. Kozan, L. Ferreira,Optimal Scheduling of Trains on a Single Line Track. *Transportation Research, Part B*, **30**(2), 1996, 147-161.

[5] D. Jovanovic, Improving Railroad In-time Performance: Models, Algorithms and Applications. PhD thesis, Decision Science Department, Wharton School, University of Pennsylvania, 1989.

[6] E. Petersen , A. Taylor, C. Martland,An Introduction to Computer-Assisted Train Dispatch. *Journal of Advanced Transportation*, **20**, 1986, 63-72.

[7] G. Sahin, R. Ahuja,C. Cunha , New approaches for the train dispatching problem. submitted to *Transportation Research Part B.*, 2005.

[8] R. Sauder, W. Westerman , Computer Aided Train Dispatching: Decision Support Through Optimization. *Interfaces*, **13**, 1983, 24-37.

[9] B. Szpigel, Optimal Train Scheduling on a Single Track Railway. In *Proceedings of the IFORS Conference on Operational Research'72*, 1973, 343-361.

[1] Faculty of Mathematics
University of Belgrade,
Studentski Trg 16
11000 Belgrade, Serbia

E-Mails: djape@sbb.rs, maric.m@sbb.rs

[2] Mathematical Institute
of Serbian Academy of Science and Arts
Kneza Mihaila 36
11001 Belgrade, Serbia

E-Mail: jkratica@mi.sanu.ac.rs

[3] Faculty of Transport and Traffic Engineering
University of Belgrade
Voivode Stepe 305
11000 Belgrade, Serbia

E-Mails: nb.bojovic@sf.bg.ac.rs, m.milenkovic@sf.bg.ac.rs

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Services on Application Level in Grid for Scientific Calculations

Radoslava Goranova

The Grid is a hardware and software infrastructure that coordinates access to distribute computational and data resources, shared by different institutes, computational centres and organizations. The Open Grid Services Architecture (OGSA) describes an architecture for a service-oriented grid computing environment, based on Web service technologies, WSDL and SOAP. In this article we investigate possibilities for realization of business process composition in grid environment, based on OGSA standard.

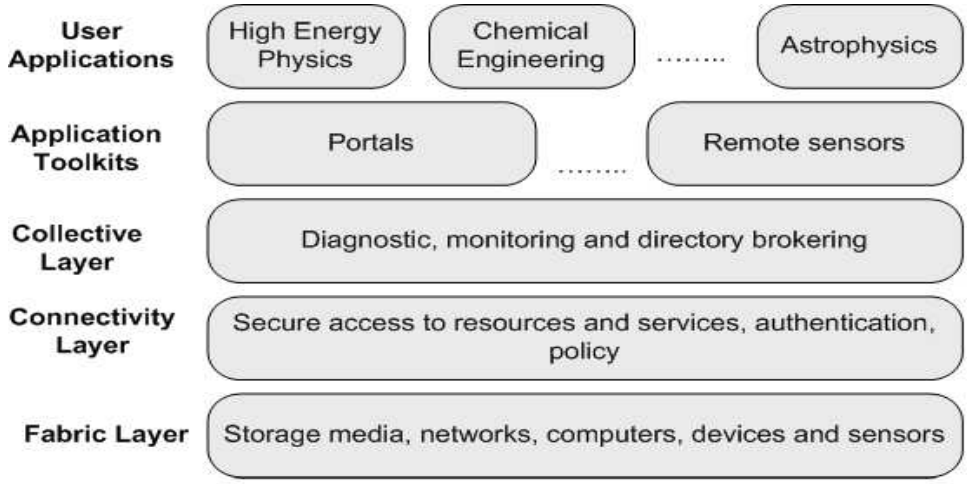
AMS Subj. Classification: 00-02, (General)

Key Words: SOA, Grid, Service orchestartion

1. Introduction

The computational Grid is a hardware and software infrastructure that coordinates access to distribute computational and data resources, shared in a large scale by different institutes, computational canters and organization. Grid resources can be storage systems, computers, software programs, even satellites and special devices. The key concept behind the Grid technology is the ability to agree the resource sharing among different participating parities (users or providers). The term sharing is used in most common sense. In Grid infrastructure sharing means direct access to computer, software, data and other recourses, not just file sharing. This sharing is coordinated by the resource provider and users who agreed about the way how the resources can be accessed. In order to organize users' rights, permission and security in the Grid, terms like virtual organizations and policies are used. Briefly, virtual organization is a set of individuals or institutions which share common goal and are spread all over the world.

From architectural point of view the Grid was defined by C. Kesselman and I. Foster in [1] as five layered protocol architecture. This architecture defines how the users can interact with the Grid resources. The five layered architecture consists of: fabric layer, connectivity layer, collective layer, application layer and user application layer.



The fabric layer provides resources - computational, storage systems, network resources, specific devices, sensors even clusters or pools. The connectivity layer defines communication and authentication protocols for exchanging data between fabric layer and secure mechanism for resource access. The collective layer contains protocols and services for resource interaction monitoring, diagnostics, scheduling, brokering, service discovering and others. The application layer provides software tools: application toolkits, API, SDK, portals and libraries for access to the services defined and implemented in collective layer. The user application layer covers user defined applications and user defined grid services.

2. Service-oriented Grid

The concept of service-orientation is very well defined in [2]. The Service-Oriented Architecture (SOA) provides terms for formal description of the system, defining system's functions, properties and interfaces. A main unit in SOA is the service. Most commonly a service can be described as a software component that realizes specific system's logic and provides it to the end client by network access. Services possess the following characteristics:

(1) Services can be individually used or integrated in composition of services.
(2) Services communicate by message transfer.
(3) Services can be part from workflows.
(4) Services can be self-contained or they may depend on the availability of outer resources or other services.
(5) Service provides information for their interfaces, policies and communication they support.

The service-oriented architecture is very suitable for constricting complex distributed systems over heterogeneous environment, providing them with advantages as easy integration, flexibility and transparency. Service-oriented Grids follow the principles of service-orientation defined in SOA.

The Open Grid Services Architecture (OGSA) [3] describes an architecture for a service-oriented grid computing environment for business and scientific use. OGSA is based on Web service technologies, WSDL and SOAP. OGSA assure interoperability on heterogeneous systems so that different types of resources can communicate and share information.

OGSA represents three major logical and abstract tiers.

The first tier includes the basic resources as CPUs, memory, disks, licenses and OS processes. These resources are usually locally owned and managed and they can high variability in their characteristics, quality of service availability. The second tier represents a high level of virtualization and logical abstraction. The service-oriented architecture of OGSA implies that virtualized resources are represented as services and that interaction with them can be initialized by any service from the architecture. The services from this tier need to use and manage resources from the bottom tier in order to deliver capabilities of individual service. At the third tier are the applications that use the OGSA capabilities to realize user functions and processes.

OGSA realized this middle layer in term of services, the interfaces, the individual and collective state of resources and the interaction between them within a service-oriented architecture.

3. **Problems definition**

The domain of our research is system services on application level in service-oriented Grids. The goal is to investigate possibilities for realization of the business process composition in grid environment, based on the OGSA standard. By business process we mean a set of services ordered into scheme for execution. The main problem which has to be solved is to define a base service for business process orchestration or choreography in Grid.

OGSA defines in general the need of services composition, but does not specify in details how exactly these services will look like or how they will interact with the other components of the architecture. OGSA specification mentions that for services composition a variety of mechanisms like choreography, orchestration or workflow can be used, but also clarifies that OGSA will not define a new mechanism in this area and will rely on the existing standards for that. The proper definition of a service for service composition in the service-oriented Grid is tightly related with proper understanding of: orchestration, choreography and workflows.

In the next section we will describe these three terms and the difference between them and will try to prove that orchestration mechanism is the most suitable one in service-oriented grid environment.

4. **Orchestration vs. Choreography**

The ability of a system to provide mechanisms for building services composition is extremely important, especially for development of complex applications. The service-oriented systems possess this ability and it is called composability. In the way it is define in SOA, the composability of a system guarantees that services defined in the system can be composed in a more complex service and this service also can be part of another service. The process of the service composition is recursive.

According to that how the process is organized we can distinguish two types of service management: orchestration and choreography.

In orchestration, the process controls all contained services and coordinates execution of their operations. In orchestration we have centralized control, which is carried out by the service coordinator. The coordinator takes care for the execution order of containing services and their operations.

In choreography there is no centralized service coordinator. Instead, the process consists of a set of services which are equal. By equal we mean that every service holds information for the services with which it has to interact. All this is based on message exchange between services.

These definitions are general and are relative for every service-oriented system. However, our point of research is service-oriented grid systems, which have additional requirements which we have to conform with. In a service-oriented Grid environment we have to deal with the dynamic nature of grid resources, with grid security and with the indefiniteness of time for execution of the process.

As we already mentioned above the nature of the Grid is dynamic. We can not rely on the fact that the Grid resources will be permanent in the environment. The resources can appear or disappear, because of many reasons: power failure, network problems, service inaccessibility etc. We must assume that various parts of a services composition will fail. This is an issue, which we have to take in mind since we want to give a definition for service composition.

Another issue which we have to deal with is the indefiniteness of time for the process execution. Namely, because of the dynamic of the Grid we have to foresee the ability some of the Grid resources (services) to be temporary unavailable. The approach in that case is to replace the unavailable service with a similar one, which provides the same functionality. But we have to take

into account, that although the service provides the same functionality, the time for task execution can be different.

Also, another critical issue is the ability to quickly bind the part of the process (task) to the appropriate Grid resources. And last, but not least we have to take into account all security issues that can arise during the execution of the process. For example if a grid user have rights to access one grid resources but does not have rights to access another grid resources and those resources are part of the process. The execution of this process most probably will fail due to security reasons.

If we conclude, management of service composition has to be service, which has to be well defined and described. Standards for description of service composition in Grid have to be extensible and to provide entities for exception handling.

The workflow is another, often mentioned term for service composition. In the terms of Grid, the workflow is defined by [4] as "The automation of the process, which involves the orchestration of a set of Grid services, agents and actors that must be combined together to solve a problem or to define a new service". Obviously, the workflow is a special case of orchestration, which concerns not only services but agents and actors. The actor is an external source, who influences the workflow. Or more concrete, the actor can be a person, who interrupts the workflow. The process can not continue without person's decision or person's choice.

The main dilemma which is the most suitable mechanism for a service-oriented grid environment, orchestration or choreography is still remaining. In order to answer this question, we will focus on existing standards for choreography and for orchestration - WS-CDL and WS-BPEL.

Historically viewed, for web services orchestration, the different services providers as IBM, BEA and Microsoft used different approaches. For example IBM used WSFL [5] language for business process description and Microsoft used XLANG [6]. In 2002, OASIS [7] defined a specification called BPEL4WS, which combined graphical process orientation from WSFL and the structure process construction from XLANG, in a new BPEL standard. The role of BPEL4WS [8] standard is to define new web services on the base of existing ones. It is a language for implementation of process orchestration, which is supported from the main web services vendors as IBM, BEA, Oracle and Microsoft.

The Web service business process execution language (WS-BPEL) is an XML based programming language for description of high level business processes. WS-BPEL provides methods for processes description and web services interaction. BPEL is the industry standard for orchestration, supported by the leading providers.

WS-BPEL provides methods for modelling business processes like: sequences, parallel threads, choices, external service invocations, exception handling and etc.

The Web Services Choreography Description Language (WS-CDL) [9] is an XML-based language that describes peer-to-peer collaborations of parties by defining, their common and complementary observable behaviour; where ordered message exchanges result in accomplishing a common business goal.

The WS-CDL model consists of entities, which describe interactions, like channels, participants, roles and work units and activities.

From grid point of view, WS-BPEL as a standard is a good candidate for service composition because of the following reasons:

(1) possibility language notation to be extended in order to meet the Grid requirements;
(2) the language specification is still evolving and supported
(3) the language provide mechanism for process execution, which is the main difference between WS-BPEL and WS-CDL

If we take into account, that WS-BPEL is a language for process execution and the reasons mentioned above we can logically conclude that orchestration as a mechanism is more suitable for grid environment, so far as it concerns existing web services standards.

5. **Conclusion and Future work**

In this article we describe two mechanisms for service compositions and choose a standard for formal description of the orchestration service, which will be most suitable in Grid environment. Still we have to specify the service, taking into account OGSA requirements and to deploy proposed Grid service into grid environment which implements OGSA.

References

[1] I. Foster, K. Kesselman, *The Grid2: Blueprint for a New Computing Infrastructure*, 2004.
[2] T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*, 2005.
[3] I. Foster, Keshimoto, *The Open Grid Service Architecture, Version 1.5* [http://forge.gridforum.org/projects/ogsa-wg], 2006.
[4] G. Fox, D. Gannon, *Workflow in Grid System,* [http://grids.ucs.indiana.edu/ptliupages/publications/Workflow-overview.pdf], 2004.
[5] *Web Services Flow Language*, [http://xml.coverpages.org/wsfl.html].
[6] *XLANG,* [http://xml.coverpages.org/xlang.html].
[7] *OASIS,* [http://www.oasis-open.org/home/index.php].
[8] *WSBPEL,* [http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html].
[9] *WS-CDL,* [http://www.w3.org/TR/2004/WD-ws-cdl-10-20041217].

Department of Computer Science,
Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski",
5, James Baucher Str.
1164 Sofia, Bulgaria

E-Mail: radoslava@fmi.uni-sofia.bg

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Nonlinear Spectral Theories and Solvability of Nonlinear Hammerstein Equations

Sanela Halilović

We study some possibilities of nonlinear spectral theories for solving nonlinear operator equations. The main aim is to research a spectrum and establish some kind of nonlinear Fredholm alternative for Hammerstein operator KF.

AMS Subj. Classification: 47J10, 47H30, 47H10

Key words: spectrum, Hammerstein equation, solvability, superposition operator, nonlinear Fredholm alternative

It is well-known that the major methods for studying solvability of linear or nonlinear equations in literature are: the variational method, the method of a vector field rotation and the fixed point methods.

"The brachistochrone problem" is usually considered to be the beginning of variational calculus and nonlinear analysis. It was first introduced by J. Bernoulli in the 17th century and was first solved by Isaac Newton. Another method for obtaining existence and uniqueness results has been built on a topological method known as the degree theory, the index theory or rotation of the vector fields. The founders of the index theory were: M. Atiyah, R. Bott, F. Hirzebruch and Is. Singer. The first claim, which is an equivalent to the Brouwer fixed-point theorem, was given and proved by H. Poincaré in 1883 and the next by P. Bole in 1904 (the first one-dimensional equivalent is the well-known Bolcano's theorem on the zeros of a continuous function, proved in 1817). In 1909 Brouwer proved the theorem on fixed points in the case of the three-dimensional space. In 1910 Adamar proved a similar statement for an arbitrary finite-dimensional space, by using a Cronecker index. The same statement was proved by Brouwer in 1912 by using simplex approximation and the notion of mapping degree.

Although Poincaré and Bole gave direct applications of their results in the theory of differential equations, as well as in spacial and analytical mechanics,

since then there have not been serious applications of the Brouwer theorem in the mathematical analysis, except one Shauder's result in 1927 about the existence of solution of elliptic partial equations. The situation reversed when John von Neumann applied this theorem to proving the existence of solution for matrix games. These results, which present a base of the classic game theories, increased the mathematicians´s interest in studying the applications of this theorem in various areas of analysis.

Modern research in contractive type conditions started with the Banach fixed point theorem which is one of the classic statements in functional analysis. The following two facts have enabled broad applications of this theorem:

1. solving many kinds of numerical and functional equations can be done by finding fixed points of some mappings;

2. the Banach theorem provides effective calculation (construction) of fixed point and also gives the possibility for estimating error i.e. finding maximal distance from approximative to the accurate solution.

Given a Banach space X over the field $\mathbb{C}$ and a bounded linear operator $A : X \to X$. If some $x_0 \neq 0$ is a fix point of the operator A, i.e. $Ax = x$ has a nontrivial solution $x = x_0$, we can also say

$$(\exists x \neq 0)(I - A)x = 0.$$

It means that operator $I - A$ is not a bijection. More generally, we may consider whether the operator $\lambda I - A$ is a bijection and it leads us to the notion of spectrum which is the set

$$\sigma(A) = \{\lambda \in \mathbb{C} : \lambda I - A \quad \text{is not a bijection}\}.$$

The importance of the spectral theory for linear operators is well-known ([3],[5]). Various attempts have been made to define and study spectra also for nonlinear operators. In the very beginning, the term *spectrum* was used for nonlinear operators just in the sense of point spectrum (i.e. the set of eigenvalues) e.g. by Nemytskij [7] , [8] or Krasnosel'skij [6]. Later it became clear that a more complete description requires, as in the linear case, other ("non-discrete") spectral sets. Starting from the late sixties, this led to a number of definitions of nonlinear spectra which are all different. It was assumed that a reasonable definition of a spectrum of a continuous nonlinear operator should satisfy some minimal requirements, namely:

- It should be reduced to the familiar spectrum in case of linear operators
- It should share some of the usual properties with the linear spectrum (e.g. compactness)
- It should contain the eigenvalues of the operator involved

- It should have nontrivial applications, i.e. those which may not be obtained by other known means

From the viewpoint of these four requirements, any definition of a spectrum should focus on its analytical and topological properties and, of course on applications. Important special classes of continuous (nonlinear) operators are: the Fréchet differentiable operators, the Lipschitz continuous operators, the quasibounded operators and the linearly bounded operators. These are the operators for which many important results have been proved in the nonlinear spectral theory. Spectra have something to do with the "lack of invertibility" of operators. For some class of continuous nonlinear operators $\mathfrak{M}(X)$ we can define a resolvent set

$$\rho(F) = \left\{\lambda \in \mathbb{K} : \lambda I - F \text{ is bijective and } (\lambda I - F)^{-1} \in \mathfrak{M}(X)\right\}$$

and a spectrum

$$\sigma(F) = \mathbb{K} \backslash \rho(F), \quad F \in \mathfrak{M}(X).$$

If we take $\mathfrak{M}(X) = \mathfrak{C}(X)$ we get Rhodius spectrum ([14]), and if we take $\mathfrak{M}(X) = \mathfrak{C}^1(X)$ we get Neuberger spectrum([15]). The Rhodius spectrum (for continuous operators) may be noncompact or empty, while the Neuberger spectrum(for $\mathfrak{C}^1$ operators) is always nonempty (in the complex case), but it needs be neither closed nor bounded. The Dörfner and Kachurovski spectrum is defined for Lipschitz continuous operators ([16],[17]). In contrast to the Neuberger spectrum, the Kachurovski spectrum is compact, but it may be empty. the Dörfner spectrum in turn is always closed, but it may be unbounded or empty. All four spectra considered so far are reduced to the familiar spectrum in the linear case, and they all contain the eigenvalues of the operator involved. A spectrum for certain special continuous operators (stably solvable operators) was introduced by Furi, Martelli and Vignoli in 1978 ([18]). That spectrum is always closed, sometimes even compact and it has many interesting applications. It need not contain the point spectrum in the nonlinear case. A certain modification of this spectrum has been recently given by Appell, Giorgieri and Väth. In 1997 Feng introduced a new spectrum of nonlinear operator (for epi and k-epi operators)([19],[2]). Roughly speaking, one may say that the Furi-Martelli-Vignoli spectrum takes into account the "asymptotic" properties of an operator, while the Feng spectrum reflects its "global" properties. This is also one reason why the latter contains the eigenvalues, but the former does not.

The range of applications of spectral theory is vast. It is a useful tool for solving nonlinear operator equations. When we want to apply some spectral theory to specific nonlinear problem, we have to choose carefully a spectrum which has at least some of the needed features for solving the problem.

For each spectral theory there is some associated eigenvalue theory dealing, to be more precise, with nontrivial solutions of the equation

$$F(x) = \lambda x.$$

The spectral theory for homogeneous nonlinear operators may be used to derive a certain *nonlinear Fredholm alternative* which provides existence and perturbation results for the p-Laplace equation ([1]).

Below we consider the nonlinear Hammerstein integral equations

$$x(t) = \int_{\Omega} k(s,t) f(s, x(s)) d\mu(s) + g(t), \quad (t \in \Omega) \tag{1}$$

where $\Omega \subset \mathbb{R}^N$ is a bounded domain, $k(s,t) : \Omega \times \Omega \to \mathbb{R}$ is a measurable kernel and $f(s,u) : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Caratheodory function (measurable in s, for all $u \in \mathbb{R}$ and continuous in u, for almost all $s \in \Omega$). A discrete analogon of (1) is nonlinear Hammerstein system of equations:

$$x(t) = \sum_{s=1}^{\infty} k(s,t) f(s, x(s)) + g(t), \quad (t \in \mathbb{N}) \tag{2}$$

with kernel $k(s,t)(s,t \in \mathbb{N})$. In system (2) $k : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ defines a linear bounded operator

$$Kx(t) = \sum_{s=1}^{\infty} k(s,t) x(s) \tag{3}$$

and $f : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ is a real function which generates a nonlinear operator superposition F

$$Fx(s) = f(s, x(s)) \quad (s \in \mathbb{N}) \tag{4}$$

The methods of studying solvability of nonlinear Hammerstein equations in the literature are: the variational method (for example, the Golomb functional $\Phi; grad\Phi = F$), the rotation of the vector field and the fixed point methods ([4],[6]). The necessary and sufficient conditions have been found for the nonlinear superposition operator F, for its action from one Banach space to another ($L_p \to L_q$; $l_p \to l_q$), as well as the conditions for its boundedness, continuity, absolute boundedness ([9]). Some results of the solvability of the Hammerstein integral equations (1) in Banach spaces L_p are given in [2]. The conditions of solvability and uniqueness of the Hammerstein system of equations (2) in Banach spaces l_p are given in [10]. For the nonlinear operator F the function

$$\mu_F(r) = \sup_{\|x\| \leq r} \|Fx\|$$

is called the growth function of the operator F in normed space X. The function

$$\nu_F(r) = \inf\{\|a\|_q + br^{\frac{p}{q}} : |f(s,u)| \le a(s) + b|u|^{\frac{p}{q}}, \quad |u| \le r\}$$

also gives us some useful information about the operator F.

If F is a linear operator, then

$$\mu_F(r) = \|F\|r$$

Since the growth function $\mu_F(r)$ and $\nu_F(r)$ are logarithmically convex functions on the convex set $\mathcal{L}(F, act.)$([11]), we could define a norm of the superposition operator F as

$$\|F\| = \frac{\mu_F(r)}{r} \quad \text{or} \quad \|F\| = \frac{\nu_F(r)}{r}.$$

In that case, these norms would be topologically equivalent norms (see Theorem 3, in [9]). We note here that the reviewer of the paper [11] has given one counterexample that $\mathcal{L}(F, act.)$ is not the convex set. But in that example, the set

$$\mathcal{L}(F, act.) = \{(1/p, 1/q) : 4 \le p < \infty, q \ge p/2\} \cup \{(1/p, 1/q) : 1 \le p \le 4, q \ge 2\}$$

is convex (see [13]), and therefore it is not in contradiction with the Theorem 2 given in [11].

The conception of the spectral radius for the linear operator L and relation $r_s(L) \le \|L\|$, could be carried over onto the spectrum of the nonlinear superposition operator F with $r_s(F) \le \|F\|$.

As for the equations (2), recently we have extended the results from [10] on the weighted Banach spaces $l_{p,\sigma}(1 \le \sigma \le \infty)$. Applying the fixed point theorem for monotone operators, we have gained (see [10],[12]) the conditions for the solvability of the system (2) in the spaces $l_{p,\sigma}$ and they are:

Theorem 1. *Let the operator K, defined by (3), be $\mathbb{P}$-positive. Suppose that the generator $f : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ of superposition operator F given by (4), satisfies condition $(u-v)(f(s,u) - f(s,v)) \le c(u-v)^2$ for some $c_f > 0$ and $f(s,0) = 0$ for all $s \in \mathbb{N}$. If $c_f \mu_K < 1$ then, for arbitrary $g \in N(l_{2,\tau})$ the equation:*

$$x = KFx + g$$

has a solution $\hat{x} \in N(l_{2,\tau})$. If $g = Nl$ for some $l \in l_{2,\tau}$, then there exists $\hat{h} \in l_{2,\tau}$, such that $\hat{x} = N\hat{h}$, and

$$\|\hat{h}\| \le \frac{\|l\|}{1 - c_f \mu_K}$$

Moreover,the solution $\hat{x}$ is unique in the space $l_{p,\sigma}$.

Theorem 2. *Let the operator K, defined by (3), be $\mathbb{P}$-quasi-positive in $l_{2,\tau}$. Suppose that the generator $f : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ of superposition operator F given by (4), satisfies $(u-v)(f(s,u)-f(s,v)) \leq c(u-v)^2$ for some $c_f > 0$ and $f(s,0) = 0$ for all $s \in \mathbb{N}$. If $c_f \nu_K < -1$, where*

$$\nu_K = sup\{\nu > 0, \quad \|Nh\| \geq \sqrt{\nu} Ph \quad (h \in l_{2,\tau})\},$$

then, for arbitrary $g \in N(l_{2,\tau})$ the equation:

$$x = KFx + g$$

has a solution $\hat{x} \in N(l_{2,\tau})$. If $g = Nl$ for some $l \in l_{2,\tau}$, then there exists $\hat{h} \in l_{2,\tau}$, such that $\hat{x} = N\hat{h}$, and

$$\|\hat{h}\| \leq -\frac{\|l\|}{1 + c_f \nu_K}$$

Moreover,the solution $\hat{x}$ is unique in the space $l_{p,\sigma}$.

The problem of solvability of the equations (1) or (2) is equivalent to the problem of solvability of the operator equation

$$x = KFx + g, \quad \text{or} \quad (I - KF)x = g. \tag{5}$$

Since the solvability of the equation (5) is closely related to the properties of the spectrum of the nonlinear Hammerstein operator KF, from the theorems that are given in [9] and [10], we can conclude that a spectrum of the operator KF is nonempty. On the other side, we could considered the spectral radius to be $r_s(F) \leq \frac{\mu_F(r)}{r} = \|F\|$.

We are looking forward to finding some kind of nonlinear Fredholm alternative for this Hammerstein operator. The aims of our research are: extending nonlinear spectral theories, finding new statements and new applications.

References

[1] J.Appell, E.Pascale and A.Vignoli, *Nonlinear Spectral Theory*, Walter de Gruyter Berlin, New York, 2004.

[2] J.Appell (Editor), *Recent Trends in Nonlinear Analysis,* Birkhäuser, Basel, 2000.

[3] V.S.Sunder, *Functional Analysis: Spectral Theory*, Birkhäuser Advanced Texts, Hardcover,1998.

[4] H.Brezis, K.C.Chang, S.J.Li, P.Rabinowitz, *Topological Methods, Variational Methods, and Their Applications: ICM 2002 Satelite Conference on Nonlinear Functional Analysis Taiyuan, Shan Xi*, 2002.

[5] W. Arveson, *A Short Course on Spectral Theory*, Hardcover, 2001.

[6] M.A.Krasnosel´skij, On a topological method in the problem of eigenfunctions for nonlinear operators, *Doklady Akad. Nauk SSSR*, **74**, 1950, 5-7.

[7] V.V.Nemytskij, Some problems concerning the structure of the spectrum of completely continuous nonlinear operators, *Doklady Akad. Nauk SSSR*, **80**, 1951, 161-164.

[8] V.V.Nemytskij, Structure of the spectrum of completely continuous nonlinear operators, *Mat.Sb.*, **33** , 1953, 545-558.

[9] F. Dedagic and P.P. Zabrejko, Ob operatorah superpozicii v prostranstvah l_p, *Sibirskij matematieskij žurnal*, **XXVIII**, N^{o}1, 1987.

[10] F. Dedagic, On the discrete nonlinear Hammerstein systems with nonsymetric kernels, *Sarajevo Journal of Mathematics*, **5**, N^{o}2, 2009.

[11] F. Dedagic, N. Okicic, On the $\mathcal{L}$- -characteristic of nonlinear superposition operators in $l_{p,\sigma}$-spaces, *Mathematica Balkanica*, **19**, 2005.

[12] F.Dedagic, S.Halilovic, E.Barakovic, On the solvability of discrete nonlinear Hammerstein systems, *MICOM 2009*, Ohrid , 2009.

[13] J.Appell, P.P.Zabrejko, Über die $\mathcal{L}$-charakteristik nichtlinearer operatoren in räumen integrierbarer funktionen, *Manuscripta Math.*, 1988, 355-367.

[14] A.Rhodius, Über numerische Werterbereiche und Spektralwertabschätzungen, *Acta Sci. Math.* **47**, 1984, 465-470.

[15] J.W. Neuberger, Existence of a spectrum for nonlinear transformations, *Pacific.J. Math.* **31**, 1969, 157-159.

[16] R.I. Kachurovskij, Regular points, spectrum and eigenfunctions of nonlinear operators, *Dokl. Akad. Nauk SSSR*, **188**, 1969, 274-277.

[17] M. Dörfner, *Spektraltheorie für nichtlineare Operatoren, Ph.D.thesis,* Universität Würzburg, 1997.

[18] M. Furi, M. Martelli, A. Vignoli, Contributions to the spectral theory for nonlinear operators in Banach spaces, *Ann. Mat. Pura Appl.*, **118**, 1978, 229-294.

[19] W.Feng, A new spectral theory for nonlinear operators and its applications, *Abstr.Appl.Anal.*, **2**, 1997, 163-183.

Department of Mathematics
Faculty of Science and Mathematics
University of Tuzla
Bosnia and Herzegovina

email: sanela.halilovic@untz.ba

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

New Vacuum Solutions for Quadratic Metric-Affine Gravity - a Metric Affine Model for the Massless Neutrino?

Vedad Pasic

In this paper we present an overview of our research that was presented at the MASSEE International Congress on Mathematics MICOM 2009 in Ohrid, Macedonia. We deal with *quadratic metric–affine gravity*, which is an alternative theory of gravity. We present *new vacuum solutions* for this theory and an attempt to give their *physical interpretation* on the basis of comparison with existing classical models. These new explicit vacuum solutions of quadratic metric–affine gravity are constructed using *generalised pp-waves*. A classical pp-wave is a 4-dimensional Lorentzian spacetime which admits a non–vanishing parallel spinor field. We *generalise* this definition to metric compatible spacetimes with torsion, describe basic properties of such spacetimes and eventually use them to construct new solutions to the field equations of quadratic metric–affine gravity. The physical interpretation of these solutions we propose is that these new solutions represent a *conformally invariant metric–affine model for the massless neutrino*. We give a comparison with a classical model describing the interaction of gravitational and massless neutrino fields, namely *Einstein-Weyl theory*. Future research topics are briefly discussed.

AMS Subj. Classification: 83C15, 83C35

Key Words: quadratic metric-affine gravity, pp-waves, torsion, exact solution, neutrino

1. Introduction

There are a number of different alternative theories of gravity that try to further the completion of Einstein's theory of gravity. One such theory, propagated by Einstein himself for some time, is the *metric–affine gravity*.

A number of developments in physics in the last several decades have evoked the possibility that the treatment of spacetime might involve more than just the Riemannian spacetime of Einstein's general relativity. The smallest departure from a Riemannian spacetime of Einstein's general relativity would

consist of admitting *torsion*, arriving thereby at a Riemann–Cartan spacetime, and, furthermore, a possible nonmetricity, resulting in a '*metric–affine*' spacetime.

The metric–affine gravity is a natural generalisation of Einstein's general relativity, which is based on a spacetime with a Riemannian metric g of a Lorentzian signature. Similarly, in the metric–affine gravity we consider spacetime to be a connected real 4–manifold M equipped with a Lorentzian metric g and an affine connection Γ. Note that the characterisation of the spacetime manifold by an *independent* linear connection Γ initially distinguishes metric–affine gravity from general relativity. The connection incorporates the inertial properties of spacetime and it can be viewed, according to Hermann Weyl [28], as the guidance field of spacetime. The metric describes the structure of spacetime with respect to its spacio-temporal distance relations.

The 10 independent components of the (symmetric) metric tensor $g_{\mu\nu}$ and the 64 connection coefficients $\Gamma^{\lambda}{}_{\mu\nu}$ are the unknowns of metric–affine gravity.

We mostly deal with *quadratic* metric-affine gravity. In the quadratic metric-affine gravity, we define our action as

$$S := \int q(R) \tag{1}$$

where q is a quadratic form on curvature R. The coefficients of this quadratic form are assumed to depend only on the metric, and the form itself is assumed to be $\mathrm{O}(1,3)$ invariant.

An independent variation of (1) with respect to the metric g and the connection Γ produces the system of Euler–Lagrange equations which we will write symbolically as

$$\partial S/\partial g = 0, \tag{2}$$

$$\partial S/\partial \Gamma = 0. \tag{3}$$

The objective of our work was the study of the combined system of field equations (2), (3). This is a system of $10 + 64$ real nonlinear partial differential equations with $10+64$ real unknowns. the quadratic curvature Lagrangians were first discussed by Weyl [28], Pauli [17], Eddington [6] and Lanczos [10, 11, 12] in an attempt to include the electromagnetic field in the Riemannian geometry.

Our motivation comes from the Yang–Mills theory. The Yang–Mills action for the affine connection is a special case of (1) with

$$q(R) = q_{\mathrm{YM}}(R) := R^{\kappa}{}_{\lambda\mu\nu}\, R^{\lambda}{}_{\kappa}{}^{\mu\nu}\,. \tag{4}$$

With this choice of $q(R)$, equation (3) is the Yang–Mills equation for the affine connection, which was analysed by Yang [29].

The idea of using a purely quadratic action in the General Relativity goes back to Hermann Weyl, who argued that the most natural gravitational action should be quadratic in curvature and involve all possible invariant quadratic combinations of curvature, like the square of Ricci curvature, the square of scalar curvature, etc. By choosing a purely quadratic curvature Lagrangian we are hoping to describe phenomena whose characteristic wavelength is sufficiently small and curvature sufficiently large.

As presented in [16], we were able to obtain a new class of solutions for quadratic metric–affine gravity.

2. **A short introduction to pp-waves**

PP-waves are well known spacetimes in the general relativity, first discovered by Brinkmann [2] in 1923, and subsequently rediscovered by several authors, for example Peres [18] in 1959. We used them as the basis for constructing new solutions for quadratic metric–affine gravity. Hence, an introduction to classical pp-waves is required in order to fully understand this construction.

Recall first the well-known notion of a pp-wave.

Definition 1. A (classical) *pp-wave* is a connected 4-manifold M equipped with Lorentzian metric g and Levi-Civita connection Γ which admits a nonvanishing parallel spinor field.

Another way of characterising a pp-wave is by its restricted holonomy group Hol^0. Definition 1 is equivalent to

Definition 2. A (classical) *pp-wave* is a connected 4-manifold M equipped with Lorentzian metric g and Levi-Civita connection Γ whose holonomy Hol^0 is, up to conjugation, a subgroup of the group

$$B^2 := \left\{ \begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix} \middle| \quad q \in \mathbb{C} \right\}. \tag{5}$$

The group (5) is, up to conjugation, the unique nontrivial Abelian subgroup of $\mathrm{SL}(2,\mathbb{C})$, where "non-trivial" is understood as "weakly irreducible and not 1-dimensional" and dimension understood as real dimension. "Weak irreducibility" means that the only non-degenerate invariant subspaces of the tangent space are $\{0\}$ and the tangent space itself.

Yet another equivalent way of characterising a pp-wave is via an explicit formula for the metric.

Definition 3. A (classical) *pp-wave* is a connected 4-manifold M equipped with Lorentzian metric g and Levi-Civita connection Γ whose metric can be written locally in the form

$$ds^2 = 2\,dx^0\,dx^3 - (dx^1)^2 - (dx^2)^2 + f(x^1, x^2, x^3)\,(dx^3)^2 \tag{6}$$

in some local coordinates (x^0, x^1, x^2, x^3).

PP-waves are well known in general relativity for their beautiful and amazing properties. For example, the curvature tensor R of a pp-wave is linear in f (in special local coordinates (6)) and is given by a simple explicit formula

$$R_{\alpha\beta\gamma\delta} = -\frac{1}{2}(l \wedge \partial)_{\alpha\beta}\,(l \wedge \partial)_{\gamma\delta} f \tag{7}$$

where l is a parallel null light–like vector and $(l \wedge \partial)_{\alpha\beta} := l_\alpha \partial_\beta - \partial_\alpha l_\beta$. See Section 3 in [16] for more details.

The main aim of our research was to extend the classical notion of a pp-wave to metric compatible spacetimes with torsion, i.e. with

$$\Gamma^\lambda{}_{\mu\nu} \neq \frac{1}{2} g^{\lambda\kappa}(\partial_\mu g_{\nu\kappa} + \partial_\nu g_{\mu\kappa} - \partial_\kappa g_{\mu\nu})\,,$$

and to do it in such a manner that all the nice properties are preserved and they are still easy to work with in practical applications, as presented in [16]. One natural way of generalising the notion of a pp-wave is simply to extend it to general metric–compatible spacetimes. However, this would give us a class of spacetimes which is too wide and difficult to work with. We choose to extend the classical definition in a more special way, and in this section we do it by introducing torsion explicitly.

Let A be a complex vector field defined by

$$A = h(x^3)m + k(x^3)l \tag{8}$$

where l is a parallel null light–like vector and m is a complex isotropic vector field orthogonal to l. We choose the set of local coordinates for which $l^\mu = (1, 0, 0, 0)$ and $m^\mu = (0, 1, \mp i, 0)$. The functions $h, k : \mathbb{R} \to \mathbb{C}$ are arbitrary.

We can then define a *generalised pp-wave* as a metric–compatible spacetime with pp–metric and torsion

$$T := \frac{1}{2}\mathrm{Re}(A \otimes \mathrm{d}A). \tag{9}$$

Torsion can be expressed more explicitly in our local coordinates as

$$T^\alpha{}_{\beta\gamma} = \frac{1}{2}\mathrm{Re}\left[(k(x^3)h'(x^3)l^\alpha + h(x^3)h'(x^3)m^\alpha)\,(l \wedge m)_{\beta\gamma}\right].$$

Torsion is purely tensor and it has 4 non-zero independent components. The formula for curvature in our local coordinates is

$$R_{\alpha\beta\gamma\delta} = -\frac{1}{2}(l\wedge\partial)_{\alpha\beta}(l\wedge\partial)_{\gamma\delta}f + \frac{1}{4}\mathrm{Re}\left((h(x^3)^2)''\,(l\wedge m)_{\alpha\beta}\,(l\wedge m)_{\gamma\delta}\right). \tag{10}$$

Curvature only has two irreducible pieces, namely symmetric trace-free Ricci and Weyl and it can be written down as

$$R_{\kappa\lambda\mu\nu} = \frac{1}{2}(g_{\kappa\mu}Ric_{\lambda\nu} - g_{\lambda\mu}Ric_{\kappa\nu} + g_{\lambda\nu}Ric_{\kappa\mu} - g_{\kappa\nu}Ric_{\lambda\mu}) + \mathcal{W}_{\kappa\lambda\mu\nu}.$$

The Ricci and Weyl curvatures are given by

$$\begin{aligned} Ric_{\mu\nu} &= \frac{1}{2}(f_{11}+f_{22})\,l_\mu l_\nu, \\ \mathcal{W}_{\kappa\lambda\mu\nu} &= \sum_{j,k=1}^{2} w_{jk}(l\wedge m_j)\otimes(l\wedge m_k), \end{aligned}$$

where $m_1 = \mathrm{Re}(m), m_2 = \mathrm{Im}(m)$, $f_{\alpha\beta} := \partial_\alpha\partial_\beta f$ and w_{jk} are real scalars given by

$$\begin{aligned} w_{11} = \frac{1}{4}[-f_{11}+f_{22}+\mathrm{Re}((h^2)'')], \quad w_{22} = -w_{11}, \\ w_{12} = \pm\frac{1}{2}f_{12} - \frac{1}{4}\mathrm{Im}((h^2)''), \quad w_{21} = w_{12}. \end{aligned}$$

Note that our generalised pp-waves have the same irreducible pieces of curvature as classical pp-waves and that their curvature has all the usual symmetries of curvature in the Riemannian case.

3. **The main result**

The main result of our research thus far is the following

Theorem 1. *Generalised pp-waves of parallel Ricci curvature are solutions of the system of equations (2), (3).*

Note that when using Theorem 1 it does not really matter whether the condition 'parallel Ricci curvature' is understood in the non-Riemannian sense $\nabla Ric = 0$, the Riemannian sense $\{\nabla\}\{Ric\} = 0$, or any combination of the two ($\{\nabla\}Ric = 0$ or $\nabla\{Ric\} = 0$). Here curly brackets refer to the Levi–Civita connection.

In special local coordinates, the condition that Ricci curvature is parallel is written as $f_{11}+f_{22} = \text{const}$, where $f_{\alpha\beta} := \partial_\alpha\partial_\beta f$. Hence, generalised pp-waves of parallel Ricci curvature admit a simple explicit description.

The proof of the main theorem is done by 'brute force'. We write down our field equations (2), (3) explicitly under certain assumptions on the properties of the spacetime, which generalised pp-waves automatically posses. The proof of the theorem is then quite straightforward, as we explicitly show that the field equations are satisfied by inserting the formulae for the irreducible pieces of curvature and torsion of generalised pp-waves.

For the proof of Theorem 1, see [16].

4. **Physical interpretation of generalised pp-waves**

Our analysis of vacuum solutions of quadratic metric–affine gravity shows, see Theorem 1, that classical pp-spaces of parallel Ricci curvature should not be viewed on their own. They are a particular (degenerate) representative of a wider class of solutions, namely, generalised pp-spaces of parallel Ricci curvature. The latter appear to admit a sensible physical interpretation. Indeed, according to formula (10) the curvature of a generalised pp-space is a sum of two curvatures: the curvature

$$-\frac{1}{2}(l \wedge \{\nabla\}) \otimes (l \wedge \{\nabla\})f \tag{11}$$

of the underlying classical pp-space and the curvature

$$\frac{1}{4}\mathrm{Re}\left((h^2)''\,(l \wedge m) \otimes (l \wedge m)\right) \tag{12}$$

generated by a torsion wave traveling over this classical pp-space. Our torsion (9), (8) and corresponding curvature (12) are waves traveling at speed of light. The underlying classical pp-space of parallel Ricci curvature can now be viewed as the 'gravitational imprint' created by a wave of some massless matter field. Such a situation occurs in the Einstein–Maxwell theory[1] and the Einstein–Weyl theory[2]. The difference with our model is that Einstein–Maxwell and Einstein–Weyl theories contain the gravitational constant which dictates a particular relationship between the strengths of the fields in question, whereas our model is conformally invariant and the amplitudes of the two curvatures (11) and (12) are totally independent.

The physical interpretation of the solution from Theorem 1 we proposed is that these new solutions represent a *conformally invariant metric–affine*

[1]The Einstein–Maxwell theory is a classical model describing the interaction of gravitational and electromagnetic fields

[2]The Einstein–Weyl theory is a classical model describing the interaction of gravitational and massless neutrino fields

model for a massless elementary particle by comparing them to solutions of the *Einstein-Weyl theory.*

In the Einstein–Weyl theory the action is given by

$$S_{EW} := 2i \int \left(\xi^a \, \sigma^\mu{}_{a\dot{b}} \, (\{\nabla\}_\mu \overline{\xi}^{\dot{b}}) \; - \; (\{\nabla\}_\mu \xi^a) \, \sigma^\mu{}_{a\dot{b}} \, \overline{\xi}^{\dot{b}} \right) + k \int \mathcal{R}, \tag{13}$$

where the constant k can be chosen so that the non-relativistic limit yields the usual form of Newton's gravity law.

In the Einstein–Weyl theory the connection is assumed to be Levi-Civita, so we only vary the action (13) with respect to the metric and the spinor to obtain the well known Einstein–Weyl field equations

$$\frac{\delta S_{EW}}{\delta g} = 0, \tag{14}$$

$$\frac{\delta S_{EW}}{\delta \xi} = 0. \tag{15}$$

We pointed out the fact that the nonlinear system of Einstein–Weyl field equations has solutions in the form of pp-waves. The main difference between the two models is that in the metric–affine model our generalised pp-waves solutions have parallel Ricci curvature, whereas in the Einstein–Weyl model the pp-wave type solutions do not necessarily have parallel Ricci curvature. However, when we look at monochromatic pp-wave type solutions in the Einstein–Weyl model their Ricci curvature also becomes parallel and we conclude that while in the metric–affine case the Laplacian of f can be *any* constant, in the Einstein–Weyl case it is required to be a *particular* constant. This should not be surprising as our metric–affine model is conformally invariant, while the Einstein–Weyl model is not.

We pointed out a very interesting fact that that generalised pp-waves of parallel Ricci curvature are sufficiently similar to pp-type solutions of the Einstein–Weyl model, which is a classical model describing the interaction of massless neutrino and gravitational fields, to suggest that *generalised pp-waves of parallel Ricci curvature represent a metric–affine model for the massless neutrino.*

5. **Planned future research**

The main aims of the research we plan to do in the immediate future would be the following:

(i) *Physical interpretation of previous results.* We would first try to solidify and expand the physical interpretation of the new solutions obtained thus far. As our generalised pp-waves of parallel Ricci curvature clearly have interesting properties, the main objective of this part of research would be to further investigate the possibility that these solutions represent a metric–affine (and thus conformally invariant) model for some massless particle. This would be done in collaboration with several people who are perhaps more involved with the physical aspects of this area. I expect this research to produce publishable results fairly soon, as the majority of the work has already been done.

(ii) *Further comparison with existing solutions.* There are several results from this area that can be compared to our solution in order to see if additional solutions can be obtained or not. The two papers of Singh [21, 22] are an example of this. In [21] Singh presents solutions of the vacuum field equations with *purely axial torsion*, which is a class of solutions unobtainable by the double duality ansatz of [1, 13].

In the second paper [22], Singh also constructs solutions unobtainable by the double duality ansatz, but this time that have *purely trace torsion*. These solutions are similar in many ways, as the metric and the hence the Riemannian pieces of curvature are the same - which leads the author to stipulate that it might be possible to combine these two solutions but he however shows that this is unfortunately not possible.

It should be pointed out that in [21, 22] Singh was not working within the setting of the most general purely quadratic action and the solutions were obtained for the Yang–Mills case (4). It is clear that these solutions differ from the ones presented in our work, as our torsion is purely tensor. It would however be of interest to us to see whether this construction of Singh's can be expanded to our most general $\mathrm{O}(1,3)$-invariant quadratic form q.

One other and more recent non-metric–compatible result comes from Obukhov [15]. The quadratic form on curvature considered is the most general, and identical to the quadratic form used in our work and in [16, 24, 27]. However, unlike the solutions presented in these works, Obukhov

constructs new solutions that have non-zero nonmetricity, which are generalisations of pp-waves. Obukhov presents solutions that have not only torsion waves present but the nonmetricity has a non-trivial wave behaviour as well, which is different from the generalised pp-waves presented in this thesis. Moreover, Obukhov suggests that his solutions provide a minimal generalisation of the pseudoinstanton, see [23] for definition of a pseudoinstanton. However, it should be pointed out that solutions presented in [15] are *not* non-metric–compatible generalisations of solutions presented in this thesis.
It would be of great interest to us to respond to this work of Obukhov's, for example by seeing what the relaxation of our condition on metric-compatibility would produce and to investigate a possible combination of these solutions.

(iii) *Teleparallelism.* The last, but definitely the most interesting part of the research we plan to do in the near future would be in the field of *teleparallelism.*

Teleparallelism is a very interesting alternative theory of gravity and it can be considered as a special case of Cosserat elasticity initially investigated by the Cosserat brothers in [5] and used by Einstein and Cartan to try to unify electromagnetism and gravity, i.e. as a candidate for the *theory of everything.* The subject of teleparallelism has a long history and its origins lie in the pioneering works of Eugène and Francois Cosserat, Élie Cartan, Albert Einstein and Roland Weitzenböck. Modern reviews of the physics of teleparallelism are given in [7, 9, 14, 20].

The basic idea of teleparallelism is to work with a Lorentzian metric, vanishing curvature and *non-vanishing torsion*, so it could be viewed as a special case of metric–affine gravity. However, in practice instead of using the metric as the unknown of this theory, one uses a quartet of covectors (a *coframe*).

An interesting recent result in teleparallel gravity related to our previous result was done by Vassiliev in [25, 26], where a new (teleparallel) representation for the Weyl Lagrangian is given. The advantage of the teleparallel approach is that it does not require the use of spinors, Pauli matrices or covariant differentiation, as we did in our work. The only geometric concepts used are those of a metric, differential form, wedge product and exterior derivative. It would be interesting to see whether this can be applied to our previous research.

Another interesting result that we plan to investigate and further build on can be found in [3] where the authors suggest an alternative mathematical model for the electron using the teleparallel approach and where the electron mass and external electromagnetic field are incorporated into the model by means of a *Kaluza–Klein extension*.

One of the main topics of research interest for us in this field would be the calculation the ground energy state of the hydrogen atom based on the model presented in [3]. The model presented in [4] would be interesting in comparison with our research thus far as [4] deals with a teleparallel model for the massless neutrino, while we dealt with the metric–affine model for the massless neutrino.

References

[1] P. Baekler, F. W. Hehland, E. W. Mielke, Nonmetricity and torsion: facts and fancies in gauge approaches to gravity, in: *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity* edited by Ruffini R (Amsterdam: Elsevier Science Publishers B.V.), 1986, 277–316

[2] M. W. Brinkmann, On Riemann spaces conformal to Euclidean space. *Proceedings of the National Academy of Sciences of USA* **9**, 1923, 1–3

[3] J. Burnett, O. Chervova, D. Vassiliev, Dirac equation as a special case of Cosserat elasticity, In: *"Analysis, Partial Differential Equations and Applications - The Vladimir Maz'ya Anniversary Volume"* (ed. A.Cialdea, F.Lanzara and P.E.Ricci), series Operator Theory: Advances and Applications, **193**, Birkhaeuser Verlag, 2009, 15–29

[4] O. Chervova , D. Vassiliev, Massless Dirac equation as a special case of Cosserat elasticity, to appear in *Applied Mathematics & Information Sciences (Proceedings of the International Conference on Recent Trends in Mathematical Sciences, Bahrain, 10-12 November 2008)*. Available as preprint arXiv:0902.1268, 2009

[5] E. Cosserat, F. Cosserat, Théorie des corps déformables, *Librairie Scientifique A. Hermann et fils, Paris*, 1909 . Reprinted by Cornell University Library.

[6] A. S. Eddington, *The Mathematical Theory of Relativity* Cambridge 1952

[7] F. Gronwaldand, F. W. Hehl, On the gauge aspects of gravity, In: *Proc. of the 14th Course of the School of Cosmology and Gravitation*

on 'Quantum gravity' (Erice, Italy 1995) World Scientific, Singapore, 1996, 148–198, gr-qc/9602013

[8] F. W. Hehl , J. D. McCrea, E. W. Mielke and Y. Ne'eman, Metric–affine gauge theory of gravity: field equations, Noether identities, world spinors, and breaking of dilation invariance, *Phys. Rep.* **258**, 1995 , 1–171

[9] F. W. Hehl, J. Nitsch and P. von der Heyde, In: *General Relativity and Gravitation, Vol. 1*, Plenum Press, New York, 1980, 329–355

[10] C. Lanczos, A remarkable property of the Riemann–Christoffel tensor in four dimensions, *Ann. Math.* **39**, 1938, 842–850

[11] C. Lanczos , Lagrangian multiplier and Riemannian spaces, *Rev. Mod. Phys.* **21**, 1949, 497–502

[12] C. Lanczos, Electricity and general relativity. *Rev. Mod. Phys.* **29**, 1957, 337–350

[13] E. W. Mielke, On pseudoparticle solutions in Yang's theory of gravity, *Gen. Rel. Grav.* **13**, 1981, 175–187

[14] U. Muench, F. Gronwald and F. W. Hehl, A small guide to variations in teleparallel gauge theories of gravity and the Kaniel–Itin model, *Gen. Rel. Grav.* **30**, 1998, 933–961

[15] Yu. N. Obukhov, Plane waves in metric–affine gravity, *Phys. Rev. D* **73**, 2006, 024025 [6 pages]

[16] V. Pasic and D. Vassiliev, PP–waves with torsion and metric–affine gravity, *Class. Quantum Grav.* **22**, 2005, 3961–3975

[17] W. Pauli, Zur Theorie der Gravitation und der Elektrizität von Hermann Weyl, *Physik. Zaitschr.* **20**, 1919, 457–467

[18] Peres, Some gravitational waves, *Phys. Rev. Lett.* **3**, 1959 , 571

[19] Peres, abstract to preprint hep-th/0205040, 2002 (reprinting of [18])

[20] T. Sauer, Field equations in teleparallel spacetime: Einstein's fernparallelismus approach towards unified field theory, *preprint* physics/0405142v1, 2004

[21] P. Singh, On axial vector torsion in vacuum quadratic Poincaré gauge field theory, *Phys. Let.* **145A**, 1990, 7–10

[22] P. Singh, On null tratorial torsion in vacuum quadratic Poincaré gauge field theory, *Class. Quantum Grav.* **7**, 1990, 2125–2130

[23] D. Vassiliev, Pseudoinstantons in metric–affine field theory, *Gen. Rel. Grav.* **34**, 2002, 1239–1265

[24] D. Vassiliev, Quadratic non–Riemannian gravity, *Journal of Nonlinear Mathematical Physics*, **11**, Supplement, 2004, 204–216

[25] D. Vassiliev, Teleparallel model for the neutrino, *Phys. Rev. D*, **75**, 025006, 2007, [6 pages]

[26] D. Vassiliev, A teleparallel representation of the Weyl Lagrangian, *International Journal of Geometric Methods in Modern Physics*, **4**, no.2, 2007, 325–332

[27] D. Vassiliev, Quadratic metric–affine gravity, *Ann. Phys. (Lpz.)*, **14**, 2005, 231–252

[28] H. Weyl, Eine neue Erweiterung der Relativitätstheorie, *Ann. Phys. (Lpz.)* **59**, 1919, 101–133

[29] C. N. Yang, Integral Formalism for Gauge Fields, *Phys. Rev. Lett.*, **33**, 1974, 445–447.

Department of Mathematics
Faculty of Science
University of Tuzla
Univerzitetska 4
75000 Tuzla, Bosnia and Herzegovina

E-Mail: vedad.pasic@untz.ba

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Canonical Objects in Classes of $(n, \mathcal{V})$-Groupoids

Vesna Celakoska-Jordanova

Free algebras are very important in studying classes of algebras, especially varieties of algebras. Any algebra that belongs to a given variety of algebras can be characterized as a homomorphic image of a free algebra of that variety. Describing free algebras is an important task that can be quite complicated, since there is no general method to resolve this problem. The aim of this work is to investigate classes of groupoids, i.e. algebras with one binary operation, that satisfy certain identities or other conditions, and look for free objects in such classes.

AMS Subj. Classification: 03C05, 08B20

Key Words: groupoid, free groupoid, $(n, \mathcal{V})$-groupoid, power $\mathcal{V}$-groupoid, free $(n, \mathcal{V})$-groupoid, injective $(n, \mathcal{V})$-groupoid

1. Introduction

This paper is a review of a part of my doctoral thesis "*Free and injective objects in some classes of n-groupoids*". The thesis was prepared during the last three years at the Institute of Mathematics in the Faculty of Natural Sciences and Mathematics, "Ss. Cyril and Methodius" University, Skopje, Macedonia, and some of its parts were supported by the Macedonian Academy of Sciences and Arts through the project "Algebraic Structures".

We introduce the basic idea of this work. For the notation and basic notions of universal algebra the reader is referred to [12] and [13].

Let X be an arbitrary nonempty set whose elements are called variables and $\mathbf{T}_X = (T, \cdot)$ be the set of all groupoid terms over X in signature $\cdot$. The terms are denoted by $t, u, v, w \ldots$. Note that $\mathbf{T}_X$ is an absolutely free groupoid over X, where the operation is defined by $(t, u) \mapsto tu$. The groupoid $\mathbf{T}_X$ is injective, i.e. if $x, y, v, w \in T$, then $xy = vw \Rightarrow x = v, y = w$. The set X is the set of primes in $\mathbf{T}_X$ and generates $\mathbf{T}_X$. (An element a of a groupoid $\boldsymbol{G} = (G, \cdot)$ is said to be *prime* in $\boldsymbol{G}$ if and only if $a \neq xy$, for all $x, y \in G$.) These two properties of $\mathbf{T}_X$ characterize all absolutely free groupoids ([1]; Lemma 1.5): A

groupoid $\boldsymbol{H} = (H, \cdot)$ *is an absolutely free groupoid if and only if it satisfies the following two conditions:* $\boldsymbol{H}$ *is injective and the set of primes in* $\boldsymbol{H}$ *is nonempty and generates* $\boldsymbol{H}$. We refer to this proposition as *Bruck Theorem* for the class of all groupoids.

Let $\mathcal{V}$ be a variety of groupoids, i.e. a class of groupoids defined by a certain set of identities (or, equivalently, a class of groupoids that is hereditary and closed under homomorphic images and direct products). For a given variety $\mathcal{V}$ of groupoids, a free groupoid of a special form, called canonical form, is constructed. Namely, if X is a non-empty set and $\mathbf{T}_X$ is the term groupoid over X, then a $\mathcal{V}$-*canonical groupoid* $\boldsymbol{R} = (R, *)$ *over* X is a groupoid that satisfies the following conditions:

- (c_0) $X \subseteq R \subseteq T$ and $t \in R \Rightarrow P(t) \subseteq R$, where $P(t)$ is the set of subterms of the term t defined by:
  $t \in X \Rightarrow P(t) = \{t\}$ and $t = t_1 t_2 \Rightarrow P(t_1 t_2) = \{t_1 t_2\} \cup P(t_1) \cup P(t_2)$;
- (c_1) $tu \in R \Rightarrow t * u = tu$ and
- (c_2) $\boldsymbol{R}$ is a $\mathcal{V}$-free groupoid over X.

Using suitable properties of the obtained $\mathcal{V}$-canonical groupoid, we introduce the notion of $\mathcal{V}$-injective groupoid that is defined separately for each particular variety $\mathcal{V}$ [6]. Then the class of $\mathcal{V}$-free groupoids can be characterized by the class of $\mathcal{V}$-injective groupoids in the following way: *A groupoid* $\boldsymbol{H} = (H, \cdot)$ *is a* $\mathcal{V}$*-free groupoid if and only if* $\boldsymbol{H}$ *is* $\mathcal{V}$*-injective and the set of* $\mathcal{V}$*-prime elements in* $\boldsymbol{H}$ *is non-empty and generates* $\boldsymbol{H}$. (An element $a \in G$ is said to be $\mathcal{V}$*-prime* if and only if any equation of the form $a = bc$ is a consequence of the axioms in $\mathcal{V}$.) We call this property "*Bruck Theorem for the variety* $\mathcal{V}$".

Such characterizations are given for some classes of $(n, \mathcal{V})$-groupoids.

2. $(n, \mathcal{V})$-Groupoids

Let $\mathcal{V}$ be a variety of groupoids. A groupoid $\boldsymbol{G} = (G, \cdot)$ is said to be $(n, \mathcal{V})$-*groupoid* if and only if any subgroupoid generated by n elements of $\boldsymbol{G}$ belongs to the variety $\mathcal{V}$. The class of $(n, \mathcal{V})$-groupoids is denoted by $(n, \mathcal{V})$. If $n = 1$, then $(1, \mathcal{V})$-groupoids are called *power* $\mathcal{V}$*-groupoids*. In that case, the variety $\mathcal{V}$ is a subclass of the class $(1, \mathcal{V})$, and more generally $\mathcal{V}$ is a subclass of the class $(n, \mathcal{V})$. For any positive integers n, k, the class $(n + k, \mathcal{V})$ is a subclass of the class $(n, \mathcal{V})$. We give a description of canonical objects in the classes of power-commutative groupoids, power left and right idempotent groupoids, power-slim groupoids and biassociative groupoids. Also, a characterization by injective objects for some of this classes is given.

Throughout the paper we will use the concept of groupoid power and some of its properties stated in [4]. By $\mathbf{E} = (E, \cdot)$ we will denote the term groupoid over the set $\{e\}$. The elements of E are called *groupoid powers* and will be denoted by $f, g, h, \ldots$. For any groupoid $\boldsymbol{G} = (G, \cdot)$, each element $f \in E$ induces a transformation $f^{\boldsymbol{G}} : G \to G$, called an *interpretation* of f in $\boldsymbol{G}$, defined by:

$$e^{\boldsymbol{G}}(x) = x, \quad (gh)^{\boldsymbol{G}}(x) = g^{\boldsymbol{G}}(x)h^{\boldsymbol{G}}(x)$$

for any $g, h \in E$ and $x \in G$. We will write $f(x)$ instead of $f^{\boldsymbol{G}}(x)$ when $\boldsymbol{G}$ is understood.

In the sequel we will present without proofs some of the main results of this part of the thesis.

The class of commutative groupoids, i.e. groupoids that satisfy the identity $xy \approx yx$, is a variety of groupoids, here denoted by $\mathcal{C}om$. We investigate a class of groupoids larger than $\mathcal{C}om$, called ***the class of power-commutative groupoids***. It will be denoted by $\mathcal{P}_c$.

If $\boldsymbol{G}$ is a groupoid, then any subgroupoid of $\boldsymbol{G}$ generated by an element $a \in G$ (denoted by $\langle a \rangle$) is called *cyclic subgroupoid* of $\boldsymbol{G}$ with a generator a [2]. The cyclic subgroupoids are characterized in [5]: if $a \in G$, then $\langle a \rangle = \{f(a) : f \in E\}$. A groupoid $\boldsymbol{G}$ is said to be *power-commutative* if and only if every cyclic subgroupoid of $\boldsymbol{G}$ is commutative. Clearly, every commutative groupoid is power-commutative. The set of all 2×2 matrices under the multiplication is a nontrivial example of a power-commutative groupoid. Moreover, all semigroups are power-commutative groupoids. Directly from the definition we obtain that $\boldsymbol{G} \in \mathcal{P}_c$ if and only if $\boldsymbol{G}$ is a union of commutative cyclic subgroupoids of $\boldsymbol{G}$. This result enables to obtain an axiom system for $\mathcal{P}_c$, i.e. the class of power-commutative groupoids $\mathcal{P}_c$ is a variety of groupoids defined by the system of identities $\{f(x)g(x) \approx g(x)f(x) : f, g \in E\}$.

In order to give a description of free objects in the variety $\mathcal{P}_c$, we will introduce an ordering of terms. Namely, let X be a linearly ordered set and let that relation be denoted by $\leq$. An extension of the relation $\leq$ from X to T is defined as follows.

Let $t, u \in T$. (0) If $t, u \in X$, then $t \leq u$ in X implies that $t \leq u$ in T; (1) If $|\,t\,| < |\,u\,|$, then $t < u$, where $|\,t\,|$ is the length of the term t defined by $|\,t\,| = 1$, if $t \in X$, $|\,uv\,| = |\,u\,| + |\,v\,|$, if $t = uv$; (2) If $|\,t\,| = |\,u\,| \geq 2$ and $t \neq u$, where $t = t_1t_2$, $u = u_1u_2$, then $t < u \;\Leftrightarrow [\,t_1 < u_1 \;\vee\; (t_1 = u_1 \;\wedge\; t_2 < u_2)\,]$. *The relation $\leq$ is a linear ordering in T.*

A term $t \in T$ is said to be *order-regular* if and only if

$$t \in X \;\vee\; (t = t_1t_2 \in T \setminus X \;\wedge\; t_1 \leq t_2).$$

Specially, a groupoid power $f \in E$ is order-regular if and only if

$$f = e \ \vee \ (f = f_1 f_2 \ \wedge \ f_1 \leq f_2).$$

We will use canonical commutative groupoids constructed as follows. Define a subset T_c of T by

$$T_c = \{t \in T : \text{every subterm of } \ t \ \text{ is order-regular}\}$$

and an operation $\odot$ on T_c by

$$t, u \in T_c \ \Rightarrow \ t \odot u = \begin{cases} tu, \text{ if } \ t \leq u \\ ut, \text{ if } \ u < t. \end{cases} \tag{2.1}$$

Then $\mathbf{T}_c = (T_c, \odot)$ is a canonical commutative groupoid over X.

Specially, $\mathbf{E}_c = (E_c, \odot)$ is a canonical commutative groupoid over $\{e\}$, where $E_c = \{f \in E : \text{every subterm of } f \text{ is order-regular}\}$ and $\odot$ is defined by (2.1).

A term t is said to be *primitive* in $\mathbf{T}_X$ if and only if $t \neq f(u)$ for any $u \in T$ and any $f \in E \setminus \{e\}$; and t is said to be *potent* (or *nonprimitive*) in $\mathbf{T}_X$ if and only if $t = f(u)$ for some $u \in T$ and $f \in E \setminus \{e\}$. The following proposition is true ([3]): *For any potent term t there is a unique primitive term u and a unique groupoid power $f \in E \setminus \{e\}$ such that $t = f(u)$.* In that case we say that: *u is the base of t, f is the power of t* and denote $\underline{t}$, $t^{\sim}$, respectively.

Define the carrier of a free groupoid in $\mathcal{P}_c$ by

$$R = \{t \in T : u \in P(t) \ \Rightarrow \ u^{\sim} \in E_c\}, \tag{2.2}$$

and an operation $*$ on R by

$$t, u \in R \ \Rightarrow \ t * u = \begin{cases} tu, \text{ if } \ tu \in R \\ ut, \text{ if } \ \underline{t} = \underline{u} \ \text{ and } \ u^{\sim} < t^{\sim}. \end{cases} \tag{2.3}$$

One can obtain that $\boldsymbol{R} = (R, *)$ *defined by* (2.2) *and* (2.3) *is a free power-commutative groupoid over X in canonical form.* We will use the properties of the canonical groupoid $\boldsymbol{R} = (R, *)$ in $\mathcal{P}_c$ related to the elements of $\boldsymbol{R}$ that are not prime, to define a subclass of the class $\mathcal{P}_c$ that is larger than the class of $\mathcal{P}_c$-free groupoids, called the class of $\mathcal{P}_c$-injective groupoids. The class of $\mathcal{P}_c$-injective groupoids will be successfully defined if the following two conditions are satisfied. Firstly, the class of $\mathcal{P}_c$-injective groupoids should unable the characterization of $\mathcal{P}_c$-free groupoids: any $\mathcal{P}_c$-injective groupoid $\boldsymbol{H}$ whose set of primes is nonempty and generates $\boldsymbol{H}$, to be $\mathcal{P}_c$-free. Secondly, the class of $\mathcal{P}_c$-free groupoids has to be a proper subclass of the class of $\mathcal{P}_c$-injective groupoids. This is done for the class of power-commutative groupoids, i.e. the *Bruck Theorem for $\mathcal{P}_c$ holds* and

the class of $\mathcal{P}_c$-free groupoids is a proper subclass of the class of $\mathcal{P}_c$-injective groupoids.

In [7] a variety $\mathcal{U}$ of left and right idempotent groupoids, i.e. $\mathcal{U} = Var(x^2y^2 \approx xy)$, is investigated. We investigate a larger class, called ***the class of power left and right idempotent groupoids***, that will be denoted by $\mathcal{P}_\mathcal{U}$. A groupoid $\boldsymbol{G} = (G, \cdot)$ is *power left and right idempotent* if and only if every cyclic subgroupoid of $\boldsymbol{G}$ is left and right idempotent, i.e. belongs to $\mathcal{U}$. The elements of any groupoid in the class $\mathcal{P}_\mathcal{U}$ have almost trivial powers, i.e. *if $f(x)$ is a power of x, then either $f(x) = x$ or $f(x) = x^2$, for any nontrivial groupoid power f.* As a consequence we obtain that *the class $\mathcal{P}_\mathcal{U}$ is a variety of groupoids defined by the identities $x^2 \approx x^2x \approx xx^2 \approx x^2x^2$.* For details the reader is referred to [3].

Define the carrier of the desired $\mathcal{P}_\mathcal{U}$-canonical groupoid $\boldsymbol{R}$ by

$$R = \{t \in T : (\forall u \in P(t)) \mid u^\sim \mid \leq 2\}, \tag{2.4}$$

and an opperation $*$ on R by

$$t, u \in R \Rightarrow t * u = \begin{cases} tu, & \text{if } tu \in R \\ v^2, & \text{if } \underline{t} = \underline{u} = v, \ \mid t^\sim \mid + \mid u^\sim \mid \geq 3. \end{cases} \tag{2.5}$$

One can show that *the groupoid $\boldsymbol{R} = (R, *)$ defined by* (2.4) *and* (2.5) *is a free power left and right idempotent groupoid over X in canonical form.*

We use the properties of the obtained $P_\mathcal{U}$-canonical groupoid $(R, *)$ that are related to the elements in $(R, *)$ that are not prime. Namely, *if $t \in R$, then $t * t$ is an idempotent element in $\boldsymbol{R}$; t is idempotent in $\boldsymbol{R}$ if and only if t is a square in $\boldsymbol{R}$ and if t is idempotent in $\boldsymbol{R}$, then there is a unique nonidempotent $u \in R$, i.e. $u \neq u * u$, such that $t = u * u$.* Also, *for every $t \in R \setminus X$ there is a unique pair $(u, v) \in R \times R$ such that $t = uv = u * v$.* We say that (u, v) is the *pair of divisors* of t in $\boldsymbol{R}$. If $u = v$, then u is a *divisor* of t.

Define a $P_\mathcal{U}$-injective groupoid in the following way. A groupoid $\boldsymbol{H} = (H, \cdot)$ is said to be *$\mathcal{P}_\mathcal{U}$-injective* if and only if the following conditions are satisfied:

(0) $\boldsymbol{H} \in \mathcal{P}_\mathcal{U}$

(1) If $a \in H$ is idempotent, then there is a unique nonidempotent $c \in H$, such that $a = c^2$ and the equality $a = xy$ holds if and only if $\{x, y\} \subseteq \{c, c^2\}$. (In that case c is the *divisor* of a or c is the *base* of a.)

(2) If $a \in H$ is nonidempotent and nonprime in $\boldsymbol{H}$, then there is a unique pair $(c, d) \in H \times H$, such that $a = cd$ and $\underline{c} \neq \underline{d}$.

(Note thet c, d can be both idempotents; one idempotent and the other nonidempotent; both nonidempotents.)

It is proved in [3] that the *Bruck Theorem for $\mathcal{P}_{\mathcal{U}}$ holds*, that *neither of the classes $\mathcal{P}_{\mathcal{U}}$-free and $\mathcal{P}_{\mathcal{U}}$-injective groupoids is hereditary* and that *the class of $\mathcal{P}_{\mathcal{U}}$-free groupoids is a proper subclass of the class of $\mathcal{P}_{\mathcal{U}}$-injective groupoids.*

An Evans' result ([8]) is used to show that the word problem is solvable for the variety $\mathcal{P}_{\mathcal{U}}$. Note that if a partial groupoid A is strongly embeddable into a power left and right idempotent groupoid, then it satisfies the following condition:

(j_0) if $a \in A$ is such that a^2 is defined, then a^2a, aa^2 and a^2a^2 are also defined and $a^2a = aa^2 = a^2a^2 = a^2$.

For a partial groupoid A satisfying (j_0) we define a groupoid $(G, \circ)$ as follows:

(j_1) if xy is defined in A, then $x \circ y = xy$

(j_2) if x^2 is not defined in A, then $x \circ x = x$

(j_3) if xy is not defined in A and $x \neq y$, then $x \circ y = c$, where c is a fixed element in A.

It is shown that *if A is a partial groupoid satisfying (j_0), then $(G, \circ)$ defined above by (j_1) – (j_3) is a power left and right idempotent groupoid.* As a special case of the Evans' Theorem we obtain the following theorem: *if every partial $\mathcal{P}_{\mathcal{U}}$-groupoid is embeddable into a $\mathcal{P}_{\mathcal{U}}$-groupoid, then the word problem is solvable for the variety $\mathcal{P}_{\mathcal{U}}$.* As a corollary, we have that the *word problem for the variety $\mathcal{P}_{\mathcal{U}}$ is solvable.*

The variety of groupoids that satisfy the identity $x(yz) \approx xz$ is called the variety of *slim groupoids*. We investigate ***the class of power-slim groupoids***, i.e. the class of groupoids such that every cyclic subgroupoid satisfies the identity $x(yz) \approx xz$. Our purpose is to construct free objects in that class. First we will give a description of the free slim groupoids (slightly different then the given description in [11]). The variety of slim groupoids will be denoted by $\mathcal{V}_s$. Define a subset F_s of T by

$$F_s = \{t \in T : \ (\forall u, v, w \in T)\ u(vw) \notin P(t)\} \tag{2.6}$$

and an operation $*$ on F_s by

$$t, u \in F_s \Rightarrow t * u = \begin{cases} tu, & \text{if } \ u \in X \\ tu_2, & \text{if } \ u = u_1u_2 \ \wedge \ u_2 \in X. \end{cases} \tag{2.7}$$

The groupoid $\boldsymbol{F}_s = (F_s, *)$ defined by (2.6) *and* (2.7) *is a canonical slim groupoid over X.* Specially, if $X = \{e\}$, then the canonical slim groupid over $\{e\}$ is denoted by $\mathbf{E}_s = (E_s, *)$, where $E_s = \{f \in E : (\forall g, h, j \in E)\ g(hj) \notin P(f)\}$, i.e. $E_s = \{e^n : \ n \geq 1\}$, and $f, g \in E_s \ \Rightarrow \ f * g = fe$.

A groupoid $\boldsymbol{G} = (G, \cdot)$ is said to be a *power-slim groupoid* if and only if

every cyclic subgroupoid of $\boldsymbol{G}$ is a slim groupoid. The class of such groupoids will be denoted by $\mathcal{P}_s$. Using the characterization of cyclic groupoids, we obtain that *$\mathcal{P}_s$ is a variety of groupoids defined by the set of identities* $\{f(x)\,(\,g(x)h(x)\,) \approx f(x)h(x) : f, g, h \in E\}$.

Define a subset R of T by

$$R = \{t \in T :\ (\forall u \in P(t))\ u^{\sim} \in E_s\} \tag{2.8}$$

and an operation $*$ on R by

$$t, u \in R \Rightarrow\ t * u = \begin{cases} tu, \text{ if } (tu)^{\sim} \in E_s \\ t\,\underline{u}, \text{ if } \underline{t} = \underline{u} \ \wedge\ |\,u^{\sim}\,| \geq 2. \end{cases} \tag{2.9}$$

One can show that *the groupoid $\boldsymbol{R} = (R, *)$ defined by* (2.8) *and* (2.9) *is a canonical power-slim groupoid over X. The groupoid $\boldsymbol{R} = (R, *)$ is right cancellative and it is not left cancellative.*

In the paper [9] ***the variety of biassociative groupoids***, denoted by $\mathcal{B}ass$ is considered. A groupoid $\boldsymbol{G}$ is said to be *biassociative* if and only if every subgroupoid generated by at most two elements of $\boldsymbol{G}$ is a subsemigroup. Free objects are constructed using a chain of partial biassociative groupoids that satisfy certain properties. The obtained free objects are not canonical. In [10] the obtained free objects have canonical form.

Let $\boldsymbol{G} = (G, \cdot)$ be a groupoid and $a, b \in G$. We denote by $\langle\, a, b \,\rangle$ the subgroupoid of $\boldsymbol{G}$ generated by a, b and by $\langle\, a \,\rangle$ the subgroupoid generated by a. Clearly, $\langle\, a \,\rangle \subseteq \langle\, a, b \,\rangle$ and if $b \in \langle\, a \,\rangle$, then $\langle\, a, b \,\rangle = \langle\, a \,\rangle$; specially, $\langle\, a, a \,\rangle = \langle\, a \,\rangle$. The subgroupoids $\langle\, a, b \,\rangle$ and $\langle\, b, a \,\rangle$ are equal.

Let $a_1, a_2, \ldots, a_n$ be a finite sequence of elements in a groupoid $\boldsymbol{G}$. We denote by $a_1 a_2 \ldots a_n$ the product of the sequence $a_1, a_2, \ldots, a_n$ in $\boldsymbol{G}$ defined as follows:

i) if $n = 3$, then $a_1 a_2 a_3 \stackrel{\text{df}}{=} a_1(a_2 a_3)$ and

ii) if $n \geq 3$, then $a_1 a_2 \ldots a_n \stackrel{\text{df}}{=} a_1(a_2 \ldots a_n)$.

We call $a_1 a_2 \ldots a_n$ the *main product* of the sequence $a_1, a_2, \ldots, a_n$. If $n = 1$ and $n = 2$, then a_1 and $a_1 a_2$ will also be called the main products of the sequences a_1 and a_1, a_2 respectively.

Let $t, u \in T$ and $\langle\, t, u \,\rangle$ be the subgroupoid of $\mathbf{T}_X$ generated by t, u. Each element x of $\langle\, t, u \,\rangle$ is a product of a finite sequence of elements $x_1, \ldots, x_n$ $(n \geq 1)$, where each x_i is either t or u, i.e. $\{x_1, x_2, \ldots, x_n\} \subseteq \{t, u\}$. Any such product is constructed by the two generators t, u and therefore we call it a *binary product* or shortly *biproduct*. Thus, if a term $x \in T$ is an element of $\langle\, t, u \,\rangle$, then we say that x *has a representation as a biproduct* (or shortly, x *is a biproduct*)

with the generating pair $\{t,u\}$ and denote it by $x_{\langle t,u\rangle}$. (In this case we also say that x is the *carrier* of the biproduct $x_{\langle t,u\rangle}$.) *If* $t,u,x \in T$, *where* $x \in \langle\, t,u\,\rangle$, $t \notin \langle\, u\,\rangle$ *and* $u \notin \langle\, t\,\rangle$, *then* x *has a unique representation as a biproduct with the generating pair* $\{t,u\}$.

A biproduct $x_{\langle t,u\rangle}$ of a term x is said to be *maximal* in $\mathbf{T}_X$ if and only if for any biproduct $x_{\langle \alpha,\beta\rangle}$ of x, the hierarchy $\chi_{\langle \alpha,\beta\rangle}(x)$ does not exceed the hierarchy $\chi_{\langle t,u\rangle}(x)$, i.e. $\chi_{\langle \alpha,\beta\rangle}(x) \leq \chi_{\langle t,u\rangle}(x)$. (For details the reader is referred to [10].)

Let $x = x_1x_2\dots x_m$ be the main product of $x_1, x_2, \dots, x_m$ in $\mathbf{T}_X$.

If $\{x_1, x_2, \dots, x_m\} \subseteq \{t,u\}$, for some terms t,u of T, then we call $x_1x_2\dots x_m$ the *main biproduct* of x in $\mathbf{T}_X$ with the generating pair $\{t,u\}$ and denote it by $x_{t,u}$. (If $u = t$, i.e. the generating "pair" is $\{t,t\}$, we write x_t instead of $x_{t,t}$.)

If $x = x_1x_2\dots x_m$ *and* $x = x'_1x'_2\dots x'_n$ *are main biproducts of* x *in* $\mathbf{T}_X$ *with the same generating pair* $\{t,u\}$, *then* $m = n$ *and* $x_i{=}x'_i$, *for* $i = 1,2,\dots,m$. *Specially, any maximal biproduct of* $x \in \mathbf{T}_X$, *that is a main biproduct, is uniquely determined.*

We define the desired groupoid $\boldsymbol{R} = (R, *)$ by:

(2.10)

$$R = \{x \in T : \text{every biproduct of any subterm of } \ x \ \text{ is a main biproduct}\}$$

and an operation $*$ on R as follows.

Let $x, y \in R$, $x = x_1x_2\dots x_m$, $y = y_1y_2\dots y_n$ be maximal biproducts and put

$$x * y = \begin{cases} xy, \text{ if } xy \in R \\ x_1x_2\dots x_my_1y_2\dots y_n, \text{ if } xy \notin R. \end{cases} \tag{2.11}$$

The groupoid $\boldsymbol{R} = (R, *)$, *defined by* (2.10) *and* (2.11) *is a canonical biassociative groupoid over* X.

The problem of power $\mathcal{V}$-groupoids can be expanded to power $\mathcal{V}$-ternary groupoids or power $\mathcal{V}$-n-ary groupoids. For instance, we can investigate power-commutative ternary groupoids and power-semicommutative ternary groupoids, since a canoninical description of free objects in the varieties of commutative ternary groupoids and semicommutative ternary groupoids are obtained in the thesis.

References

[1] R. H. Bruck, *A Survey of Binary Systems*, Springer-Verlag 1958

[2] V. Celakoska-Jordanova, Cyclic subgroupods of an absolutely free groupoid, *Proc. of III Congesss of SMM 2005,* 2007, 217–225

[3] V. Celakoska-Jordanova, Free groupoids in the class of power left and right idempotent groupoids, *International Journal of Algebra*, **2**, no. 10, 2008, 451 - 461

[4] G. Čupona, N. Celakoski, S. Ilić, Groupoid powers, *Bulletin Mathematique*, Skopje, Macedonia, **25**, 2001, 5–12

[5] G. Čupona, N. Celakoski, S. Ilić, On monoassociative groupoids, *Bulletin Mathematique*, Skopje, Macedonia, **26**, (LII), 2002, 5–16

[6] G. Čupona, N. Celakoski, B. Janeva, Injective groupoids in some varieties of groupoids, *Proc. of II Congesss of SMIM 2000*, 2003, 47–55

[7] G. Čupona, N. Celakoski, On groupoids with the identity $x^2y^2 = xy$, *Contributions Sec. Math. Tech. Sci., Macedonian Academy of Sciences and Arts*, **XVIII**, 1–2, 1997, 5 – 15

[8] T. Evans, Embeddability and the word problem, *J. London Math. Soc.*, **28**, 1953, 76 – 80

[9] S. Ilić, B. Janeva, N. Celakoski, Free Biassociative Groupoids, *Novi Sad J. Math.*, **35**, No. 1, 2005, 15 – 23

[10] B. Janeva, S. Ilić, V. Celakoska-Jordanova, Canonical Biassociative Groupoids, *Publications de'lInstitute Mathematique, SANU, Nouvelle serie*, **81(95)**, 2007, 103 – 109

[11] J. Ježek, Slim Groupoids, *Czechoslovak Math. J., MATH-alg-2005/22* (13 pages)

[12] J. Ježek, *Universal Algebra*, First edition, April 2008, http://www.karlin.mff.cuni.cz/ jezek/

[13] R. N. McKenzie, G. F. McNulty, W. F. Taylor, *Algebras, Lattices, Varieties*, Wadsworth & Brooks/Cole, 1987

Institute of Mathematics
Faculty of Natural Sciences and Mathematics
"Ss. Cyril and Methodius" University
Skopje, Macedonia

E-Mail: vesnacj@pmf.ukim.mk

M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Splines in Numerical Integration

Zlatko Udovičić

We gave a short review of several results which are related to the role of splines (cardinal, centered or interpolating) in numerical integration. Results deal with the problem of approximate computation of the integrals with spline as a weight function, but also with the problem of approximate computation of the integrals without weight function. Besides, we presented an algorithm for calculation of the coefficients of the polynomials which correspond to the cardinal B-spline of arbitrary order and described five methods for calculation of the moments in the case when cardinal B-spline of order $m, m \in \mathbb{N}$, is a weight function.

AMS Subj. Classification: 65D07, 65D30.

Key Words: cardinal B-spline, coefficients, moments, rectangular rule, interpolating quadratic spline, hat function, cubic B-spline.

1. Introduction

In the approximation theory, cardinal B-splines have a very important place (in different methods for solving initial and boundary value problems, spline interpolation, multiresolution approximation...). The main topic of the author's PhD thesis is the use of the cardinal B-splines in numerical integration. The obtained results can be grouped in two ways, depending on the type of integral under consideration or the order of the spline applied.

The first grouping refers to the results where the problem of approximate computation of the integral

$$\int_0^m \varphi_m(x) f(x) dx \tag{1.1}$$

(integrals with cardinal B-spline of order m as a weight function) is considered and those where the classical problem of numerical analysis, i.e. the problem of approximate computation of the integral

$$\int_a^b f(x) dx \tag{1.2}$$

is considered.

In organizing this paper (as well as in the PhD thesis) we opt for the second possibility. In the next section, we deal with splines of arbitrary order, while the final one presents the results related to the splines of a low order.

Let us recall here the definition of the cardinal B-spline and list its basic properties.

Definition 1. *Cardinal B-spline of the first order, denoted by* $\varphi_1(\cdot)$, *is the characteristic function of the interval* $[0,1)$, *i.e.*

$$\varphi_1(x) = \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases}.$$

Cardinal B-spline of order $m, m \in \mathbb{N}$, *denoted by* $\varphi_m(\cdot)$, *is defined as a convolution*

$$\begin{aligned} \varphi_m(x) &= (\varphi_{m-1} * \varphi_1)(x) = \int_{\mathbb{R}} \varphi_{m-1}(x-t)\varphi_1(t)dt \\ &= \int_0^1 \varphi_{m-1}(x-t)dt. \end{aligned}$$

Theorem 1. *Cardinal B-spline of order* $m, m \in \mathbb{N}$, *has the following properties*

(1) $supp\varphi_m(\cdot) = [0,m]$;

(2) $\varphi_m(\cdot) \in C^{m-2}[0,m]$;

(3) *at each interval* $[k,k+1], 0 \le k \le m-1$, *cardinal B-spline of order* m *is a polynomial of degree equal to* $m-1$;

(4)

$$(\forall t \in [0,m])\, \varphi_m(t) = \frac{t}{m-1}\varphi_{m-1}(t) + \frac{m-t}{m-1}\varphi_{m-1}(t-1), m \ge 2; \tag{1.3}$$

(5)

$$(\forall t \in [0,m])\, \varphi_m'(t) = \varphi_{m-1}(t) - \varphi_{m-1}(t-1), m \ge 2; \tag{1.4}$$

(6) *Cardinal B-spline is symmetric at the interval* $[0,m]$, *i.e.* $(\forall t \in [0,m])\, \varphi_m(t) = \varphi_m(m-t)$;

(7)

$$(\forall a \in \mathbb{R}) \sum_{i \in \mathbb{Z}} \varphi_m(i-a) = 1; \tag{1.5}$$

(8) *For any* m *time differentiable function* $g(\cdot)$ *holds*

$$\int_{\mathbb{R}} \varphi_m(x) g^{(m)}(x)dx = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} g(k);$$

(9) *The cardinal B-spline is a solution of the so-called dilatation equation*

$$(\forall t \in \mathbb{R})\, \varphi_m(t) = \frac{1}{2^{m-1}} \sum_{k=0}^{m} \binom{m}{k} \varphi_m(2t-k).$$

The proof of the theorem, and more details about cardinal B-splines one can find in [1] or [2].

2. **Splines of arbitrary order**

2.1. **Calculation of the coefficients of the cardinal B-spline.** The first result presented in this paper is an algorithm for calculation the coefficients of the polynomials which correspond to the cardinal B-spline of order m. By using equalities (1.3) and (1.4) it is easy to obtain differential equation

$$(m-x)\varphi_m'(x) + (m-1)\varphi_m(x) = m\varphi_{m-1}(x).$$

Evidently, the solution of this equation is a cardinal B-spline of order m. Furthermore, let $x \in [k, k+1]$ and let $\varphi_m(x) = \sum_{i=0}^{m-1} a_i^{(m,k)} x^i$. By using the previous differential equation and the symmetry of the cardinal B-spline one can prove that the following reccurence relations hold:

$$\begin{aligned}
a_{m-1}^{(m,0)} &= \frac{1}{m-1} a_{m-2}^{(m-1,0)}, \\
a_i^{(m,0)} &= 0, m-2 \geq i \geq 0, \\
a_{m-1}^{(m,k)} &= \frac{1}{(m-1)(m-k)} \left[m a_{m-2}^{(m-1,k)} - k(m-1) a_{m-1}^{(m,k-1)} \right. \\
&\quad \left. - a_{m-2}^{(m,k-1)} \right], \\
a_i^{(m,k)} &= \frac{m}{i+1-m} \left[(i+1) a_{i+1}^{(m,k)} - a_i^{(m-1,k)} \right], \\
&\quad m-2 \geq i \geq 0, \\
a_{m-1}^{(m,m-k-1)} &= (-1)^{m-1} a_{m-1}^{(m,k)}, \\
a_j^{(m,m-k-1)} &= \frac{(-1)^j}{j!} \sum_{i=0}^{m-j-1} (i+1)(i+2)\dots(i+j) \cdot a_{i+j}^{(m,k)} m^i, \\
&\quad 0 \leq j \leq m-2.
\end{aligned}$$

Those relations enable us to construct an algorithm for calculating the coefficients of the cardinal B-spline of order m.

2.2. **Calculation of the moments of the cardinal B-spline.** Calculating moments of the given weight function has essential significance in construction of the orthogonal polynomials and quadrature rules, as well as in other fields of the approximation theory. On the other hand, a very frequent weight function is exactly the cardinal B-spline (finite elements method, multiresolution approximation,...). In the paragraph below, five methods (without proofs) for calculation of moments

$$\mathbb{M}_{n,m} = \int_0^m \varphi_m(t) t^n dt, n \in \mathbb{N}_0.$$

are provided.

(1)

$$\mathbb{M}_{n,m} = \frac{1}{2^m(2^n-1)} \sum_{k=1}^{m} \sum_{l=0}^{n-1} \binom{m}{k} \binom{n}{l} k^{n-l} \mathbb{M}_{l,m};$$

(2)

$$\begin{aligned} \mathbb{M}_{n,m} &= \sum_{k=0}^{n} \binom{n}{k} \mathbb{M}_{k,m-1} - \frac{1}{m-1} \sum_{k=0}^{n-1} \binom{n}{k} \mathbb{M}_{k+1,m-1} \\ &= 1 + \sum_{k=1}^{n} \left[\binom{n}{k} - \frac{1}{m-1} \binom{n}{k-1} \right] \mathbb{M}_{k,m-1}; \end{aligned}$$

(3)

$$\mathbb{M}_{n,m} = \frac{1}{n+1} \sum_{k=0}^{n} \binom{n+1}{k} \mathbb{M}_{k,m-1};$$

(4)

$$M_{n,m} = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} \frac{k^{m+n}}{(m+n)(m+n-1)\dots(n+1)}.$$

The last method is, of course, the most important result of this paragraph. This method generalizes equality (1.5) and we will formulate it as a theorem.

Theorem 2. *For any $a \in \mathbb{R}$ and every $m \in \mathbb{N}, m \geq 2$, the following equality holds*

$$\mathbb{M}_{n,m} = \sum_{i \in \mathbb{Z}} \varphi_m(i-a)(i-a)^n, 0 \leq n \leq m-1.$$

In the case when a is an integer and m is odd, one can easily check that the previous equality also holds for $n = m$. Those results were recently published [8].

2.3. **One point quadrature rule with cardinal B-spline.** In this paragraph we are presenting probably the most important result of our research which was published in [7]. The basic formula is a classical one point quadrature rule

$$\int_a^b f(x)dx \approx (b-a)f(X), \tag{2.6}$$

with choosen point $X = (1-\lambda)a + \lambda b$, for some $\lambda \in [0,1]$.

Let $x_1 < x_2 < \ldots < x_{p-1}$ be an arbitrary points from the interval $(0,1)$ and let the partition

$$\begin{array}{ccccc} 0, & x_1, & x_2, & \ldots, & x_{p-1}, \\ 1, & x_1+1, & x_2+1, & \ldots, & x_{p-1}+1, \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m-1, & x_1+m-1, & x_2+m-1, & \ldots, & x_{p-1}+m-1, \\ m, & & & & \end{array}$$

of the interval $[0,m]$ be given. We call this kind of mesh "quaziuniform". Furthermore, let the points $X_k+i, 0 \le i \le m-1$, where $X_k = (1-\lambda_k)x_k + \lambda_k x_{k+1}$, for some $\lambda_k \in [0,1], 0 \le k \le p-1$, at each interval of the given partition, be chosen (naturally, $x_0 = 0$ and $x_p = 1$).

At each interval of partition, we will use the formula (2.6) with chosen points $X_k + i, 0 \le i \le m-1, 0 \le k \le p-1$, to approximately compute the integral (1.1). After summation, we will obtain

$$\int_0^m \varphi_m(x)f(x)dx \approx \sum_{j=0}^{p-1}(x_{j+1}-x_j)\sum_{i=0}^{m-1}\varphi_m(X_j+i)f(X_j+i). \tag{2.7}$$

Theorem 3. *If $f(x) = x^n, 0 \le n \le m-1$, then the quadrature rule* (2.7) *is exact, for each $m \in \mathbb{N}$.*

In accordance with the previous theorem formula (2.7) has, conditionally speaking, algebraic degree of exactness equal to $m-1$. In some special cases this degree of exactness can be improved up to m.

2.4. **Numerical integration using cardinal B-splines.** One of the classical problems in the mathematical analysis is, of course, the calculation of the integral (1.2). Since this integral cannot, in general, be calculated exactly, there is a large number of methods for its approximate computation. We are presenting here a new method which is obtained by using cardinal B-splines. The main idea is to approximate the function $f(\cdot)$ by its projection on the cardinal B-spline space

$$\widehat{f}(x) = \sum_{k\in\mathbb{Z}} c_k \varphi_m^{(j,k)}(x),$$

where $\varphi_m^{(j,k)}(x) = 2^{\frac{j}{2}}\varphi_m(2^j x - k)$. The coefficients of the approximation are solution of the corresponding system of linear equations. This approach leaves $m-2$ coefficients free, so additional conditions can be chosen depending on the concrete situation. One possibility is to choose some integer nodes to have multiplicity two. Finally, one obtains

$$\begin{aligned}\int_0^m f(x)dx \quad &\approx \quad \int_0^m \widehat{f}(x)dx = 2^{-\frac{j}{2}}\left(\sum_{k=-m+1}^{-1} c_k\left(1-\mu_{0,m}^{-k}\right)\right.\\ &\left. + \sum_{k=0}^{(2^j-1)m} c_k + \sum_{k=(2^j-1)m+1}^{2^j m-1} c_k \mu_{0,m}^{2^j m-k}\right),\end{aligned}$$

where

$$\mu_{n,m}^x = \int_0^m \varphi_m(t)t^n dt$$

are the so-called shortened spline moments. Those moments, for example, can be calculated recursively

$$\begin{aligned}\mu_{n,1}^x &= \int_0^x \varphi_1(t)t^n dt = \begin{cases} 0, & x<0 \\ \frac{x^{n+1}}{n+1}, & x\in[0,1] \\ \frac{1}{n+1} & x>1 \end{cases},\\ \mu_{n,m}^x &= \int_0^x \varphi_m(t)t^n dt\\ &= \frac{x^{n+1}\varphi_m(x)}{n+1} - \frac{1}{n+1}\int_0^x \varphi_{m-1}(t)t^{n+1}dt\\ &\quad + \frac{1}{n+1}\int_0^{x-1}\varphi_{m-1}(t)(t+1)^{n+1}dt\\ &= \frac{1}{n+1}\left(x^{n+1}\varphi_m(x) - \mu_{n+1,m-1}^x + \sum_{k=0}^{n+1}\binom{n+1}{k}\mu_{k,m-1}^{x-1}\right).\end{aligned}$$

This result was inspired by the paper [5].

3. Splines of given (low) order

3.1. **Some modifications of the trapezoidal rule.** This section begins with the result recently published in [6]. The main idea is based on the fact that the construction of the quadratic interpolating spline leaves one parameter free. By an appropriate choice of the free parameter we obtain reproduction of some well known quadrature rules, but also a wide class of new quadrature rules.

Except the reproduction of the Simpsons and Ermits quadrature rule, we obtain the following formulas:

$$\begin{aligned}
\int_a^b f(x)dx &\approx \frac{h}{2}\left(f_0 + 2\sum_{k=1}^{2m-1} f_k + f_{2m}\right) - \frac{h^3}{6}\sum_{k=1}^{m} f''_{2k-1};\\
&\approx \frac{h}{2}(f_0+f_1) - \frac{h^3 f''(X)}{12} + Q_S[f, x_1, b, 2m];\\
&\approx \frac{h}{2}\left(f_0 + 2\sum_{k=1}^{2m} f_k + f_{2m+1}\right) - \frac{h^3}{6}\left(\frac{f''(X)}{2} + \sum_{k=1}^{m} f''_{2k}\right);\\
&\approx \frac{h}{2}(f_0+f_1) - \frac{h^2}{12}(f'_1 - f'_0) + Q_S[f, x_1, b, 2m];\\
&\approx Q_T[f, x_0, x_1] - \frac{h^3}{12}\left[f''(X) + f'''(X)\left(x_0 + \frac{h}{2} - X\right)\right]\\
&\quad + Q_S[f, x_1, b, 2m].
\end{aligned}$$

In the second and the third formulas X is any point from the interval $[x_0, x_1]$ such that $f''(X) = \lambda f''_0 + (1-\lambda)f''_1$, where $\lambda \in [0,1]$, while in the last formula X is any point from the interval $[x_0, x_1]$. Furthermore, Q_T and Q_S denote a classical trapezoidal and the Simpsons rule respectively. The accuracy of the obtained formulas is $O(h^4)$.

3.2. **A certain class of quadratures with specific weight function.** The last result was inspired by the papers [3] and [4], and it is about the so-called quadrature formulas of "practical type". Hence, we are considering the quadrature formula

$$\int_0^m \varphi_m(x) f(x)dx \approx \sum_{i=1}^{5} A_i f(x_i).$$

This formula is of "practical type" if the following conditions hold:

- coefficients $A_i, 1 \le i \le 5$, are symmetric, i.e. ($A_1 = A_5$ and $A_2 = A_4$);
- nodes $x_i, 1 \le i \le 5$, are symmetric, rational numbers from the interval $[0, m]$.

We proved that in the case when the weight function is cardinal B-spline of order two (hat function), i.e. in the case when the weight function is a cardinal B-spline of order four (cubic B-spline), the maximal algebraic degree of exactness of such type of quadratures is equal to five.

References

[1] C. Chui, *An Introduction to Wavelets*, Academic Press, inc., Boston San Diego New York London Sydney Tokyo Toronto, 1992.

[2] C. Chui, *Wavelets: A mathematical tool for signal analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[3] E. Constantinescu, A certain class of quadratures, *Acta Univ. Apulensis Math. Inform.*, **7**, 2004, 119-116.

[4] E. Constantinescu, A certain class of quadratures, *Gen. Math.*, **10**, No. 1-2, 2002, 75-84.

[5] H. Hashish, S. H. Behiry, N. A. El-Shamy , Numerical Integration Using Wavelets, *Appl. Math. Comput.*, 2009

[6] Z. Udovičić, Some modifications of the trapezoidal rule, *Sarajevo J. Math.*, **2**(15), 2006, 237-245.

[7] Z. Udovičić,One point quadrature rule with cardinal B-spline, *J. Math. Anal. Appl.*, **360**, 2009, 432-438.

[8] Z. Udovičić, Calculation of the moments of the cardinal B-spline, *Sarajevo J. Math.*, accepted.

[9] Z. Udovičić, A certain class of quadratures with hat function as a weight function, *Acta Univ. Apulensis Math. Inform.*, accepted.

Faculty of Sciences,
Department of Mathematics
Zmaja od Bosne 35
71000 Sarajevo
Bosnia and Herzegovina

E-Mail: zzlatko@pmf.unsa.ba

O r d e r

F o r m

Mathematica Balkanica

Date:_____

Bulgarian Academy of Sciences
National Committee for Mathematics
Editor-in-Chief: **acad. Petar Kenderov**

□ **LIBRARIES & INSTITUTIONS** □ **INDIVIDUALS**

Institutions:_____ **Name:_____**
_____ **_____**
E-mail:_____

Contact Person:_____
_____
E-mail:_____

Full Postal Address: **_____**
Volume to be sent **_____**
_____

Please send us/me □ **One copy** □ **Two copies** □ **More:_____**

of the journal **Mathematica Balkanica, ISSN 0205-3217**

PRICE : 95 EUR, postage included

DETAILS OF PAYMENT:

□ Encl,. **Check,** payable to **Mathematica Balkanica** for EUR**_____** **(**Send to the Editors!)

□ **Bank Transfer** for USD_____ to (Send Order Form and copy of payment document to the Editors!)

UNICREDIT BULBANK
BIC UNCRBGSF
IBAN BG65UNCR96603148580710
IPP-BAS
Mathematica Balkanica

MATHEMATICA BALKANICA
Phone +359-2-979 63 11; Fax +359-2-870 72 73 **E-mail**: **balmat@bas. bg**

MATHEMATICA BALKANICA

BULGARIAN ACADEMY OF SCIENCES
(NATIONAL COMMITTEE FOR MATHEMATICS)

COPYRIGHT TRANSFER AGREEMENT

This is to verify that my / our article:

TITLE OF THE ARTICLE: _____

(ALL) AUTHORS: _____

DATE OF SUBMISSION: _____-_____

is my /our own original work, and has **not been / and will be not / published elsewhere**, in the same or substantially the same form without acknowledging prior publication in the journal "Mathematica Balkanica".

We agree to transfer the copyright of this article to journal "Mathematica Balkanica" **(MB).** The author(s) hereby grant(s) this journal the license to publish and distribute this article in this journal in any and all forms of actual or future media.

The author(s) reserve(s) the right to distribute the final version of this article (exactly as published in the MB journal) and to place it on their own homepage(s) or in a public digital repository, always referring to its coordinates in MB.

The corresponding author warrants that he/she has the full power and authority to enter into this Agreement and to grant the rights granted in this Agreement. All the authors of this work authorize the corresponding author to sign this agreement on their behalf.

THE CORRESPONDING AUTHOR:
Name: .....
Signature: ..... Date: .....

This copyright transfer form is essential for publication of your article in the MB journal**,** and should be sent (by fax, e-mail or air mail) to:

MATHEMATICA BALKANICA – EDITORIAL OFFICE
„ACAD. G. BONTCHEV" STR. BLOCK 25A,
SOFIA 1113, BULGARIA
FAX +359-2-8707273 , TEL. +359-2-9796311
E-MAIL BALMAT@BAS.BG (SECRETARY: MRS. VOLYA ALEXANDROVA)

Editorial Board [1]

[1] The Editorial Board is in a process of constituting.

Contents (Continued from the back cover page)

ISSN 0205-3217

The present Volume 24 (2010), Fasc. 3-4, *of* **MATHEMATICA BALKANICA** *is supported by* **TEMPUS Project SEE Doctoral Studies in Mathematical Sciences (144703-TEMPUS-2008-BA-TEMPUS-JPCR)**

Contents

(Continued on page 364)