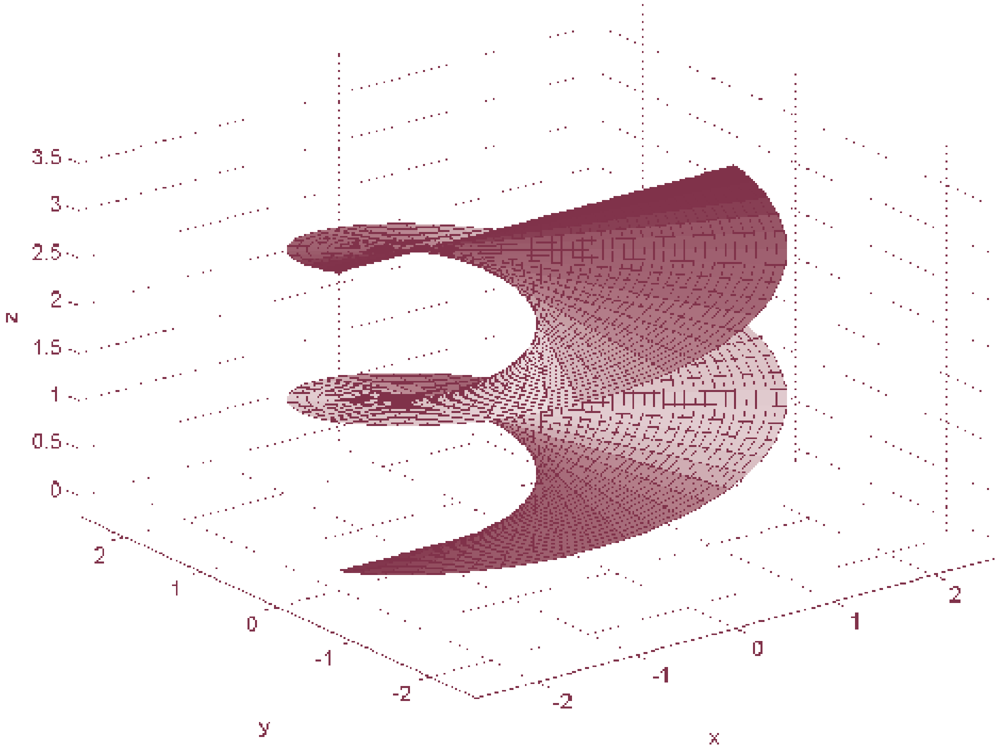
STUDIA UNIVERSITATIS BABEŞ-BOLYAI



MATHEMATICA

4/2014

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
MATHEMATICA

4/2014

YEAR **(LIX) 2014**
MONTH **DECEMBER**
ISSUE **4**

S T U D I A
UNIVERSITATIS BABEŞ-BOLYAI

MATHEMATICA
4

Redacţia: 400084 Cluj-Napoca, str. M. Kogălniceanu nr. 1
Telefon: 0264 405300

CONTENTS

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 399–418

Mathematics education in Romanian at Babeş-Bolyai University Cluj-Napoca

Dorel I. Duca and Adrian Petruşel

Abstract. In this paper, we will present the most important moments of the evolution and development of the mathematical education and research activities in Romanian at Babeş-Bolyai University Cluj-Napoca. The main figures of the mathematical university staff are also presented.

Mathematics Subject Classification (2010): 01A72, 01A73, 01A70.

Keywords: University of Cluj, Babeş-Bolyai University, Faculty of Sciences, Faculty of Mathematics.

1. A short walk through the history of the university education in Cluj

On May 12, 1581 the prince Ştefan Bathory decided to set up at Cluj a college with three faculties: Theology, Philosophy and Law. This is the first official attestation of a higher education institution in our city.

After some climbings and descents and a contradictory evolution of the higher education in Cluj, on October 12, 1872 the emperor Ferenc József approves a decision of the Hungarian Parliament for setting up the University of Cluj. This Hungarian university have had four faculties: Law and State Sciences, Medicine, Philosophy, Letters and History and, the last one, Mathematics and Natural Sciences. During this period some pre-eminent mathematicians (such as Gyula Farkas, Lipót Fejér, Frigyes Riesz or Alfréd Haar) have had essential contributions to the development of the Cluj mathematical school.

The great wish of the Romanian nation to have their own university with complete studies in Romanian was finally accomplish after the union of the province of Transylvania with the Romanian principality in 1918. On October 1st 1919, by a decree of the King Ferdinand of Romania, the Romanian University was set up under the same name as before, the University of Cluj. The faculties of the new university were:

This paper was presented at the 8th Conference on History of Mathematics & Teaching of Mathematics, Cluj-Napoca, May 21-25, 2014.

Letters and Philosophy, Sciences, Medicine and Law. Professor Dim. Călugăreanu was appointed the first dean of the Faculty of Sciences, while the first Rector of the new university was elected Professor Sextil Puşcariu. The official opening of the new university took place on February 1-3, 1920 in the presence of King Ferdinand and Queen Mary. The University of Cluj starts its activity with 171 Professors and 2034 students enrolled in four faculties. Meanwhile, the Hungarian University moved to Szeged, Hungary.

The activities of the university were suddenly interrupted, since in August 1940, the nord-west part of Romania (including Cluj) was surrended to Hungary, by a decision took in Vienna under the pression of the German Third Reich. Very fast, the patrimony of the university was moved to Alba-Iulia and Turda. Moreover, the Faculty of Sciences was moved to Timişoara, while the rest of the faculties were moved to Sibiu. All of them start their activities in November 1940.

In 1944, after the defeat of the Third Reich and its allies, the Vienna Dictate was abolish and the nord-west part was re-integrated to Romania. It was also the time of a new start of the University of Cluj. Actually, starting with 1945 two universities will operate in Cluj: a Hungarian university, *Bolyai University* and a Romanian one, re-entitled *King Ferdinand*. In 1948, the Romanian university took its name after the great Romanian biologist *Victor Babeş* and, finally, in 1959, the two universities were unified in a Romanian-Hungarian university called as today *Babeş-Bolyai University*. The Faculty of Sciences was divided in four new faculties: Mathematics and Physics, Chemistry, Geology and Geography and, the last one, Natural Sciences. In 1962, the Faculty of Mathematics and Physics is separated in two faculties, the Faculty of Mathematics-Mechanics and the Faculty of Physics. From 1973 our faculty was re-named the Faculty of Mathematics and, finally, from 1994 it is called the Faculty of Mathematics and Computer Science.

Despite of the vicissitudes of the life, the university education and the research activities in Cluj were permanently on an ascendent slope. The Professors of the Cluj University, regardless of nationality, always worked, with abnegation and responsibility, on the development of the university and for of the perennial values promoted by it.

2. The Faculty of Sciences

In 1919, in the moment of its founding, the Romanian university of Cluj have had five sections:

- Mathematics
- Physics
- Chemistry
- Geography
- Natural Sciences

The Mathematics Section was also divided in several chairs:

- Analytical and Descriptive Geometry
- General Mathematics

- Mechanics
- Function Theory
- Mathematical Analysis
- Algebra
- Astronomy

Between the professors of the Mathematics Section of that years let us mention some important names: Dimitrie Pompeiu, Gheorghe Nichifor, Aurel Angelescu, Gheorghe Bratu, Nicolae Abramescu, George Iuga, Theodor Angheluţă, Petre Sergescu, Dumitru V. Ionescu, Gheorghe Călugăreanu. The Honorary Director of the Mathematical Seminar was Paul Montel, while the Director of the Astronomical Observatory was Gheorghe Bratu.

During the academic year 1938/39 the Mathematics Section of the Faculty of Sciences was performed by the following professors: *Nicolae Abramescu* (Geometry Chair), *Dimitrie Pompeiu* and *Theodor Angheluţă* (Algebra Chair), *Gheorghe Bratu* (Astronomy Chair), *Dumitru V. Ionescu* (Rational Mechanics Chair), *Petre Sergescu* (Differential and Integral Calculus), *Radu Bădescu* (Mechanics Chair), *Gheorghe Călugăreanu* (General Mathematics and Geometry Chair).

From the beginning and up today the mathematical studies in Romanian knew a continuous development not only because the increasing number of students, but mainly based on the performing research and scientific achievements of the mathematical school from Babeş-Bolyai University.

3. The research activity in Mathematics

During the years, the most important research actvities in mathematics were materialized in the following directions:

- Differential and Integral Equations (Gh. Bratu, Th. Angheluţă, P. Sergescu, D.V. Ionescu, Gh. Micula)
- Functional and Difference Equations (Th. Angheluţă, A. Angelescu, G. Iuga, T. Popoviciu, F. Radó)
- Function Theory and Topology (D. Pompeiu, Th. Angheluţă, N. Abramescu, P. Sergescu, G. Călugăreanu, T. Popoviciu)
- Mathematical Analysis and Optimization (T. Popoviciu, E. Popoviciu, I. Muntean, I. Maruşciac)
- Algebra and Number Theory (Th. Angheluţă, P. Sergescu, A. Angelescu, T. Popoviciu, Gh. Călugăreanu, Gh. Pic)
- Numerical Analysis and Approximation Theory (T. Popoviciu, D.V. Ionescu, Gh. Micula, E. Popoviciu, D.D. Stancu)
- Geometry (N. Abramescu, Tib. Mihăilescu, J. Gergely, Gh. Călugăreanu, M. Ţarină)
- Mechanics (C. Iacob, A. Angelescu, D.V. Ionescu, D. Pompeiu, P. Brădeanu)
- Astronomy and Astrophysics (Gh. Bratu, I. Armeanca, C-tin Pârvulescu, Gh. Chiş)
- Computer Science (T. Popoviciu, D.D. Stancu, E. Muntean)

- History and Philosophy of Mathematics (P. Sergescu, V. Marian, D.V. Ionescu, M. Ţarină, Gh. Micula)
- Didactics of Mathematics (D.V. Ionescu, T. Popoviciu, E. Popoviciu, I. Muntean, I. Maruşciac)

4. The scientific journals on Mathematics from Cluj

The scientific life of the mathematical section of the University of Cluj was roused by the publication of several scientific journals on Mathematics. Here are the most important ones.

1. Mathematica

The first volume of the journal Mathematica appears in 1929. " *MATHEMATICA est une nouvelle publication scientifique qui a pour bat d'établie des relations entre l'activité mathématique de la Roumanie et celle des autres pays...* it is written in the Preface of the first issue.

The Editorial board was composed by:

Directors: G. Ţiţeica and D. Pompeiu;

Editors: N. Abramescu (Cluj), A. Angelescu (Cluj), Th. Angheluţă (Cluj), G. Bratu (Cluj), A. Davidoglu (Bucureşti), D.V. Ionescu (Cluj), O. Onicescu (Bucureşti), C. Popovici (Iaşi), S. Sanielevici (Iaşi), S. Stoilow (Cernăuţi), V. Vâlcovici (Timişoara)

Secretary of the Board: Petre Sergescu (Cluj)

The first article published in this journbal belongs to Paul Montel, Professeur á la Faculté des Sciences de Paris.

2. Studia Universitatis Babeş-Bolyai, series Mathematica

The first volume of this journal appears in 1957 under the name *Bulletin of the Victor Babeş and Janos Bolyai Universities Cluj-Natural Sciences Series* and then, from 1958, under the name of *Studia Universitatis Babes-Bolyai, Series Mathematica.* Starting to 1996, the new *Studia Universitatis Babes-Bolyai, Series Informatica* is edited by the Department of Computer Science.

3. Revue d'Analyse Numérique et de Théorie de l'Approximation (ANTA)

The journal was founded, in 1972, by Tiberiu Popoviciu. The editors-in chief of the journal were successively Elena Popoviciu, Caius Iacob, Ion Păvăloiu and Dimitrie D. Stancu.

Today the journal is edited by the Tiberiu Popoviciu Institute on Numerical Analysis of the Romanian Academy, Cluj branch.

4. Didactica Mathematica

The first volume of this journal appears in 1985 and since 2013 it is an electronic journal. The editor-in-chief of the journal is Professor Dorel Duca.

5. Fixed Point Theory – An International Journal on Fixed Point Theory, Computation and Applications

This specialized journal appeared in 2000 under the coordination of Professor Ioan A. Rus and from 2007 it is the first mathematical journal from Cluj indexed Web of Science (ISI) by Thomson-Reuters Products.

5. The Romanian Professors of the University of Cluj

In the last part of this work, we will present (in alphabetical order) short biographical notes of the most important Romanian mathematicians of the University of Cluj.

Professor Nicolae Abramescu (1884-1947)

Nicolae Abramescu was born at Târgovişte on March 31, 1884.

University studies: Professor Nicolae Abramescu graduated the study program Mathematics from the Faculty of Science, University of Bucharest, where he was colleague with Traian Lalescu. In 1921 Nicolae Abramescu get his Ph.D. in Mathematics at the same university with a dissertation on the *Systematization of the orthogonal polynomials technique.*

Didactical activity: In November 1919 Professor Nicolae Abramescu was appointed as an Associate Professor at the University of Cluj, following the recommendation of Gheorghe Ţiţeica. Thus, Professor Nicolae Abramescu was a founder member of the Faculty of Science of our university. Here, together with Aurel Angelescu and Gheorghe Bratu constitute a strong and valuable kernel around Professor Dimitrie Pompeiu-Director of the Mathematical Seminar.

On October 1st 1926, Nicolae Abramescu is appointed full professor of Descriptive Geoemetry at the University of Cluj, position that he will keep until the end of his career.

Research activity: Nicolae Abramescu puts a lot of effort in organizing the Cluj Scientific Society, and the First Congress of Romanian Mathematicians. He was also a founder member of the journal *Mathematica* and member of the Romanian Academy.

Professor Abramescu passed away on February 11, 1947 at Cluj.

Professor Aurel Angelescu (1886-1938)

Aurel Angelescu was born at Ploieşti on April 15, 1886.

University studies: Professor Aurel Angelescu graduated his bachelor studies at Sorbonne, Paris. On April 7, 1916 he get (also at Sorbonne) the Ph.D. in Mathematics, under the guidance of Paul Appel. The title of his thesis was *Sur les polynômes généralisant les polynômes de Legendre et d'Hermite et sur le calcul approché des integrals multiples.*

Didactical activity: After his return to Romania, Aurel Angelescu is appointed, in 1919, professor at the Function Theory Chair of the University of Cluj. From now on, Professor Aurel Angelescu is fully dedicated to the intense work of organizing the mathematics education, being also one of the mentors of the new journal *Mathematica.*

Professor Aurel Angelescu was director of the Geometry and Mechanics Seminar and, for one year, between 1927 and 1928, he was appointed as Dean of the Faculty of Sciences at the University of Cluj. Starting to January 1st, 1930 Aurel Angelescu becomes full professor of Algebra and Number Theory at the University of Bucharest.

Research activity: Professor Aurel Angelescu main interest fields were generating functions for polynomials, linear differential equations, functional analysis, trigonometric series. He published more than 60 research works on the field of Algebra and Function Theory.

Very young, at the age of almost 52 de ani, Professor Aurel Angelescu tragically passed away on April 6, 1938.

Professor Theodor Angheluţă (1882-1964)

Professor Theodor Angheluţă was born on April 28, 1882 in the small village Adam from the (former) Tutova county.

University studies: After the primary and secondary studies at Bârlad, he becomes, between 1902-1905, a student of the Faculty of Sciences at the University of Bucharest getting the bachelor in Mathematics. From 1910, Professor Theodor Angheluţă is enrolled at Sorbonne, working mainly on the guidance of Emile Picard. On June 16, 1922 Professor Theodor Angheluţă defended his Ph.D. thesis *On a general class of trigonometric polynomials and the approximation of a continuous function.*

Didactical activity: In 1919, Professor Theodor Angheluţă is appointed associate professor at the Faculty of Sciences from the University of Bucharest, while five years later he get a full professor position on Algebra at the Faculty of Sciences from the University of Cluj.

Professor Theodor Angheluţă was the dean of the Faculty of Sciences from the University of Cluj between 1931 and 1932. He is retired starting to September 1st, 1947, but then, at the end of 1950, Professor Theodor Angheluţă is appointed again as full professor at the Faculty of Mathematics and Physics from Victor Babeş University of Cluj.

From October 1st, 1955 to September 1962 Professor Theodor Angheluţă was full professor at the Math Department of the Technical Institute of Cluj.

Research activity: Theodor Angheluţă has important contributions to the Function Theory, Differential and Integral Equations, Functional and Algebraic Equations. A special kind of functional equations carry even today his name: *Angheluţă type functional equations.*

On May 30, 1964 Professor Theodor Angheluţă passed away at Cluj.

Professor Ion Armeanca (1899-1954)

Ion Armeanca was born at Săcărâmb, Hunedoara county.

University studies: Professor Ion Armeanca took his secondary studies at Deva and then the university studies at the University of Cluj. His Ph.D. thesis (defended on July 26, 1933) entitled *Photographische und photovisuelle Helligkeiten von pohlnahen Sternen* was written under the guidance of Professor H. Kienle from the Astronomical Observatory in Göttingen. The thesis was published in one of the issue of the journal *Zeitschrift für Astrophysik*, a proof of its incontestable value.

Research activity: Starting to January 1st 1922, Ion Armeanca was secretary-librarian at the Astronomical Observatory. Then, from February 1928, he get a research position at the research center of the Astronomical Observatory in Cluj. Professor Ion Armeanca was an excellent specialist in photoelectronic photometry. Starting with 1939, Ion Armeanca made systematic observations at the Astronomical Observatory in Cluj, using the Guthnick photometer. From 1945, Ion Armeanca was the director of the Astronomical Observatory in Cluj and he made new research activities on the variable stars problem.

Professor Ion Armeanca was a member of the following societies: Gazeta Matematică, Astronomische Gesellschaft, Société Astronomique de France, National Committee for Astronomy.

Professor Gheorghe Bratu (1881-1941)

Gheorghe Bratu was born at Bucharest in 1881.

University studies: Gheorghe Bratu graduated the university studies at the University of Iaşi. His Ph.D. thesis *Sur l'equilibre des fils soumis á des forces intérieures*, written under the guidance of Professor Paul Appell from the Astronomical Observatory of Paris, was defended in 1914. Moreover, between 1909 and 1914, Professor Gheorghe Bratu get an Adamachi fellowship at the Astronomical Observatory of Paris.

Didactical activity: Gheorghe Bratu started his didactical activity in 1914, when he was appointed as assistent professor at the Astronomical Observatory of Iaşi. In 1918, Professor Gheorghe Bratu was appointed associate professor of Mathematical Analysis at the University of Iaşi and, then, from 1919 to the end of his life, Gheorghe Bratu was full professor of Astronomy at the Faculty of Sciences from the University of Cluj. Professor Gheorghe Bratu also wrote a very interesting course of Astronomy, published at Cluj.

Professor Gheorghe Bratu was the Dean of the Faculty of Sciences during the following academic years: 1923/1924, 1938/1939 and 1939/1940.

Research activity: Professor Gheorghe Bratu is the founder of the (modern times) Astronomical Observatory in Cluj. He was also the Director of the Observatory between 1920-1923 and 1928-1941.

Professor Gheorghe Bratu was a member of the following societies: Gazeta Matematică, Société Mathématiques de France, Societé Astronomique de France, Societa Astronomica Italiana, Circolo Matematica di Palermo, Romanian Scientific Academy, Astronomical National Committee. Professor Gheorghe Bratu also founded the Alliance Française, Cluj branch and he was decorate with Legion of Honour at rank of knight.

Professor Gheorghe Călugăreanu (1902-1976)

Gheorghe Călugăreanu was born at Iaşi, on July 16, 1902.

University studies: Gheorghe Călugăreanu started his studies at Bucharest at the Gheorghe Lazăr High School. Then, between 1921 and 1924, he was a student at the Faculty of Science from King Ferdinand University of Cluj, in the Mathematics-Physics study program.

In 1926, Gheorghe Călugăreanu leaves Cluj for Paris, with a fellowship of the Romanian goverment. He get the bachelor in Mathematics and, then, in 1928, the Ph.D. in Mathematics with a thesis entitled *Sur les fonctions polygènes d'une variable complexe*.

Didactical activity: The entire activity of Professor Gheorghe Călugăreanu is related to our university. Gheorghe Călugăreanu was assistent professor (1930-1934), associate professor (1934-1942) and then, from 1942, full professor at the University of Cluj. Later on, as a Dean of the Faculty of mathematics and Physics (1953-1957) or as a head of the Function Theory Chair, Professor Gheorghe Călugăreanu have had important contributions to the organization of the mathematics education in Cluj.

Research activity: The scientific activity of Professor Gheorghe Călugăreanu was focused on the study of the main problems of the theory of complex variable functions, geometry, algebra and topology. His first papers, including the Ph.D. thesis are valuable contributions to the theory of complex variable functions, in such a way that Gheorghe Călugăreanu can be seen as a prestigious follower of Dimitrie Pompeiu. Other remarkable results were obtained in domain of geometric theory of univalent functions. Professor Gheorghe Călugăreanu also has very important contributions in knots theory.

In 1963 Professor Gheorghe Călugăreanu becomes a member of the Romanian Academy. He passed away on November 15, 1976.

Professor Gheorghe Chiş (1913-1981)

Gheorghe Chiş was born in the village of Santău, Satu Mare county, on August 8, 1913.

University studies: Gheorghe Chiş attended the primary and secondary school in his village and then the high school in Carei, Satu Mare county. In 1935, Gheorghe Chiş get the bachelor in Mathematics-Physiscs at the University of Cluj. He get the Ph.D. also from the University of Cluj, in 1949.

Didactical activity: Gheorghe Chiş started his didactical career in 1936 at the Astronomical Observatory of the University of Cluj. Starting to 1960, Gheorghe Chiş is full professor of Astronomy and Astrophysics at the same University of Cluj. For six years, between 1962 and 1968, Professor Gheorghe Chiş was the dean of the Faculty of Mathematics-Mechanics from the University of Cluj. Moreover, from 1954 he was the Director of the Astronomical Observatory of the University of Cluj.

Research activity: Professor Gheorghe Chiş published more than 100 scientific papers and books. He was also the initiator of variable stars observations and of the permanent observation point of the artificial satellite with the code COASPAR 1132.

Professor Gheorghe Chiş was a member of the International Astronomical Union, of COSPAR and of the Romanian Astronomical Committee.

Professor Gheorghe Demetrescu (1885-1969)

Gheorghe Demetrescu was born in 1885 at Bucharest.

University studies: Gheorghe Demetrescu attended the courses of the University of Bucharest and get his Ph.D. also in Bucharest with a thesis *On a computation method to predict the Sun eclipses* on March 13, 1915.

Didactical activity: After a short stage at the Astronomical Observatory in Paris (1908-1912), Professor Gheorghe Demetrescu is appointed at the Astronomical Observatory in Bucharest. Starting to the academic year 1922/1923 and until 1927/1928 Professor Gheorghe Demetrescu get a new position at the Astronomical Observatory of the Faculty of Sciences in Cluj. Later on, Professor Gheorghe Demetrescu moved agaid to Bucharest, where he was the Director of the Astronomical Observatory.

Research activity: Professor Gheorghe Demetrescu published relevant scientific works on Astronomy and he was a member of the Romanian Academy.

Professor Caius Iacob (1912-1992)

Caius Iacob was born at Arad on March 29, 1912. His father, Lazăr Iacob, was a member of the Romanian mission to the Great Assembly from Alba Iulia on December, 1st, 1918.

University studies: Caius Iacob attended the primary and secondary school in Arad and Oradea. Then, he attended the university studies at the University of Bucharest, the Faculty of Mathematics (1928-1931). Caius Iacob get his Ph.D. in 1935, at the Faculty of Sciences of the University of Paris with a thesis entitled *Sur la determination des fonctions harmoniques conjuguees par certaines conditions aux limites. Applications a l'hydrodynamique*, under the guidance of Professor Henri Villat.

Didactical activity: Professor Caius Iacob starts his didactical career in 1935 as assistant professor at Technical Institute of Timişoara. From March 15, 1938 Caius Iacob is appointed as assistent professor at the Mathematical Section of the Faculty of Sciences from the University of Cluj. In 1939, he moves to Bucharest where he get a position at the Mechanics Laboratory of the University of Bucharest. In 1942, Caius Iacob returns to Cluj University as associate professor and, from December 30, 1943 (at the age of 31) Caius Iacob is appointed full professor of Mechanics. Later on, Professor Caius Iacob worked both in Cluj and in Bucharest University with a special emphasis on the courses of Fluid Mechanics and Aerodynamics.

Research activity: Professor Caius Iacob organized at Cluj University the research seminar on Fluid Mechanics. Caius Iacob was a laureate of the prize *Henri de Parville - for Mechanics* in 1940, awarded by the Science Academy in Paris. At the age of 43, on July 2, 1955, Professor Caius Iacob was elected correspondent member of the Romanian Academy, while on March 21, 1963 he becomes full member of the Romanian Academy. Professor Caius Iacob was the president of the Mathematics Section of the Romanian Academy and he was the founder of the Applied Mathematics Institute of the Romanian Academy which today carries his name. Professor Caius Iacob is the father of the Romanian School of Mechanics. He published more than 120 scientific works. His main book is *A Mathematical Introduction to Fluid Mechanics.*

Professor Caius Iacob passed away on February 6, 1992 at Bucharest.

Professor Dumitru V. Ionescu (1901-1985)

Dumitru V. Ionescu was born at Bucharest on May 14, 1901.

University studies: Dumitru V. Ionescu was, between 1919-1922, a student of the Mathematics study program at the Faculty of Sciences of the University of Bucharest, having as professors some very important Romanian mathematicians: Anton Davidoglu, David Emmanuel, Gh. Ţiţeica, Traian Lalescu, Dimitrie Pompeiu, ... Between 1923-1927 Dumitru V. Ionescu attended the courses of the famous École Normale Supérieure de Paris. Some of his professors were Emile Goursat, H. Lebesque, Paul Montel, Emile Picard. Between his colleagues: H. Cartan, J. Dieudonné, P. Dubreil and other great mathematicians of that time. The Ph.D. thesis, entitled *Sur une classe d'équations fonctionnelles* was defeated at Paris on June 7, 1927.

Didactical activity: Starting with the academic year 1927/1928, Dumitru V. Ionescu is appointed as associate professor at the University of Cluj and, then, from 1930 he is appointed professor at the Rational Mechanics Chair of the University of Cluj. After a short period at the Technical University of Cluj, from 1955 until 1971 (the year of his retirement), Dumitru V. Ionescu is full professor at the Differential Equations Chair.

Professor Dumitru V. Ionescu took many high scientical level courses, such as: Ordinary Differential Equations, Partial Differential Equations, Variational Calculus, Integral Equations, Numerical Analysis. Dumitru V. Ionescu published several courses, such as:

1. *Ecuaţii diferenţiale şi integrale*, Editura Didactică şi Pedagogică, Bucureşti 1965; 1972 .

2. (with C. Kalik) *Ecuaţii diferenţiale ordinare şi cu derivate parţiale*, Editura Didactică şi Pedagogică, Bucureşti, 1965.

3. (with Gh. Călugăreanu) *Curs de Analiză Matematică*, Universitatea din Cluj, 1956.

Dumitru V. Ionescu was the Dean of the Faculty of Sciences of the University of Cluj between 1941-1945) and head of the Chair of Differential Equations (between 1955-1971).

Research activity: The research topics of Professor Dumitru V. Ionescu were Differential Equations, Numerical Analysis, History of Mathematics, Didactical Mathematics. Professor Dumitru V. Ionescu published more than 200 scientific papers and the following monographs:

1. *Cuadraturi numerice*, Editura Tehnică, Bucureşti, 1957, (340 pp.)

2. *Diferenţe divizate*, Editura Academiei, Bucureşti, 1978, (303 pp.)

One of the most important achievement in his research activity was the so-called *the method of the function* ϕ.

Professor Dumitru V. Ionescu passed away on January 20, 1985 at Cluj-Napoca.

Professor George Iuga (1871-1958)

George Iuga was born on October 13, 1871 at Braşov.

Professor George Iuga was one of the first Romanian mathematicians who obtained the Ph.D. in Mathematics (1896) in France at Strasbourg.

He was a professor of the Faculty of Sciences of the University of Cluj between 1923 and 1938.

Professor Ioan Maruşciac (1925-1987)

Ioan Maruşciac was born at Crăciuneşti, Maramureş county on March 27, 1925.

University studies: After the primary school in his native village, Ioan Maruşciac attended the secondary school in the city of Sighetul Marmaţiei. Without a financial support from his family, he must work until 1947, as an ordinary worker at the Railways Company and then, at the Mayor House of Crăciuneşti. After his military stage, he is enrolled at the Ukrainean High School in Sighetul Marmaţiei and finally, in 1951 he finish the studies. Between 1951 and 1954, Ioan Maruşciac is a student of the Faculty of Mathematics and Physics of the Victor Babeş University in Cluj.

Didactical activity: Just after the faculty, Ioan Maruşciac get a didactical position in the university, at the Chair of Function Theory. From 1972, Ioan Maruşciac get a full professor position at the Babeş-Bolyai University Cluj-Napoca. During his activity in our university, Professor Ioan Maruşciac teached several courses, such as: Mathematical Analysis, Operational Research, Algorithm Theory, Mathematical Programming, Numerical Methods in Optimization.

Research activity: Professor Ioan Maruşciac has important contributions in Approximation Theory by Polynomials and Infrapolynomials, Optimization Theory. Professor Ioan Maruşciac published three monographs and more than 85 scientific papers. He was also Ph.D. supervisor in Operational Research.

Professor Ioan Maruşciac passed away in 1987.

Professor Gheorghe Micula (1943-2003)

Gheorghe Micula was born in the small village of Delureni, Bihor county on April 23, 1943.

University studies: After the high school attended at Vadu Crişului, Gheorghe Micula was a student (between 1960-1965) of the Mathematics specialization of the Faculty of Mathematics and Physics from Babeş-Bolyai University Cluj. He also get a Humboldt fellowship in Germany (1974-1976) and a Fulbright fellowship in USA in 1971. Gheorghe Micula wrote his Ph.D. thesis under the guidance of Professor Dumitru V. Ionescu and defended it in 1971 at the Faculty of Mathematics-Mechanics from Babeş-Bolyai University Cluj.

Didactical activity: The entire didactical activity of Professor Gheorghe Micula took place at the University of Cluj. He was full professor at the Differential Equations Chair since 1992. Professor Gheorghe Micula teached several courses as: Differential Equations, Spline Functions, Finite Elements Methods, etc. He also wrote several books on Differential Equations and Spline Functions.

Research activity: The main research interests of Professor Gheoreghe Micula were focused on differential equations, numerical analysis and spline functions. He published more than 90 scientific papers and two monographs on spline functions. Gheorghe Micula was also invited professor at several important universities from Germany, USA, China, South Korea, New Zeeland, Israel, Italy, Czech Republik, Switzerland, etc.

Professor Gheorghe Micula passed away at Cluj-Napoca on December 24, 2003.

Professor Emil Muntean (1933-2009)

Emil Muntean was born at Măgura, Hunedoara county on July 31, 1933.

University studies: Emil Muntean graduated the Faculty of Mathematics from the University of Cluj in 1957. Then, he get the Ph.D. in Mathematics in 1964 at the University of Saint Petersburg, Soviet Union.

Didactical and research activity: Emil Muntean worked for the construction of the first Romanian computers: MARICA (1959), DACICC-1 (1961) and DACICC-200 (1969). DACICC-200 was the most efficient Romanian computer of the second generation, being capable to do more than 200,000 arithmetical operation/second. In 1968 Emil Muntean becomes the Director of the Institute for Computing in Cluj.

Since 1990, Emil Muntean get a full professor position at the Faculty of Mathematics from Babeş-Bolyai University. From 2000 Emil Muntean was full professor of Computer Science at the Faculty of Economics from Dimitrie Cantemir University of Cluj-Napoca. He was a fruitful Ph.D. supervisor in the field of Computer Science. He is also the founder of the publication series "MicroInformatica".

Emil Muntean passed away at Cluj-Napoca on November 29, 2009.

Professor Ioan Muntean (1931-1996)

Ioan Muntean was born at the village of Sântimbru, Alba county, on May 27, 1931.

University studies: After the primary school in the village of Sântimbru, Ioan Munteanu moved at the Petroşani High School, Hunedoara county. Then, Ioan Muntean started the university studies at the Faculty of Mathematics-Physics of the Babeş University. After two years of studies in Cluj (1950-1952), Ioan Muntean moved to the Faculty of Mathematics-Mechanics of the Lomonosov University. Here he graduated the studies in 1955. In 1976, Ioan Muntean get his Ph.D. in Mathematics with a thesis entitled *Contributions to the qualitative study of the nonlinear oscilations*, under the scientific coordinations of academician Tiberiu Popoviciu.

Didactical activity: The didactical activity of Professor Ioan Muntean was entirely sustained at the Faculty of Mathematics from the University of Cluj, between 1976 and 1996. Professor Ioan Muntean teached several courses on Mathematical Analysis, Optimal Control, Functional Analysis, He was also very much involved in the Didactic of Mathematics having many presentations in the high schools in Transylvania and in the Mathematics Didactic Conference. Professor Ioan Muntean was Vice-Dean of the Faculty of Mathematics and Head of the Chair of Mathematical Analysis.

Research activity: The research activity of Professor Ioan Muntean was oriented to several topics as: Qualitative Theory of Differential Equations (he was initiated, during his stage in Moscow, by the Russian mathematicians Nemytski and Stepanov), Optimal Control, Approximation Theory, Functional Analysis, Real Analysis. Professor Ioan Muntean published more than 100 scientific papers. Since 1976, he was the leader of the Mathematical Analysis research group. Professor Ioan Muntean was a prolific supervisor in the field of Mathematics.

Professor Ioan Muntean passed away in August 1996 at Cluj-Napoca.

Professor Constantin Pârvulescu (1890-1945)

Constantin Pârvulescu was born at Ploieşti in 1890, on July 21.

University studies: After the university studies in Romania at Bucharest, Constantin Pârvulescu defended his Ph.D. entitled *Sur les amas globulaires d'étoiles et leurs relations dans l'espace* in 1925 at Sorbonne.

Didactical activity: Constantin Pârvulescu started his activity at the Astronomical Obesrvatory in Paris between 1921 and 1924. Then, he was professor of Astronomy and Rational Mechanics (1925-1940) at the Faculty od Sciences from the University of Cernăuţi. After a short period in Bucharest, Constantin Pârvulescu becomes, starting from 1941, full professor at the University of Cluj.

Research activity: Professor Constantin Pârvulescu was the Director of the Astronomical Observatory in Cluj (1941-1945) and founder of the Astronomical National Committee. Professor Constantin Pârvulescu was decorate with Legion of Honour at rank of knight and get, post-mortem, a honorary position in the Romanian Academy.

Professor Pârvulescu passed away on July 2, 1945.

Professor Dimitrie Pompeiu (1873-1954)

Dimitrie Pompeiu was born in the village of Broscăuţi, Dorohoi county on September 22, 1873.

University studies: After his primary and secondary schools in Dorohoi and Bucharest, Dimitrie Pompeiu went (1898) to Paris for the university studies. In 1905, Dimitrie Pompeiu get his Ph.D. with a thesis entitled *Sur la continuité des fonctions de variables complexes* under the guidance of Henri Poincaré. The motivation for such a study was an open problem concerning the singularities of uniform analytic functions, open problem posed by Painlevé in 1897.

Ludovic Zoritti wrote, also in 1905, a Ph.D. thesis in which he claimed to have proved that a uniform analytic function cannot be continuously extended on the set of its singularities. On the other hand, Pompeiu in his doctoral thesis proved the existence of certain analytic functions which could be extended continuously on their set of singularities even though this set had positive measure. Since both results could not be correct, the problem was "Where is the mistake ?" The mistery was resolved in 1909 when Denjoy confirmed that Pompeiu's results were correct, and he found the error in Zoritti's theorems. In 1907, in his paper *Sur les fonctions dérivées*, Dimitrie Pompeiu had clarified the whole situation by constructing simpler examples of functions, functions which are now called "Pompeiu functions". There was another important idea in Pompeiu's Ph.D. thesis, namely the distance between two sets, which he called the "écart" and "écart mutuel". Consequently, Dimitrie Pompeiu could be also considered as one of the founders of the theory of hyperspaces.

Didactical activity: In the autumn of 1905, Dimitrie Pompeiu comes back in Romania and get a position of associate professor on Mathematical Analysis and then of Mechanics at the University of Iaşi. In 1912, he moves to the University of Bucharest, as a successor of Spiru Haret. Starting to 1930, Dimitrie Pompeiu is appointed as a full professor of Function Theory, after the retirement of Professor David Emmanuel. Starting from the beginning of the academic year 1920, Dimitrie Pompeiu was appointed (for two academic years) full professor at the Faculty of

Sciences of the new Romanian University of Cluj and the Head of the Mathematical Seminar. Actually, Professor Dimitrie Pompeiu have had an important role in the organization, not only of the Mathematical Seminar (following the model lauched at College de France), but also of the whole mathematics education in the University of Cluj.

Research activity: Concerning the research activity of Professor Pompeiu, his main contributions were in the field of Complex Analysis. Academician Petru Mocanu described very well the contributions of Pompeiu to the field of Function Theory. "There is no doubt that Pompeiu's preferred area was Analysis, especially Complex Analysis, but he achieved remarkable results in other areas such as Mechanics. Pompeiu initiated the theory of polygenus functions as a natural extension of analytic functions. He introduced the notion of a special type of derivative, *the areolar derivative of a complex function*, extending the Cauchy formula which today is sometimes called the Cauchy-Pompeiu formula. In a short paper in 1929 entitled *Sur certains systemes d'équations linéires et sur une propriété intégrale des fonctions de plusieurs variables*, he proved that if the double integral of a continuous function takes the same value over any square of given side, then the function is constant. This simple remark has led to many interesting problems in Analysis and it is known today as the problem of Pompeiu." Let us also mention that more than 1000 papers cite this 1929 paper by Pompeiu. Among other topics on which Pompeiu published research articles we mention Interpolation Theory and Mechanics.

Dimitrie Pompeiu published around 150 research papers. In 1929, together with Petre Sergescu, Dimitrie Pompeiu founded the scientific journal *Mathematica (Cluj)*, one of the most influential journal of that period. In 1934 Dimitrie Pompeiu was elected member of the Romanian Academy.

Professor Dimitrie Pompeiu passed away on October 8, 1954 at Bucharest.

Professor Elena Popoviciu (1924-2009)

Elena Moldovan (married Popoviciu in 1974) was born on August 26, 1924 in Cluj-Napoca.

University studies: All her studies were attended in Cluj. Then, Elena Moldovan get her Ph.D. in Mathematics in 1965, with a thesis (supervised by academician Tiberiu Popoviciu) entitled *Set of interpolation functions and the concept of convex function.*

Didactical activity: The entire didactical activity of Professor Elena Popoviciu took place at the Faculty of Mathematics from Babeş-Bolyai University Cluj-Napoca. From 1969, Elena Popoviciu get a full professor position at the Mathematical Analysis Chair. She teached several courses such as: Mathematical Analysis, Abstract Algebra, Functional Analysis, Linear Programming, Distribution Theory, Approximation Theory, Operatorial Calculus.

Research activity: Elena Popoviciu started the research activity under the coordination of academician Grigore Călugăreanu, but then fascinated by the remarkable personality of Tiberiu Popoviciu, her research topic moved to convex function theory and interpolation function theory. Starting with 1974, Elena Popoviciu becomes Ph.D. supervisor and finally she have had 23 doctoral students. Elena Popoviciu founded, in

1960, the research *Seminar on Best Approximation and Mathematical Programming* and, in 1974, the *Interdisciplinary Research Lab.*

Elena Popoviciu was also very much involved in the editorial work of the following journals: *Revue Numérique et d'Analyse et de Théorie de l'Approximation* and *Annals of the Tiberiu Popoviciu Seminar of Functional Equations, Approximation and Convexity.*

Elena Popoviciu passed away on June 24, 2009 at Cluj-Napoca.

Professor Tiberiu Popoviciu (1906-1975)

Tiberiu Popoviciu was born on February 15, 1905 at Arad.

University studies: After the primary and secondary schools in Arad, Tiberiu Popoviciu attended (between 1924 and 1927) the courses of the Faculty of Sciences, the specialization Mathematics at the University of Bucharest.

His professors were some famous mathematicians of that time, such as: David Emmanuel, Gheorghe Ţiţeica, Dimitrie Pompeiu, Anton Davidoglu. In 1927, after a strong competition, Tiberiu Popoviciu is admitted at École Normale Superieure de Paris. Between 1927 and 1930, he attended simultaneous the mathematics courses from Sorbonne. During his stage at Paris, he attended the courses of great mathematicians such as: Emile Picard, Edouard Goursat, Jacques Hadamard, Elie Cartan, Paul Montel, Emile Vessiot, Gaston Julia, Jean Chazy. In 1928 Tiberiu Popoviciu get the bachelor in Mathematics and also starts the preparation of his Ph.D. thesis under the guidance of Paul Montel. On June 12, 1933 Tiberiu Popoviciu defended, with great success, the Ph.D. thesis *Sur quelques propriétes des fonctions d'une ou de deux variables réelles.*

Didactical activity: After his return to Romania, Tiberiu Popoviciu starts his university activities at Cluj, Cernăuţi and Iaşi. In 1948, Tiberiu Popoviciu comes back to Cluj and is appointed professor, first at the Chair of Algebra and then at the Chair of Mathematical Analysis.

Research activity: Professor Tiberiu Popoviciu has important contributions in Mathematical Analysis, Approximation Theory, Convexity, Numerical Analysis, Functional Equations, Algebra and Number Theory, etc. One of his most important scientific contribution was the concept of convex functions of higher order (as a generalization of the notion of convex function) given in his Ph.D. thesis and then published in Mathematica, 8(1934), pp. 1-85. Most of the results concerning the theory of convex functions of higher order are contained in his famous book *Les fonctions convexes*, Actualitées Scientifique et Industrielles, Paris, 1944.

Professor Tiberiu Popoviciu is the founder of the Cluj School on Numerical Analysis. Because of his efforts, in 1957 it was created the Computing Institute of Cluj. In this institute, in 1961 is produced one of the first Romanian computers DACICC-1 (Dispozitiv Automat de Calcul al Institutului de Calcul din Cluj). Then, in 1969, also in Cluj, it is realized DACICC-200 - one of the most performant Romanian computer of the Sixties.

Tiberiu Popoviciu was, since 1948, corresponding member and from 1963 full member of the Romanian Academy. He was also for more than 30 years the president of the Cluj branch of the Romanian Mathematical Society.

Some other achievements of Tiberiu Popoviciu were: the reactivation, in 1958, of the journal *Mathematica (Cluj)*, the founding, in 1972, of the journal *Revue d'Analyse Numerique et de Theorie de l'Approximation*, the opening, in 1967, of a research seminar: *The Itinerant Seminar on Functional Equations*, later transfomed in *The Itinerant Seminar on Functional Equations, Approximation and Convexity.*

Professor Tiberiu Popoviciu was a very active, creative and prolific mathematician until his unexpected death in 1975, on October 29, after just half year from the moment of the abolition of his Institute of Computing by the communist regime.

Professor Petre Sergescu (1893-1954)

Petre Sergescu was born at Turnu-Severin in December 17, 1893.

University studies: After the primary and secondary schools attended at Turnu-Severin, between 1912 and 1916, Petre Sergescu is enrolled at the Mathematics section of the University of Bucharest.

He attended, in the same period, the courses of Faculty of Philosophy and the Music Academy from Bucharest. In 1919 Petre Sergescu leaves Romania for doing studies at Paris. In 1923 Petre Seregescu get his Ph.D. in Mathematics with a thesis *Sur les noyaux symétrisables* at the University of Bucharest.

Didactical activity: Professor Petre Sergescu starts his didactical activity in 1924, as assistent professor in Bucharest. In 1926, he is appointed associate professor and then, in 1938, full professor at the Analitical Geometry Chair and respectively the Mathematical Analysis Chair of the Faculty of Sciences from the University of Cluj. He also was Rector of the Technical University of Bucharest.

Because of the communist regime, he is forced to leave Romania and from 1948, Professor Petre Sergescu and his wife Marya Kasterska lived in Paris. Working in Paris, Petre Sergescu was for many years the secretary of the International Academy of the History of Sciences and founder and general secretary of the International Union for the History of Sciences. Petre Sergescu was also Director of the journal *Archives Internationales d'Histoire des Sciences.*

Research activity: Professor Petre Sergescu was one of the founder of the journal *Mathematica (Cluj)* being also the secretary of the editorial staff until 1948. Professor Petre Sergescu was also the initiator of the first two Congresses of the Romanian Mathematicians (Cluj 1929 and Turnu-Severin 1932).

In 1940, when the North-West part of Transylvania was surrended to Hungary and the Faculty of Science moved to Timişoara, Professor Petre Sergescu was an active member of the Mathematical Seminar. Professor Sergescu published more than 160 scientific papers and took part to numerous international congresses and conferences. Professor Petre Sergescu was a corresponding member of the Romanian Academy.

Professor Petre Sergescu passed away at Paris on December 21, 1954.

Professor Dimitrie D. Stancu (1927-2014)

Dimitrei D. Stancu was born at the village of Călacea, Timiş county on February 11, 1927.

University studies: The life and the activity of the academician Dimitrie D. Stancu overlapped with the life of the Faculty of Mathematics from Babes-Bolyai University Cluj-Napoca, where he was admitted, in a pre-eminent way, in 1947.

Because of his remarkable results during the faculty, D.D. Stancu is appointed in 1951 assistant professor at the Mathematical Analysis Chair, conducted at that time by academician Tiberiu Popoviciu. In the same time, D.D. Stancu starts the work on his Ph.D. thesis and get the Ph.D. in Mathematics in 1956 with a thesis entitled *A study of the polynomial interpolation of several variables functions: applications to the derivative and the numerical integartion* under the guidance of Tiberiu Popoviciu. During the academic year 1961-1962 Dimitrie D. Stancu gets a fellowship from the Romanian Ministry of Education for a research stage in U.S.A. at the University of Wisconsin, research stage which will be very important for the future development of his career. After his return in Romania, he obtain in 1968, a full professor position at the Numerical and Statistical Calculus Chair from the Faculty of Mathematics.

Didactical activity: Professor Dimitrie D. Stancu teached high level courses on Mathematical Analysis, Numerical Analysis, Approximation Theory, Probability Theory, etc. Professor Dimitrie D. Stancu was Vice-Dean of the Faculty of Mathematics and, for many years, Head of the Numerical and Statistical Calculus Chair.

Research activity: Professor Dimitrie D. Stancu research activity was decisive influenced by his scientific cooperation with academician Tiberiu Popoviciu. His main research topics were: interpolation theory, derivative and numerical integration, orthogonal polynomials, spline functions, approximation of the functions by linear and positive operators, probabilistic and combinatoric methods in approximation theory. Professor Dimitrie D. Stancu dedicated part of his research work to Numerical Analysis in connection to Computer Science. Academician D.D. Stancu was the scientific coordinator of 41 Ph.D. students in the field of Numerical Analysis and Approximation Theory. Professor Dimitrie D. Stancu published more than 120 research papers with a strong nternational impact. More than 50 papers have the name of Dimitrie D. Stancu in their title and the concept of *Stancu operator* is nowadays a very well-known notion in the mathematics literature.

Professor Dimitrie D. Stancu was elected in 1999 honorary member of the Romanian Academy. He also was an active collaborator of the Tiberiu Popoviciu Institute on Computing of the Romanian Academy and editor-in-chief of the journal *Revue d'Analyse Numérique et de Théorie de l'Approximation.*

Professor Dimitrie D. Stancu passed away at Cluj-Napoca on April 17, 2014.

Professor Marian Ţarină (1932-1992)

Marian Ţarină was born at Turda on August 15, 1932.

University studies: Marian Ţarină graduated the high school *Regele Ferdinand*, now *Mihai Viteazul National Colleage* in Turda. Then, he was addimited at the Faculty of Matehmatics and Physics from the University of Cluj, getting a *Magna Cum Laude* Diploma in 1954.

Under the guidance of academician Gheorghe Vrânceanu, he get the Ph.D. In Mathematics in 1964 with a thesis entitled *Partial Projective Spaces with Maximal Group of Motion* at the University of Bucharest.

Didactical activity: The entire didactical activity of Professor Marian Ţarină was at the University of Cluj. He obtained a full professor position at the Geometry Chair in 1990. Professor Marian Ţarină teached several courses such as: Differential Geometry, Foundaments of Algebraic Topology, Symmetric Spaces, Lie Groups, History of Mathematics, etc.

Research activity: Professor Marian Ţarină published more than 50 research papers and presented almost 200 scientific communications. His research topics were: Noneuclidean Geometry, Motion Groups in Riemann Spaces, Recurent Spaces, G-structures on Differentiable Manifolds, Finsler Spaces.

Professor Marian Ţarină unexpected passed away on May 31, 1992 at Oradea.

Professor Gheorghe Ţiţeica (1873-1939)

Gheorghe Ţiţeica was born at Drobeta Turnu-Severin on October 4, 1873.

University studies: After the secondary school attended in Craiova, Gheorghe Ţiţeica choose, for the university studies, the University of Bucharest. He get the bachelor in Mathematics in 1895.

Then, Gheorghe Ţiţeica leaves the country and get another bachelor (on the first position) in Mathematics at École Normale Supérieure de Paris. Gheorghe Ţiţeica also get the Ph.D. in Mathematics at Sorbonne (under the scientific coordination of Professor Gaston Darboux), being the fifth Romanian mathematician with doctoral studies at Sorbonne (after Spiru Haret, David Emanuel, Constantin Gogu and Nicolae Coculescu).

Didactical activity: In 1900, after his return to Romania, Gheorghe Ţiţeica was appointed as a professor of Geometry at the University of Bucharest. Starting with 1913 Gheorghe Ţiţeica becomes a member of the Romanian Academy. He was also the Dean of the Faculty of Sciences at the University of Bucharest and Doctor Honoric Causa of the Warsaw University.

Research activity: The scientific work of Professor Gheorghe Ţiţeica counts more than 400 scientific works, most of then in the area of differential geometry. Professor Gheorghe Ţiţeica discovered a new category of surfaces and a new category of curves which now carry his name. He also studied R-networks in the n-dimensional space, defined through some Laplace type equations. He is today recognized as the founder of the Romanian School of Differential Geometry.

References

[1] Andonie, Şt. G., *Istoria matematicii în România*, vol. II, Editura Ştiinţifică, Bucureşti, 1966.

[2] Bogdan-Duică, Gh., *Raportul rectorului*, în "Anuarul Universităţii" (1927/1928), Cluj, 1929.

[3] Both, N., Iancu, C., *Momente din istoria matematicii clujene*, Universitatea "Babeş-Bolyai", Facultatea de Matematică şi Informatică, Cluj-Napoca, 1997.

[4] Călugăreanu, D., *Dare de seamă*, în "Anuarul Universităţii", Cluj, 1923.

[5] Drăganu, N., *Discurs inaugural*, în "Anuarul Universităţii Regele Ferdinand I Cluj pe anul şcolar 1931/32", Cluj, 1932.

[6] Ghibu, O., *Universitatea Daciei Superioare*, Bucureşti, 1929.

[7] Ghibu, O., *La a douăzecea aniversare a Universităţii Daciei Superioare*, Cluj, 1939.

[8] Bănescu, N., *Raportul rectorului*, în "Anuarul Universităţii", 1923/1924, Cluj, 1925.

[9] Haţieganu, E., *Raportul rectorului*, în "Anuarul Universităţii", 1928/1929, Cluj, 1929.

[10] Haţieganu, I., *Universitate şi Naţiune*, în "Anuarul Universităţii Regele Ferdinand I Cluj-Sibiu".

[11] Neagoe, S., *Viaţa universitară clujeană interbelică*, vol. I, Editura Dacia, Cluj-Napoca, 1980.

[12] Neagoe, S., *Viaţa universitară clujeană interbelică*, vol. II, Editura Dacia, Cluj-Napoca, 1980.

[13] Pascu, Şt., *Universitatea "Babeş-Bolyai" din Cluj*, Editura Dacia, Cluj, 1972.

[14] Popescu-Spineni, M., *Instituţii de înaltă cultură. Învăţământul superior al Ardealului* 1581-1930, Vălenii de Munte, 1932.

[15] Puşcariu, S., *Raportul Rectorului*, în "Anuarul Universităţii din Cluj, Anul I, 1919-1920", Cluj, 1921.

[16] Racoviţă, E., *Raportul rectorului*, în "Anuarul Universităţii", 1929/1930, Cluj, 1930.

[17] Sergescu, P., *Le développement des sciences mathématiques en Roumanie*, Extrait de "La vie scientifique en Roumanie: I. Sciences Pures", Bucharest, 1937.

[18] Sergescu, P., *Un episod din bătălia pentru triumful calculului diferenţial: polemica Rolle-Saurin* 1702-1705, Monitorul Oficial şi Imprimeria Statului, Imprimeria Naţională, Bucureşti, 1942.

[19] Sergescu, P., *Gândirea matematică*, Cluj, 1928.

[20] Sergescu, P., *Matematica la români*, Conferinţă ţinută la "Liga culturală" din Braşov, Aşezământul tipografic "Datina Românească", Vălenii-de-Munte, 1934.

[21] Ştefănescu-Goangă, Fl., *Dare de seamă*, în "Anuarul Universităţii" (1932/1933), Cluj, 1933.

[22] Ştefănescu-Goangă, Fl., *Discurs inaugural*, în "Anuarul Universităţii Regele Ferdinand I Cluj" (1933/1934), Cluj, 1934.

[23] Ştefănescu-Goangă, Fl., *Discurs inaugural şi dare de seamă*, în "Anuarul Universităţii Regele Ferdinand I Cluj" (1934/1935), Cluj, 1935.

[24] Ştefănescu-Goangă, Fl., *Dare de seamă*, în "Anuarul Universităţii Regele Ferdinand I Cluj" (1936/1937), Cluj, 1936.

[25] Ştefănescu-Goangă, Fl., *Dare de seamă*, în "Anuarul Universităţii Regele Ferdinand I Cluj" (1937/1938), Cluj, 1939.

[26] Toth, Al., *Primele manuscrise matematice româneşti din Transilvania*, Editura Dacia, Cluj, 1974.

[27] *** *Procesele verbale ale Consiliului profesoral al Facultăţii de Ştiinţe de la Universitatea din Cluj.*

[28] *** *Acte solemne*, în "Anuarul Universităţii" (1934/1935), Cluj, 1935.

[29] *** Anuarul Universităţii" (1938/1939), Cluj, 1940.

[30] *** A.U.C., registrul cu procesele verbale ale şedinţelor Senatului universitar (1939/1940).

[31] *** A.U.C., registrul cu procesele verbale ale şedinţelor Senatului universitar (1943/1944).

[32] *** *The MacTutor History of Mathematics archive*, http://www-history.mcs.st-and.ac.uk/.

Dorel I. Duca
Babeş-Bolyai University of Cluj-Napoca
Department of Mathematics
Kogălniceanu Street, No.1, 400084 Cluj-Napoca, Romania
e-mail: `dduca@math.ubbcluj.ro`

Adrian Petruşel
Babeş-Bolyai University of Cluj-Napoca
Department of Mathematics
Kogălniceanu Street, No.1, 400084 Cluj-Napoca, Romania
e-mail: `petrusel@math.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 419–433

Remarkable Hungarian mathematicians at the Cluj University

Ferenc Szenkovits

Abstract. We provide a brief overview of the life and activity of the most remarkable Hungarian mathematicians who worked at the University of Cluj, from the beginnings to the present day.

Mathematics Subject Classification (2010): 01A55, 01A60, 01A73.

Keywords: University of Cluj, Hungarian mathematicians.

1. Introduction

The first higher education institution in Cluj (Kolozsvár, Claudiopolis), a Jesuit colleage with three faculties: Theology, Philosophy and Law, was set up on May 12, 1581 by Stephen Báthory, the prince of Transylvania and king of Poland.

Over the centuries astronomy and mathematics had an important role between the subjects taught at this catholic school. The most remarkable professors of astronomy and mathematics of this school were Miklós Jánosi (1700–1741) and Maximilian Hell (1720–1792). Jánosi and Hell published the first mathematical textbooks in Cluj:

Miklós Jánosi: *Trigonometria plana et sphaerica cum selectis ex geometria et astronomia problematibus, sinuum canonibus et propositionibus ex Euclide magis necessariis*. Claudiopoli, 1737.

Maximilian Hell: *Compendia varia praxesque omnium operationum arithmeticarum*. Claudiopoli, 1755.

Elementa mathematicae naturalis philosophiae ancillantia ad praefixam in scholis normam concinnata. Pars I., Elementa arithmeticae numericae et litteralis seu algebrae. Claudiopoli, 1755.

Exercitationum mathematicarum Partes Tres. Claudiopoli. 1760.

A new era begins in the Cluj university education on October 12, 1872, when the emperor Franz Joseph I of Austria approves a decision of the Hungarian Parliament for setting up the University of Cluj. This Hungarian university was between the

This paper was presented at the 8th Conference on History of Mathematics & Teaching of Mathematics, Cluj-Napoca, May 21-25, 2014.

first universities which have had in structure a separate Faculty of Mathematics and Natural Sciences.

At this university, moved to Szeged (Hungary) in 1919, worked world renowned mathematicians who founded a remarkable mathematical school. In the next section we present the most outstanding Hungarian mathematicians who contributed to the development of this important mathematical centre.

2. Hungarian professors of the University of Cluj

Professor Samu Borbély (1907–1884)

Samu Borbély was born in Torda (today Turda in Cluj country, Ramania) on April 23, 1907.

University studies: After his high school studies in Torda, Kolozsvár and Kecskemét, Samu Borbély started the university studies in mechanical engineering and mathematics in Budapest and continued in Berlin where he had personal contacts with *Albert Einstein*. He graduated as engineer-mathematician the Technical University of Berlin in 1933, and get his Ph.D. in 1938, at the same university under the supervision of professor *Rudolf Rothe*.

Didactical activity: In 1941 Borbély moves from Berlin to Cluj, where he is assistant professor and associate professor at the Ferenc József University. In 1944 he is taken to Berlin, where refusing to collaborate in the development of the V2 rackets he is incarcerated. He escapes in December 1944 and is hiding in Budapest until the end of the second World War, when he came back to Cluj, as professor of the Hungarian Bolyai University. Until 1949, when he moved back to Hungary, he contributed essential to the scientific developpement of the new university and to the rise of new generations of mathematicians at the Bolyai University.

After 1949 Samu Borbély activates at the University of Miskolc as head of the department of mathematics. Starting with 1955 he is the head of the department of mathematics at the Technical University in Budapest. In 1960 he is moving to Magdeburg, where he leads the department of mathematics of the Otto von Guericke University until 1964, when he came back to his position as head of the department of mathematics at the Technical University in Budapest. He is retired in 31 december, 1977.

Research activity: Between 1933 and 1941 Borbély, as associate researcher at the German Institute of Aeronautics in Berlin, studied problems of aerodinamics and technical mathematics.

His researches are oriented toward different problems of the applied mathematics. Main results concern ballistic problems, aerodynamics and nonlinear heat transfer.

In 1946 professor Samu Borbély becomes a corresponding member and in 1979 full member of the Hungarian Academy of Sciences.

Samu Borbély passed away in Budapest (Hungary) on August 14, 1984 and is buried in Târgu Mureş (Romania).

Professor Sámuel Brassai (1800–1897)

The information about the date and place where Sámuel Brassai was born is uncertain. We can find: June 15, 1800 in Torockószentgyörgy, Hungary (now Colţeşti, Alba country, Romania), or February 13, 1800, Torockó (Râmetea, Alba, Romania).

University studies: Brassai has never attended any university, he was an autodidact. Initiated by his father, a Unitarian minister in the mysteries of knowledge, he studied from the age of 12 at the Unitarian High School in Cluj. After the high school he completed his knowledge by traveling through Transylvania and Hungary, and as tutor in several families.

Didactical activity: After a long teaching experience at the Unitarian High School of Cluj and at Budapest, Brassai is appointed in 1872 professor of the elementary mathematics at the new founded University in Cluj (Magyar Királyi Tudományegyetem). Sámuel Brassai gives various courses of the elementary mathematics, focussing especially on the future teachers training. He also deliver courses on general linguistics and Sanskrit language. In academic year 1879-80 Brassai holds the position of rector of the university.

Research activity: Brassai was not a mathematician in the ordinary sense of the word today. He was a polymath, often remembered as the last Transylvanian polymath, attracted by mathematics, among many other sciences (history, geography, astronomy, linguistics, statistics, economy, theory of music, ...).

He published many scientific papers and articles to promote sciences, in various journals and newspapers of the time. In the area of mathematics he published textbooks for school education, articles about didactics of mathematics, and his most significant result has been the first translation in Hungarian of Euclid's Elements.

In recognition of his scientific merits, Brassai was received in the Hungarian Academy of Sciences a corresponding member of the department of mathematics and natural sciences in 1837, a full member of the department of history and philosophy in 1865, and honorary member in 1887.

Sámuel Brassai passed away in Kolozsvár on June 24, 1897.

Professor Vilmos Cseke (1915–1983)

Vilmos Cseke was born in Hátszeg, Hungary (now Haţeg, Hunedoara country, Romania) on May 15, 1915.

University studies: Vilmos Cseke graduated mathematics at the Cluj University in 1936 and received Ph.D. at the same university in 1947 advised by professor Teofil Vescan.

Didactical activity: After four years spent in the Catholic High School in Cluj, Vilmos Cseke is appointed assistant professor at the Cluj University in 1941, where he activates as associate professor (from 1948) and professor until his retirement. Professor Cseke contributed (1957–1979) to the development of the mathematical teaching as member and leader of the editorial board of the mathematical and physical magazines: *Matematikai és Fizikai Lapok*, and *Matematikai Lapok*.

Research activity: His research is focused on problems of the theory of probability, mathematical logics, mathematical statistics and applications of the mathematics in economy.

Professor Vilmos Cseke passed away in Kolozsvár on March 10, 1983.

Professor Lajos Dávid (1881–1962)

Lajos Dávid was born in Kolozsvár (Cluj) on May 28, 1881.

University studies: Lajos Dávid studied in Cluj, Göttingen and Paris. He attended courses taught by Gyula Farkas, Lajos Schlesinger and Frigyes Riesz in Cluj, David Hilbert and Felix Klein in Göttingen. He obtained the Ph.D. in mathematics at the Cluj University, advised by Lajos Schlesinger (1903).

Didactical activity: Gyula Dávid was a tutor at the Cluj University (1910) and at the Budapest University (1916), privatdozent in Budapest (1919) and professor from 1925 at the University of Debrecen. He leads a chair of mathematics in Cluj (1940–1944).

Research activity: Gyula David has published results concerning problems of algebra, theory of functions and history of mathematics. His research concerning the life and the activity of the two Bolyai is materialized in two books: *A két Bolyai élete es munkássága* (Budapest, 1923), *Bolyai-geometria az Appendix alapján* (Kolozsvár, 1944).

Professor Lajos Dávid passed away in Budapest on January 9, 1962.

Professor Gyula Farkas (1847–1930)

Gyula Farkas was born in Sárosd, Fejér Country, Hungary on March 28, 1847.

University studies: Gyula Farkas graduated the high school at the Benedictines gymnasium of the famous Pannonhalma abbey, founded in 969 by Prince Géza. After completing his schooling by the Benedictines, Farkas went to the Pest University with the intention of studying law and music. Soon he changed the direction of his studies and graduated chemistry in 1870. Later he continued his studies in chemistry and physics, obtaining the doctoral degree in 1876.

Didactical activity: Farkas worked as a private tutor for a while before returning to university to study physics and chemistry. He then returned to his native county of Fejér, teaching at the Modern School in the county town of Székesfehérvár. In 1874 he went to work for Géza Batthyány, the Count of Polgárdi, teaching his children mathematics and physics. Farkas now had time to undertake research both in mathematics and physics. Farkas was also given the opportunity to make visits abroad to broaden his background in mathematics and physics.

By 1880 Farkas had an impressive publication record in Comptes Rendus and was appointed as a dozent in function theory at the university in Pest. His career continued to flourish and on January 1887 he was appointed as an extraordinary professor at the University of Kolozsvár (Cluj), and in the following year he became an ordinary professor of theoretical physics there. Not only did Farkas serve the University of Kolozsvár as a professor, but he also served as Dean and as Rector of the University. In 1915 he resigned his position at the University since his eyesight was deteriorating. He retired to Budapest where he lived in retirement for 15 years.

Research activity: Gyula Farkas is known in mathematics for Farkas Theorem (or lemma) which is used in linear programming and also for his work on linear inequalities. In 1881 Gyula Farkas published a paper on Farkas Bolyai's iterative solution to the trinomial equation, making a careful study of the convergence of the algorithm. In a paper published three years later, Farkas examined the convergence of more general iterative methods. He also made major contributions to applied mathematics and physics, particularly in the areas of mechanical equilibrium, thermodynamics, and electrodynamics.

The Hungarian Academy of Science elected him corresponding member in 1898 and full member in 1914. For his contribution on developing Italian–Hungarian scientific collaborations he was elected member of the Circolo Matematico di Palermo and awarded with the title of Doctor Honoris Causa of the Padova University (1892).

Professor Gyula Farkas passed away in Pestszentlőrinc (today part of Budapest), Hungary on December 27, 1930.

The Hungarian professors of the Departmant of Mathematics and Informatics of the Babeş-Bolyai University of Cluj has named their association, founded in 2001, the *Gyula Farkas Association for Mathematics and Informatics.*

Professor Lipót Fejér (1880–1959)

Lipót (Leopold) Fejér was born in Pécs, Hungary on February 9, 1880.

University studies: Fejér graduated from the high school in Pécs in 1897. In the same year he won second prize in the national Eötvös Mathematics Competition and entered the Polytechnic University of Budapest. Here he studied mathematics and physics until 1902, except for the year 1899–1900 which he spent at the University of Berlin, where he attended courses by Georg Frobenius, Lazarus Fuchs and Hermann Schwarz.

Fejér presented his doctoral thesis focusing on his fundamental summation theorem for Fourier series to the Eötvös Loránd University in Budapest in 1902. He spent the winter of 1902-3 on a visit to Göttingen, attending lectures by David Hilbert and Hermann Minkowski, and the summer of 1903 in Paris where he attended lectures by Émile Picard and Jacques Hadamard.

Didactical activity: Lipót Fejér started his universitary career at the University of Budapest (1903-1905) and continued as privatdozent in Kolozsvár (Cluj) from 1905 to 1911.

In the years he spent in Cluj he offered different courses, like: Differential and integral calculus, Differential equations, Partial differential equations, Theory of functions, Exercises for beginners, New results on integer transcendental functions.

In 1911 Fejér was appointed to the chair of mathematics at the University of Budapest and he held that post until his death. During his period in the chair at Budapest, Fejér led a highly successful Hungarian school of analysis. According to the Mathematics Genealogy Project current on-line database, Leopold Fejér has 20 students and 6336 descendants. Among his PhD students we can find proeminent mathematiciens as Paul Erdős, George Polya and John von Neumann.

Research activity: Discussions with Hermann Schwarz in Berlin led Fejér to look at the convergence of Fourier series and prove the highly significant "Fejér's theorem": *The Fourier series is summable* $(C, 1)$ *to the value of the function at each point of continuity*, result submitted to be published to the Paris Academy of Sciences on 10 December 1900, in a paper titled *Sur les fonctions bornées et intégrables.* During the years spent in Cluj professor Fejér produced many high quality beautifully written papers: six in 1906, three in 1907, five in 1908, four in 1909 and six in 1910. After mowing to Budapest Fejér continued tu publish important works such as *Über die Konvergenz der Potenzreihe an der Konvergenzgrenze in Fallen der konformen Abbildung auf die schlichte Ebene* (1914), *Über Interpolation* (1916), and *Interpolation und konforme Abbildung*, (1918).

Fejér's main work was in harmonic analysis. He worked on power series and on potential theory. Much of his work is on Fourier series and their singularities but he also contributed to approximation theory. He collaborated to produce important papers, one with Carathéodory on entire functions in 1907 and another major work with Riesz in 1922 on conformal mapping.

Lipót Fejér was honoured with election to the Hungarian Academy of Sciences in 1911 and being a vice-president of the International Congress of Mathematicians held in Cambridge, England, in August 1912. He was elected to the Göttingen Academy of Sciences (1925), the Bavarian Academy of Sciences (1954), and the Polish Academy of Sciences (1957). He was elected an honorary member of the Calcutta Mathematical Society (1930), and awarded an honorary doctorate by Brown University in Providence, USA (1933) and by Eötvös Loránd University of Budapest (1950). He also served as an editor of the *Rendiconti del Circolo Matematico di Palermo* and of the *Mathematische Zeitschrift*. In addition, he received the highest state awards: the *Kossuth Prize*, first grade (1948), the *People's Order of Merit* (1950), and the *Labour Red Flag of Merit* (1953).

Professor Lipót Fejér passed away in Budapest, Hungary on October 15, 1959.

Professor Jenő Gergely (1896–1974)

Jenő (Eugen) Gergely was born in Kolozsvár, Hungary on March 4, 1896.

University studies: He attended all studies, from primary to university level in hometown. He graduated from the Faculty of Science of the University (Hungarian at the time) in Cluj. Jenő Gergely presented his doctoral thesis entitled: *Variations of double integrals*, supervised by of Alfred Haar, in 1921 at the University of Szeged, Hungary.

Didactical activity: Jenő Gergely served from 1918 to 1948 as professor of mathematics at Marianum, the Catholic school for girls, in Cluj, publishing a handbook of Algebra in 1937. Later, in the last 15 years of his activity, he followed all the steps of the academic career at universities in Cluj, first Bolyai University (1948–1959), then, after the unification of the two universities, the Babeş-Bolyai University (1959–1966). He published two university courses, one of Ordinary Differential Equations (1951), and another of Differential Geometry, with Árpád Kiss in 1957.

Research activity: Jenő Gergely worked in several areas of geometry and topology. He studied the classification of areas based on their intrinsic geometry, the geometry of Bolyai-Lobacevski, the polar theory of ovals and ovaloids based on their intrinsic equations and problems related on practical applications of geometry. In the last period of activity, he was interested in n-dimensional manifolds in separable Hilbert spaces and their applications in particle physics.

Professor Jenő Gergely passed away in Kolozvár (Cluj), Romania on May 15, 1974.

Professor Alfréd Haar (1885–1933)

Alfréd Haar was born in Budapest, Hungary on October 11, 1885.

University studies: Alfréd Haar attended the *Fasori Evangélikus Gimnázium* in Budapest. When he was a high school student has worked for the magazine for students *Középiskolai Matematikai Lapok* and won *Eötvös Loránd* national mathematics contest. Haar has started his universitary studies at the department of chemical engineering at the Technical University of Budapest, but in the same year moved to the University of Budapest, and after one year he becomes a student at the University of Göttingen. His professors were Carathéodory, David Hilbert, Felix Klein, Ernest Zermelo. Alfréd Haar presented his PhD thesis, written under the direction of David Hilbert, in June 1909.

Didactical activity: At age 24 Alfred Haar was appointed as a privatdozent at the University of Göttingen. He then moved to Zürich, where he taught mathematics at the famous technical university. In 1912 he was invited at the Franz Joseph University in Cluj, in the place of Fejér, called at the University of Budapest. He remains at the Hungarian University of Cluj until 1919, when he went to the University of Szeged. Haar, together with Riesz, rapidly made in Szeged a major mathematical centre from the new university. With support from the *Society of Friends of the Franz Joseph University*, they had founded the famous journal *Acta Scientiarum Mathematicarum* in 1930.

Research activity: Most of Haar's work was in analysis. The main results of his thesis, entitled *Zur Theorie der orthogonalen Funktionensysteme* appeared in a paper which he published in *Mathematische Annalen* in 1910. After the work of his thesis, he went on to study partial differential equations with applications to elasticity theory. He also wrote on Chebyshev approximations of functions, linear inequalities, analytic functions, and discrete groups. Between 1917 and 1919 he worked on the variational calculus, proving *Haar's Lemma*, and applying his results to problems like *Plateau's problem.* Haar introduced a system of orthogonal functions, a measure in mathematical analysis, with special properties, which today bears his name.

He was honoured in 1931 by election to the Hungarian Academy of Sciences.

Professor Alfréd Haar passed away in Szeged, Hungary on March 16, 1933.

Professor Lipót Klug (1855–1945)

Leopold (in German) or Lipót Klug (the Hungarian version) was born in Gyöngyös, Hungary on January 23, 1854.

University studies: Lipót Klug attended the gymnasium in his hometown and entered the University of Budapest in 1872. He graduated from the University on July 1874 with a teaching diploma. Later he undertook research, first for his diploma which was awarded in 1882, then for his habilitation (1897).

Didactical activity: After he graduated from the University, he was appointed on 25 September, 1874 as a high school teacher in the Science High School in Pozsony (Bratislava). He taught there between 1874 and 1893, writing his first books on geometry, after which he taught at a secondary school in Budapest. He also taught as a privatdocent in Synthetic Geometry at the University of Budapest from 1891. During these years, he was greatly influenced by Gyula Kőnig who taught at the Technical University of Budapest from the early 1870s. In 1897 Klug was appointed to the Franz Joseph University of Kolozsvár (now Cluj) as an extraordinary professor in Descriptive Geometry. After two years he was appointed to the chair of Descriptive Geometry at the University of Kolozsvár, a position he held for nearly twenty years, until 1917, when he retired and moved back to Budapest.

Research activity: Research areas: descriptive and synthetic projective geometry. In this topic has been one of the most influential and prolific Hungarian mathematician.

Professor Lipót Klug passed away in Budapest, Hungary on March 24, 1945.

Professor Lajos Martin (1827–1897)

Lajos Martin was born in Buda, Hungary on August 30, 1827.

University studies: His school education took place in Buda where he attended the Roman Catholic Secondary School. He then began studies at the university in Pest, taking courses in the Faculty of Arts in his first two years of study. However he then turned towards engineering taking courses at the university's Institutum Geometrico-Hydrotechnicum. The revolutions that swept Europe in 1848 disrupted his studies. Due to his active participation to the revolution, at the end of this, he was imprisoned and later enrolled to the army of the Austrian Empire, where he continued his technical studies, graduating in 1854 the military engineering academy in Vienna.

Didactical activity: Starting with 1855 he was teaching in the artillery school. Leaving the army, he started to work in 1860 as engineer, later being the chief Engineer of Buda town (1861). Then he taught first in a school in Selmecbnya (now Banská Štiavnica, Slovakia), and from 1862 at a school in Bratislava. He also worked as the director of a telegraph office at one stage in his career during the 1860s. During these years he wrote some textbooks for secondary schools. In 1872 he was appointed as Professor of Mathematics in the Department of Advanced Mathematics at the new University of Kolozsvár. Here, Martin took also on the reorganization and the management of the Observatory of the University, which had been founded by Jesuits in 1755 but had been very neglected. He served this university until 1897, being rector in 1895/96.

Research activity: In army Martin started to undertake research in ballistics, an later continuated this, both making theoretical calculations and carrying out practical experiments. He became interested in hydraulics, undertaking research on ships propellers, and giving an early formulation of the principle of the steam turbine. Later Matrin published his studies of the theory of the best propeller and windmill. Lloyd's of London, the shipping insurance firm, were interested in Martin's screw propellers which they tested but Martin refused to sell patents for his ideas.

He was honoured with election to the Hungarian Academy of Sciences as a corresponding member in 1859, becoming a full member in 1861.

Professor Lajos Martin passed away in Kolozsvár, Hungary (now Cluj, Romania) on March 4, 1897.

Professor Árpád Pál (1929–2006)

Árpád Pál was born in Hodgya (Hoghia, Harghita country, Romania) on June 25, 1929.

University studies: After finishing elementary school in his native village, he continued the high school studies in Gyergyószentmiklós (Gheorgheni) and graduated in 1949 from Udvarhely Mixed High School (now Tamási Áron High School, Odorhei). In the same year gave admission to Bolyai University, Faculty of Mathematics and Physics in Cluj. He graduates as chief of promotion in 1952, and was sent to study in Moscow at the VM Lomonosov Astronomical Institute, Stenberg Department of Celestial Mechanics. Here he completed post-graduate studies with a thesis entitled *Analytical theory of interpolation of small planets (55) Astrea*, advised by professors Duboshin and Moiseev (1957).

Didactical activity: His entire academic career, starting from 1957, is linked to the University of Cluj, where he attended all functions of hierarchy even after his retirement (1995), when he became a consultant professor and continued the work with his Ph.D. students until his death.

He was particularly well regarded as teacher, as researcher and as manager: he was Dean of the Faculty of Mathematics (3 legislations), scientific secretary of the Senate, Vice-Rector of the University. A remarcable result was the construction of the modern buildings of the Astronomical Observatory (1982). As a doctoral supervisor gave the country 23 PhDs in astronomy (celestial mechanics) some of which itself became doctoral supervisors.

Research activity: He wrote (alone or jointly) courses and textbooks and more than 150 scientific papers in different journals. He founded as editor in chief the Romanian Astronomical Journal. He presented his scientific results at dozens of national and international scientific conferences. His central theme of research was celestial mechanis.

He was a member of the International Astronomical Union, the European Astronomical Union, the Romanian National Council of Astronomy, the Romanian Mathematical Society, the Academy of Science of America. As president and later honorary chairman of the Romanian National Astronomical Committee, represented Romania at several general assemblies and promoted most valuable Romanian astronomers to become members of the International Astronomical Union.

Professor Árpád Pál passed away in Kolozsvár, on July 21, 2006. In recognition for his outstanding scientific results, the international astronomical community honored him in 2012 by naming an asteroid (Arpadpal, 257,005) in his memory.

Professor Ferenc Radó (1921–1990)

Ferenc (Francisc) Radó was born in Timişoara on May 21, 1921.

University studies: After undergraduate studies in his hometown, he studied at the Engineering School in Bucharest and at the University of Cluj, where he graduated mathematics at the end of the second Word War, in 1945. He got his PhD at the same university in 1959, under the supervision of professor Tiberiu Popoviciu.

Didactical activity: He worked as a teacher in Timisoara, then was appointed assistant professor in 1950 at the Bolyai University in Cluj. After the unification of the Bolyai and Babeş universities, he was associate professor and later full professor at the Babeş-Bolyai University until his retirement in 1985. For many years he worked as a researcher at the Institute of Computing of the Academy. In the academic year 1969–70 was a visiting professor at the University of Waterloo (Canada).

Research activity: He has published articles in the country and abroad in the fields of: functional equations, nomograms transformation, algebraic and geometric structures, about the foundations of algebraic geometry, isometries in metric spaces, convex sets, geometries over rings (Barbilian type structures) etc. In 1981 he published the monograph in Hungarian: "A geometria mai szemmel" (Geometry seen today - with Béla Orbán).

Professor Ferenc Radó passed away in Kolozsvár, on November 27, 1990. His name was given to the Ferenc Rado Mathematical Association (established in 1993), which publish the mathematical journal for undergraduate students MatLap. Each year, in Cluj is organized the Ferenc Rado Memorial Mathematical Contest for High School students.

Professor Mór Réthy (1846–1925)

Mór Réthy was born in Nagykőrös, Hungary on November 9, 1846.

University studies: After following the primary and secondary school in his native town Nagykőrös, Mór Réthy attended the Technical Universities of Vienna and Budapest. He graduated with a degree in mathematics and descriptive geometry from the Technical University of Budapest in 1870.

Thanks to a state bursary suggested by Baron Loránd Eötvös, he had the opportunity to continue his studies at the famous universities of Göttingen and Heidelberg. In Heidelberg Kirchhoff, Königsberger and Schering assured him lifelong mental munition. He obtained his doctoral degree from Heidelberg University in 1874.

Didactical activity: Following his graduation, Réthy worked for two years as a teacher of mathematics and descriptive geometry at the Modern Technical School of Körmöcbánya (now Kremnica, Slovakia). Returning home after the award of his doctorate, he was appointed extraordinary professor at the University of Kolozsvár (1874). His seminars in mathematics – on elliptic functions, complex functions and determinants – gave a new colour to contemporary mathematical life. In 1876 he was promoted professor in the Mathematical and Theoretical Physics Department at Kolozsvár. He was also promoted as a Dean of the Faculty of Mathematics and Natural Sciences, serving in this role on two separate occasions. From 1884 to 1886 he was the Head of the Department of Elementary Mathematics. In 1886 Mór Réthy was invited to the Technical University of Budapest, where he first lectured on geometry. Then his interest turned to theoretical problems of physics and mechanics. From 1892 he was professor of the Analytical Mechanics and Theoretical Physics Department.

Research activity: His first paper on the diffraction of light was presented at Göttingen in 1872. During his stay at Kolozsvár problems concerning of navigation, including the question of constructing the most efficient propeller, were the focus of interest. Mór Réthy took part in very fruitful debates between outstanding mathematicians of his age and soon enriched the literature with two papers on the topic. His whole life was interwoven with analysing, communicating and developing the work of the two Bolyais.

In 1878 Professor Mór Réthy was elected a corresponding member of the Hungarian Academy of Sciences and in 1900 became a full-member of the Academy. On 24 July 1924, he was awarded with a Jubilee doctorate from Heidelberg University.

Professor Mór Réthy passed away in Budapest, on October 16, 1925.

Professor Frigyes Riesz (1880–1956)

Frigyes Riesz was born in Győr, Hungary on January 22, 1880.

University studies: Frigyes Riesz studied at several universities: Technical University in Zurich (1897–99), Technical University in Budapest (1899–1901) and University of Göttingen (1901–1902). He obtained his doctorate from the Eötvös Loránd University in Budapest, in 1902. His doctoral dissertation was on geometry, supervised by Gyula Vályi.

Didactical activity: He spent a few years teaching in high schools in Lőcse (now Levoča, Slovakia) and Budapest before being appointed to a university post. Riesz was appointed to a chair in Kolozsvár in 1911. Starting with 1920 he continud to work at the Franz Joseph University moved to Szeged, Hungary.

In Szeged in 1922 Riesz set up the Bolyai Mathematical Institute in a joint venture with Haar and founded the journal of the Institute: *Acta Scientiarum Mathematicarum.*

In 1945 Riesz was appointed to the chair of mathematics in the University of Budapest.

Research activity: Riesz was a founder of functional analysis and his work has many important applications in physics.

By using Fréchet's ideas of distance, in his dissertation Riesz constructed a link between Lebesgue's work on real functions and the integral equations developed by Hilbert and his student Schmidt.

Many of Riesz's fundamental findings in functional analysis were incorporated with those of Banach. His theorem, now called the Riesz-Fischer theorem, which he proved in 1907, is fundamental in the Fourier analysis of Hilbert space. It was the mathematical basis for proving that matrix mechanics and wave mechanics were equivalent. This is of fundamental importance in early quantum theory.

Riesz made many contributions to other areas including ergodic theory where he gave an elementary proof of the mean ergodic theorem in 1938. He also studied orthonormal series and topology.

Frigyes Riesz received many honours for his work. He was elected to the Hungarian Academy of Sciences (corresponding member in 1916, full member in 1936) and, in 1949, he was awarded its Kossuth Prize. He was elected to the Paris Academy of Sciences and to the Royal Physiographic Society of Lund in Sweden. He received honorary doctorates from the universities of Szeged, Budapest and Paris.

Professor Frigyes Riesz passed away in Budapest, on February 28, 1956.

Professor Ludwig Schlesinger (1864–1933)

Ludwig (Lajos) Schlesinger was born in Nagyszombat, Hungary (now Trnava, Tyrnau, Slovakia) on November 1, 1864.

University studies: Ludwig Schlesinger started elementary school in Trnava and followed the high school in Presburg, now Bratislava (Slovakia). He then studied mathematics and physics at the universities of Heidelberg and Berlin, and he received a doctorate from the University of Berlin in 1887 for a thesis on differential equations, advised by Lazarus Immanuel Fuchs and Leopold Kronecker.

Didactical activity: In 1889 Schlesinger became an associate professor at the University of Berlin; in 1897 he was an invited professor at the University of Bonn, and in the same year he was appointed professor of mathematics at the University of Kolozsvár, Hungary (now Cluj, Romania). Here he served as head of the department of higher mathematics and, in 1906–07, he was the dean of the Faculty of Mathematics and Sciences. At the Franz Joseph University he was one of the most dedicated organisers of the centenary festivities dedicated to the hundredth anniversary of János Bolyai (1902). He identified the house in which János Bolyai was born and he held an excellent conference on the centenary festivity. During his stay in Kolozsvár (Cluj), Schlesinger contributed significantly – together with Gyula Farkas and Gyula Vályi – to the advancement of mathematics in the city. They also had a decisive role in the establishment of an excellent mathematics library within the university. He wrote in Kolozsvár 16 undergraduate courses in Hungarian, in several areas of mathematics. In 1911 he left Kolozsvár and moved to the University of Giessen, Germany, where he continued to teach until he retired in 1930.

Research activity: Like his professor Fuchs, he worked primarily on linear ordinary differential equations. His two-volume *Handbuch der Theorie der Linearen Differentialgleichungen* was published from 1895 to 1898 in Teubner in Leipzig (Vol.2 in two parts). He also published *Einführung in die Theorie der gewöhnlichen Differentialgleichungen auf funktionentheoretischer Grundlage* (Auflage, 1922), *Vorlesungen über lineare Differentialgleichungen* (1908) and *Automorphe Funktionen* (Gruyter, 1924). In 1909 he wrote a long report for the annual report of the *German Mathematical Society* on the history of linear differential equations since 1865. He also studied differential geometry, and wrote a book of lectures on Einstein's general relativity theory.

Today, his best known work is *Über eine Klasse von Differentialsystemen beliebiger Ordnung mit festen kritischen Punkten* (*Crelle's Journal*, 1912). There he considered the problem of isomonodromy deformations for a certain matrix Fuchsian equation; this is a special case of Hilbert's 21st Problem (existence of differential equations with prescribed monodromy). The paper introduced what are today called *Schlesinger transformations* and *Schlesinger equations*.

Schlesinger was also a historian of mathematics. He wrote an article on the function theory of Carl Friedrich Gauss and translated René Descartes' *La Géomtrie* into German (1894). From 1904 to 1909 with R. Fuchs he collected the works of his professor Lazarus Fuchs, who was also his father-in-law.

From 1929 until his death he was co-editor of *Crelle's Journal.*

In 1902 Schlesinger was elected as a corresponding member of the Hungarian Academy of Sciences, and in 1909 he was honoured with the award of the Lobachevsky Prize.

Professor Ludwig Schlesinger passed away in Giessen, Germany on December 16, 1933.

Professor Gyula Szőkefalvi-Nagy (1887–1953)

Gyula Szőkefalvi-Nagy was born in Erzsébetváros, Hungary (now Dumbrăveni, Sibiu country, Romania) on April 11, 1887.

University studies: Gyula Szőkefalvi-Nagy studied mathematics and physics at the University of Cluj, where he received a doctorate in 1909, for a thesis on arithmetic properties of algebraic curves, advised by Gyula Schlesinger. In 1911–12, he made postgraduate studies, supported by a state stipendium in Gottingen, where he had contacts with the best mathematicians of the moment, like David Hilbert.

Didactical activity: Following his graduation, Gyula Szőkefalvi-Nagy worked for two years in high schools at Privigye (now Prividza, Slovakia) and Csíkszereda (Miercurea-Ciuc, Romania). Returning home after his studies in Germany, he was appointed extraordinary professor at the University of Kolozsvár (1915) and in the same year, director at the Catholic School Marianum in Cluj. He left Kolozsvár in 1929, and continued to work at the Ferenc József University moved to Szeged (Hungary). During the second world war he was appointed to the chair of geometry of the University of Kolozsvár (1940–1945). Starting with 1945 he was leading the chair of geometry in Szeged, until his death in 1953.

Research activity: The main results of Gyula Szőkefalvi-Nagy are in the geometrical applications of algebra and number theory. His results were published in the most prestigious journals, like *Bulletin of the American Mathematical Society*, *Archiv der Mathematic und Physik*, *Mathematische Annalen*, *Acta Scientiarum Mathematicarum*, ...

Professor Gyula Szőkefalvi-Nagy was elected to the Hungarian Academy of Sciences (corresponding member in 1934, full member in 1946). He passed away in Szeged, Hungary on October 14, 1953.

Professor Gyula Vályi (1855–1913)

Gyula Vályi was born in Marosvásárhely, Hungary (now Târgu-Mureş, Mureş country, Romania) on January 25, 1855.

University studies: After graduating the high school in his hometown in 1873, he went to Kolozsvár, the capital of Transylvania, where he attended the recently established university. After qualifying as a teacher of mathematics and physics in 1877, Vályi was awarded a scholarship to allow him to study for two years at the University of Berlin, where he attended lectures of Kummer, Borchardt, Weierstrass and Kronecker. A few months after his return to Cluj, in 1880 Valy receive his Ph.D., with a thesis titled: *On the theory of partial differential equations of the second order.*

Didactical activity: Gyula Vályi become a dozent at the University of Kolozsvár in 1881, is appointed professor of theoretical physics in 1884, and in the following year he also became professor of mathematics, lecturing on analysis, geometry and number theory. He was lecturing also on non euclidean geometry following the *Appendix* of János Bolyai. Vályi remained in Kolozsvár all his life despite being offered a professorship in Budapest. He retired in 1911 because of his deteriorating eyesight.

Research activity: His research were focused on partial differential equations, analytic and projective geometry, elementary mathematics and number theory. His doctoral dissertation on the theory of the propeller, which led to his developing on the theory of partial differential equations of the second order, published originally in Hungarian, was republished in 1906 and published also in German in 1909. He published in *Matematikai és Természettudományi Értesítő*, in *Crelle's Journal*, and other domestic and foreign journals. He also contributed essentially to the research of the Bolyai-legacy.

Professor Gyula Vályi was elected a corresponding member of the Hungarian Academy of Sciences in 1891.

Professor Gyula Vályi passed away in Kolozsvár, Hungary (now Cluj-Napoca, Romania) on October 13, 1913.

References

[1] ***, *A Kolozsvári Magyar Királyi Ferenc József tudományegyetem tanrendje*, 1872-1919 között, Kolozsvári Egyetemi Könyvtár.

[2] Andonie, Gh. Şt., *Istoria matematicii din România*, Ed. Ştiinţifică, Bucureşti, 1967.

[3] Balázs, M., Szenkovits, F., *Az erdélyi magyar matematikusok, csillagászok és informatikusok tudományos munkássága az 1945–1990 időszakban*, Műszaki Szemle, Historia Scientiarium, **37**(2007), no. 4, 22–37.

[4] Császár, Á., *Ötven éve halt meg Szőkefalvi Nagy Gyula*, A Szent István Tudományos Akadémia székfoglaló előadásai, Új Folyam. 8. szm, Budapest, 2005.

[5] Filep, L., *Life and work of Gyula Farkas* (Hungarian), Mat. Lapok, **29**(4)(1977/81), 231-244.

[6] Gaal, Gy., *Egyetem a Farkas utcában*, Erdélyi Magyar Műszaki Tudományos Tárdsaság, Kolozsvár, 2001.

[7] Kása, Z., *The beginning and ending of the Franz Joseph University*, Műszaki Szemle, Historia Scientiarium, **46**(2009), no. 6, 24-27.

[8] Kása, Z., *Kolozsvári temetőkben nyugvó matematikusok (Mathematicians resting in Cluj cemeteries)*, Természet Világa, **141**(2010), 265-267.

[9] Kása, Z., *Negyven éve hunyt el Gergely Jenö matematikus (Forty years after death of the mathematician Eugen Gergely)*, Historia Scientiarium, **12**(2014), 14-24.

[10] Kolumbán, J., *A kolozsvári matematikai iskola kialakulása (Formation of the Cluj mathematical school)*, in "125 éves a Kolozsvári egyetem (Cluj University at 125 years)", Péter Cseke and Melinda Hauer (eds.), Kom-Press Kiadó, Kolozsvár, 1999, 65–75.

[11] Maurer, I. Gy., *Az erdélyi magyar matematikai élet vázlatos története 1945-ig*, manuscript, 2004.

[12] Nagydobai Kiss, S., *A matematika szolgálatában. Radó Ferenc matematikus életpályája*, Szatmárnémeti, 2011.

[13] Oláh-Gál, R., *The list of persons obtaining doctor's degree in Mathematics at Franz Joseph University* (in Hungarian), Műszaki Szemle, Historia Scientiarium, **46**(2009), no. 6, 28-33.

[14] Szénássy, B., *History of Mathematics in Hungary until the 20th Century*, Berlin-Heidelberg-New York, 1992.

[15] Weszely T., *Vályi Gyula élete és munkássága*, Kriterion, Bukarest, 1978.

[16] ***, *Magyar életrajzi lexikon*, http://mek.oszk.hu/00300/00355/html/index.html

[17] ***, *Mathematics Genealogy Project*, http://genealogy.math.ndsu.nodak.edu/.

[18] ***, *The MacTutor History of Mathematics archive*, http://www-history.mcs.st-and.ac.uk/.

[19] ***, *Profesori care au onorat şcoala clujeană de matematică şi informatică*, http://www.cs.ubbcluj.ro/despre-facultate/istoric/magistri/

[20] ***, *Romániai Magyar Irodalmi Lexikon.* http://lexikon.kriterion.ro

Ferenc Szenkovits
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street
400084 Cluj-Napoca, Romania
e-mail: `fszenko@math.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 435–442

Reconstructibility of trees from subtree size frequencies

Dénes Bartha and Péter Burcsi

Abstract. Let T be a tree on n vertices. The subtree frequency vector (STF-vector) of T, denoted by $\text{stf}(T)$ is a vector of length n whose kth coordinate is the number of subtrees of T that have exactly k vertices. We present algorithms for calculating the subtree frequencies. We give a combinatorial interpretation for the first few and last few entries of the STF-vector. The main question we investigate – originally motivated by the problem of determining molecule structure from mass spectrometry data – is whether T can be reconstructed from $\text{stf}(T)$. We show that there exist examples of non-isomorphic pairs of unlabeled free (i.e. unrooted) trees that are STF-equivalent, i.e. have identical subtree frequency vectors. Using exhaustive computer search, we determine all such pairs for small sizes. We show that there are infinitely many non-isomorphic STF-equivalent pairs of trees by constructing infinite families of examples. We also show that for special kinds of trees (e.g. paths, stars and trees containing a single vertex of degree larger than 2), the tree is reconstructible from the subtree frequencies. We consider a version of the problem for rooted trees, where only subtrees containing the root are counted. Finally, we formulate some conjectures and open problems and outline further research directions.

Mathematics Subject Classification (2010): 05C05.

Keywords: Tree reconstruction, subtree size frequencies.

1. Introduction

Reconstruction of combinatorial structures from partial information is a widely discussed topic in the literature, full of intriguing and notoriously hard problems. Our present paper falls in the domain of reconstructibility investigations. Similar problems include reconstructibility of strings from factors or subsequences [2, 4], reconstructibility of graphs from vertex- or edge-deleted subtrees [6, 7], reconstructibility of matrices [3, 5], reconstruction of strings from Parikh vectors [1] and others.

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

The problem we investigate is the possibility of reconstruction of an unlabeled free (i.e. unrooted) tree with n vertices, given the number of subtrees of size $1, 2, \dots, n$, which we call the STF-vector of the tree. The motivation of the questions comes from the interpretation of mass spectrometry data.

The paper is structured as follows: in Section 2, we give the definition of the subtree frequency vector, and discuss some of its properties. In Section 3, we introduce methods for calculating the STF-vectors. Our two main tools are a version of the STF-vector for rooted trees and a polynomial representation of the STF-vector. In Section 4, we show that in some cases, the STF-vector uniquely determines the tree (up to isomorphism). In Section 5, we present examples where the STF information is insufficient for reconstructing the tree. In the Conclusion we present open questions and propose new research directions.

All symbols – if not stated otherwise – represent nonnegative integers, x is used for the variable of univariate polynomials and n usually denotes the number of vertices in a tree.

2. Basic definitions and properties

Definition 2.1. *Let T be a tree on n vertices. The subtree frequency vector (STF-vector) of T, denoted by* $\mathrm{stf}(T)$ *is a vector of length n whose entry at position k is the number of subtrees of T that have exactly k vertices.*

Remark 2.2. *Note that* stf *is clearly invariant for isomorphism. Thus in the reconstruction problem mentioned later, we are only interested in reconstructing the (unlabeled) tree up to isomorphism. Note however, that in the calculation of the STF-vector, all subtrees are considered, and isomorphic subtrees are counted with multiplicity.*

For example, if P_5 denotes a path of length 5 and S_4 a star with 4 leaves, then we have $\mathrm{stf}(P_5) = [6, 5, 4, 3, 2, 1]$ and $\mathrm{stf}(S_4) = [5, 4, 6, 4, 1]$.

Proposition 2.3. *Let T be a tree on n vertices with* $\mathrm{stf}(T) = [a_1, a_2, \dots, a_n]$. *Then the following holds:*

i) $a_1 = n$,
ii) $a_2 = n - 1$,
iii) $a_3 = \sum_{v \in V} \binom{d(v)}{2}$, *where V is the set of vertices and $d(v)$ denotes the degree of v.*
iv) a_{n-1} *equals the number of leaves,*
v) $a_n = 1$.

Proof. For iii) note that a tree with 3 vertices is a path, so we can calculate such subtrees by counting how many of them are centered at each vertex of T, giving the formula. For iv) note that omitting a vertex v from T is connected if and only if v is a leaf. The other statements are trivial. □

We also introduce the rooted version of the STF-vector, partly because it is interesting on its own, but it also helps in calculating the unrooted STF-vector.

Definition 2.4. *Let T be a tree on n vertices and v a vertex of T. The rooted subtree frequency vector (RSTF-vector) of T with root v, denoted by* $\mathrm{rstf}(T, v)$ *is a vector of length n whose entry at position k is the number of subtrees of T that contain v and have exactly k vertices.*

For example if T is a path on 5 vertices, and v is its center, then $\mathrm{rstf}(T, v) = [1, 2, 3, 2, 1]$. If v' is a leaf in T, then $\mathrm{rstf}(T, v') = [1, 1, 1, 1, 1]$.

Proposition 2.5. *Let T be a rooted tree on n vertices, v the root of T, and for all vertices v' denote by $T_{v'}$ the subtree rooted at v'. Then* $\mathrm{stf}(T) = \sum_{v'} \mathrm{rstf}(T_{v'}, v')$.

Proof. Simply observe that each subtree has a unique node v' highest up in the tree, and is thus counted exactly once on the right side. □

3. Methods for calculating subtree frequencies

One possible solution to calculate the STF-vector of an unlabeled free (i.e. unrooted) tree with n vertices is to generate all the subtrees of the given tree and count their sizes in a vector. Since there can be exponentially many subtrees, this is not always applicable.

Another possibility is to use Proposition 2.5 and apply recursion. The problem then reduces to calculating RSTF-vectors for arbitrary $T_{v'}$ and v'. RSTF vectors can also be calculated using recursion. We could give a combinatorial description of the process, but it would be essentially equivalent to the polynomial method given below. We introduce polynomials for representing STF-vectors. It turns out that they are useful in both calculation of STF-vectors and in proving results about reconstructibility.

Definition 3.1. *Let T be a tree, v a vertex of T. Let* $\mathrm{stf}(T) = [a_1, a_2, \ldots, a_n]$ *and* $\mathrm{rstf}(T, v) = [b_1, b_2, \ldots, b_n]$. *The STF-polynomial of T, denoted by $s(T)$ is defined by $s(T) = a_1 + a_2x + a_3x^2 + \cdots + a_nx^{n-1}$. The RSTF-polynomial of T with root v, denoted by $r(T, v)$ is defined by $r(T, v) = b_1 + b_2x + b_3x^2 + \cdots + b_nx^{n-1}$.*

Remark 3.2. *Note that the degree k coefficient of the polynomial corresponds to the number of subtrees with k edges rather than k vertices and is a degree $n-1$ polynomial. This will yield simpler formulas later.*

We prove a few results which together allow a recursive calculation of s and r.

Lemma 3.3. *Let T_1, T_2 be rooted trees with roots v_1 and v_2 respectively. Let T be the rooted tree obtained by joining the two trees by identifying v_1 and v_2 as a new vertex v. Then $r(T, v) = r(T_1, v_1)r(T_2, v_2)$.*

Proof. A subtree of T containing v and exactly k edges is obtained by joining a a subtree of T_1 containing v_1 and i edges with a subtree of T_2 containing v_2 and j edges, where $i+j = k$. The number of such pairs is $\sum_{i+j=k} \mathrm{rstf}(T_1, v_1)[i+1] \cdot \mathrm{rstf}(T_2, v_2)[j+1]$, which is exactly the kth coefficient in the polynomial product. (We denote by $v[i]$ the ith component of vector v). □

Example 3.4. *Let T_1 and T_2 be paths of length 2 with v_1 and v_2 leaves of T_1 and T_2 respectively. Then T is a path of length 4 rooted at its center v. The polynomials $r(T_1, v_1) = r(T_2, v_2) = 1 + x + x^2$, while $r(T, v) = 1 + 2x + 3x^2 + 2x^3 + x^4 = r(T_1, v_1)r(T_2, v_2)$.*

Lemma 3.5. *Let T_1, T_2 be rooted trees with roots v_1 and v_2 respectively. Let T be the rooted tree obtained by joining the two trees by identifying v_1 and v_2 as a new vertex v. Then $s(T) = r(T_1, v_1)r(T_2, v_2) + s(T_1) - r(T_1, v_1) + s(T_2) - r(T_2, v_2)$.*

Proof. Observe that a subtree not containing v is either a subtree of T_1 not containing v_1, or a subtree of T_2 not containing v_2. The number of such subtrees is counted by the polynomials $s(T_1) - r(T_1, v_1)$ and $s(T_2) - r(T_2, v_2)$, respectively. This gives the desired result. □

This latter statement allows one to calculate r and s of a rooted tree recursively if the root is not a leaf. Take the subtrees that are obtained by taking the root and all nodes below one child of the root, and join them at the root. Since these subtrees are smaller than the original tree, the calculation can proceed recursively using the following proposition. Note that the base cases of the recursion are trees with 1 or 2 vertices for which the calculation is trivial.

Lemma 3.6. *Let T be a tree and v a leaf of T. Denote by v' the only neighbor of v and by T' the subtree obtained by removing v. Then $r(T, v) = 1 + xr(T', v')$ and $s(T) = s(T') + r(T)$.*

Proof. Apart from the single-node subtree consisting of v itself, all subtrees containing v also contain v', and such subtrees of T of size k are in bijection with subtrees of T' containing v' and of size $k-1$. This proves the first statement. The second statement is trivial. □

4. Reconstructibility results

We present a few results which show that in some cases, $\mathrm{stf}(T)$ uniquely determines T. The first two are trivial observations, the third one requires deeper analysis.

Proposition 4.1. *If $\mathrm{stf}(T) = [n, n-1, \ldots, 1]$ for some n, then T is a path on n vertices.*

Proof. From the vector we deduce that the tree has n vertices and that it contains $\binom{n}{2}$ subtrees. Every tree on n vertices contains at least $\binom{n}{2}$ subtrees, namely the paths between pairs of vertices. The only tree that does not contain any further subtrees is a path. □

Proposition 4.2. *Let S_k be a star with $k \geq 2$ leaves. If $\mathrm{stf}(T) = \mathrm{stf}(S_k)$, then T is isomorphic to S_k.*

Proof. By Proposition 2.3, the number of vertices is $k + 1$ and the number of leaves is k, which implies the claim. □

Definition 4.3. *Let $a_1, a_2, \dots, a_k \geq 1$ and let $SL(a_1, a_2, \dots, a_k)$ denote the star-like graph obtained by joining paths of length $a_1, a_2, \dots, a_k$ at their endpoints. See Figure 5 for an illustration.*

Theorem 4.4. *Let $k, l \geq 3$, and $1 \leq a_1 \leq a_2 \leq \dots \leq a_k$, $1 \leq b_1 \leq b_2 \leq \dots \leq b_l$. If $\mathrm{stf}(SL(a_1, a_2, \dots, a_k)) = \mathrm{stf}(SL(b_1, b_2, \dots, b_l))$, then $k = l$ and $a_i = b_i$ for $i = 1, 2, \dots, k$.*

Proof. Let $T_1 = SL(a_1, \dots, a_k)$ and $T_2 = SL(b_1, \dots, b_k)$. By Proposition 2.3, the number of leaves is the same in the two graphs, which implies $k = l$. By Lemma 3.5 we obtain for the polynomials (which by the conditions are equal):

$$s(T_1) = \prod_{i=1}^{k}(1 + x + x^2 + \cdots + x^{a_i}) + \sum_{i=1}^{k}\left(a_i + (a_i - 1)x + \cdots + x^{a_i - 1}\right)$$
$$s(T_2) = \prod_{i=1}^{k}(1 + x + x^2 + \cdots + x^{b_i}) + \sum_{i=1}^{k}\left(b_i + (b_i - 1)x + \cdots + x^{b_i - 1}\right)$$

Assume by contradiction that for some i, $a_i \neq b_i$ holds and i is the smallest such index. Denote by c the constant term in the polynomials and note that $c = a_1 + a_2 + \cdots + a_k + 1 = b_1 + b_2 + \cdots + b_k + 1$, thus $i < k$. Wlog., we may assume $a_i < b_i$. Compare the coefficients of degree $c - a_i - 2$ in the expansion of the two polynomials. Note that $c - 1$ is the degree of the polynomials. The sums $a_i + (a_i - 1)x + \cdots + x^{a_i - 1}$ do not contribute to this term (because $k \geq 3$ and $i < k$), so we only have to compare the expansion of the products. The expansion of the product gives a reciprocal polynomial, thus it is enough to show that the degree $a_i + 1$ term differs in the products. This coefficient can be calculated if we consider the products modulo $x^{a_i + 2}$. Then the first $i - 1$ factors coming from $s(T_1)$ and $s(T_2)$ are identical, but in the ith factor, $s(T_2)$ has the additional term $x^{a_i + 1}$, which contributes to the product. For the remaining factors, $s(T_2)$ has always at least as many terms as $s(T_1)$. □

A similar statement holds for rooted STF-vectors which, however, is easier to prove.

Proposition 4.5. *Let $k, l \geq 3$, and $1 \leq a_1 \leq a_2 \leq \dots \leq a_k$, $1 \leq b_1 \leq b_2 \leq \dots \leq b_l$. Let $T_1 = SL(a_1, a_2, \dots, a_k)$ and $T_2 = SL(b_1, b_2, \dots, b_l)$, with the vertices of degree larger than 2: v_1 and v_2 as roots. If $r(T_1, v_1) = r(T_2, v_2)$, then $k = l$ and $a_i = b_i$ for $i = 1, 2, \dots, k$.*

Proof. We have

$$f = r(T_1, v_1) = \prod_{i=1}^{k}(1 + x + x^2 + \cdots + x^{a_i})$$
$$g = r(T_2, v_2) = \prod_{i=1}^{l}(1 + x + x^2 + \cdots + x^{b_i})$$

Assume by contradiction that $a_k \neq b_l$, say $a_k > b_l$. If we look at the polynomials as complex polynomials, then a primitive (a_k)th root of unity is a root of f but not a root of g. So $a_k = b_l$, and a primitive (a_k)th root of unity is a root of both f and g. We deduce that the factor $(1 + x + \cdots + x^{a_k})$ is present in both products. After simplifying, we proceed by induction on $\max(k, l)$ and the claim follows. □

5. STF-equivalent trees

Definition 5.1. *We say that trees T_1 and T_2 are STF-equivalent if* $\mathrm{stf}(T_1) = \mathrm{stf}(T_2)$.

In this section we consider non-isomorphic STF-equivalent trees. We performed computer experiments in order to determine STF-equivalent pairs of trees for $n \le 21$ (n is the number of vertices). We found that for $n \le 9$ no such pairs exist and for $10 \le n \le 21$, there always exist non-isomorphic STF-equivalent trees. This means that in general, unique reconstruction from STF-vectors is impossible. We show the computational results in Table 1.

n	#trees	#classes	largest class	#dog's bone
0	1	0	0	0
⋮	⋮	⋮	⋮	⋮
9	47	0	0	0
10	106	1	2	1
11	235	4	2	0
12	551	5	2	1
13	1301	12	2	1
14	3159	32	2	0
15	7741	62	2	0
16	19320	139	3	3
17	48629	298	3	0
18	123867	649	3	0
19	317955	1441	4	2
20	823065	3330	3	2
21	2144505	7932	4	0
22	5623756	?	?	3
24	39299897	?	?	2
25	104636890	?	?	3
28	2023443032	?	?	7
31	40330829030	?	?	4

Table 1. The number of STF-equivalence classes containing at least two trees and the maximal size of a class for $n \le 21$. The last column shows the number of classes that contain a special kind of graph which we call dog's bone graphs – all such examples are shown for $n \le 31$.

Based on computational investigation, we tried to construct general examples of non-isomorphis STF-equivalent pairs. We present two infinite families of non-isomorphic STF-equivalent pairs, showing that for sizes $n = 3k + 1$, there always exist such pairs. We introduce a notation for a special kind of graph, which – based on its shape – we call dog's bone graphs.

Definition 5.2. *Let $a, b, c, d, e \geq 1$. The dog's bone tree $DB(a, b, c, d, e)$ is a tree that contains two vertices v, v' of degree 3 connected by a path of length c, and two paths of length a and b starting at v, and two other paths of length d and e starting at v'. See Figure 5 for an illustration.*

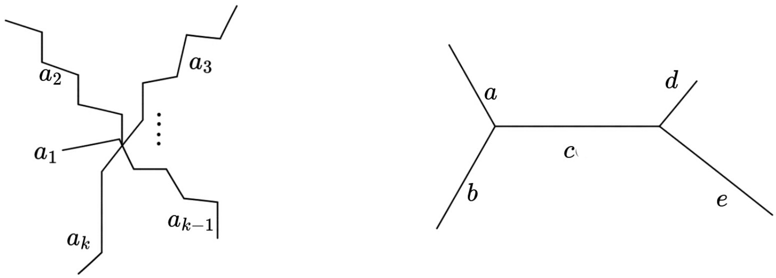


Figure 1. On the left: the star-like graph $SL(a_1, a_2, \dots, a_k)$.
On the right: the dog's bone $DB(a, b, c, d, e)$.

Theorem 5.3. *Let $k \geq 1$. The trees $T_1 = DB(k, 2k+1, 1, k, 2k+1)$ and $T_2 = DB(k, k, 1, 2k, 2k+2)$ are STF-equivalent.*

Proof. Using Lemma 3.5, and applying summation for geometric series, after some calculation we have the following polynomials $f = s(T_1), g = s(T_2)$.

$$f = \frac{x\left(x^{k+1}-1\right)^2\left(x^{2k+2}-1\right)^2}{(x-1)^4} + 2\left(\frac{x\left(x^{3k+2}-1\right)}{x-1} - 3k - 2\right)(x-1)^{-1}$$

$$g = \frac{x\left(x^{k+1}-1\right)^2\left(x^{2k+1}-1\right)\left(x^{2k+3}-1\right)}{(x-1)^4}$$
$$+ \left(\frac{x\left(x^{2k+1}-1\right)}{x-1} + \frac{x\left(x^{4k+3}-1\right)}{x-1} - 6k - 4\right)(x-1)^{-1}$$

Their equality would be tedious to check by hand, but can readily be verified on a computer algebra system: if we replace all occurrences of x^k by a new variable y, then the difference of the resulting bivariate polynomials simplifies to 0. □

The following theorem can be proved similarly.

Theorem 5.4. *Let $k \geq 1$. The trees $T_1 = DB(k, 2k+2, 1, k+1, 2k+2)$ and $T_2 = DB(k, k+1, 1, 2k+1, 2k+3)$ are STF-equivalent.*

Corollary 5.5. *There exist non-isomorphic pairs of STF-equivalent trees for $6k+1$ and $6k+4$ vertices, for any $k \geq 1$.*

6. Summary and further work

In this paper we introduced the concept of STF-vectors and investigated the problem of reconstructibility of trees from subtree frequencies. We pose some open questions.

- Find more families of non-isomorphic STF-equivalent pairs and prove that such pairs exist for all $n \geq 10$.
- Find more types of graphs which are reconstructible from their STF-vectors.
- Are STF-equivalence class sizes unbounded as n grows?
- Calculate the STF-vector of a tree together with all RSTF-vectors. Are these $n+1$ vectors already sufficient for reconstruction up to isomorphism?
- Investigate the relationship of STF-vectors with other graph invariants, e.g. spectrum.

Besides these, our ongoing research will mainly focus on the labeled version of the problem, where each vertex or edge of the tree has a label from a finite set of colors.

Acknowledgement. The research of the second author was partially supported by a special contract No. 18370-9/2013/TUDPOL with the Ministry of Human Recources.

References

[1] Acharya, J., Das, H., Milenkovic, O., Orlitsky, A., Pan, S., *String reconstruction from substring compositions*, arXiv:1403.2439v1.

[2] Dudik, J., Schulman, L.J., *Reconstruction from subsequences*, J. Combin. Theory Ser. A, **103**(2003), no. 2, 337–348.

[3] Kós, G., Ligeti, P., Sziklai, P., *Reconstruction of matrices from submatrices*, Math. Comput., **78**(2009), 1733–1747.

[4] Krasikov, I., Roddity, Y., *On a reconstruction problem for sequences*, J. Combin. Theory Ser. A, **77**(1997), no. 2, 344–348.

[5] Manvel, B., Stockmeyer, P.K., *On reconstruction of matrices*, Math. Mag., **44**(1971), no. 4, 218–221.

[6] Manvel, B., *Reconstruction of trees*, Canad. J. Math., **22**(1970), 55–60.

[7] Manvel, B., *On reconstructing graphs from their sets of subgraphs*, J. Combin. Theory Ser. B, **21**(1976), 156–165.

Dénes Bartha
Eötvös Loránd University
Department of Computer Algebra
Budapest, Hungary
e-mail: `denesb@gmail.com`

Péter Burcsi
Eötvös Loránd University
Department of Computer Algebra
Budapest, Hungary
e-mail: `bupe@compalg.inf.elte.hu`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 443–455

Reconstructing graphs from a deck of all distinct cards

Miklós Bartha and Amitesh S. Shuva

Abstract. The graph reconstruction conjecture is looked at from a new perspective. Given a graph G, three equivalence relations are considered on $V(G)$: card equivalence, automorphism equivalence, and the equivalence of having the same behavior. A structural characterization of card equivalence in terms of automorphism equivalence is worked out. Relative degree-sequences of subgraphs of G are introduced, and a new conjecture aiming at the reconstruction of G from the multiset of relative degree-sequences of its induced subgraphs is formulated. Finally, it is shown that graphs having a deck free from duplicate cards are reconstructible.

Mathematics Subject Classification (2010): 05C60, 18D10.

Keywords: Graph reconstruction, card equivalence, automorphism equivalence, relative degree sequences, traced monoidal categories.

1. Introduction

For a graph G and vertex $v \in V(G)$, $G - v$ is the graph obtained from G by deleting the vertex v and its incident edges. We call $G - v$ a vertex-deleted subgraph of G, or the card associated with vertex v in G. We do not distinguish between isomorphic cards, though. The multiset of cards collected from G in this way is called the deck of G, denoted $D(G)$.

Perhaps the most well-known unsolved problem of graph theory asks whether an arbitrary graph G having at least three vertices can be reconstructed in a unique way (up to isomorphism) from its deck. The likely positive answer to this question is commonly known as the Reconstruction Conjecture (R.C., for short), and it was formulated by Kelly and Ulam as early as 1942. Ever since its inception, this problem has remained a mystery. Trying to solve it is similar to conducting a criminal investigation. There is a suspect, the graph G, who leaves plenty of evidence (i.e., the deck $D(G)$) on the crime scene. Yet, no brilliant detective has been able to track

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

down the suspect for over 70 years, and the number of works on the case is rapidly decreasing year by year. The reconstruction problem was, however, very popular in the past. According to [15], more than 300 research papers had been published on graph reconstruction between 1950 and 2004.

One of the last true champions of graph reconstruction was F. Harary. He suggested a natural analogue [7] of the R.C., which says that every graph having at least four vertices is uniquely reconstructible from the deck of its edge-deleted subgraphs. Others have come up with similar conjectures for directed graphs, cf. [14, 16], and have obtained partial results proving or disproving them. The reader is referred to [8] and [13] for two excellent surveys on graph reconstruction.

In this paper we propose an original new approach to the study of the reconstruction problem. This approach is structural, rather than combinatorial. It is deeply rooted in algebra and category theory, despite the fact that the proofs of our present results are completely elementary. The results themselves, however, have been distilled from an entirely independent study focusing on the completeness of the traced monoidal category axioms [1, 10] in different well-known mathematical structures satisfying these axioms. We shall elaborate on this study to some extent in Section 4.

2. Definitions, and some easily recoverable data

Let G be a graph having at least three vertices, fixed for the rest of the paper. As usual, $V(G)$ and $E(G)$ will denote the set of vertices and edges of G, respectively. We assume that G is *simple* in the sense that it does not contain loops or multiple edges. In general, we rely on the terminology of [12] to deal with graphs.

Two vertices $u, v \in V(G)$ are called *hypomorphic* or *card-equivalent* (c-equivalent, for short) if the card associated with u is identical with the one associated with v, i.e., $G-u \cong G-v$. (Remember that we do not distinguish between isomorphic cards.) Clearly, c-equivalence is an equivalence relation on $V(G)$. Two graphs G and H are hypomorphic if $D(G)$ and $D(H)$ are identical as multisets, that is, each card appears in $D(G)$ and $D(H)$ the same number of times. (Recall that $D(G)$ denotes the deck of G.) If G and H are hypomorphic, then a *hypomorphism* of G onto H is a bijection $\phi : V(G) \to V(H)$ such that $G - v \cong H - \phi(v)$ holds for every $v \in V(G)$. A *reconstruction* of G is a graph G' such that G and G' are hypomorphic, or, equivalently, there exists a hypomorphism of G onto G'. Using this terminology, the R.C. simply says that two graphs G and H are hypomorphic iff they are isomorphic. In other words, all reconstructions of G are isomorphic (to G, of course). Clearly, every isomorphism of G onto H is a hypomorphism, but the converse is not true, even if the R.C. holds.

Graph G is called *card-minimal* if $D(G)$ is a set, that is, each card is unique in $D(G)$. Our aim in this paper is to show that the R.C. holds true for all card-minimal graphs. (Note that any graph on two vertices has two identical cards.) One might think that this result is trivial, since there is a unique hypomorphism between any two hypomorphic card-minimal graphs G and H. While this is certainly true, we have no direct information on $E(G)$ and $E(H)$, therefore the given unique hypomorphism may not be an isomorphism. Reconstructing G from $D(G)$ is still a very complex issue for such graphs. As we shall see, any duplication of cards in $D(G)$ indicates a kind of symmetry in the internal structure of G. Consequently, the class of card-minimal graphs is really large. Our result is therefore in accordance with the observation in [6] saying that the probability that a randomly chosen graph on n vertices is not reconstructible goes to 0 as n goes to infinity.

In general, it is trivial that $|V(G)|$, the number of vertices of G, is recoverable from $D(G)$. It is still easy to see that $|E(G)|$ is also recoverable. Indeed, add up the numbers of edges appearing on the cards of $D(G)$, and observe that this sum is equal to

$$(|V(G)| - 2) \cdot |E(G)|.$$

See [13, Theorem 2.1] for the details of this simple combinatorial argument.

Once $|E(G)|$ is given, calculating the degree $d(v)$ of vertex v for card $G - v$ is straightforward:

$$d(v) = |E(G)| - |E(G - v)|.$$

Clearly, the degree of any vertex c-equivalent with v is the same as that of v. We thus have managed to recover the degree-sequence of G from $D(G)$. Recall that the *degree-sequence* of G is the sequence of degrees of G's vertices in a non-decreasing order.

A similar combinatorial argument leads to the following result, known as Kelly's Lemma [11], see also [13, Theorem 2.4].

Proposition 2.1. *For any graph Q, let $s_Q(G)$ denote the number of subgraphs of G isomorphic to Q. Then $s_Q(G) = s_Q(H)$ whenever G and H are hypomorphic and $|V(Q)| < |V(G)|$.*

Nash-Williams [13] has also shown that the so-called degree-sequence sequence of G is recoverable from $D(G)$. Essentially this means that, not only $d(v)$ can be read from the card $G-v$ as above, but also the degrees of the neighbors of v are recoverable in this way. We shall reformulate Nash-Williams' proof in Section 4 in terms of relative degree-sequences. A natural question to ask at this point is whether the degrees of the neighbors of the neighbors of v are also recoverable, and so on, moving away further and further from vertex v. This question is already a lot more difficult to answer, mainly because the desired degrees or degree-sequences are no longer c-equivalence invariant. In other words, the answer depends on the representant vertex v chosen for card $G - v$.

3. Characterizing card equivalence

The simple results discussed in Section 2 are of a strictly combinatorial nature, and they do not even touch on the structural properties of card equivalence. In this section we present a real structural characterization of c-equivalence, which is our first main result. In this characterization, card equivalence is compared to two other important equivalence relations on $V(G)$, namely automorphism equivalence and the equivalence of having the same behavior. Card equivalence will be denoted by $\sim_c$.

Definition 3.1. Two vertices $u, v \in V(G)$ are *automorphism-equivalent* (a-equivalent, for short) if there exists an automorphism of G taking u to v.

Automorphism equivalence will be denoted by $\sim_a$. It is obvious that $\sim_a$ is an equivalence relation, but its relationship to $\sim_c$ is not clear for the first sight.

Example 3.2. Let G be the graph in Fig. 1a, and consider the vertices u_1, u_2, u_3 in G. It is easy to see that $u_i \sim_c u_j$ and $u_i \sim_a u_j$ both hold for any $1 \le i, j \le 3$.

In general, it is clear by the definitions that $\sim_a \subseteq \sim_c$. Example 3.3 below shows, however, that $\sim_c \not\subseteq \sim_a$.

Example 3.3. Let G be the graph of Fig. 2, and consider again the vertices u_1, u_2, u_3. As it turns out, $u_1 \sim_c u_3$, but $u_1 \not\sim_a u_3$. Furthermore, G has no automorphisms other than the identity.

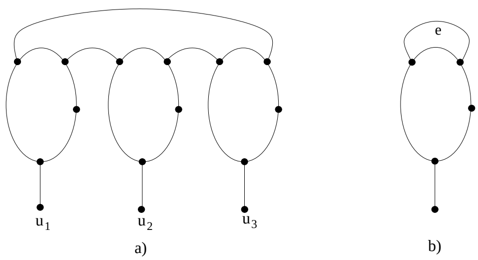


Figure 1. The graph of Example 3.2.

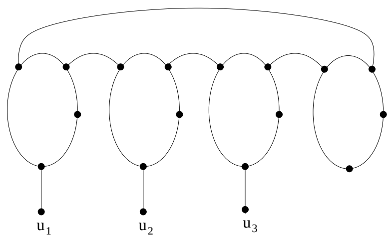


Figure 2. The graph of Example 3.3.

The reader familiar with flowchart schemes and their behaviors [1, 2, 4, 5] will notice that the graphs in Figures 1a and 2 have been inspired by appropriate flowcharts. To recover these flowcharts, make each edge bidirectional in the graphs and supply the degrees with appropriate input-output port distinctions at each vertex. The resulting flowcharts will have no entry or exit vertices, though. Also, no two lines (edges) will be joined at any input or output port. The characteristic feature of such connected "injective" flowcharts is that their proper automorphisms do not have fixed-points. The automorphisms themselves can be neatly characterized by Ésik's *commutativity axioms* [5, 3] for iteration theories. Regarding the graph G in Fig. 1a this means that G can be constructed by taking three copies of the *minimal* graph (scheme) M – shown in Fig. 1b as a multigraph – and turn the edge $e \in E(M)$ into a sequence of edges running through the three copies of $M - e$ in an appropriate way, following a cyclic permutation. This is of course a very simplistic interpretation of the otherwise truly complex commutativity axioms, but it is right to the point. On the other hand, graphs that are not scheme-like, e.g. the simple graph in Fig. 3, do have proper automorphisms with fixed-points, and the concept of minimal graph is meaningless for them.

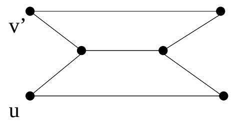


Figure 3. A non-scheme-like graph

Yet another important equivalence relation on $V(G)$ closely related to $\sim_a$ and $\sim_c$ is that of having the same behavior. The reader is referred again to [1, 4, 5] for the original definition of this concept in flowchart schemes.

Definition 3.4. The relation $\sim$ of *having the same behavior* is defined on $V(G)$ as the largest equivalence having the following two properties.

1. If $u \sim v$, then $d(u) = d(v)$.
2. If $u \sim v$ and $\{u_1, \ldots u_k\}$ ($\{v_1 \ldots v_k\}$) is the set of vertices adjacent to u (respectively, v), then the multiset of $\sim$-equivalence groups defined by the set of representants $\{u_1, \ldots u_k\}$ is the same as the one determined by $\{v_1 \ldots v_k\}$.

It is easy to see that two vertices u and v have the same behavior iff G unfolds to isomorphic infinite rooted trees starting from u and v. For example, any two vertices of a regular graph have the same behavior.

Clearly, $\sim_a \subseteq \sim$, but $\sim \not\subseteq \sim_a$. Indeed, not every two vertices of a regular graph are a-equivalent in general. On the other hand, $\sim_c$ is not comparable with $\sim$. The regular graph counterexample shows that $\sim \not\subseteq \sim_c$, and vertices u_1, u_3 in the graph of Fig. 2 demonstrate that $\sim_c \not\subseteq \sim$.

The practical importance of the equivalence $\sim$ is that it is computable in polynomial time. The algorithm to isolate the equivalence groups of $\sim$ is completely analogous to Hopcroft's [9] well-known algorithm for minimizing finite state automata. Even though $\sim_a$ is a lot more costly to compute because of the isomorphism check involved, it still helps to know that $\sim_a$ is a refinement of $\sim$.

The above comparison with the relations $\sim$ and $\sim_a$ shows that $\sim_c$ is rather inconvenient to deal with in a direct way. We need to find a characterization of $\sim_c$ that brings it in line with the much better structured equivalence $\sim_a$. The basis of this characterization is the following lemma.

Lemma 3.5. *Let u and v be two distinct vertices of G. Then $u \sim_c v$ iff there exists a sequence of vertices $x_0, x_1, \ldots, x_n$ ($n \geq 1$) in G satisfying the conditions (i) and (ii) below.*

(i) *$x_0 = v$ and $x_n = u$;*
(ii) *there exists an isomorphism ϕ of $G - u$ onto $G - v$ such that $\phi(x_i) = x_{i+1}$ for every $0 \leq i < n$.*

Proof. Notice first that the graphs $G-u$ and $G-v$ are not separated in the lemma, they both use the vertices of the common supergraph G. The lemma therefore establishes a link between two c-equivalent vertices u and v in G through a sequence of (necessarily distinct) vertices $x_1,\ldots,x_{n-1}$ in $G-u-v$. These vertices, however, need not be c-equivalent with u or each other in G. For example, in the graph of Fig. 2, if $v=u_1$ and $u=u_3$, then $n=2$, $x_1=u_2$, and ϕ can be derived from the automorphism of $G-\{u_1,u_2,u_3\}$ that determines a cyclic permutation of the four small cycles of G from left to right. Clearly, $u_1 \not\sim_c u_2$.

Sufficiency of condition (ii) alone for having $u\sim_c v$ is trivial. Assuming that $u\sim_c v$, choose an arbitrary isomorphism $\phi: G-u\to G-v$. Let $x_1=\phi(v)$, $x_2=\phi(x_1)$, and so on, until $u=x_n=\phi(x_{n-1})$ is reached. Vertex u must indeed be encountered at some point along this line, since ϕ, being an isomorphism, is an injective mapping $V(G)\setminus\{u\}\to V(G)\setminus\{v\}$. Consequently, the vertices $x_1,\ldots,x_{n-1}$ in $V(G)\setminus\{u,v\}$ will all be different until $x_n=u$ stops this necessarily finite sequence. (Mind that $x_{i+1}=\phi(x_i)\neq v$, since v is not a vertex of $G-v$.) The proof is complete. □

Theorem 3.6. *Let u and v be two distinct vertices of G. Then $u\sim_c v$ iff there exists a sequence of pairwise distinct vertices $x_0,x_1,\ldots,x_n$ ($n\geq 1$) satisfying the following conditions.*

(i) *$x_0=v$ and $x_n=u$;*
(ii) *for $X=\{x_0,x_1,\ldots,x_n\}\subseteq V(G)$ there exists an automorphism ψ of $G-X$ such that:*
  (iia) *for every $0\leq i<n$ and vertex $w_i\in V(G-X)$ adjacent to x_i in G (or, equivalently, in $G-u$), the vertex $w_{i+1}=\psi(w_i)$ is also in $V(G-X)$ and is adjacent to x_{i+1} in G (i.e., in $G-v$);*
  (iib) *for every $0\leq i<j<n$,*

$$x_i \text{ is adjacent to } x_j \text{ iff } x_{i+1} \text{ is adjacent to } x_{j+1}$$

(in G, of course).

Vertices u and v are a-equivalent iff the assignments $x_i\mapsto x_{i+1}$, $u\mapsto v$ extend the automorphism ψ in (ii) to one of G.

Proof. Intuitively, condition (iia) says that for every $0\leq i<n$, the neighbors of x_i in $G-X$ are matched up with those of x_{i+1} in $G-X$ by the automorphism ψ. Condition (iib) settles the issue of how the vertices x_i themselves are connected in G. Notice that the question whether u is connected to v is irrelevant. Indeed, it can easily happen that $u\sim_c v$ and $u\sim_c v'$ both hold, while u is adjacent to v but not to v'. See Fig. 3.

The first statement of the theorem is in fact a simple consequence of Lemma 3.5. Regarding sufficiency, if ψ is an automorphism of $G-X$ satisfying (iia) and (iib), then it can be extended to an isomorphism ϕ of $G-u$ onto $G-v$ satisfying (ii) of Lemma 3.5. Thus, $u\sim_c v$. Conversely, if $u\sim_c v$, then the required automorphism ψ can be derived in a unique way from the isomorphism ϕ guaranteed by Lemma 3.5. Notice that the subgraph $G-X$ may turn out to be empty. The second statement of the theorem is obvious. □

At this point the reader may want to have a second look at Examples 3.2 and 3.3, and identify the underlying automorphism ψ in the graphs of Fig. 1 and Fig. 2. One important point is that, given the fact $v \sim_a u$ (and therefore $u \sim_c v$), one must not jump to the conclusion saying that $x_0 = v$ and $x_1 = u$ will do for $X = \{x_0, x_1\}$ in (ii) of Theorem 3.6, and then be taken by surprise that the desired automorphism ψ cannot be located in $G - X$. For example, in the graph G of Fig. 1, if $v = u_1$ and $u = u_2$, then $x_1 = u_3$! Consequently, $X = \{u_1, u_2, u_3\}$, and the automorphism ψ is just the one taking the three small cycles into one another following a cyclic permutation with offset 2 from left to right.

4. Relative degree-sequences

Recall from Section 2 that the degree-sequence of graph G is the sequence of degrees of its vertices in a non-decreasing order. Let Q be a subgraph of G. The degree of a vertex $v \in V(Q)$ relative to G is a pair (r, d), where d (r) is the degree of v in G (respectively, Q). We shall use the notation r^d for the pair (r, d), and say that v has relative degree r out of d. Then the *relative degree-sequence* of Q (with respect to G) is the sequence of relative degrees of its vertices in an order that is non-decreasing regarding the superscripts d and also non-decreasing in r among those degrees that have the same superscript d.

The degree-sequence of G and the relative degree-sequence of Q with respect to G will be denoted by $ds(G)$ and $rds_G(Q)$, respectively. In order to ensure that $ds(G)$ and $rds_G(Q)$ have the same length, we shall include a relative "degree" $\emptyset^d$ in $rds_G(Q)$ for each vertex $v \in V(G) \setminus V(Q)$ with degree d. The "number" $\emptyset$ is treated as 0, but the notation $\emptyset$ will distinguish between a vertex that has been deleted and one that is still present but isolated. This distinction is purely technical, however, because one can easily fill in the $\emptyset^d$ relative degrees in $rds_G(Q)$ once $ds(G)$ is known.

Example 4.1. *Consider the graph G and its subgraph Q in Fig. 4. The degree-sequence of G is* $2, 2, 3, 3$*, while the relative degree sequence of Q with respect to G is* $1^2, 1^2, 1^3, 3^3$.

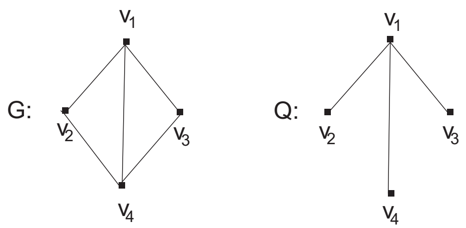


Figure 4. Graph G and its subgraph Q

The following simple combinatorial observation is equivalent to Nash-Williams' result [13, Corollary 3.5] on degree-sequence sequences.

Proposition 4.2. *For every vertex $v \in V(G)$, $rds_G(G - v)$ is recoverable from $D(G)$.*

Proof. We have seen in Section 2 that $d(v)$ and $ds(G)$ are recoverable from $D(G)$. Write the sequence $ds(G-v)$ underneath $ds(G)$ by inserting the "degree" $\emptyset$ in $ds(G-v)$ right under the position of the first occurrence of $d(v)$ in $ds(G)$. For example:

$ds(G)$:	2	2	2	3	3	4	4
$ds(G-v))$:	1	1	2	2	3	$\emptyset$	3
$rds_G(G-v)$:	1^2	1^2	2^2	2^3	3^3	$\emptyset^4$	3^4

Observe that the "true" degrees in $ds(G-v)$ will lag behind those in $ds(G)$, so that the difference between two degrees in aligned positions is at most 1. Therefore it is trivial to fill out the missing superscripts in $ds(G-v)$, so that the resulting sequence becomes $rds_G(G-v)$. □

Proposition 4.2 basically says that, for every card $G-v$, the degrees of the vertices adjacent to v in G are uniquely determined by $ds(G)$ and $ds(G-v)$. Indeed, these are exactly the degrees $r+1$ appearing in $rds_G(G-v)$ as r^{r+1}. Of course, we still have no information about the actual position of v's neighbors in $G-v$.

We immediately generalize Proposition 4.2 to find out the relative degree-sequence of all 2-vertex-deleted subgraphs of G. Notice that, for two distinct vertices $u, v \in V(G)$, the subgraph $G-u-v$ is no longer determined by the cards $G-u$ and $G-v$ in a unique way, since the cards themselves do not uniquely identify the vertices u, v. Moreover, the subgraph $G-u-v$, too, can be isomorphic to other subgraphs $G-u'-v'$ in which u' and v' are associated with some different cards.

Theorem 4.3. *Let u and v be two distinct vertices of G. Given the degree-sequence of the subgraph $G-u-v$, $rds_G(G-u-v)$ is uniquely determined by the data $ds(G)$, $ds(G-u)$, and $ds(G-v)$. Moreover, the question whether u and v are adjacent in G or not turns out from the data $ds(G)$, $ds(G-u-v)$, $d(u)$ and $d(v)$.*

Proof. We use the same alignment argument as in the proof of Proposition 4.2. Write the degree-sequences $ds(G)$, $ds(G-u)$, and $ds(G-v)$ under each other, inserting the $\emptyset$ symbol in the appropriate positions of $ds(G-u)$ and $ds(G-v)$. Furthermore, insert two $\emptyset$'s in $ds(G-u-v)$ aligned with the ones already inserted in $ds(G-u)$ and $ds(G-v)$. If $d(u) = d(v) = d$, then insert two consecutive $\emptyset$'s aligned with the beginning of the block marked by degree d in $ds(G)$. For example:

$ds(G)$:	2	2	2	3	3	4	4	...	...	...	...
$ds(G-u)$:	1	1	2	2	3	4	4	...	$\emptyset$	...	...
$ds(G-v)$:	1	2	2	2	2	3	4	...	...	$\emptyset$	...
$ds(G-u-v)$:	0	1	1	2	2	3	4	...	$\emptyset$	$\emptyset$	...
			$\rightarrow$	$\leftarrow$							
$rds_G(G-u-v)$:	0^2	1^2	2^2	1^3	2^3	3^4	4^4	...	$\emptyset$	$\emptyset$	...

Let $n_G(d)$ ($n_{G,Q}(r^d)$) denote the number of occurrences of d (r^d) in $ds(G)$ (respectively, $rds_G(Q)$). Assume, for simplicity, that the smallest degree in G is $d_0 \geq 2$. Then, clearly:

$$n_{G,Q}((d_0-2)^{d_0}) = n_Q(d_0-2).$$

It follows that:

$$n_{G,Q}((d_0-1)^{d_0}) = n_{G-u}(d_0-1) + n_{G-v}(d_0-1) - 2 \cdot n_Q(d_0-2), \quad \text{and}$$

$$n_{G,Q}(d_0^{d_0}) = n_G(d_0) - n_{G,Q}((d_0-2)^{d_0}) - n_{G,Q}((d_0-1)^{d_0}),$$

provided that neither of the degrees $d(u)$ and $d(v)$ equals d_0. If either or both does, then the above calculation changes in a straightforward way regarding the numbers $n_{G,Q}((d_0-1)^{d_0})$ and $n_{G,Q}(d_0^{d_0})$. One can then carry on in the same way, calculating the numbers $n_{G,Q}((d_0-1)^{d_0+1})$, $n_{G,Q}(d_0^{d_0+1})$, $n_{G,Q}((d_0+1)^{d_0+1})$, and so on. Details are left to the reader.

As to the second statement of the theorem, if

$$|E(G)| - |E(G-u-v)| = d(u) + d(v),$$

then u and v are not connected in G, otherwise they are. The numbers $|E(G)|$ and $|E(G-u-v)|$ are determined by $ds(G)$ and $ds(G-u-v)$, respectively. The proof is complete. □

Proposition 4.2 and Theorem 4.3 show that the concept of relative degree-sequence is rather fundamental in the study of graph reconstruction. To provide yet another evidence for this observation, let $Rds(G)$ denote the multiset

$$\{rds_G(Q) | Q \text{ is an induced subgraph of } G\}.$$

Thus, relative degree-sequences of subgraphs count with multiplicity in $Rds(G)$. We put forward the following conjecture, which is very closely related to the R.C..

Conjecture 4.4. *For every graph G, $Rds(G)$ identifies G up to isomorphism.*

Conjecture 4.4 is especially useful for several reasons.

1. It appears to hold for all graphs with no exceptions.
2. It provides a characterization of graph isomorphism, which has been sought for a very long time.
3. Algebraically, if $G = G_1 + G_2$, then

$$Rds(G) = Rds(G_1) \times Rds(G_2). \qquad (4.1)$$

In equation 4.1 above, $\times$ stands for concatenation of sets of relative degree-sequences in the formal language sense (taking the quotient of the product by commutativity). In terms of polynomials, we can think of a relative degree r^d as a formal variable. Let X denote the set of all such variables. Then $Rds(G)$ becomes a polynomial of the variables X over the integer ring $\mathbf{Z}$, in which all coefficients are non-negative. (Treat union of multisets as addition in this polynomial.) Let $\mathbf{Z}[X]$ denote the commutative $\mathbf{Z}$-module (in fact algebra) of X-polynomials over $\mathbf{Z}$. (Mind that addition of polynomials is commutative in $\mathbf{Z}[X]$.) Our fundamental observation is that the operation $\times$ in (4.1) translates naturally into product of polynomials in the algebra $\mathbf{Z}[X]$. This product makes the algebra $\mathbf{Z}[X]$ associative and commutative, therefore a commutative ring.

Conjecture 4.4 was the starting point of the present study, and the fundamental observation in the previous paragraph served as a motivation for it. In the language of category theory this observation suggests that the traced monoidal category of graphs (flowchart schemes), in which tensor is disjoint union and trace is feedback (i.e., creating an internal edge by merging two external ones, see [1, 3]) can be embedded in a natural way into the compact closed category of free modules over the commutative algebra (ring) $\mathbf{Z}[X]$, in which tensor and trace are the standard matrix operations. There is a clear analogy in this statement with the *Int* construction, cf. [10], for the "scalar" connection between graphs and polynomials is lifted to the level of traced monoidal and compact closed categories by observing that the given translation of graphs into polynomials is compatible with the trace operation at the higher level.

Naturally enough, Conjecture 4.4 also has an "edge" version, in which $Rds(G)$ is defined as the set of relative degree-sequences of *all* subgraphs of G. This version, too, appears to hold for all graphs G with no exceptions, even for multigraphs as one would expect after the flowchart scheme analogy.

The connection between Conjecture 4.4 and the R.C. is the following. If we could compute $rds(G)$ from $D(G)$, then Conjecture 4.4 would imply the R.C.. As our main result in Section 5 shows, however, computing the whole multiset $Rds(G)$ is far too much work in order to reconstruct G. Therefore this reconstruction argument probably does not hold much water, indicating that Conjecture 4.4 is even tougher than the R.C..

On the other hand, if, given $Rds(G)$, we could isolate $Rds(G-v)$ for each vertex-deleted subgraph of G, then the R.C. would imply Conjecture 4.4 through a straightforward induction argument. Since our concern is eventually Conjecture 4.4, and the construction of the multiset of multisets

$$\{Rds(G-v)|v \in V(G)\}$$

from $Rds(G)$ looks promising, we definitely must prove the R.C. first.

5. The reconstruction of card-minimal graphs

In this section we present our second main result, which aims at the reconstruction of card-minimal graphs. Temporarily, we are going to assume a further technical condition in order to keep the reconstruction simple. Dropping this condition will be the subject of a forthcoming paper. The condition is formally defined as follows.

Definition 5.1. Graph G is *2-card reconstructible* if it is connected, and for every $u, v, x, y \in V(G)$, the isomorphism

$$G - u - v \cong G - x - y$$

implies that u, v, x, and y cannot all be distinct.

To shed some light on the intuition behind Definition 5.1, let G be card-minimal, and Q be an arbitrary graph having $|V(G)| - 2$ vertices. Consider the set C of cards in $D(G)$ in which Q is isomorphic to at least one vertex-deleted subgraph. Construct the graph G_Q which has C as its set of vertices, and any two cards $G - u$, $G - v$ are connected in G_Q iff $G - u - v \cong Q$. (Remember that G is card-minimal, therefore the definition of G_Q is correct.) Then G is 2-card reconstructible iff G_Q is either a triangle or a star graph for every 2-vertex-deleted subgraph Q of G. In other words, if $|C| > 2$, then the following two conditions are met:

1. the subgraph Q occurs $k \geq 2$ times as a vertex-deleted subgraph in some card $G - u \in C$;
2. $|C| = k + 1$ and the cards in C different from $G - u$ all have a single occurrence of Q in them, with the possible exception that $k = 2$ and all the three cards in C have two occurrences of Q in them.

See Fig. 5a for a card-minimal graph G which is, and Fig. 5b for one which is not 2-card reconstructible.

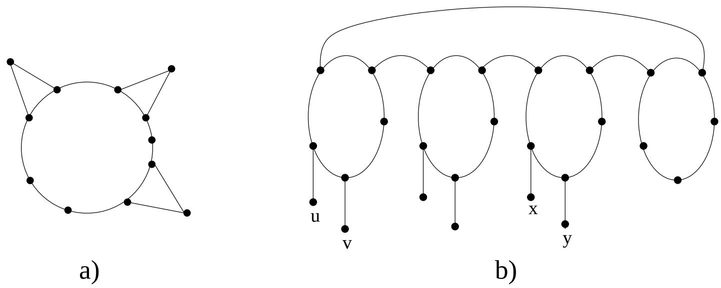


Figure 5. The 2-card reconstructibility condition

Theorem 5.2. *Every card-minimal and 2-card reconstructible graph G is reconstructible.*

Proof. Let Q be an arbitrary graph having $|V(G)| - 2$ vertices, and find the set C of cards in which Q is isomorphic to at least one vertex-deleted subgraph. If $C = \emptyset$, then drop Q as uninteresting. Otherwise C has at least two elements. If there are exactly two cards $G - u$ and $G - v$ in C, then conclude that $Q \cong G - u - v$, and use Theorem 4.3 to decide if u and v are adjacent in G or not. If C has more than two elements, then the condition of 2-card reconstructibility implies that either $|C| = 3$ and each card in C has two subgraphs isomorphic to Q, or there is exactly one card $G - u \in C$ that contains more than one subgraph isomorphic to Q. In the first case $Q \cong G - u - v$ for any pair $G - u$, $G - v$ of distinct cards in C, while in the latter $Q \cong G - u - v$ for all vertices $v \neq u$ such that $G - v \in C$. Furthermore, in this case Q is not isomorphic to any other 2-vertex-deleted subgraph of G. (In other words, $Q \ncong G - u_1 - u_2$, where $G - u_1$ and $G - u_2$ are both in C but $u_i \neq u$ for either $i = 1$ or 2.) Again, use Theorem 4.3 to find out if u is adjacent to v in G, knowing that $Q \cong G - u - v$. It is evident that the above procedure will decide for each pair of cards $G - u$, $G - v$ in $D(G)$ if the vertices u and v are adjacent in G or not. The proof is now complete. □

At this point the reader might have the impression that the condition of 2-card reconstructibility is overly restrictive. In fact it is not, and a fairly simple analysis based on the combination of Proposition 2.1 and Theorem 4.3 shows that whenever

$$G - u - v \cong G - x - y$$

holds for four distinct vertices u, v, x, y, then each possible correspondence of these vertices to appropriate cards in $D(G)$ can be identified in a consistent way. This analysis is technically complicated, however, therefore we do not include it in the present introductory paper.

6. Conclusion

Motivated by an independent research on traced monoidal categories, we have presented a structural analysis of graphs with the aim of being able to reconstruct them from some partial information. The basis of the reconstruction of graph G could either be the classical multiset of G's vertex-deleted subgraphs, or the multiset of relative degree-sequences of all induced subgraphs of G.

We have introduced three equivalence relations on $V(G)$ for the better understanding of the reconstruction problem. Card equivalence is the one directly related to the reconstruction conjecture. Our examples have shown, however, that this equivalence is rather inconvenient to deal with. Automorphism equivalence and having the same behavior have been adopted from the study of flowchart schemes and their behaviors. These relations have a much more transparent structure, and both have turned out to be very closely related to card equivalence. For an evidence, we have worked out a characterization theorem for card equivalence to bring it in line with automorphism equivalence.

With respect to relative degree sequences, we have provided a generalization of an earlier observation by Nash-Williams on the degree-sequence sequence of graphs. As an application of this result we have shown that every card-minimal graph G satisfying a further simple condition is reconstructible from the deck of G. However, the condition of 2-card reconstructibility used in the proof of this result appears to be purely technical, and could be replaced by a thorough analysis of G's 2-vertex-deleted subgraphs on the basis of our characterization theorem for card equivalence.

References

[1] Bartha, M., *A finite axiomatization of flowchart schemes*, Acta Cybernetica, **8**(1987), no. 2, 203–217, `http://www.cs.mun.ca/~bartha/linked/flow.pdf`.

[2] Bartha, M., *An algebraic model of synchronous systems*, Information and Computation **97**(1992), 97–131.

[3] Bartha, M., *The monoidal structure of Turing machines*, Mathematical Structures in Computer Science, **23**(2013), no. 2, 204–246.

[4] Bloom, S.L., Ésik, Z., *Axiomatizing schemes and their behaviors*, J. Comput. System Sci., **31**(1985), 375–393.

[5] Bloom, S.L., Ésik,Z., *Iteration Theories: The Equational Logic of Iterative Processes*, Springer-Verlag, 1993.

[6] Bollobás, B., *Almost every graph has reconstruction number three*, J. Graph Theory, **14**(1990), 1–4.

[7] Harary, F., *On the reconstruction of a graph from a collection of subgraphs*, In: Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963). Publ. House Czechoslovak Acad. Sci., Prague, 1964, 47–52.

[8] Harary, F., *A survey of the reconstruction conjecture*, Lecture Notes in Mathematics, **406**(1974), 18–28.

[9] Hopcroft, J., *An n log n algorithm for minimizing states in a finite automaton*, Theory of machines and computations (Proc. Internat. Sympos., Technion, Haifa, 1971), New York, Academic Press, 1971, 189–196.

[10] Joyal, A., Street, R., Verity, D., *Traced monoidal categories*, Math. Proc. Camb. Phil. Soc., **119**(1996), 447–468.

[11] Kelly, P.J., *A congruence theorem for trees*, Pacific J. Math., **7**(1957), 961–968.

[12] Lovász,L., Plummer, M.D., *Matching Theory*, North Holland, 1986.

[13] Nash-Williams, C.St.J.A., *The Reconstruction Problem*, In: L. Beineke, R.J. Wilson, eds., Selected topics in graph theory, Academic Press, 1978, 205–236.

[14] Ramachandran, S., *On a new digraph reconstruction conjecture*, Journal of Combinatorial Theory, **B31**(1981), no. 2, 143–149,

[15] Ramachandran, S., *Graph Reconstruction–Some New Developments*, AKCE J. Graphs. Combin., **1**(2004), no. 1, 51–61.

[16] Stockmeyer, P.K., *The falsity of the reconstruction conjecture for tournaments*, J. Graph Theory, **1**(1977), 19–25.

Miklós Bartha
Memorial University of Newfoundland
Department of Computer Science
St. John's, NL, A1B 3X5, Canada
e-mail: `bartha@mun.ca`

Amitesh S. Shuva
Memorial University of Newfoundland
Department of Computer Science
St. John's, NL, A1B 3X5, Canada
e-mail: `ss4545@mun.ca`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 457–462

The largest known Cunningham chain of length 3 of the first kind

Gábor Farkas, Gábor E. Gévay, Antal Járai and Emil Vatai

Abstract. Cunningham chains of length n of the first kind are n long sequences of prime numbers $p_1, p_2, \ldots, p_n$ so that $p_{i+1} = 2p_i + 1$ (for $1 \leq i < n$). In [3] we have devised a plan to find large Cunningham chains of the first kind of length 3 where the primes are of the form $p_{i+1} = (h_0 + cx) \cdot 2^{e+i} - 1$ for some integer x with $h_0 = 5\,775$, $c = 30\,030$ and $e = 34\,944$. The project was executed on the non-uniform memory access (NUMA) supercomputer of NIIF in Pécs, Hungary. In this paper we report on the obtained results and discuss the implementation details. The search consisted of two stages: sieving and the Fermat test. The sieving stage was implemented in a concurrent manner using lockfree queues, while the Fermat test was trivially parallel. On the 27th of April, 2014 we have found the largest known Cunningham chain of length 3 of the first kind which consists of the numbers $5110664609396115 \cdot 2^{34944+j} - 1$ for $j = 0, 1, 2$.

Mathematics Subject Classification (2010): 11Y11.

Keywords: Cunningham chains, primality, computational number theory.

1. Cunningham chain search project

Cunningham chains of length n of the first kind are n long sequences of prime numbers $p_1, p_2, \ldots, p_n$ so that $p_{i+1} = 2p_i + 1$ (for $1 \leq i < n$). In 2013 we set out to find the largest primes which form a Cunningham chain of the first kind of length 3. The first stage of the plan was to take a large number of candidates, each representing one chain, i.e. three primes, and eliminate most of them using a sieve similar to the sieve of Eratosthenes. In the second stage the Fermat test was used to check if the remaining candidates are probable primes. Finally, that the numbers were actually primes was proven using the OpenPFGW open source implementation of the Brillhart-Lehmer-Selfridge test.

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

The program was looking for primes of the form

$$p_{i+1} = p_{i+1}(x) = (h_0 + cx) \cdot 2^{e+i} - 1 \text{ for } i = 0, 1, 2, \tag{1.1}$$

therefore the parameters of the program were e which determined the magnitude of the primes and h_0 and c which were required to ensure the deterministic Riesel test would prove primality fast enough.

1.1. The sieve

The candidates were different values of x, and these candidates were represented by bits in a large sieve table H which was sieved with a set of primes. The size of the sieve table, denoted by h, and the upper bound of primes P were also parameters specific to the sieve program, so the values of $0 \leq x < h$ were sieved with primes $p < P$.

Sieving with a prime $p < P$ means finding the first $0 \leq x_i < h$ for which $p|p_i(x_i)$ and eliminating x_i (for $i = 1, 2, 3$). Then the candidates $x_i + kp$ (for any $k \in \mathbb{Z}$) can also be eliminated since

$$\begin{aligned} p_i(x_i + kp) &= (h_0 + c(x_i + kp)) \cdot 2^{e+i} - 1 \\ &= (h_0 + cx_i) \cdot 2^{e+i} - 1 + ckp \cdot 2^{e+i} \\ &= p_i(x_i) + ckp \cdot 2^{e+i} \end{aligned}$$

that is, $p|p_i(x_i)$ implies $p|p_i(x_i + kp)$, therefore $p_i(x_i + kp)$ is composite (for $k \in \mathbb{Z}$ and $i = 1, 2, 3$). More details can be found in [3].

1.2. Probabilistic and deterministic primality tests

For the probabilistic primality test the Fermat test was used with base 3. For each candidate x, not eliminated by the sieve, the program checked if

$$3^{p_i - 1} \equiv 1 \pmod{p_i} \text{ where } p_i = p_i(x) \text{ for } i = 1, 2, 3. \tag{1.2}$$

The probability of a false positive result of the Fermat test for all three p_i's is close to none, so practically, after finding a candidate x for which (1.2) is true, the search would be over.

However this would not prove that these numbers were prime, but this was not a problem, because the Riesel test provides a very fast and efficient way to verify the primality of the numbers of the form $k \cdot 2^e - 1$ so it can be used for (1.1).

2. Implementation

As described in [3], the above mentioned parameters were chosen as follows:

- the number of initial candidates, i.e. the size of the sieve table $H = 2^{37}$;
- the upper bound of the sieving primes $P = 2^{48}$;
- the parameters for the $p_i(x)$ polynomials were $h_0 = 5\,775$, $c = 30\,030$ and $e = 34\,944$.

2.1. The implementation of the sieve

The sieve implementation was written by Gábor E. Gévay in C++ and compiled with GCC version 4.8.1. To implement concurrent execution of the sieve program, OpenMP and lock-free queues from the Boost [1] library were used.

The primes up to $\sqrt{P} = 2^{24}$ were generated as the initial step, then each CPU core generated a "chunk" of the remaining sieving primes up to $P = 2^{48}$. The size of one chunk was 2^{31} bits, without representing the even numbers, so the effective chunk size was actually 2^{32}. Ergo for each chunk a sieve of Eratosthenes was executed on an interval of 2^{32} numbers and because all the primes up to $P = 2^{48}$ were to be sieved with, the number of chunks was $2^{16} = 65\,536$.

After generating the primes for the given chunk, each CPU core proceeded with sieving the main sieve table H representing the candidates with the primes found. Thus, several hundred cpu cores were doing bit operations on a shared 16 GB bitset at the same time, which required some synchronization. There were 32 special threads that were actually writing to the bitset. Each of these was responsible for doing operations on a 1/32th chunk of the bitset, and each had a queue (`boost::lockfree::queue`) to which the computing threads pushed the bit operations. Furthermore, each computing thread had a thread-local proxy object of the bitset, and used a method of this proxy object to request the bit operations. These objects were responsible for calculating the index of the writer thread to which the operation is to be forwarded, and also for buffering the operations to prevent the synchronization of the queue from incurring too much overhead. (Note that sequential consistency of the bit operations was not required.) The supercomputer used has a NUMA architecture. The above scheme requires remote memory accesses only for the queue operations, while both writing to the buffers in the proxy objects and executing the bit operations on the sieve table involves only local memory access. Each prime sieved three times into H, once for each polynomial $p_i(x_i)$.

Finally the sieve program converted the sieve table into a `vector` of 64bit `long long ints` and wrote them into a (binary) file.

2.2. The Fermat test

The Fermat test was written as a function with two parameters: x the candidate and i the index of the polynomial. It calculated the value of $p = p_i(x)$, and then checked if $3^{p-1} \equiv 1 \pmod{p}$. So when $3^{p-1} \bmod p = 1$ the function returned `true`, indicating that p was a probable prime, otherwise it returned `false` which meant p was certainly composite.

The output of the sieve program, containing the x candidates as `long long ints`, was the input for the Fermat test, which was implemented using the *GNU Multiple Precision (GMP)* library. Concurrent execution was not a problem: each thread read a different candidate x, calculated $p = p_1(x)$ and checked if $3^{p-1} \equiv 1 \pmod{p}$, and if the test returned `true`, the test is executed for $p_2(x)$ and if it was still `true` then for $p_3(x)$ also.

2.3. Execution

The programs were written for and executed on the supercomputers of NIIF institute [2] in Pécs. The NIIF institute provided us with access to other supercomputers, including the ones in Budapest, Debrecen and Szeged, but the supercomputer of the University in Pécs was targeted, because of it's shared memory which could hold the sieve table of $H = 2^{37}$ bits i.e. 16GiB in size.

To provide us with the advantages of C++11, we used GCC version 4.8.1. Performance of the available GMP library was suboptimal, so we compiled a newer, 5.1.2 version, which provided better performance. We also tried to use the `tuneup` utility of GMP to optimize it, but it did not improve performance.

3. Results

3.1. The first run

The sieve was executed on the 8th October 2013 and ran for about 41 hours with 352 threads. After sieving, 88 573 926 candidates were left. In [3] the number of remaining candidates after sieving was estimated using the Bateman-Horn conjecture and was to be approximately 88 570 684. The number of the actual candidates not eliminated during sieving came very close to this value, the error was only about 0.003%.

The generated output file, `fermat_in` was about 708MB in size. This file was divided into smaller parts using the `split` command to be processed by the Fermat test. The Fermat tests running on these parts of the `fermat_in` file were started immediately after the sieve program finished on the 10th of October. The estimate of the time it took to finish the Fermat tests was roughly 4 weeks, however there were some unanticipated slowdowns in the execution, which were only later solved.

On November 16th 2013. the Fermat test finished, without finding any Cunningham chains and the project came to a temporary halt.

3.2. The second run

The estimated number of Cunningham chains found (based on the above mentioned parameters) should have been ≈ 1.3, and it implied that we might have just been out of luck, i.e. we needed to continue with more candidates, with an extension of the sieve table. The project was resumed in March of 2014. The program had to be modified, because in the first run the candidates were $0 \leq x < 2^{37}$, and now the search was to be extended to the candidates $2^{37} \leq x < 3 \cdot 2^{37}$. The upper bound for the primes P was also modified from 2^{48} to 2^{50} to save a little time on the Fermat tests. The number of candidates not eliminated after sieving was 156 743 147. Using calculations similar to the ones described in [3], this value was expected to be 156 722 877. Again, the actual value was very close to the expected one, the error was only about 0.013%.

The sieve was again run on the NIIF supercomputer in Pécs, but to find the primes more quickly, the Fermat test was executed on all available computers of the NIIF institute, including the ones in Debrecen and Szeged.

On Friday, April 25th 2014. at 02:59:14 on the supercomputer in Pécs the Fermat test finished with a positive result. The result was verified with a quick reimplementation of the Fermat test in the Maple computer algebra system, and there was no mistake, we have found three probable primes for the candidate $x = 170\,185\,301\,678$. Afterward we verified the results using the OpenPFGW program, which is an open source implementation of the very fast and, most importantly deterministic, BrillhartLehmerSelfridge test for primes of the form $k \cdot 2^e - 1$.

So the largest know Cunningham chains of the first kind of length 3, according to "The Top Twenty: Cunningham Chains (1st kind)" [4], where we submitted our findings, consist of the following three primes:

$$p_{i+1}(x) = (5775 + 30030 \cdot x) \cdot 2^{34944-i} - 1 \text{ for } x = 170185301678$$

for $i = 1, 2, 3$, that is:

$$p_1(x) = 5110664609396115 \cdot 2^{34944} - 1 \qquad \text{10535 digits}$$

$$p_2(x) = 5110664609396115 \cdot 2^{34945} - 1 \qquad \text{10536 digits}$$

$$p_3(x) = 5110664609396115 \cdot 2^{34946} - 1 \qquad \text{10536 digits}$$

Acknowledgment. We are thankful to the operators at NIIF, who wrote some very useful wiki pages, which were of great help to us, and we are thankful to Gábor Kőszegi, for having the idea to use the supercomputer at the University of Pécs and helping us obtain access to it.

References

[1] *Chapter 18. Boost.Lockfree*, http://www.boost.org/doc/libs/1_55_0/doc/html/lockfree.html

[2] *National Information Infrastructure development Institute* (Nemzeti Információs Infrastruktúra fejlesztési Intézet), http://www.niif.hu/

[3] Farkas, G., Vatai, E., *Sieving for large Cunningham chains of length 3 of the first kind*, Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae Sectio Computatorica, **40**(2013), 215–222.

[4] *The Top Twenty: Cunningham Chains (1st kind)*, http://primes.utm.edu/top20/page.php?id=19

Gábor Farkas
"Eötvös Loránd" University
Faculty of Informatics
Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
e-mail: `farkasg@compalg.inf.elte.hu`

Gábor E. Gévay
"Eötvös Loránd" University
Faculty of Informatics
Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
e-mail: `ggab90@gmail.com`

Antal Járai
"Eötvös Loránd" University
Faculty of Informatics
Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
e-mail: ajarai@moon.inf.elte.hu

Emil Vatai
"Eötvös Loránd" University
Faculty of Informatics
Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
e-mail: vatai@inf.elte.hu

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 463–468

Bilateral inequalities for harmonic, geometric and Hölder means

Mira-Cristiana Anisiu and Valeriu Anisiu

Abstract. For $0 < a < b$, the harmonic, geometric and Hölder means satisfy $H < G < Q$. They are special cases $(p = -1, 0, 2)$ of power means M_p. We consider the following problem: find all $\alpha, \beta \in \mathbb{R}$ for which the bilateral inequalities

$$\alpha H(a,b) + (1-\alpha)Q(a,b) < G(a,b) < \beta H(a,b) + (1-\beta)Q(a,b)$$

hold $\forall\ 0 < a < b$. Then we replace in the bilateral inequalities the mean Q by M_p, $p > 0$ and address the same problem.

Mathematics Subject Classification (2010): 26E60.

Keywords: Means, power means, bilateral inequalities.

1. Introduction

We consider bivariate means $m : \mathbb{R}_+^2 \to \mathbb{R}$ which are symmetric $(m(b,a) = m(a,b)$ for all $a, b > 0)$ and homogeneous $(m(\lambda a, \lambda b) = \lambda m(a,b)$ for all $a, b, \lambda > 0)$.

For two means m_1 and m_2 we write $m_1 \leq m_2$ if and only if $m_1(a,b) \leq m_2(a,b)$ for every $a, b > 0$, and $m_1 < m_2$ if and only if $m_1(a,b) < m_2(a,b)$ for all $a, b > 0$ with $a \neq b$.

Since we are dealing with strict inequalities, we may and shall assume in the following that $0 < a < b$.

We consider the bivariate means

$$A(a,b) = \frac{a+b}{2};\ \ G(a,b) = \sqrt{ab};\ \ H(a,b) = \frac{2ab}{a+b};\ \ Q(a,b) = \left(\frac{a^2+b^2}{2}\right)^{1/2}; \tag{1.1}$$

$$M_p(a,b) = \begin{cases} \left(\frac{a^p+b^p}{2}\right)^{1/p}, & \text{for } p \neq 0 \\ \sqrt{ab}, & \text{for } p = 0, \end{cases} \tag{1.2}$$

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

which are known as the *arithmetic*, *geometric*, *harmonic*, *Hölder* and *power means*, respectively. Properties and comparison of standard means can be found in [3].

The means from (1.1) are comparable:

$$\min < H < G < A < Q < \max,$$

where min and max are the trivial means given by $(a, b) \mapsto \min(a, b)$ and $(a, b) \mapsto \max(a, b)$. The power means are monotonic in p, and $M_{-1} = H$, $M_0 = G$, $M_1 = A$, and $M_2 = Q$.

Recently, many bilateral inequalities between means have been proved ([1], [2], [4], [5], [6]). We mention one of them, which was the starting point for this paper, and refers to the means G, A and Q.

Theorem 1.1. [2] *The double inequality*

$$\alpha G(a, b) + (1 - \alpha)Q(a, b) < A(a, b) < \beta G(a, b) + (1 - \beta)Q(a, b), \ \forall\, 0 < a < b$$

holds if and only if $\alpha \geq 1/2$ *and* $\beta \leq 1 - \sqrt{2}/2$.

In what follows we shall prove a similar result for the means H, G and Q. Afterwards we consider the more general case of the means H, G and M_p, $p > 0$. We show that for $p = 5/2$ the auxiliary function f is still monotone and we formulate an open problem.

2. Main result

For $0 < a < b$, the geometric, harmonic and Hölder means satisfy $H < G < Q$. We shall find all the values of α and β in order that the geometric mean to be strictly between the combination of H and Q with parameters α, respectively β. Due to the homogeneity of all the means considered here, we may denote $t = b/a$, $t > 1$, and write in the following $m(t)$ instead of $m(1, t) = (1/a)\, m(a, b)$. For any three means $m_1 < m_2 < m_3$, the double inequality

$$\alpha m_1(t) + (1 - \alpha)m_3(t) < m_2(t) < \beta m_1(t) + (1 - \beta)m_3(t) \tag{2.1}$$

is equivalent to

$$\beta < f(t) < \alpha, \tag{2.2}$$

where

$$f(t) = \frac{m_3(t) - m_2(t)}{m_3(t) - m_1(t)}. \tag{2.3}$$

We shall prove the following result.

Theorem 2.1. *The double inequality*

$$\alpha H(t) + (1 - \alpha)Q(t) < G(t) < \beta H(t) + (1 - \beta)Q(t), \ \forall t > 1$$

holds if and only if $\alpha \geq 1$ *and* $\beta \leq 2/3$. *The function*

$$f_1(t) = \frac{Q(t) - G(t)}{Q(t) - H(t)}$$

is strictly increasing on $(1, \infty)$.

Proof. The function f_1 is given by

$$f_1(t) = \frac{((2t^2+2)^{1/2} - 2t^{1/2})(t+1)}{(2t^2+2)^{1/2}t + (2t^2+2)^{1/2} - 4t}. \tag{2.4}$$

We substitute $t = s^2$, $s > 1$ and get

$$f_1(s^2) = \frac{((2s^4+2)^{1/2} - 2s)(s^2+1)}{(2s^4+2)^{1/2}s^2 + (2s^4+2)^{1/2} - 4s^2}.$$

The numerator of the derivative of this expression is

$$\begin{aligned} &4\left(s^8 - 4s^7 + 2s^6 + 2(2s^4+2)^{1/2}s^4 - 2(2s^4+2)^{1/2}s^2 - 2s^2 + 4s - 1\right) \\ &= 4(s^2-1)(s^6 - 4s^5 + 3s^4 - 4s^3 + 3s^2 - 4s + 1 + 2(2s^4+2)^{1/2}s^2) \end{aligned}$$

and the denominator is obviously positive. We shall prove that

$$g_1(s) = s^6 - 4s^5 + 3s^4 - 4s^3 + 3s^2 - 4s + 1 + 2(2s^4+2)^{1/2}s^2$$

is positive for $s > 1$, hence f_1 is strictly increasing. We write $g_1(s) = 0$ as

$$s^6 - 4s^5 + 3s^4 - 4s^3 + 3s^2 - 4s + 1 = -2(2s^4+2)^{1/2}s^2, \tag{2.5}$$

square both sides and get

$$(s^8 - 4s^7 - 4s^5 + 6s^4 - 4s^3 - 4s + 1)(s-1)^4 = 0.$$

Denoting by $h_1(s) = s^8 - 4s^7 - 4s^5 + 6s^4 - 4s^3 - 4s + 1$ we get

$$h_1(s+4) = s^8 + 28s^7 + 336s^6 + 2236s^5 + 8886s^4 + 20956s^3 + 26640s^2 + 12604s - 2831,$$

which has only one change of sign. We apply Descartes' rule of signs for $h_1(s+4)$ and we obtain that the polynomial $h_1(s)$ has a single root greater than 4. We denote by $k_1(s)$ the 6th degree polynomial in the left hand side of (2.5) and get

$$k_1(s+4) = s^6 + 20s^5 + 163s^4 + 684s^3 + 1523s^2 + 1620s + 545. \tag{2.6}$$

Then the polynomial (2.6) is positive on $s > 4$, hence $g_1(s) = 0$ has no solutions on $s > 1$. It follows that f_1 is strictly increasing on $(1, \infty)$. Since $\lim_{t\to 1} f_1(t) = 2/3$ and $\lim_{t\to\infty} f_1(t) = 1$, the theorem is proved. □

We try to see if a similar result can be obtained by taking instead of $M_2 = Q$ another power mean. For $p = 5/2$ we can prove

Theorem 2.2. *The double inequality*

$$\alpha H(t) + (1-\alpha)M_{5/2}(t) < G(t) < \beta H(t) + (1-\beta)M_{5/2}(t), \ \forall t > 1$$

holds if and only if $\alpha \geq 1$ *and* $\beta \leq 5/7$. *The function*

$$f_2(t) = \frac{M_{5/2}(t) - G(t)}{M_{5/2}(t) - H(t)}$$

is strictly increasing on $(1, \infty)$.

Proof. We have

$$f_2(t) = \frac{(\frac{1}{2}t^{5/2} + \frac{1}{2})^{2/5} - t^{1/2}}{(\frac{1}{2}t^{5/2} + \frac{1}{2})^{2/5} - \frac{2t}{t+1}}. \tag{2.7}$$

By substituting $t = s^2$, $s > 1$ we get

$$f_2(s^2) = \frac{((16s^5 + 16)^{2/5} - 4s)(s^2 + 1)}{(s^2 + 1)(16s^5 + 16)^{2/5} - 8s^2}.$$

We differentiate the above function and obtain its numerator

$$32(s-1)(2s^8 - 6s^7 - 2s^6 - 2s^5 - 2s^3 - 2s^2 - 6s + 2 + s^2\,(s+1)\,(16s^5 + 16)^{2/5}),$$

the denominator being positive. We denote

$$g_2(s) = 2s^8 - 6s^7 - 2s^6 - 2s^5 - 2s^3 - 2s^2 - 6s + 2 + s^2\,(s+1)\,(16s^5 + 16)^{2/5}$$

and we write $g_2(s) = 0$ as

$$\frac{2(s^8 - 3s^7 - s^6 - s^5 - s^3 - s^2 - 3s + 1)}{s^2\,(s+1)} = -(16s^5 + 16)^{2/5}. \tag{2.8}$$

We apply the 5th power to both sides of (2.8) and get $h_2(s) = 0$, where

$$\begin{aligned} h_2(s) &= s^{30} - 10s^{29} + 25s^{28} + 20s^{27} - 50s^{26} - 196s^{25} - 150s^{24} + 320s^{23} \\ &+1305s^{22} + 2090s^{21} + 2439s^{20} + 2320s^{19} + 2550s^{18} + 3460s^{17} + 4760s^{16} \\ &+5240s^{15} + 4760s^{14} + 3460s^{13} + 2550s^{12} + 2320s^{11} + 2439s^{10} + 2090s^{9} \\ &+1305s^{8} + 320s^{7} - 150s^{6} - 196s^{5} - 50s^{4} + 20s^{3} + 25s^{2} - 10s + 1. \end{aligned}$$

Using the Sturm sequence, we obtain that $h_2(s)$ has no roots in $(1, \infty)$. It follows that $h_2(s) > 0$ on $(1, \infty)$, and the derivative of $f_2(t)$ is positive on this interval, hence $f_2(t)$ is strictly increasing. Since $\lim_{t\to 1} f_2(t) = 5/7$, $\lim_{t\to\infty} f_2(t) = 1$, the theorem is proved. □

Remark 2.3. We can consider the function

$$f_3(t) = \frac{M_p(t) - G(t)}{M_p(t) - H(t)}$$

for arbitrary $p > 0$. It is easy to check that $\lim_{t\to 1} f_3(t) = p/(p+1)$ and $\lim_{t\to\infty} f_3(t) = 1$. It remains to study the monotonicity of f_3. In the following theorem we prove that, for $p > 5/2$, the function f_3 is not monotone on $(1, \infty)$.

Theorem 2.4. *For $p > 5/2$, the infimum of the function f_3 on $(1, \infty)$ satisfies the inequality*

$$\inf_{t>1} f_3(t) < \frac{p}{p+1}.$$

Proof. Let $p > 5/2$. The function f_3 is given by

$$f_3(t) = \frac{(\frac{1}{2}t^{p} + \frac{1}{2})^{1/p} - t^{1/2}}{(\frac{1}{2}t^{p} + \frac{1}{2})^{1/p} - \frac{2t}{t+1}},$$

and after the substitution $t = s^2$, $s > 1$ we get

$$f_3(s^2) = \frac{((\frac{1}{2}s^{2p} + \frac{1}{2})^{1/p} - s)(s^2+1)}{(s^2+1)(\frac{1}{2}s^{2p} + \frac{1}{2})^{1/p} - 2s^2}.$$

The Taylor series for $s_0 = 1$ reads

$$\frac{p}{p+1} - \frac{p(2p-5)}{12(p+1)}(s-1)^2 + \frac{p(2p-5)}{12(p+1)}(s-1)^3 + O((s-1)^4), \text{ for } s \to 1$$

and its derivative will be

$$-\frac{p(2p-5)}{6(p+1)}(s-1) + O((s-1)^2).$$

It follows that the derivative is negative at least for $s > 1$ close to 1, hence f_3 decreases and $\inf\limits_{t>1} f_3(t) < p/(p+1)$. □

Based on the results in theorems 2.1 and 2.2, we formulate the following **Open problem.** Prove that the function f_3 is strictly increasing on $(1, \infty)$ for each $p \in (0, 5/2]$. Then, for each $p \in (0, 5/2]$, the double inequality

$$\alpha H(t) + (1-\alpha)M_p(t) < G(t) < \beta H(t) + (1-\beta)M_p(t), \ \forall t > 1$$

will be true if and only if $\alpha \geq 1$ and $\beta \leq p/(p+1)$.

References

[1] Alzer, H., Qiu, S.-L., *Inequalities for means in two variables*, Arch. Math. (Basel), **80**(2003), 201-215.

[2] Anisiu, M.-C., Anisiu, V., *Bilateral inequalities for means*, Revue d'Analyse Numérique et de Théorie de l'Approximation, **42**(2)(2013), 85-93.

[3] Bullen, P.S., *Handbook of Means and Their Inequalities*, 2nd edn., Mathematics and Its Applications, Springer, Berlin, 1987.

[4] Chu, Y.-M., Wang, M.-K., Gong, W.-M., *Two sharp double inequalities for Seiffert mean*, J. Inequal. Appl. 2010, 2010: Article ID 44.

[5] Neuman, E., Sándor, J., *Companion inequalities for certain bivariate means*, Appl. Anal. Discr. Math., **3**(2009), 46–51.

[6] Xia, W.-F., Chu, Y.-M., *Optimal inequalities related to the logarithmic, identric, arithmetic and harmonic means*, Revue d'Analyse Numérique et de Théorie de l'Approximation, **39**(2)(2010), 176-183.

Mira-Cristiana Anisiu
"Tiberiu Popoviciu" Institute of Numerical Analysis
Romanian Academy
P.O. Box 68, 400110 Cluj-Napoca, Romania
e-mail: `mira@math.ubbcluj.ro`

Valeriu Anisiu
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: anisiu@math.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 469–477

A note on elliptic problems on the Sierpinski gasket

Brigitte E. Breckner and Csaba Varga

Abstract. Using a method that goes back to J. Saint Raymond, we prove the existence of infinitely many weak solutions of certain nonlinear elliptic problems defined on the SG.

Mathematics Subject Classification (2010): 35J20, 28A80, 35J25, 35J60, 47J30, 49J52.

Keywords: Sierpinski gasket, weak Laplace operator, nonlinear elliptic equation on the Sierpinski gasket.

1. Introduction

In the last two decades there has been an increasing interest in studying PDEs on fractals, also motivated and stimulated by the considerable amount of literature devoted to the definition of a Laplace-type operator for functions on fractals. A particular concern has been devoted to PDEs on the Sierpinski gasket. The framework for the study of elliptic equations on the Sierpinski gasket goes back to J. Kigami's pioneering paper [4]. This paper has considerably influenced subsequent papers devoted to PDEs on the Sierpinski gasket. A list of them, including also several recent contributions, may be found in the introduction of [2].

The present paper is devoted to the nonlinear elliptic equation

$$\Delta u(x) + \gamma(x)u(x) = g(x)f(u(x)),$$

defined on the Sierpinski gasket and with zero Dirichlet boundary condition. By imposing that the nonlinearity $f\colon \mathbb{R} \to \mathbb{R}$ has an oscillating behavior at ∞, the results of the paper complete those obtained in our previous article [1], where we have studied the same problem, but under the assumption that f oscillates at 0^+. We use, as in [1], a method that goes back to J. Saint Raymond in order to prove that this problem has infinitely many weak solutions. This method has also been used to prove, in

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

the context of certain Sobolev spaces, the existence of infinitely many solutions for Dirichlet problems on bounded domains [6], for one-dimensional scalar field equations and systems [3], and for homogeneous Neumann problems [5]. The aim of the present paper is to show that the methods used in [3] can be successfully adapted to the fractal case.

Notations. We denote by $\mathbb{N}$ the set of natural numbers $\{0, 1, 2, \dots\}$, by $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$ the set of positive naturals, and by $|\cdot|$ the Euclidean norm on the spaces $\mathbb{R}^n$, $n \in \mathbb{N}^*$. The spaces $\mathbb{R}^n$ are endowed, throughout the paper, with the Euclidean topology induced by $|\cdot|$.

2. Preliminaries

We briefly recall some notations which will be used in the sequel, and refer to sections 2–4 in [1] for a more detailed presentation of these aspects. Throughout the paper, the letter V stands for the the *Sierpinski gasket* (SG for short) in $\mathbb{R}^{N-1}$, where $N \geq 2$ is a fixed natural number. There are two different approaches that lead to V, starting from given points $p_1, \dots, p_N \in \mathbb{R}^{N-1}$ with $|p_i - p_j| = 1$ for $i \neq j$, and from the similarities $S_i \colon \mathbb{R}^{N-1} \to \mathbb{R}^{N-1}$, defined by

$$S_i(x) = \frac{1}{2}\, x + \frac{1}{2}\, p_i,$$

for $i \in \{1, \dots, N\}$. While in the first approach the set V appears as the unique nonempty compact subset of $\mathbb{R}^{N-1}$ satisfying the equality

$$V = \bigcup_{i=1}^{N} S_i(V),$$

in the second one V is obtained as the closure of the set $V_* := \bigcup_{m \in \mathbb{N}} V_m$, where

$$V_0 := \{p_1, \dots, p_N\} \text{ and } V_m := \bigcup_{i=1}^{N} S_i(V_{m-1}), \text{ for } m \in \mathbb{N}^*.$$

In what follows V is considered to be endowed with the relative topology induced from the topology on $\mathbb{R}^{N-1}$. The set V_0 is called the *intrinsic boundary* of the SG. The natural measure μ associated with V is the normalized restriction of the $\frac{\ln N}{\ln 2}$-dimensional Hausdorff measure on $\mathbb{R}^{N-1}$ to the subsets of V. Thus $\mu(V) = 1$. The Lebesgue spaces $L^p(V, \mu)$, with $p \geq 1$, are equipped with the usual $||\cdot||_p$ norm.

The analog, in the case of the SG, of the Sobolev spaces is the real Hilbert space $H_0^1(V)$, equipped with the inner product $\mathcal{W} \colon H_0^1(V) \times H_0^1(V) \to \mathbb{R}$ which induces the norm $||\cdot||$ (see Section 3 in [1]). The space $H_0^1(V)$ can be compactly embedded in a space of continuous functions. More exactly, if one denotes by $C(V)$ the space of real-valued continuous functions on V, by $C_0(V) := \{u \in C(V) : \ u|_{V_0} = 0\}$, and consider both spaces being endowed with the usual supremum norm $||\cdot||_{\sup}$, then the following Sobolev-type inequality holds for $H_0^1(V)$

$$||u||_{\sup} \leq c||u||, \text{ for every } u \in H_0^1(V), \tag{2.1}$$

where c is a positive constant depending on N. Moreover, the embedding

$$(H_0^1(V), ||\cdot||) \hookrightarrow (C_0(V), ||\cdot||_{\sup}) \tag{2.2}$$

is compact.

As in [1], $\Delta\colon D \to L^2(V,\mu)$ denotes the *weak Laplacian* on V, where D is a certain linear subset of $H_0^1(V)$ which is dense in $L^2(V,\mu)$ (and dense also in $(H_0^1(V), ||\cdot||)$). Thus Δ is bijective, linear, self-adjoint, and satisfies

$$-\mathcal{W}(u,v) = \int_V \Delta u \cdot v \mathrm{d}\mu, \text{ for every } (u,v) \in D \times H_0^1(V).$$

We recall the following useful property of the space $H_0^1(V)$, stated in Lemma 3.1 of [1].

Lemma 2.1. *Let $h\colon \mathbb{R} \to \mathbb{R}$ be a Lipschitz mapping with Lipschitz constant $L \geq 0$ and such that $h(0) = 0$. Then, for every $u \in H_0^1(V)$, we have $h \circ u \in H_0^1(V)$ and $||h \circ u|| \leq L \cdot ||u||$.*

3. The main results

Let $\gamma, g \in L^1(V,\mu)$ and let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. We are concerned with the following nonlinear elliptic problem, with zero Dirichlet boundary condition, on the SG

$$(P) \begin{cases} \Delta u(x) + \gamma(x)u(x) = g(x)f(u(x)), \ \forall\, x \in V \setminus V_0, \\ \\ u|_{V_0} = 0. \end{cases}$$

We recall from [1] that a function $u \in H_0^1(V)$ is called a *weak solution of* (P) if

$$\mathcal{W}(u,v) - \int_V \gamma(x)u(x)v(x)\mathrm{d}\mu + \int_V g(x)f(u(x))v(x)\mathrm{d}\mu = 0, \ \forall\, v \in H_0^1(V).$$

Define $F\colon \mathbb{R} \to \mathbb{R}$ by

$$F(t) = \int_0^t f(\xi)\mathrm{d}\xi. \tag{3.1}$$

We know from Proposition 5.3 in [1] that the functional $T\colon H_0^1(V) \to \mathbb{R}$, given, for every $u \in H_0^1(V)$, by

$$T(u) = \frac{1}{2}||u||^2 - \frac{1}{2}\int_V \gamma(x)u^2(x)\mathrm{d}\mu + \int_V g(x)F(u(x))\mathrm{d}\mu, \tag{3.2}$$

is Fréchet differentiable on $H_0^1(V)$, and that it is an energy functional of problem (P), i.e., $u \in H_0^1(V)$ is a weak solution of (P) if and only if u is a critical point of T.

Remark 3.1. Assume that $\gamma \leq 0$ and $g \leq 0$ a.e. in V. Consider $u \in H_0^1(V)$ and $d, b \in \mathbb{R}$ such that $d \leq u(x) \leq b$ for every $x \in V$. According to the fact that $g \leq 0$ a.e. in V, we then have

$$\int_V g(x)F(u(x))\mathrm{d}\mu \geq \max_{s\in[d,b]} F(s) \cdot \int_V g(x)\mathrm{d}\mu. \tag{3.3}$$

We state, for later use, the following relations about the functional $T\colon H_0^1(V) \to \mathbb{R}$ defined by (3.2): The inequalities (3.3) and $\gamma \le 0$ a.e. in V imply that

$$T(u) \ge \max_{s\in[d,b]} F(s) \cdot \int_V g(x)\mathrm{d}\mu \tag{3.4}$$

and

$$\frac{1}{2}||u||^2 \le T(u) - \max_{s\in[d,b]} F(s) \cdot \int_V g(x)\mathrm{d}\mu. \tag{3.5}$$

We recall the definition of the coercivity of a functional, respectively, the subsequent standard result concerning the existence of minimum points of sequentially weakly lower semicontinuous functionals.

Definition 3.2. Let X be a real normed space and let M be a nonempty subset of X. A functional $L\colon M \to \mathbb{R}$ is said to be *coercive* if, for every sequence (x_n) in M such that $\lim_{n\to\infty} ||x_n|| = \infty$, it follows that $\lim_{n\to\infty} L(x_n) = \infty$.

Proposition 3.3. *Let X be a reflexive real Banach space, M a nonempty sequentially weakly closed subset of X, and $L\colon M \to \mathbb{R}$ a sequentially weakly lower semicontinuous and coercive functional. Then L possesses at least one minimum point.*

We derive now from Proposition 3.3 the following key result for our approach.

Proposition 3.4. *Let $\gamma, g \in L^1(V,\mu)$ be so that $\gamma \le 0$ and $g \le 0$ a.e. in V, let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous, and let $a, b, c, d \in \mathbb{R}$ be so that $d < c < 0 < a < b$. Furthermore, assume that the map F, defined by* (3.1)*, satisfies the conditions*

$$F(s) \le F(c), \forall\ s \in [d,c], \tag{3.6}$$

and

$$F(s) \le F(a), \forall\ s \in [a,b]. \tag{3.7}$$

Denoting by

$$M := \{u \in H_0^1(V) \mid d \le u(x) \le b,\ \forall x \in V\},$$

there exists an element $u \in H_0^1(V)$ with the properties:

(i) $T(u) = \inf T(M)$,

(ii) $c \le u(x) \le a$, *for every* $x \in V$,

where the functional $T\colon H_0^1(V) \to \mathbb{R}$ is defined by (3.2)*.*

Proof. Obviously the set M is non-empty (it contains the constant 0 function) and convex. Since the inclusion (2.2) is continuous, M is closed in the norm topology on $H_0^1(V)$. It follows that M is also closed in the weak topology on $H_0^1(V)$, thus M is sequentially weakly closed. It follows from (3.5) that the restriction of T to M is coercive. Proposition 3.3 implies now that there exists $\tilde{u} \in M$ such that $T(\tilde{u}) = \inf T(M)$. Define $h\colon \mathbb{R} \to \mathbb{R}$ by

$$h(t) = \begin{cases} c,\ t < c \\ t,\ t \in [c,a] \\ a,\ t > a. \end{cases}$$

Note that $h(0) = 0$ and that h is a Lipschitz map with Lipschitz constant $L = 1$. According to Lemma 2.1, the map $u := h \circ \tilde{u}$ belongs to $H_0^1(V)$ and

$$||u|| \leq ||\tilde{u}||. \tag{3.8}$$

Moreover, u belongs to M and obviously satisfies condition (ii) to be proved. We next show that (i) also holds true. For this set

$$V_1 := \{x \in V \mid \tilde{u}(x) < c\}, \quad V_2 := \{x \in V \mid \tilde{u}(x) > a\}.$$

Then

$$u(x) = \begin{cases} c, \ x \in V_1 \\ \tilde{u}(x), \ x \in V \setminus (V_1 \cup V_2) \\ a, \ x \in V_2. \end{cases}$$

It follows that

$$u^2(x) \leq \tilde{u}^2(x), \ \text{for every } x \in V. \tag{3.9}$$

Furthermore, if $x \in V_1$, then $\tilde{u}(x) \in [d, c[$, hence, by (3.6), $F(\tilde{u}(x)) \leq F(c) = F(u(x))$. Analogously, if $x \in V_2$, then (3.7) yields $F(\tilde{u}(x)) \leq F(a) = F(u(x))$. Thus

$$F(\tilde{u}(x)) \leq F(u(x)), \ \text{for every } x \in V. \tag{3.10}$$

The inequalities (3.8), (3.9) and (3.10) imply, together with the fact that $\gamma \leq 0$ and $g \leq 0$ a.e. in V, that

$$\begin{aligned} T(\tilde{u}) - T(u) &= \frac{1}{2}||\tilde{u}||^2 - \frac{1}{2}||u||^2 - \frac{1}{2}\int_V \gamma(x)(\tilde{u}^2(x) - u^2(x))\mathrm{d}\mu \\ &+ \int_V g(x)(F(\tilde{u}(x)) - F(u(x)))\mathrm{d}\mu \geq 0. \end{aligned}$$

Thus $T(\tilde{u}) \geq T(u)$. Since $T(\tilde{u}) = \inf T(M)$ and since $u \in M$, we conclude that $T(u) = \inf T(M)$, thus (i) is also fulfilled. □

The main result of the paper is contained in the following theorem concerning the existence of multiple weak solutions of problem (P).

Theorem 3.5. *Assume that the following conditions hold:*

(C1) $\gamma \in L^1(V, \mu)$ *and* $\gamma \leq 0$ *a.e. in* V*.*

(C2) $f \colon \mathbb{R} \to \mathbb{R}$ *is continuous such that*

 (1*) *there exist two sequences* (a_k) *and* (b_k) *in* $]0, \infty[$ *with* $a_k < b_k < b_{k+1}$, $\lim_{k\to\infty} b_k = \infty$ *and such that* $f(s) \leq 0$ *for every* $s \in [a_k, b_k]$,

 (2*) *there exist reals* $d < c < 0$ *with* $f(s) \geq 0$ *for every* $s \in [d, c]$.

(C3) $F \colon \mathbb{R} \to \mathbb{R}$*, defined by* (3.1)*, is such that*

 (3*) $-\infty < \liminf_{s\to\infty} \frac{F(s)}{s^2}$,

 (4*) $\limsup_{s\to\infty} \frac{F(s)}{s^2} = \infty$.

(C4) $g \colon V \to \mathbb{R}$ *is continuous, not identically* 0*, and with* $g \leq 0$.

Then there exists a sequence (u_k) *of pairwise distinct weak solutions of problem* (P) *such that* $\lim_{k\to\infty} \|u_k\| = \infty$.

Remark 3.6. According to Example 4 in [3], the function $f\colon \mathbb{R} \to \mathbb{R}$, defined by $f(s) = s^2 \sin^2 s - 1$, satisfies the conditions (C2) and (C3) of Theorem 3.5.

In what follows we assume that the conditions (C1)–(C4) in the hypotheses of Theorem 3.5 are satisfied. For every $k \in \mathbb{N}$ set now

$$M_k := \{u \in H_0^1(V) \mid d \leq u(x) \leq b_k, \ \forall\, x \in V\}. \tag{3.11}$$

The proof of Theorem 3.5 includes the following main steps contained in the next results:

1. we show that the functional $T\colon H_0^1(V) \to \mathbb{R}$, defined by (3.2), has at least one critical point in each of the sets M_k,
2. since T is an energy functional of Problem (P), each of these critical points is a weak solution of Problem (P),
3. we show that there are infinitely many pairwise distinct such weak solutions.

Lemma 3.7. *For every $k \in \mathbb{N}$, there is an element $u_k \in M_k$ such that the following conditions hold:*

(i) $T(u_k) = \inf T(M_k)$,
(ii) $c \leq u_k(x) \leq a_k$, *for every $x \in V$.*

Proof. Note that, while condition (1^*) in the hypotheses of Theorem 3.5 yields

$$F(s) \leq F(a_k), \forall\, s \in [a_k, b_k],$$

condition (2^*) implies (3.6). Applying Proposition 3.4, we finish the proof. □

Lemma 3.8. *For every $k \in \mathbb{N}$, let $u_k \in M_k$ be a function satisfying the conditions* (i) *and* (ii) *of Lemma* 3.7*. The functional T has then in u_k a local minimum (with respect to the norm topology on $H_0^1(V)$), for every $k \in \mathbb{N}$. In particular, (u_k) is a sequence of weak solutions of problem (P).*

Proof. Fix $k \in \mathbb{N}$. Suppose to the contrary that u_k is not a local minimum of T. This implies the existence of a sequence (w_n) in $H_0^1(V)$ converging to u_k in the norm topology such that

$$T(w_n) < T(u_k), \text{ for every } n \in \mathbb{N}.$$

In particular, $w_n \notin M_k$, for all $n \in \mathbb{N}$. Choose a real number ε such that

$$0 < \varepsilon \leq \frac{1}{2} \min\{b_k - a_k, c - d\}.$$

In view of (2.1) the sequence (w_n) converges to u_k in the supremum norm topology on $C(V)$. Hence there is an index $m \in \mathbb{N}$ such that

$$||w_m - u_k||_{\sup} \leq \varepsilon.$$

For every $x \in V$ we then have, according to condition (ii) of Lemma 3.7,

$$w_m(x) = w_m(x) - u_k(x) + u_k(x) \leq \varepsilon + u_k(x) \leq \frac{b_k - a_k}{2} + a_k < b_k$$

and

$$w_m(x) = w_m(x) - u_k(x) + u_k(x) \geq -\varepsilon + u_k(x) \geq \frac{d - c}{2} + c > d.$$

Thus $w_m \in M_k$, a contradiction. We conclude that T has in u_k a local minimum, so u_k is a critical point of T. The last assertion of the lemma follows now from the fact that T is an energy functional of problem (P). □

Lemma 3.9. *For every $k \in \mathbb{N}$, put $\gamma_k := \inf T(M_k)$. Then $\lim_{k\to\infty} \gamma_k = -\infty$.*

Proof. Observe first that the inclusions $M_k \subseteq M_{k+1}$, for all $k \in \mathbb{N}$, imply that the sequence (γ_k) is decreasing.

Condition (C4) in Theorem 3.5 yields the existence of a nonempty open subset U of $V \setminus V_0$ such that $g|_U < 0$. By the same arguments as those used in the proof of statement (2.1) in [1] we may conclude that there exists a compact set $K \subseteq U$ with $\mu(K) > 0$. Hence we get that

$$\int_K g(x)\mathrm{d}\mu < 0. \tag{3.12}$$

We show next that we can find a function $v \in H_0^1(V)$ such that

$$0 \le v \le 1 \text{ and } v|_K = 1. \tag{3.13}$$

Indeed, by Urysohn's Lemma, there exists a continuous function $\phi\colon V \to [0,1]$ such that $\phi(x) = 0$, for $x \in V_0$, and $\phi(x) = 1$, for $x \in K$. According to Theorem 1.4.4 in [7], there exists a function $u \in H_0^1(V)$ with $||\phi - u||_{\sup} < 1$. In particular, $u(x) \neq 0$ for all $x \in K$. Hence $|u(x)| > 0$ for every $x \in K$. Note that $|u| \in H_0^1(V)$, by Lemma 2.1. Let

$$\xi := \min_{x\in K} |u(x)|.$$

Then $\xi > 0$. Define $h\colon \mathbb{R} \to \mathbb{R}$ by $h(t) = \min\{t, \xi\}$. Since h is a Lipschitz map with $h(0) = 0$, Lemma 2.1 yields that $h \circ |u| \in H_0^1(V)$. We have that $(h \circ |u|)(x) = \xi$ for every $x \in K$. Thus $v := \frac{1}{\xi}(h \circ |u|)$ satisfies (3.13).

By condition (3*) in the requirements of Theorem 3.5, there exist $m \in \mathbb{R}$ and $\delta > 0$ such that

$$m \le \frac{F(s)}{s^2}, \text{ for all } s > \delta.$$

Denote by $\widetilde{m} := \min\{F(s) - ms^2 \mid s \in [0,\delta]\}$. In particular, $\widetilde{m} \le 0$. So we obtain that

$$\widetilde{m} + ms^2 \le F(s), \text{ for all } s \ge 0. \tag{3.14}$$

Condition (4*) in the hypotheses of Theorem 3.5 implies the existence of a sequence (r_n) of positive reals with

$$\lim_{n\to\infty} r_n = \infty \text{ and } \lim_{n\to\infty} \frac{F(r_n)}{r_n^2} = \infty. \tag{3.15}$$

Using (3.2) and (3.13), we compute, for every $n \in \mathbb{N}$,

$$\begin{aligned} T(r_n v) =& \frac{1}{2} r_n^2 ||v||^2 - \frac{r_n^2}{2} \int_V \gamma(x) v^2(x)\mathrm{d}\mu + F(r_n) \int_K g(x)\mathrm{d}\mu \\ &+ \int_{V\setminus K} g(x) F(r_n v(x))\mathrm{d}\mu. \end{aligned}$$

On the other hand, by (3.14) and the fact that $g \leq 0$, we get

$$\int_{V\setminus K} g(x)F(r_n v(x))\mathrm{d}\mu \leq \widetilde{m}\int_{V\setminus K} g(x)\mathrm{d}\mu + mr_n^2\int_{V\setminus K} g(x)v^2(x)\mathrm{d}\mu.$$

Thus

$$\begin{aligned}\frac{T(r_n v)}{r_n^2} \leq & \frac{\|v\|^2}{2} - \frac{1}{2}\int_V \gamma(x)v^2(x)\mathrm{d}\mu + \frac{F(r_n)}{r_n^2}\int_K g(x)\mathrm{d}\mu \\ & + \frac{\widetilde{m}}{r_n^2}\int_{V\setminus K} g(x)\mathrm{d}\mu + m\int_{V\setminus K} g(x)v^2(x)\mathrm{d}\mu.\end{aligned} \tag{3.16}$$

Involving (3.12) and (3.15), we obtain from (3.16) that $\lim\limits_{n\to\infty}\dfrac{T(r_n v)}{r_n^2} = -\infty$, so

$$\lim_{n\to\infty} T(r_n v) = -\infty. \tag{3.17}$$

Recall from condition (1*) in the statement of Theorem 3.5 that $\lim\limits_{k\to\infty} b_k = \infty$. Thus we may find a subsequence (b_{k_n}) of the sequence (b_k) such that $r_n \leq b_{k_n}$, for every $n \in \mathbb{N}$. Since $0 \leq v \leq 1$, we get that

$$0 \leq r_n v \leq b_{k_n}, \text{ for all } n \in \mathbb{N}.$$

By (3.11), we hence conclude that $r_n v \in M_{k_n}$, for every $n \in \mathbb{N}$, so

$$\gamma_{k_n} \leq T(r_n v), \text{ for all } n \in \mathbb{N}.$$

In view of (3.17) we thus obtain that $\lim\limits_{n\to\infty}\gamma_{k_n} = -\infty$. Since (γ_k) is decreasing we finally conclude that $\lim\limits_{k\to\infty}\gamma_k = -\infty$. □

Proof of Theorem 3.5 *concluded*. From Lemma 3.8 we know that there is a sequence (u_k) of weak solutions of problem (P) such that $u_k \in M_k$ and $\gamma_k = T(u_k)$, where $\gamma_k = \inf T(M_k)$, for every natural k. Assume, by contradiction, that $\lim\limits_{k\to\infty}\|u_k\| \neq \infty$. Then there exists a bounded subsequence (u_{k_n}) of the sequence (u_k). According to (2.1) and to the fact that $\lim\limits_{k\to\infty} b_k = \infty$, we may find $p \in \mathbb{N}$ such that $u_{k_n} \in M_p$, for every $n \in \mathbb{N}$. This yields that $\gamma_p \leq \gamma_{k_n}$, for every $n \in \mathbb{N}$, contradicting the statement of Lemma 3.9. Thus $\lim\limits_{k\to\infty}\|u_k\| = \infty$. Hence we can find a subsequence (u_{k_j}) of the sequence (u_k) consisting of pairwise distinct elements. □

References

[1] Breckner, B.E., Rădulescu, V. and Varga, Cs., *Infinitely many solutions for the Dirichlet problem on the Sierpinski gasket*, Anal. Appl. (Singap.), **9**(2011), 235-248.

[2] Breckner, B.E. and Varga, Cs., *Multiple solutions of Dirichlet problems on the Sierpinski gasket*, to apper in J. Optim. Theory Appl., DOI 10.1007/s10957-013-0368-7.

[3] Faraci, F. and Kristály, A., *One-dimensional scalar field equations involving an oscillatory nonlinear term*, Discrete Contin. Dyn. Syst., **18**(2007), no. 1, 107-120.

[4] Kigami, J., *A harmonic calculus on the Sierpinski spaces*, Japan J. Appl. Math., **6**(1989), 259-290.

[5] Kristály, A. and Motreanu, D., *Nonsmooth Neumann-type problems involving the p-Laplacian*, Numer. Funct. Anal. Optim., **28**(2007), no. 11-12, 1309-1326.

[6] Saint Raymond, J., *On the multiplicity of solutions of the equation* $-\triangle u = \lambda \cdot f(u)$, J. Differential Equations, **180**(2002), 65-88.

[7] Strichartz, R.S., *Differential Equations on Fractals. A Tutorial*, Princeton University Press, Princeton, NJ, 2006.

Brigitte E. Breckner
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: brigitte@math.ubbcluj.ro

Csaba Varga
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: csvarga@cs.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 479–495

On systems of semilinear hyperbolic functional equations

László Simon

Abstract. We consider a system of second order semilinear hyperbolic functional differential equations where the lower order terms contain functional dependence on the unknown function. Existence of solutions for $t \in (0,T)$ and $t \in (0,\infty)$, further, examples and some qualitative properties of the solutions in $(0,\infty)$ are shown.

Mathematics Subject Classification (2010): 35L71, 35L53, 35B40.

Keywords: Semilinear hyperbolic equations, functional partial differential equations, second order hyperbolic systems, qualitative properties of solutions.

1. Introduction

In the present work we shall consider weak solutions of initial-boundary value problems of the form

$$u_j''(t) + Q_j(u(t)) + \varphi(x)D_j h(u(t)) + H_j(t,x;u) + G_j(t,x;u,u') = F_j, \tag{1.1}$$

$$t > 0, \quad x \in \Omega, \quad j = 1,...,N$$

$$u(0) = u^{(0)}, \quad u'(0) = u^{(1)} \tag{1.2}$$

where $\Omega \subset \mathbb{R}^n$ is a bounded domain and we use the notations $u(t) = (u_1(t),...,u_N(t))$, $u(t) = (u_1(t,x),...,u_N(t,x))$, $u' = (u_1',...,u_N') = D_t u = (D_t u_1,...,D_t u_N)$, $u'' = D_t^2 u$, Q_j is a linear second order symmetric elliptic differential operator in the variable x; h is a C^1 function having certain polynomial growth, H_j and G_j contain nonlinear functional (non-local) dependence on u and u', with some polynomial growth.

There are several papers on semilinear hyperbolic differential equations, see, e.g., [3], [4], [10], [14] and the references there. Semilinear hyperbolic functional equations were studied, e.g. in [5], [6], [7], with certain non-local terms, generally in the form of particular integral operators containing the unknown function. First order quasilinear evolution equations with non-local terms were considered, e.g., in [13] and [15], second

---

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

order quasilinear evolution equations with non-local terms were considered in [11], by using the theory of monotone type operators (see [2], [9], [16]).

This work was motivated by the classical book [9] of J.L. Lions on nonlinear PDEs where a single equation was considered in a particular case (semilinear hyperbolic differential equation). We shall use ideas of the above work.

Semilinear hyperbolic functional equations were considered in a previous work of the author (see [12]).

2. Existence in $(0,T)$

Denote by $\Omega \subset \mathbb{R}^n$ a bounded domain with sufficiently smooth boundary, and let $Q_T = (0,T) \times \Omega$. Denote by $W^{1,2}(\Omega)$ the Sobolev space with the norm

$$\|u\| = \left[\int_\Omega \left(\sum_{j=1}^n |D_j u|^2 + |u|^2 \right) dx \right]^{1/2} .$$

Further, let $V_j \subset W^{1,2}(\Omega)$ be closed linear subspaces of $W^{1,2}(\Omega)$, $V_j^\star$ the dual space of V_j, $V = (V_1, ..., V_N)$, $V^\star = (V_1^\star, ..., V_N^\star)$, $H = L^2(\Omega) \times ... \times L^2(\Omega)$, the duality between $V_j^\star$ and V_j (and between $V^\star$ and V) will be denoted by $\langle \cdot, \cdot \rangle$, the scalar product in $L^2(\Omega)$ and H will be denoted by $(\cdot, \cdot)$. Denote by $L^2(0,T;V_j)$ and $L^2(0,T;V)$ the Banach space of measurable functions $u : (0,T) \to V_j$, $u : (0,T) \to V$, respectively, with the norm

$$\|u_j\|_{L^2(0,T;V_j)} = \left[\int_0^T \|u_j(t)\|_{V_j}^2 dt \right]^{1/2}, \quad \|u\|_{L^2(0,T;V)} = \left[\int_0^T \|u(t)\|_V^2 dt \right]^{1/2},$$

respectively.

Similarly, $L^\infty(0,T;V_j)$, $L^\infty(0,T;V)$, $L^\infty(0,T;L^2(\Omega))$, $L^\infty(0,T;H)$ is the set of measurable functions $u_j : (0,T) \to V_j$, $u : (0,T) \to V$, $u_j : (0,T) \to L^2(\Omega)$, $u : (0,T) \to H$, respectively, with the $L^\infty(0,T)$ norm of the functions $t \mapsto \|u_j(t)\|_{V_j}$, $t \mapsto \|u(t)\|_V$, $t \mapsto \|u_j(t)\|_{L^2(\Omega)}$, $t \mapsto \|u(t)\|_H$, respectively.

Now we formulate the assumptions on the functions in (1.1).

(A_1). $Q : V \to V^\star$ is a linear continuous operator defined by

$$\langle Q(u), v \rangle = \sum_{j=1}^N \langle Q_j(u), v_j \rangle = \sum_{j=1}^N \left[\sum_{k=1}^N \langle Q_{jk}(u_k), v_j \rangle \right],$$

$$u = (u_1, ..., u_N), \quad v = (v_1, ..., v_N),$$

where $Q_{jk} : W^{1,2}(\Omega) \to [W^{1,2}(\Omega)]^\star$ are continuous linear operators satisfying

$$\langle Q_{jk}(u_k), v_j \rangle = \langle Q_{jk}(v_j), u_k \rangle, \quad Q_{jk} = Q_{kj}, \text{ thus } \langle Q(u), v \rangle = \langle Q(v), u \rangle$$

for all $u, v \in V$ and

$$\langle Q(u), u \rangle \geq c_0 \|u\|_V^2 \text{ with some constant } c_0 > 0.$$

(A_2). $\varphi : \Omega \to \mathbb{R}$ is a measurable function satisfying

$$c_1 \leq \varphi(x) \leq c_2 \text{ for a.a. } x \in \Omega$$

with some positive constants c_1, c_2.

(A_3). $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function satisfying

$$h(\eta) \geq 0, \quad |D_j h(\eta)| \leq \mathrm{const}|\eta|^\lambda \text{ for } |\eta| > 1 \text{ where}$$

$$1 < \lambda \leq \lambda_0 = \frac{n}{n-2} \text{ if } n \geq 3, \quad 1 < \lambda < \infty \text{ if } n = 2.$$

(A_3'). $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function satisfying with some positive constants c_3, c_4

$$h(\eta) \geq c_3|\eta|^{\lambda+1}, \quad |D_j h(\eta)| \leq c_4|\eta|^\lambda \text{ for } |\eta| > 1, \quad n \geq 3 \text{ where } \lambda > \lambda_0 = \frac{n}{n-2},$$

$$|D_j h(\eta)| \leq c_4|\eta|^\lambda \quad \text{for } |\eta| > 1, \quad n = 2 \text{ where } 1 < \lambda < \infty.$$

(A_4). $H_j : Q_T \times [L^2(Q_T)]^N \to \mathbb{R}$ are functions for which $(t, x) \mapsto H_j(t, x; u)$ is measurable for all fixed $u \in H$, H_j has the Volterra property, i.e. for all $t \in [0, T]$, $H_j(t, x; u)$ depends only on the restriction of u to $(0, t)$; the following inequality holds for all $t \in [0, T]$ and $u \in H$:

$$\int_\Omega |H_j(t, x; u)|^2 dx \leq c^\star \left[\int_0^t \int_\Omega h(u(\tau)) dx d\tau + \int_\Omega h(u) dx \right].$$

Finally, $(u^{(k)}) \to u$ in $[L^2(Q_T)]^N$ and $(u^{(k)}) \to u$ a.e. in Q_T imply

$$H_j(t, x; u^{(k)}) \to H_j(t, x; u) \text{ for a.a. } (t, x) \in Q_T.$$

(A_5). $G_j : Q_T \times [L^2(Q_T)]^N \times L^\infty(0, T; H) \to \mathbb{R}$ is a function satisfying: $(t, x) \mapsto G_j(t, x; u, w)$ is measurable for all fixed $u \in [L^2(Q_T)]^N$, $w \in L^\infty(0, T; H)$, G_j has the Volterra property: for all $t \in [0, T]$, $G_j(t, x; u, w)$ depends only on the restriction of u, w to $(0, t)$ and

$$G_j(t, x; u, u') = \varphi_j(t, x; u)u_j'(t) + \psi_j(t, x; u, u')$$

where

$$\varphi_j \geq 0, \quad |\varphi_j(t, x; u)| \leq \mathrm{const} \tag{2.1}$$

if (A_3) is satisfied.

(A_5') If (A_3') is satisfied, we assume instead of the second inequality in (2.1)

$$\int_\Omega |\varphi_j(t, x; u)|^2 dx \leq \mathrm{const} \left[\int_{Q_t} |u|^{2\mu} d\tau dx + \int_\Omega |u|^{2\mu} dx \right] \tag{2.2}$$

where $\mu \leq \frac{n+1}{n-1} \frac{\lambda-1}{\lambda+1}$.

Further, on ψ_j we assume

$$\int_\Omega |\psi_j(t, x; u, u')|^2 dx \leq c_1 + c_2 \int_{Q_t} |u'|^2 dx d\tau$$

with some constants c_1, c_2.

Further, if $(u^{(\nu)}) \to u$ in $[L^2(Q_T)]^N$ then

$$\varphi_j(t, x; u^{(\nu)}) \to \varphi_j(t, x; u) \text{ for a.a. } (t, x) \in Q_T$$

and if

$$(u^{(\nu)}) \to u \text{ in } [L^2(Q_T)]^N \text{ and a.e. in } Q_T, \quad (w^{(\nu)}) \to w$$

weakly in $L^\infty(0,T;H)$ in the sense that for all fixed $g_1 \in L^1(0,T;H)$

$$\int_0^T \langle g_1(t), w^{(\nu)}(t)\rangle dt \to \int_0^T \langle g_1(t), w(t)\rangle dt,$$

then for a.a. $(t,x) \in Q_T$

$$\psi_j(t,x;u^{(\nu)},w^{(\nu)}) \to \psi_j(t,x;u,w).$$

Theorem 2.1. *Assume (A_1), (A_2), (A_3), (A_4), (A_5). Then for all $F \in L^2(0,T;H)$, $u^{(0)} \in V$, $u^{(1)} \in H$ there exists $u \in L^\infty(0,T;V)$ such that*

$$u' \in L^\infty(0,T;H), \quad u'' \in L^2(0,T;V^\star),$$

u satisfies the system (1.1) in the sense: for a.a $t \in [0,T]$, all $v \in V$

$$\langle u_j''(t), v_j\rangle + \langle Q_j(u(t)), v_j\rangle + \int_\Omega \varphi(x) D_j h(u(t)) v_j dx + \int_\Omega H_j(t,x;u) v_j dx + \tag{2.3}$$
$$\int_\Omega G_j(t,x;u,u') v_j dx = (F_j(t), v_j) \quad j = 1,...,N$$

and the initial condition (1.2) is fulfilled.

If (A_1), (A_2), (A_3'), (A_4), (A_5) are satisfied then for all $F \in L^2(0,T;H)$, $u^{(0)} \in V \cap [L^{\lambda+1}(\Omega)]^N$, $u^{(1)} \in H$ there exists $u \in L^\infty(0,T;V \cap [L^{\lambda+1}(\Omega)]^N)$ such that

$$u' \in L^\infty(0,T;H),$$
$$u'' \in L^2(0,T;V^\star) + L^\infty(0,T;[L^{\frac{\lambda+1}{\lambda}}(\Omega)]^N) \subset L^2\left(0,T;[V \cap (L^{\lambda+1}(\Omega))^N]^\star\right)$$

and u satisfies (1.1) in the sense: for a.a $t \in [0,T]$, all $v_j \in V_j \cap L^{\lambda+1}(\Omega)$ (2.3) holds, further, the initial condition (1.2) is fulfilled.

Proof. We apply Galerkin's method. Let $w_1^{(j)}, w_2^{(j)}, \ldots$ be a linearly independent system in V_j if (A_3) is satisfied and in $V_j \cap L^{\lambda+1}(\Omega)$ if (A_3') is satisfied such that the linear combinations are dense in V_j and $V_j \cap L^{\lambda+1}(\Omega)$, respectively. We want to find the m-th approximation of u in the form

$$u_j^{(m)}(t) = \sum_{l=1}^m g_{lm}^{(j)}(t) w_l^{(j)} \quad (j = 1,2,\ldots,N) \tag{2.4}$$

where $g_{lm}^{(j)} \in W^{2,2}(0,T)$ if (A_3) holds and $g_{lm}^{(j)} \in W^{2,2}(0,T) \cap L^\infty(0,T)$ if (A_3') holds such that

$$\langle (u_j^{(m)})''(t), w_k^{(j)}\rangle + \langle Q(u^{(m)}(t)), w_k^{(j)}\rangle + \int_\Omega \varphi(x) D_j h(u^{(m)}(t)) w_k^{(j)} dx \tag{2.5}$$
$$+ \int_\Omega H_j(t,x;u^{(m)}) w_k^{(j)} dx + \int_\Omega G_j(t,x;u^{(m)},(u^{(m)})') w_k^{(j)} dx = \langle F_j(t), w_k^{(j)}\rangle,$$
$$k = 1,\ldots,m, \quad j = 1,\ldots,N$$

$$u_j^{(m)}(0) = u_{j0}^{(m)}, \quad (u_j^{(m)})'(0) = u_{j1}^{(m)} \tag{2.6}$$

where $u_{j0}^{(m)}$, $u_{j1}^{(m)}$ $(j = 1,2,\ldots,N)$ are linear combinations of $w_1^{(j)}, w_2^{(j)} \ldots, w_m^{(j)}$ satisfying

$$(u_{j0}^{(m)}) \to u_j^{(0)} \text{ in } V_j \text{ and } V_j \cap L^{\lambda+1}(\Omega), \text{ respectively, as } m \to \infty \text{ and} \tag{2.7}$$

$$(u_{j1}^{(m)}) \to u_j^{(1)} \text{ in } H \text{ as } m \to \infty. \tag{2.8}$$

It is not difficult to show that all the conditions of the existence theorem for a system of functional differential equations with Carathéodory conditions are satisfied.

Thus, by using the Volterra property of G and H, we obtain that there exists a solution of (2.5), (2.6) in a neighbourhood of 0 (see [8]). Further, the maximal solution of (2.5), (2.6) is defined in $[0, T]$. Indeed, multiplying (2.5) by $[g_{lm}^{(j)}]'(t)$ and taking the sum with respect to j, and k we obtain

$$\begin{gathered}\langle (u^{(m)})''(t), (u^{(m)})'(t)\rangle + \langle Q(u^{(m)}(t)), (u^{(m)})'(t)\rangle \\ + \int_\Omega \varphi(x) \frac{d}{dt}[h(u^{(m)}(t))]dx \\ + \int_\Omega (H(t, x; u^{(m)}), (u^{(m)})'(t))dx + \int_\Omega (G(t, x; u^{(m)}, (u^{(m)})'), (u^{(m)})'(t))dx \\ = \langle F(t), (u^{(m)})'(t)\rangle.\end{gathered} \tag{2.9}$$

Integrating the above equality over $(0, t)$ we find (see, e.g., [16], [12])

$$\begin{gathered}\frac{1}{2}\|(u^{(m)})'(t)\|_H^2 + \frac{1}{2}\langle Q(u^{(m)}(t)), u^{(m)}(t)\rangle + \int_\Omega \varphi(x) h(u^{(m)}(t))dx \\ + \int_0^t \left[\int_\Omega (H(\tau, x; u^{(m)}), (u^{(m)})')dx\right] d\tau + \int_0^t \left[\int_\Omega (G(\tau, x; u^{(m)}, (u^{(m)})'), (u^{(m)})')dx\right] d\tau \\ = \int_0^t \left[\langle F(\tau), (u^{(m)})'(\tau)\rangle\right] d\tau.\end{gathered} \tag{2.10}$$

Hence, by using Young's inequality, Sobolev's imbedding theorem and the assumptions of our theorem, we obtain

$$\begin{gathered}\|(u^{(m)})'(t)\|_H^2 + \int_\Omega h(u^{(m)}(t))dx + \|u^{(m)}(t)\|_V^2 \\ \le \text{const}\left\{1 + \int_0^t \left[\|(u^{(m)})'(\tau)\|_H^2 + \int_\Omega h(u^{(m)}(\tau))dx\right] d\tau\right\}\end{gathered}$$

where the constant is not depending on t and m. Thus by Gronwall's lemma

$$\|(u^{(m)})'(t)\|_H^2 + \int_\Omega h(u^{(m)}(t))dx \le \text{const} \tag{2.11}$$

and thus

$$\|u^{(m)}(t)\|_V^2 \le \text{const} \tag{2.12}$$

Further, the estimates (2.11), (2.12) hold for all $t \in [0, T]$ and all m and in the case $\lambda > \lambda_0$, $n \ge 3$

$$\|u^{(m)}(t)\|_{V \cap [L^{\lambda+1}(\Omega)]^N} \le \text{const}. \tag{2.13}$$

By (2.11), (2.12), if (A_3) is satisfied, there exist a subsequence of $(u^{(m)})$, again denoted by $(u^{(m)})$ and $u \in L^\infty(0, T; V)$ such that

$$(u^{(m)}) \to u \text{ weakly in } L^\infty(0, T; V), \tag{2.14}$$

$$(u^{(m)})' \to u' \text{ weakly in } L^\infty(0, T; H) \tag{2.15}$$

in the following sense: for any fixed $g \in L^1(0,T;V^\star)$ and $g_1 \in L^1(0,T;H)$

$$\int_0^T \langle g(t), u^{(m)}(t)\rangle dt \to \int_0^T \langle g(t), u(t)\rangle dt,$$

$$\int_0^T (g_1(t), (u^{(m)})'(t))dt \to \int_0^T (g_1(t), u'(t))dt.$$

Similarly, in the case $\lambda > \lambda_0$, $n \geq 3$, (when (A_3') holds) there exist subsequence of $(u^{(m)})$ and $u \in L^\infty(0,T;V \cap [L^{\lambda+1}(\Omega)]^N)$ such that

$$(u^{(m)}) \to u \text{ weakly in } L^\infty(0,T;V \cap [L^{\lambda+1}(\Omega)]^N), \tag{2.16}$$

which means: for any fixed $g \in L^1(0,T;(V \cap L^{\lambda+1}(\Omega))^\star)$

$$\int_0^T \langle g(t), u^{(m)}(t)\rangle dt \to \int_0^T \langle g(t), u(t)\rangle dt.$$

Since the imbedding $W^{1,2}(\Omega)$ into $L^2(\Omega)$ is compact, by (2.14) – (2.16) we have for a subsequence

$$(u^{(m)}) \to u \text{ in } L^2(0,T;H) = [L^2(Q_T)]^N \text{ and a.e. in } Q_T. \tag{2.17}$$

(see, e.g., [9]). Finally, we show that the limit function u is a solution of problem (1.1), (1.2).

As $Q : V \to V^\star$ is a linear and continuous operator, by (2.14) for all $v \in V$ and $v \in V \cap \left[L^{\lambda+1}(\Omega)\right]^N$, respectively we have

$$\langle (Q(u^{(m)}m)(t)), v\rangle \to \langle (Q(u(t)), v\rangle \text{ weakly in } L^\infty(0,T) \tag{2.18}$$

and by (2.15)

$$\langle (u^{(m)})''(t), v\rangle = \frac{d}{dt}\langle (u^{(m)})'(t), v\rangle \to \langle u''(t), v\rangle \tag{2.19}$$

with respect to the weak convergence of the space of distributions $D'(0,T)$.

Further, by (2.17) and the continuity of $D_j h$

$$\varphi(x) D_j h(u_m(t)) \to \varphi(x) D_j h(u(t)) \text{ for a.e. } (t,x) \in Q_T. \tag{2.20}$$

Now we show that for any fixed

$$v \in L^2(0,T;V), \quad v \in L^2(0,T;V) \cap L^1(0,T;(L^{\lambda+1}(\Omega))^N),$$

respectively, the sequence of functions

$$\varphi(x) D_j h(u^{(m)}(t)) v \quad j = 1, \dots, N \tag{2.21}$$

is equiintegrable in Q_T. Indeed, if (A_3) is satisfied then by Sobolev's imbedding theorem and (2.12) for all $t \in [0,T]$

$$\|\varphi(x) D_j h(u^{(m)}(t))\|^2_{L^2(\Omega)} \leq \text{const}\|D_j h(u^{(m)}(t))\|^2_{L^2(\Omega)}$$

$$\leq \text{const}\left[1 + \int_\Omega |u^{(m)}(t)|^{2\lambda_0} dx\right] \leq \text{const}\left[1 + \|u_m(t)\|_V^{2\lambda_0}\right] \leq \text{const},$$

because $2\lambda_0 = \frac{2n}{n-2}$ and $W^{1,2}(\Omega)$ is continuously imbedded into $L^{\frac{2n}{n-2}}(\Omega)$, thus Cauchy–Schwarz inequality implies that the sequence of functions (2.21) is equiintegrable in Q_T.

If (A_3') is satisfied then for all $t \in [0,T]$

$$\int_\Omega |\varphi(x) D_j h(u^{(m)}(t))|^{\frac{\lambda+1}{\lambda}} dx \le \text{const} \int_\Omega [h(u^{(m)}(t)) + 1] dx \le \text{const}$$

thus Hölder's inequality implies that the sequence (2.21) is equiintegrable in Q_T. Consequently, by (2.20) and Vitali's theorem we obtain that for any fixed

$$v \in L^2(0,T;V), \quad v \in L^2(0,T;V) \cap L^1(0,T;L^{\lambda+1}(\Omega)),$$

respectively

$$\lim_{m\to\infty} \int_{Q_T} \varphi(x) D_j h(u^{(m)}(t)) v_j dtdx = \int_{Q_T} \varphi(x) D_j h(u(t)) v_j dtdx \tag{2.22}$$

and

$$\varphi(x) D_j h(u(t)) \in L^2(0,T;V^\star), \quad \varphi(x) D_j h(u(t)) \in L^\infty(0,T;L^{\frac{\lambda+1}{\lambda}}(\Omega)) \tag{2.23}$$

if (A_3), (A_3') holds, respectively.

Further, by (2.17) and (A_4)

$$H_j(t,x;u^{(m)}) \to H_j(t,x;u) \text{ a.e. in } Q_T \tag{2.24}$$

and by (2.11)

$$\int_{Q_T} |H_j(t,x;u_m)|^2 dxdt \le \text{const} \int_{Q_T} h(u_m(t)) dxdt \le \text{const},$$

hence, by Cauchy–Schwarz inequality, for any fixed $v \in L^2(0,T;V)$, the sequence of functions $H_j(t,x;u^{(m)})v_j$ is equiintegrable in Q_T $(j = 1,\dots,N)$, thus by (2.24) and Vitali's theorem

$$\lim_{m\to\infty} \int_{Q_T} H_j(t,x;u^{(m)}) v_j dtdx = \int_{Q_T} H_j(t,x;u) v_j dtdx \tag{2.25}$$

and

$$H(t,x;u) \in L^2(0,T;V^\star).$$

Similarly, (2.15) – (2.17) and (A_5) imply

$$\psi_j(t,x;u^{(m)},(u^{(m)})') \to \psi_j(t,x;u,u') \text{ a.e. in } Q_T \tag{2.26}$$

and for arbitrary $v \in L^2(0,T;V)$ the sequence of functions $\psi_j(t,x;u^{(m)},(u^{(m)})')v_j$ is equintegrable in Q_T by Cauchy – Schwarz inequality, because by (2.11)

$$\int_{Q_T} |\psi_j(t,x;u^{(m)},(u^{(m)})')|^2 dtdx \le \text{const} \left[1 + \int_{Q_T} |(u^{(m)})'|^2 dx\right] dt \le \text{const}.$$

Consequently, Vitali's theorem implies that for $j = 1,\dots,N$

$$\lim_{m\to\infty} \int_{Q_T} \psi_j(t,x;u^{(m)},(u^{(m)})') v_j dtdx = \int_{Q_T} \psi_j(t,x;u,u') v dtdx \tag{2.27}$$

and

$$\psi_j(t,x;u,u') \in L^2(0,T;V^\star).$$

Further, by using Vitali's theorem, we show that for arbitrary fixed $v \in L^2(0,T;V)$

$$\varphi_j(t,x;u^{(m)}) v_j \to \varphi_j(t,x;u) v_j \text{ in } L^2(Q_T), \quad j = 1,\dots,N. \tag{2.28}$$

Indeed, by (A_5) and (2.17)

$$\varphi_j(t,x;u^{(m)}) \to \varphi_j(t,x;u) \text{ for a.e. } (t,x)\in Q_T, \quad j=1,\dots,N. \tag{2.29}$$

Further, by (A_5) $|\varphi_j(t,x;u^{(m)}))|^2$ is bounded and so for fixed $v \in L^2(0,T;V)$ the sequence

$$\int_{Q_T} |\varphi_j(t,x;u^{(m)})v_j - \varphi_j(t,x;u)v_j|^2 dtdx \le \text{const}|v_j|^2$$

is equiintegrable which implies with (2.29) by Vitali's theorem (2.28). Consequently, by (2.15) we obtain

$$\lim \int_{Q_T} \varphi_j(t,x;u^{(m)})(u^{(m)})'(t)v_j dtdx = \int_{Q_T} \varphi_j(t,x;u)u'(t)v_j dtdx, \quad j=1,\dots,N \tag{2.30}$$

and $\varphi(t,x;u)u' \in L^2(0,T;V^\star)$.

If (A_5') (and (A_3')) is satisfied, then for a fixed $v \in L^2(0,T;V)\cap[L^{\lambda+1}(Q_T)]^N$ we also have

$$\varphi_j(t,x;u^{(m)})v_j \to \varphi_j(t,x;u)v_j \text{ in } L^2(Q_T), \quad j=1,\dots,N. \tag{2.31}$$

Indeed, by (2.11), (2.12) $(u^{(m)})$ is bounded in $W^{1,2}(Q_T)$, hence it is bonded in $L^{\frac{2(n+1)}{n-1}}(Q_T)$. Thus Hölder's inequality implies for any measurable $M \subset Q_T$

$$\int_M |\varphi_j(t,x;u^{(m)})v_j - \varphi_j(t,x;u)v_j|^2 dtdx \tag{2.32}$$

$$\le \text{const}\left\{\int_{Q_T} [|u^{(m)}|^{2\mu} + |u^{(m)}|^{2\mu}]^{q_1} dtdx\right\}^{1/q_1} \cdot \left\{\int_M |v_j|^{2p_1}\right\}^{1/p_1}$$

$$\le \text{const}\left\{\int_M |v_j|^{2p_1}\right\}^{1/p_1}$$

where

$$2p_1 = \lambda+1, \quad \frac{1}{p_1}+\frac{1}{q_1},$$

thus

$$2\mu q_1 = 2\mu\frac{p_1}{p_1-1} = 2\mu\frac{\lambda+1}{\lambda-1} \le \frac{2(n+1)}{n-1}$$

since

$$\mu \le \frac{n+1}{n-1}\cdot\frac{\lambda-1}{\lambda+1},$$

hence (2.29), (2.32) and Vitali's theorem imply (2.31). Consequently, by (2.15) we obtain (2.30) (when (A_5') holds).

Now let

$$v=(v_1,\dots,v_N)\in V \text{ and } \chi_j \in C_0^\infty(0,T) \quad (j=1,\dots,N)$$

be arbitrary functions. Further, let $z_j^M = \sum_{l=1}^M b_{lj} w_l^{(j)}$, $b_{lj} \in \mathbb{R}$ be sequences of functions such that

$$(z_j^M) \to v_j \text{ in } V_j \text{ and } V_j \cap L^{\lambda+1}(\Omega), \quad j=1,\dots,N, \tag{2.33}$$

respectively, as $M \to \infty$. Further, by (2.5) we have for all $m \geq M$

$$\int_0^T \langle -(u_j^{(m)})'(t), z_j^M \rangle \chi_j'(t)dt + \int_0^T \langle Q(u^{(m)}(t)), z_j^M \rangle \chi_j(t)dt \tag{2.34}$$
$$+ \int_0^T \int_\Omega \varphi(x) D_j h(u^{(m)}(t)) z_j^M \chi_j(t)dtdx + \int_0^T \int_\Omega H_j(t,x;u^{(m)}) z_j^M \chi_j(t)dtdx$$
$$+ \int_0^T \int_\Omega G_j(t,x;u^{(m)},(u^{(m)})') z_j^M \chi_j(t)dtdx$$
$$= \int_0^T \langle F_j(t), z_j^M \rangle \chi_j(t)dt.$$

By (2.15), (2.18), (2.22), (2.25), (2.27), (2.30) we obtain from (2.34) as $m \to \infty$

$$-\int_0^T \langle u_j'(t), z_j^M \rangle \chi_j'(t)dt + \int_0^T \langle Q_j(u(t)), z_j^M \rangle \chi_j(t)dt \tag{2.35}$$
$$+ \int_0^T \int_\Omega \varphi(x) D_j h(u(t)) z_j^M \chi_j(t)dtdx$$
$$+ \int_0^T \int_\Omega H_j(t,x;u) z_j^M \chi_j(t)dtdx + \int_0^T \int_\Omega G_j(t,x;u,u') z_j^M \chi_j(t)dtdx$$
$$= \int_0^T \langle F_j(t), z_j^M \rangle \chi(t)dt.$$

From equality (2.35) and (2.33) we obtain as $M \to \infty$

$$-\int_0^T \langle u_j'(t), v_j \rangle \chi_j'(t)dt + \int_0^T \langle Q_j(u(t)), v_j \rangle \chi_j(t)dt \tag{2.36}$$
$$+ \int_0^T \int_\Omega \varphi(x) D_j h(u(t)) v_j \chi_j(t)dtdx$$
$$+ \int_0^T \int_\Omega H_j(t,x;u) v_j \chi_j(t)dtdx + \int_0^T \int_\Omega G_j(t,x;u,u') v_j \chi_j(t)dtdx$$
$$= \int_0^T \langle F_j(t), v_j \rangle \chi_j(t)dt.$$

Since $v_j \in V_j$ and $\chi_j \in C_0^\infty(0,T)$ are arbitrary functions, (2.36) means that

$$u_j'' \in L^2(0,T;V_j^\star) \text{ and } u_j'' \in L^2(0,T;(V \cap L^{\lambda+1}(\Omega))^\star), \tag{2.37}$$

respectively (see, e.g. [16]) and for a.a. $t \in [0,T]$

$$u_j'' + Q_j(u(t)) + \varphi(x) D_j h(u(t)) + H_j(t,x;u) + G_j(t,x;u,u') = F_j, \quad j = 1, \dots, N, \tag{2.38}$$

i.e. we proved (1.1).

Now we show that the initial condition (1.2) holds. Since $u \in L^\infty(0,T;V)$, $u' \in L^\infty(0,T;H)$, we have $u \in C([0,T];H)$ and for arbitrary $\chi_j \in C^\infty[0,T]$ with the properties $\chi_j(0) = 1$, $\chi_j(T) = 0$, all j, k

$$\int_0^T \langle u_j'(t), w_k^{(j)} \rangle \chi_j(t)dt = -(u_j(0), w_k^{(j)})_{L^2(\Omega)} - \int_0^T \langle u_j(t), w_k^{(j)} \rangle \chi_j'(t)dt,$$

$$\int_0^T \langle (u_j^{(m)})'(t), w_k^{(j)}\rangle \chi_j(t)dt = -(u_j^{(m)}(0), w_k^{(j)})_{L^2(\Omega)} - \int_0^T \langle u_j^{(m)}(t), w_k^{(j)}\rangle \chi_j'(t)dt.$$

Hence by (2.6), (2.7), (2.8), (2.14), (2.15), we obtain as $m \to \infty$

$$(u^{(0)}, w_k^{(j)})_{L^2(\Omega)} = \lim_{m\to\infty} (u_{j0}^{(m)}, w_k^{(j)})_{L^2(\Omega)}$$

$$= \lim_{m\to\infty} (u_j^{(m)}(0), w_k^{(j)})_{L^2(\Omega)} = (u_j(0), w_k^{(j)})_{L^2(\Omega)}$$

for all j and k which implies $u(0) = u^{(0)}$.

Similarly can be shown that $u'(0) = u^{(1)}$.

3. Examples

Let the operator Q be defined by

$$\langle Q_{jk}(u_k), v_j\rangle = \int_\Omega \left[\sum_{i,l=1}^n a_{il}^{jk}(x)(D_l u_k)(D_i v_j) + d^{jk}(x)u_k v_j\right] dx$$

where $a_{il}^{jk}, d^{jk} \in L^\infty(\Omega)$, $a_{il}^{jk} = a_{li}^{jk}$, $\sum_{i,l=1}^n a_{il}^{jj}(x)\xi_i\xi_l \ge c_1|\xi|^2$, $d^{ii}(x) \ge c_0$ with some positive constants c_0, c_1; further, $a_{il}^{jk} = a_{il}^{kj}$ and for some $\tilde{c}_0 < c_1$

$$\|a_{il}^{jk}\|_{L^\infty(\Omega)} < \frac{\tilde{c}_0}{n-1}, \quad \|d^{jk}\|_{L^\infty(\Omega)} < \frac{\tilde{c}_0}{n-1} \text{ for } j \ne k.$$

Then assumption (A_1) is satisfied.

If h is a C^1 function such that $h(\eta) = |\eta|^{\lambda+1}$ if $|\eta| > 1$ then (A_3), (A_3'), respectively, are satisfied.

Further, let $\tilde{h}_j : \mathbb{R}^N \to \mathbb{R}$ be continuous functions satisfying

$$|\tilde{h}_j(\eta)| \le \text{const } |\eta|^{\frac{\lambda+1}{2}} \text{ for } |\eta| > 1, \quad j = 1, \dots, N$$

with some positive constant. It is not difficult to show that operators H_j defined by one of the formulas

$$H_j(t,x;u) = \chi_j(t,x)\tilde{h}_j\left(\int_{Q_t} u_1(\tau,\xi)d\tau d\xi, \dots, \int_{Q_t} u_N(\tau,\xi), d\tau d\xi\right),$$

$$H_j(t,x;u) = \chi_j(t,x)\tilde{h}_j\left(\int_0^t u_1(\tau,x)d\tau, \dots, \int_0^t u_N(\tau,x)d\tau\right),$$

$$H_j(t,x;u) = \chi_j(t,x)\tilde{h}_j\left(\int_\Omega u_1(t,\xi)d\xi, \dots, \int_\Omega u_N(t,\xi)d\xi\right),$$

$$H_j(t,x;u) = \chi_j(t,x)\tilde{h}_j(u_1(\tau_1(t),x), \dots, u_N(\tau_k(t),x)) \text{ where}$$

$$\tau_k \in C^1, \quad 0 \le \tau_k(t) \le t, \quad \tau_k'(t) \ge c_1 > 0, \quad k = 1, \dots, N$$

satisfy (A_4) if $\chi_j \in L^\infty(Q_T)$.

The operators φ_j, ψ_j may have forms, similar to the above forms of H_j with bounded continuous functions $\tilde{h}_j$. Then (A_5) is fulfilled.

Remark. One can show uniqueness and continuous dependence of the solution of (1.1), (1.2) if the following additional conditions are satisfied:

$$G_j(t,x;u,u') = \tilde{\varphi}_j(x)u_j'(t)$$

where $\tilde{\varphi}_j$ is measurable and $0 \le \tilde{\varphi}_j(x) \le \text{const}$, h is twice continuously differentiable and

$$|D_i D_k h(\eta)| \le \text{const}|\eta|^{\lambda-1} \text{ for } |\eta| > 1.$$

Further $H_j(t,x;u)$ satisfy some Lipschitz condition with respect to u.

4. Solutions in $(0,\infty)$

Now we formulate and prove existence of solutions for $t \in (0,\infty)$. Denote by $L^p_{loc}(0,\infty;V)$ the set of functions $u : (0,\infty) \to V$ such that for each fixed finite $T > 0$, their restrictions to $(0,T)$ satisfy $u|_{(0,T)} \in L^p(0,T;V)$ and let $Q_\infty = (0,\infty)\times\Omega$, $L^\alpha_{loc}(Q_\infty)$ the set of functions $u : Q_\infty \to \mathbb{R}^N$ such that $u_j|_{Q_T} \in L^\alpha(Q_T)$ $(j = 1,\dots,N)$ for any finite T.

Now we formulate assumptions on H_j and G_j.

(B_4) The functions $H_j : Q_\infty \times \left[L^2_{loc}(Q_\infty)\right]^N \to \mathbb{R}$ are such that for all fixed $u \in \left[L^2_{loc}(Q_\infty)\right]^N$ the functions $(t,x) \mapsto H_j(t,x;u)$ are measurable, H_j have the Volterra property (see (A_4)) and for each fixed finite $T > 0$, the restrictions of H_j to $Q_T \times \left[L^2(Q_T)\right]^N$ satisfy (A_4).

Remark. Since H_j has the Volterra property, this restriction H_j^T is well defined by the formula

$$H_j^T(t,x;\tilde{u}) = H_j(t,x;u), \quad (t,x) \in Q_T, \quad \tilde{u} \in [L^2(Q_T)]^N$$

where $u \in [L^2_{loc}(Q_\infty)]^N$ may be any function satisfying $u(t,x) = \tilde{u}(t,x)$ for $(t,x) \in Q_T$.

(B_5) The operators

$$G_j : Q_\infty \times [L^2_{loc}(Q_\infty)]^N \times L^\infty_{loc}(0,\infty;H) \to \mathbb{R}$$

are such that for all fixed $u \in L^2_{loc}(0,\infty;V)$, $w \in L^\infty_{loc}(0,\infty;H)$ the functions $(t,x) \mapsto G_j(t,x;u,w)$ are measurable, G_j have the Volterra property and for each fixed finite $T > 0$, the restrictions G_j^T of G_j to $Q_T \times [L^2(Q_T)]^N \times L^\infty(0,T;H)$ satisfy (A_5).

(B_5') It is the same as (B_5) but G_j^T satisfy (A_5').

Theorem 4.1. *Assume (A_1) – (A_3), (B_4), (B_5). Then for all $F \in L^2_{loc}(0,\infty;H)$, $u^{(0)} \in V$, $u^{(1)} \in H$ there exists*

$$u \in L^\infty_{loc}(0,\infty;V) \text{ such that } u' \in L^\infty_{loc}(0,\infty;H), \quad u'' \in L^2_{loc}(0,\infty;V^\star),$$

u satisfies (1.1) for a.a. $t \in (0,\infty)$ (in the sense, formulated in Theorem 2.1) and the initial condition (1.2).

If (A_1), $A_2)$, (A_3'), (B_4), (B_5) are fulfilled then for all $F \in L^2_{loc}(0,\infty;H)$, $u^{(0)} \in V \cap [L^{\lambda+1}(\Omega)]^N$, $u^{(1)} \in H$ there exists

$$u \in L^\infty_{loc}(0,\infty;V \cap [L^{\lambda+1}(\Omega)]^N) \text{ such that } u' \in L^\infty_{loc}(0,\infty;H),$$

$$u'' \in L^2_{loc}(0,\infty;V^\star) + L^\infty_{loc}(0,\infty;[L^{\frac{\lambda+1}{\lambda}}(\Omega)]^N) \subset L^2_{loc}(0,\infty;[V \cap (L^{\lambda+1}(\Omega))^N]^\star),$$

u satisfies (1.1) for a.a. $t \in (0,\infty)$ (in the sense, formulated in Theorem 2.1) and the initial condition (1.2).

Assume that the following additional conditions are satisfied: there exist T_0 and a function $\gamma \in L^2(T_0,\infty)$ such that for $t > T_0$

$$|G(t,x;u,u')| \le \gamma(t), |H(t,x;u)| \le \gamma(t) \text{ and } \|F(t)\|_{V^\star} \le \gamma(t). \tag{4.1}$$

Then for the above solution u we have

$$u \in L^\infty(0,\infty;V), \quad u \in L^\infty(0,\infty;V \cap [L^{\lambda+1}(\Omega)]^N), \text{ respectively and} \tag{4.2}$$

$$u' \in L^\infty(0,\infty;H).$$

Further, assume that there exists a positive constant $\tilde{c}$ such that

$$\varphi_j(t,x;u) \ge \tilde{c}, \quad (t,x) \in Q_\infty, \quad j = 1,\dots,N \tag{4.3}$$

and there exist $F_\infty \in H$, $u_\infty \in V$ such that

$$Q(u_\infty) = F_\infty, \quad F - F_\infty \in L^2(0,\infty;H), \tag{4.4}$$

$$|H_j(t,x;u)| \le \beta(t,x), \quad |\psi_j(t,x;u,u')| \le \beta(t,x), \quad |\varphi_j(t,x;u)| \le const \tag{4.5}$$

with some $\beta \in L^2(0,\infty;L^2(\Omega))$. Then for the above solution we have

$$u \in L^\infty(0,\infty;V), \quad u \in L^\infty(0,\infty;v \cap [L^{\lambda+1}(\Omega)]^N), \tag{4.6}$$

$$\|u'(t)\|_H \le const\ e^{-\tilde{c}t}, \quad t \in (0,\infty) \tag{4.7}$$

and there exists $w^{(0)} \in H$ such that

$$u(T) \to w^{(0)} \text{ in } H \text{ as } T \to \infty, \quad \|u(T) - w^{(0)}\|_H \le const\ e^{-\tilde{c}T}. \tag{4.8}$$

Finally, $w^{(0)} \in V$ and

$$Q(w^{(0)}) + \varphi Dh(w^{(0)}) = F_\infty. \tag{4.9}$$

Proof. Similarly to the proof of Theorem 2.1, we apply Galerkin's method and we want to find the m-th approximation of solution $u = (u_1,\dots,u_N)$ for $t \in (0,\infty)$ in the form (see (2.4))

$$u_j^{(m)}(t) = \sum_{l=1}^m g_{lm}^{(j)}(t) w_l^{(j)}, \quad j = 1,\dots,N$$

where $g_{lm}^{(j)} \in W^{2,2}_{loc}(0,\infty)$ if (A_3) is satisfied and $g_{lm}^{(j)} \in W^{2,2}_{loc}(0,\infty) \cap L^\infty_{loc}(0,\infty)$ if (A_3') is satisfied. Here $W^{2,2}_{loc}(0,\infty)$ and $L^\infty_{loc}(0,\infty)$ denote the set of functions $g : (0,\infty) \to \mathbb{R}$ such that for all T the restriction of g to $(0,T)$ belongs to $W^{2,2}(0,T)$, $L^\infty(0,T)$, respectively.

According to the arguments in the proof of Theorem 2.1, there exists a solution of (2.5), (2.6) in a neighbourhood of $t = 0$. Further, we obtain estimates (2.11), (2.12) and (2.13), respectively, for $t \in [0,T]$ with sufficiently small T where on the right hand side are finite constants (depending on T). Consequently, the maximal solutions of (2.5), (2.6) are defined in $(0,\infty)$ and the estimates (2.11), (2.12), (2.13) hold for all finite $T > 0$ (if $t \in [0,T]$), the constants on the right hand sides are depending only on T.

Let $(T_k)_{k\in\mathbb{N}}$ be a monotone increasing sequence, converging to $+\infty$. According to the arguments in the proof of Theorem 2.1, there is a subsequence $(u^{(m1)})$ of $(u^{(m)})$ for which (2.14), (2.15) and (2.16) hold, respectively, with $T = T_1$. Further, there is a subsequence $(u^{(m2)})$ of $(u^{(m1)})$ for which (2.14), (2.15) and (2.16) hold, respectively, with $T = T_2$, etc. By a diagonal process we obtain a sequence $(u^{(mm)})_{m\in\mathbb{N}}$ such that (2.14), (2.15), (2.16) hold for every fixed $T > 0$; further,

$$u \in L^\infty_{loc}(0,\infty;V), \quad u' \in L^\infty_{loc}(0,\infty;H), \quad u'' \in L^2_{loc}(0,\infty;V^\star) \text{ and}$$
$$u \in L^\infty_{loc}(0,\infty;V \cap [L^{\lambda+1}(\Omega)]^N), \quad u' \in L^\infty_{loc}(0,\infty;H),$$
$$u'' \in L^2_{loc}(0,\infty;V^\star) + L^\infty_{loc}(0,\infty;[L^{\frac{\lambda+1}{\lambda}}(\Omega)]^N),$$

respectively and (1.1) holds for $t \in (0,\infty)$.

Now we consider the case when (4.1) holds. Then by (2.10) we obtain for all $t \geq T_1 \geq T_0$

$$\frac{1}{2}\|(u^{(m)})'(t)\|^2_H + \frac{1}{2}\langle(Q(u^{(m)})(t), u^{(m)}(t)\rangle + c_1\int_\Omega h(u^{(m)}(t))dx$$
$$\leq \int_0^{T_1}\int_\Omega |\langle G(\tau,x;u^{(m)},(u^{(m)})'),(u^{(m)})'(\tau)\rangle|d\tau + \int_0^{T_1}\int_\Omega |\langle H(\tau,x;u^{(m)}),(u^{(m)})'(\tau)\rangle|d\tau$$
$$+\int_0^{T_1}\int_\Omega |\langle F(\tau),(u^{(m)})'(\tau)\rangle|d\tau + 3\lambda(\Omega)\left[\int_{T_1}^\infty |\gamma(\tau)|d\tau\right]\sup_{\tau\in[0,t]}\|(u^{(m)})'(\tau)\|_H.$$

Choosing sufficiently large $T_1 > 0$, since $\lim_{T_1\to\infty}\int_{T_1}^\infty |\gamma(\tau)|d\tau = 0$, we find

$$\frac{1}{4}\|(u^{(m)})'(t)\|^2_H + \frac{1}{2}\langle Q(u^{(m)}(t)), u^{(m)}(t)\rangle + c_1\int_\Omega h(u^{(m)}(t)dx \leq \text{const}$$

for all $t > 0$, m which implies (4.2).

Finally, consider the case when (4.3) – (4.5) are satisfied, too. Denoting $u^{(mm)}$ by $u^{(m)}$, for simplicity, by (2.9), $Qu_\infty = F_\infty$ we obtain for $w_m = u_m - u_\infty$ (since $(w^{(m)})' = (u^{(m)})'$):

$$\langle(w^{(m)})''(t),(w^{(m)})'(t)\rangle + \langle(Qw^{(m)})(t),(w^{(m)})'(t)\rangle + \int_\Omega \varphi(x)\frac{d}{dt}[h(u^{(m)}(t))]dx \quad (4.10)$$
$$+\int_\Omega (H(t,x;u^{(m)}),(w^{(m)})'(t))dx + \int_\Omega (G(t,x;u^{(m)},(u^{(m)})'),(w^{(m)})'(t)dx$$
$$= \langle F(t) - F_\infty,(w^{(m)})'(t)\rangle.$$

Integrating over $[0,t]$ we find (similarly to (2.10))

$$\frac{1}{2}\|(w^{(m)})'(t)\|^2_H + \frac{1}{2}\langle Q(w^{(m)}(t)), w^{(m)}(t)\rangle + c_1\int_\Omega h(u^{(m)}(t))dx \quad (4.11)$$
$$+\tilde{c}\int_0^t\left[\int_\Omega |(w^{(m)})'(\tau)|^2dx\right]d\tau$$
$$\leq \varepsilon\int_0^t\left[\int_\Omega |(w^{(m)})'(\tau)|^2dx\right]d\tau + C(\varepsilon)\int_0^t \|F(\tau) - F_\infty\|^2_H d\tau$$

$$+\frac{1}{2}\|(u^{(m)})'(0)\|_H^2+\frac{1}{2}\langle(Qu^{(m)})(0),u^{(m)}(0)\rangle+c_2\int_\Omega h(u^{(m)}(0))dx$$
$$+\varepsilon\int_0^t\left[\int_\Omega|(w^{(m)})'(\tau)|dx\right]d\tau+C(\varepsilon)\|\beta\|_{L^2(0,\infty;H)}.$$

Choosing $\varepsilon=\tilde{c}/4$ we obtain

$$\int_0^t\left[\int_\Omega|(w^{(m)})'(\tau)|^2dx\right]d\tau\le\text{const.} \tag{4.12}$$

Further, from (4.11), (4.12) we obtain

$$\|(u^{(m)})'(t)\|_H^2+\tilde{c}\int_0^t\|(u^{(m)})'(\tau)\|_H^2d\tau\le c^\star$$

with some positive constant $c^\star$ not depending on m and t. Thus by Gronwall's lemma we find

$$\|(u^{(m)})'(t)\|_H^2=\|(w^{(m)})'(t)\|_H^2\le c^\star\mathrm{e}^{-\tilde{c}t},\quad t>0$$

which implies (4.7) as $m\to\infty$ (since $(u^{(m)})'\to u'$ weakly in $L^\infty(0,T;H)$). Further, by (A_1) one obtains from (4.11) that for all $t>0$, m

$$\|w^{(m)}(t)\|_V\le\text{const},\quad\|w^{(m)}(t)\|_{V\cap[L^{\lambda+1}(\Omega)]^N}\le\text{const},$$

respectively, which implies (4.6).

Further, for arbitrary $T_1<T_2$

$$\|u(T_2)-u(T_1)\|_H^2=(u(T_2),u(T_2)-u(T_1))_H-(u(T_1),u(T_2)-u(T_1))_H$$
$$=\int_{T_1}^{T_2}\langle u'(t),u(T_2)-u(T_1)\rangle dt=\int_{T_1}^{T_2}(u'(t),u(T_2)-u(T_1))_Hdt$$
$$\le\|u(T_2)-u(T_1)\|_H\int_{T_1}^{T_2}\|u'(t)\|_Hdt$$

which implies

$$\|u(T_2)-u(T_1)\|_H\le\int_{T_1}^{T_2}\|u'(t)\|_Hdt. \tag{4.13}$$

Hence by (4.7)

$$\|u(T_2)-u(T_1)\|_H\to0\text{ as }T_1,T_2\to\infty$$

which implies (4.8) and by (4.10), (4.7) we obtain

$$\|u(T)-w_0\|_H\le\int_T^\infty\|u'(t)\|_Hdt\le\text{const }\mathrm{e}^{-\tilde{c}T}.$$

Now we show $w_0\in V$ and (4.9) holds. Since $u\in L^\infty(0,\infty;V)$,

$$(u(T_k))\to w_0^\star\text{ weakly in }V,\quad w_0^\star\in V \tag{4.14}$$

for some sequence (T_k), $\lim(T_k)=+\infty$. Clearly, (4.14) implies

$$(u(T_k))\to w_0^\star\text{ weakly in }H,$$

thus by (4.8) $w_0=w_0^\star\in V$ and (4.14) holds for arbitrary sequence (T_k) converging to $+\infty$.

In order to prove (4.9), consider arbitrary fixed $v \in V$, $v \in V \cap [L^{\lambda+1}(\Omega)]^N$, respectively and

$$\chi_T(t) = \chi(t-T) \text{ where } \chi \in C_0^\infty(\mathbb{R}),\ \operatorname{supp}\chi \subset [0,1],\quad \int_0^1 \chi(t)dt = 1.$$

Multiply (2.3) by $\chi_T(t)$ and integrate with respect to t on $(0,\infty)$ and take the sum with respect to j, then we obtain

$$\int_0^\infty \langle u''(t), v\rangle \chi_T(t)dt + \int_0^\infty \langle Q(u(t)), v\rangle \chi_T(t)dt \tag{4.15}$$

$$+ \int_0^\infty \left[\int_\Omega \varphi(x)((Dh)(u(t)), v)dx\right] \chi_T(t)dt + \int_0^\infty \left[\int_\Omega (H(t,x;u), v)dx\right] \chi_T(t)dt$$

$$+ \int_0^\infty \left[\int_\Omega (G(t,x;u,u'), v)dx\right] \chi_T(t)dt = \int_0^\infty (F(t), v)\chi_T(t)dt.$$

Let (T_k) be an arbitrary sequence converging to $+\infty$ and consider (4.15) with $T = T_k$. For the first term on the left hand side of this equation we have by (4.7) (if $T_k > 1$)

$$\int_0^\infty \langle u''(t), v\rangle \chi_{T_k}(t)dt = -\int_0^\infty \langle u'(t), v\rangle (\chi_{T_k})'(t)dt \to 0 \text{ as } k \to \infty. \tag{4.16}$$

Further, by (A_1), (4.14) and Lebesgue's dominated convergence theorem

$$\int_0^\infty \langle Q(u(t)), v\rangle \chi_{T_k}(t)dt = \int_0^\infty \langle Q(v), u(t)\rangle \chi_{T_k}(t)dt \tag{4.17}$$

$$= \int_0^1 \langle Q(v), u(T_k+\tau)\rangle \chi(\tau)d\tau \to \int_0^1 \langle Q(v), w_0\rangle \chi(\tau)d\tau = \langle Q(v), w_0\rangle$$

$$= \langle Q(w_0), v\rangle \text{ as } k \to \infty.$$

For the third term on the left hand side of (4.15) we have

$$\int_0^\infty \left[\int_\Omega \varphi(x)((Dh)(u(t)), v)dx\right] \chi_{T_k}(t)dt \tag{4.18}$$

$$= \int_0^1 \left[\int_\Omega \varphi(x)((Dh)(u(T_k+\tau)), v)dx\right] \chi(\tau)d\tau$$

$$\to \int_0^1 \left[\int_\Omega \varphi(x)((Dh)(w_0), v)dx\right] \chi(\tau)d\tau = \int_\Omega \varphi(x)((Dh)(w_0), v)dx$$

as $k \to \infty$ since by (4.8)

$$u(T_k+\tau) \to w_0 \text{ in } [L^2((0,1)\times\Omega)]^N \text{ as } k \to \infty$$

and thus for a.a. $(\tau, x) \in (0,1)\times\Omega$ (for a subsequence), consequently

$$(Dh)(u(T_k+\tau, x)) \to (Dh)(w_0(x)) \text{ for a.a. } (\tau,x) \in (0,1)\times\Omega. \tag{4.19}$$

By using Hölder's inequality, (A_3), (A_3'), respectively and Vitali's theorem, we obtain (4.18) from (4.19).

The fourth and fifth terms on the left hand side of (4.15) can be estimated by (4.5) and (4.7) as follows: for sufficiently large k

$$\left|\int_0^\infty \left[\int_\Omega (H(t,x;u),v)dx\right]\chi_{T_k}(t)dt\right| = \left|\int_0^\infty \left[\int_\Omega (H(T_k+\tau,x;u),v)dx\right]\chi(\tau)d\tau\right| \tag{4.20}$$

$$\leq \int_0^\infty \left[\int_\Omega \beta(T_k+\tau,x)|v|dx\right]|\chi(\tau)|d\tau \to 0 \text{ as } k\to\infty,$$

$$\left|\int_0^\infty \left[\int_\Omega (G(t,x;u,u'),v)dx\right]\chi_{T_k}(t)dt\right| \tag{4.21}$$

$$\leq \int_0^1 \left[\int_\Omega \{c_5|u'(T_k+\tau)| + \beta(T_k+\tau,x)\}|v|dx\right]|\chi(\tau)|d\tau \to 0.$$

Finally, for the right hand side of (4.15) we obtain by using (4.4) and the Cauchy – Schwarz inequality

$$\int_0^\infty (F(t),v)\chi_{T_k}(t)dt = \int_0^1 (F(T_k+\tau),v)\chi(\tau)d\tau \to \int_0^1 (F_\infty,v)\chi(\tau)d\tau = (F_\infty,v). \tag{4.22}$$

From (4.15) – (4.18), (4.20) – (4.22) one obtains (4.9).

Acknowledgments. This work was supported by the Hungarian National Foundation for Scientific Research under grant OTKA K 81403.

References

[1] Adams, R.A., *Sobolev spaces*, Academic Press, New York - San Francisco - London, 1975.

[2] Berkovits, J., Mustonen, V., *Topological degreee for perturbations of linear maximal monotone mappings and applications to a class of parabolic problems*, Rend. Mat. Ser. VII, **12**(1992), 597-621.

[3] Boukhatem, Y., Benabderrahmane, B., *Blow up of solutions for a semilinear Hyperbolic equation*, EJQTDE 2012, **40**, 1-12.

[4] Boulite, S., Maniar, L., N'Guérékata, G.M., *Almost automorphic solutions for hyperbolic semilinear evolution equations*, Semigroup Forum, **71**(2005), 231-240.

[5] Czernous, W., *Global solutions of semilinear first order partial functional differential equations with mixed conditions*, Functional Differential Equations, **18**2011, 135-154.

[6] Giorgi, C., Munoz Rivera, J.E., Pata, V., *Global attractors for a semilinear hyperbolic equation in viscoelasticity*, J. Math. Analysis and Appl., **260**(2001), 83-99.

[7] Guezane-Lakoud, A., Chaoui, A., *Roethe Method applied to semilinear hyperbolic integro-differential equation with integral conditions*, Int. J. Open Problems Compt. Math., **4**(2011), no. 1.

[8] Hale, J.K., *Theory of Functional Differential Equations*, Springer,1977.

[9] Lions, J.L., *Quelques méthodes de résolution des problemes aux limites non linéaires*, Dunod Gauthier-Villars, Paris, 1969.

[10] Nowakowski, A., *Nonhomogeneous boundary value problemfor semilinear hyperbolic equation. Stability*, J. Dynamical and Control Systems, **14**(2008), 537-558.

[11] Simon, L., *On nonlinear hyperbolic functional differential equations*, Math. Nachr., **217**(2000), 175-186.

[12] Simon, L., *Semilinear hyperbolic functional equations*, Banach Center Publications, **101**(2014), 205-222.

[13] Simon, L., *Application of monotone type operators to parabolic and functional parabolic PDE's*, Handbook of Differential Equations. Evolutionary Equations, Vol. 4, Elsevier, 2008, 267-321.

[14] Tataru, D., *Strichartz estimates in the hyperbolic space and global existence for the semilinear wave equation*, Trans. Amer. Math. Soc., **353**(2001), 795-807.

[15] Wu, J., *Theory and Applications of Partial Functional Differential Equations*, Springer, 1996.

[16] Zeidler, E., *Nonlinear functional analysis and its applications II A and II B*, Springer, 1990.

László Simon
Institute of Mathematics
L. Eötvös University of Budapest
e-mail: `simonl@cs.elte.hu`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 497–512

Continuous wavelet transform in variable Lebesgue spaces

Kristóf Szarvas and Ferenc Weisz

Abstract. In the present note we investigate norm and almost everywhere convergence of the inverse continuous wavelet transform in the variable Lebesgue space.

Mathematics Subject Classification (2010): Primary 42C40, Secondary 42C15, 42B08, 42A38, 46B15.

Keywords: Variable Lebesgue spaces, continuous wavelet transform, Wiener amalgam spaces, θ-summability, inversion formula.

1. Introduction

The topic of variable Lebesgue spaces is a new chapter of mathematics and it is studied intensively nowadays. Instead of the classical L_p-norm, the variable $L_{p(\cdot)}$-norm is defined by

$$\|f\|_{p(\cdot)} := \inf\left\{\lambda > 0 : \int_{\mathbb{R}^d} \left|\frac{f(x)}{\lambda}\right|^{p(x)} dx \leq 1\right\}$$

and the variable $L_{p(\cdot)}$ spaces contains all measurable functions f, for which $\|f\|_{p(\cdot)} < \infty$. The variable Lebesgue spaces have a lot of common property with the classical Lebesgue spaces (see Kováčik and Rákosník [12], Cruz-Uribe and Fiorenza [4], Diening, Hästö and Růžička [6], Cruz-Uribe, Firorenza and Neugebauer [3], Cruz-Uribe, Fiorenza, Martell and Pérez [2]).

The so called θ-summation method is studied intensively in the literature (see e.g. Butzer and Nessel [1], Trigub and Belinsky [15], Gát [9], Goginava [10], Simon [14] and Weisz [17, 18]). This summability is generated by a single function θ and

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

includes the well known Fejér, Riesz, Weierstrass, Abel, etc. summability methods. The θ-summation is defined by

$$\sigma_T^\theta f(x) = \int_{\mathbb{R}^d} f(x-t) T^d \theta(Tt)\, dt.$$

Feichtinger and Weisz [7, 8, 16] have proved that the θ-means $\sigma_T^\theta f$ converge to f almost everywhere and in norm as $T \to \infty$, whenever f is in the $L_p(\mathbb{R}^d)$ space or in a Wiener amalgam space. The points of the set of almost everywhere convergence are characterized as the Lebesgue points.

Some similar results are known in the variable Lebesgue spaces (see e.g. Cruz-Uribe and Fiorenza [4]). Under some conditions on the exponent function $p(\cdot)$ and θ, the θ-means of f converge to f almost everywhere and in norm for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$ as $T \to \infty$.

The continuous wavelet transform of f with respect to a wavelet g is defined by $W_g f(x,s) = \langle f, T_x D_s g\rangle$ $(x \in \mathbb{R}^d, s \in \mathbb{R}, s \neq 0)$, where D_s is the dilation operator and T_x is the translation operator. The inversion formula holds for all $f \in L_2(\mathbb{R}^d)$ (in case g and γ are suitable):

$$\int_0^\infty \int_{\mathbb{R}^d} W_g f(x,s) T_x D_s \gamma \,\frac{dxds}{s^{d+1}} = C_{g,\gamma} f,$$

where the equality is understood in a vector-valued weak sense (see Daubechies [5] and Gröchenig [11]).

Recently Li and Sun [13] have proved that if g and γ are radial, both have a radial majorant φ such that $\varphi(\cdot)\ln(2+|\cdot|) \in L_1(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} (g^* * \gamma)(x)\,dx = 0$, then for any $f \in L_p(\mathbb{R}^d)$ $(1 \le p < \infty)$

$$\lim_{S\to 0_+, T\to\infty} \int_S^T \int_{\mathbb{R}^d} W_g f(x,s) T_x D_s \gamma \,\frac{dxds}{s^{d+1}} = C'_{g,\gamma} f \tag{1.1}$$

at every Lebesgue point of f, where $C'_{g,\gamma}$ is a constant depending on g and γ. If $1 < p < \infty$, or if $1 \le p < \infty$ and $T = \infty$, then the convergence holds in the $L_p(\mathbb{R}^d)$-norm for all $f \in L_p(\mathbb{R}^d)$. Under some other conditions Weisz [19] has proved similar results.

In this paper we will investigate the norm and almost everywhere convergence of (1.1) in variable Lebesgue spaces. We lead back the problem to the summability of Fourier transforms, more exactly, we show that the integral on the left hand side of (1.1) can be formulated as $\sigma_{1/S}^\theta f - \sigma_{1/T}^\theta f$, where θ is a given function depending on g and γ.

2. θ-summability on the classical Lebesgue spaces

Let us fix $d \ge 1$, $d \in \mathbb{N}$. For a set $\mathbb{Y} \neq \emptyset$ let $\mathbb{Y}^d$ be its Cartesian product $\mathbb{Y} \times \ldots \times \mathbb{Y}$ taken with itself d-times. The space $L_p(\mathbb{R}^d)$ equipped with the norm

$$\|f\|_p := \left(\int_{\mathbb{R}^d} |f(x)|^p \, dx\right)^{1/p} \qquad (1 \le p \le \infty),$$

is the classical Lebesgue space. We use the notation $|I|$ for the Lebesgue measure of the set I. The set of *locally integrable functions* is denoted by $L_1^{loc}(\mathbb{R}^d)$.

A measurable function f belongs to the *Wiener amalgam space* $W(L_p,\ell_q)(\mathbb{R}^d)$ $(1 \le p,q \le \infty)$ if

$$\|f\|_{W(L_p,\ell_q)} := \left(\sum_{k\in\mathbb{Z}^d} \|f(\cdot+k)\|_{L_p[0,1)^d}^q\right)^{1/q} < \infty$$

with the usual modification for $q=\infty$. Note that for all $1\le p\le\infty$, $W(L_p,\ell_p)(\mathbb{R}^d) = L_p(\mathbb{R}^d)$ and $L_p(\mathbb{R}^d) \subset W(L_1,\ell_\infty)(\mathbb{R}^d)$.

Let $\theta \in L_1(\mathbb{R}^d)$ be a radial function. The *θ-means* of $f \in W(L_1,\ell_\infty)(\mathbb{R}^d)$ are defined by

$$\sigma_T^\theta f(x) := (f*\theta_T)(x) = \int_{\mathbb{R}^d} f(x-t)\theta_T(t)\,dt,$$

where

$$\theta_T(x) := T^d\theta(Tx) \qquad (x\in\mathbb{R}^d, T>0).$$

It is known that $\theta(t) = \widehat{\chi}_{B(0,1)}(t)$ implies $\sigma_T^\theta f = s_T f$, where $s_T f$ is the Dirichlet integral of the Fourier transform of f,

$$s_T f(x) := \int_{\{\|u\|_2<T\}} \widehat{f}(u)e^{2\pi \imath x\cdot u}\,du \qquad (T>0)$$

and

$$B(a,\delta) := \{x\in\mathbb{R}^d : \|x-a\|_2 < \delta\}.$$

Similarly, if $\theta(t) = \widehat{F}(t)$, where $F(t) = \max(1-\|t\|_2, 0)$, then we obtain the Fejér means of f.

The *classical Hardy-Littlewood maximal operator* is defined by

$$M(f)(x) := \sup_{x\in Q} \frac{1}{|Q|}\int_Q |f|\,d\lambda,$$

where $f \in L_1^{loc}(\mathbb{R}^d)$ and the supremum is taken over all cube $Q\subset\mathbb{R}^d$ with sides parallel to the axes. It is known that

$$\|Mf\|_p \le C\,\|f\|_p \tag{2.1}$$

for all $f\in L_p(\mathbb{R}^d)$ $(1<p\le\infty)$ and

$$\sup_{t>0} t\lambda\left(x\in\mathbb{R}^d : Mf(x) > t\right) \le C\,\|f\|_1$$

for all $f \in L_1(\mathbb{R}^d)$.

A point $x\in\mathbb{R}^d$ is called a *Lebesgue point* of $f\in L_1^{loc}(\mathbb{R}^d)$ if

$$\lim_{h\to 0_+}\left(\frac{1}{|B(0,h)|}\int_{B(0,h)} |f(x+u)-f(x)|\,du\right) = 0.$$

It is known that if $f \in W(L_1,\ell_\infty)(\mathbb{R}^d)$ $(1\le p\le\infty)$, then almost every $x\in\mathbb{R}^d$ is a Lebesgue point of f (see Feichtinger and Weisz [7, 8]).

We say that η is a *radial majorant* of f if η is radial, non-increasing as a function on $(0,\infty)$, non-negative, bounded, $|f| \le \eta$ and $\eta \in L_1(\mathbb{R}^d)$. If in addition $\eta(\cdot)\ln(|\cdot|+2) \in L_1(\mathbb{R}^d)$, then we say that η is a *radial log-majorant* of f.

The following results were proved in Feichtinger and Weisz [7] and [8].

Theorem 2.1. *Suppose that θ has a radial majorant η. Then for all $T > 0$*

$$|\sigma_T^\theta f(x)| \le C\|\eta\|_1 Mf(x) \qquad (x \in \mathbb{R}^d).$$

Theorem 2.2. *Suppose that θ has a radial majorant. Then*

(i) *for all $f \in W(L_1, \ell_\infty)(\mathbb{R}^d)$*

$$\lim_{T\to\infty} \sigma_T^\theta f(x) = \int_{\mathbb{R}^d} \theta(y)\,dy \cdot f(x)$$

at each Lebesgue points of f.

(ii) *for all $f \in L_p(\mathbb{R}^d)$ $(1 \le p < \infty)$*

$$\lim_{T\to 0_+} \sigma_T^\theta f(x) = 0$$

for all $x \in \mathbb{R}^d$.

Proof. The proof of the first statement can be found in Feichtinger and Weisz [8].

Consider (ii). Since θ has a radial majorant, $\theta \in L_1(\mathbb{R}^d) \cap L_\infty(\mathbb{R}^d)$. Therefore $\theta \in L_p(\mathbb{R}^d)$ $(1 \le p \le \infty)$. Let q the conjugate exponent of p i.e., $1/p + 1/q = 1$. Using Hölder's inequality

$$\begin{aligned}
\left|\sigma_T^\theta f(x)\right| &\le T^d \int_{\mathbb{R}^d} |f(x-t)|\,|\theta(Tt)|\,dt \\
&\le T^d \left(\int_{\mathbb{R}^d} |\theta(Tt)|^q\,dt\right)^{1/q} \left(\int_{\mathbb{R}^d} |f(x-t)|^p\,dt\right)^{1/p} \\
&= T^d \left(\int_{\mathbb{R}^d} |\theta(y)|^q T^{-d}\,dy\right)^{1/q} \|f\|_p \\
&= T^{d(1-1/q)} \|\theta\|_q \|f\|_p \to 0,
\end{aligned}$$

as $T \to 0_+$, because of $d(1-1/q) > 0$. □

Almost every point is a Lebesgue point of $f \in W(L_1, \ell_\infty)(\mathbb{R}^d)$, so (i) holds almost everywhere.

The next Theorem can be found in Feichtinger and Weisz [7].

Theorem 2.3. *Suppose that $\theta \in L_1(\mathbb{R}^d)$. Then*

(i) *for all $f \in L_p(\mathbb{R}^d)$ $(1 \le p < \infty)$*

$$\lim_{T\to\infty} \sigma_T^\theta f = \int_{\mathbb{R}^d} \theta(x)\,dx \cdot f \qquad \textit{in the } L_p(\mathbb{R}^d)\textit{-norm.}$$

(ii) *If in addition θ has a radial majorant, then for all $f \in L_p(\mathbb{R}^d)$ $(1 < p < \infty)$*

$$\lim_{T\to 0_+} \sigma_T^\theta f = 0 \qquad \textit{in the } L_p(\mathbb{R}^d)\textit{-norm.}$$

Proof. For the proof of (i) see Feichtinger and Weisz [7].

(ii) follows from Theorem 2.2 (ii), Theorem 2.1, (2.1) and Lebesgue dominated convergence theorem. □

The next lemma can be found in Li and Sun [13].

Lemma 2.4. *If g and γ have radial log-majorants, then $(g * \gamma)\ln(|\cdot|) \in L_1(\mathbb{R}^d)$ and $(|g| * |\gamma|)\ln(|\cdot|) \in L_1(\mathbb{R}^d)$.*

3. θ-summability on the variable Lebesgue spaces

For the variable Lebesgue spaces we can state similar theorems. A function $p(\cdot)$ belongs to $\mathcal{P}(\mathbb{R}^d)$ if $p : \mathbb{R}^d \to [1, \infty]$ and $p(\cdot)$ is measurable. Then we say that $p(\cdot)$ is an exponent function. Let

$$p_- := \inf\{p(x) : x \in \mathbb{R}^d\} \qquad \text{and} \qquad p_+ := \sup\{p(x) : x \in \mathbb{R}^d\}.$$

Set

$$\Omega_\infty := \{x \in \mathbb{R}^d : p(x) = \infty\}.$$

The modular generated by $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ is defined by

$$\varrho_{p(\cdot)}(f) := \int_{\mathbb{R}^d \setminus \Omega_\infty} |f(x)|^{p(x)}\, dx + \|f\|_{L_\infty(\Omega_\infty)},$$

where f is a measurable function. A measurable function f belongs to the $L_{p(\cdot)}(\mathbb{R}^d)$ space if there exists $\lambda > 0$ such that $\varrho_{p(\cdot)}(f/\lambda) < \infty$. We can see that the modular $\varrho_{p(\cdot)}$ is not a norm. Define the $L_{p(\cdot)}(\mathbb{R}^d)$-norm by

$$\|f\|_{p(\cdot)} := \inf\left\{\lambda > 0 : \varrho_{p(\cdot)}\left(\frac{f}{\lambda}\right) \le 1\right\}.$$

Then $\|\cdot\|_{p(\cdot)}$ is a norm and the space $(L_{p(\cdot)}(\mathbb{R}^d), \|\cdot\|_{p(\cdot)})$ is a normed space. In case $p(\cdot) = p$ is a constant, then we get back the usual $L_p(\mathbb{R}^d)$ spaces.

We say that $r(\cdot)$ is locally log-Hölder continuous if there exists a constant C_0 such that for all $x, y \in \mathbb{R}^d$, $0 < \|x - y\|_2 < 1/2$,

$$|r(x) - r(y)| \le \frac{C_0}{-\log(\|x - y\|_2)}.$$

We denote this set by $LH_0(\mathbb{R}^d)$.

We say that $r(\cdot)$ is log-Hölder continuous at infinity if there exist constants C_∞ and r_∞ such that for all $x \in \mathbb{R}^d$

$$|r(x) - r_\infty| \le \frac{C_\infty}{\log(e + \|x\|_2)}.$$

We write briefly $r(\cdot) \in LH_\infty(\mathbb{R}^d)$. Let

$$LH(\mathbb{R}^d) := LH_0(\mathbb{R}^d) \cap LH_\infty(\mathbb{R}^d).$$

The following two results were proved in Cruz-Uribe and Fiorenza [4, p.27, p.35].

Theorem 3.1 (Hölder's inequality). *Let $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, $1/p(x) + 1/q(x) = 1$. Then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$ and $g \in L_{q(\cdot)}(\mathbb{R}^d)$, $fg \in L_1(\mathbb{R}^d)$ and*

$$\int_{\mathbb{R}^d} |f(x)g(x)|\,dx \le C_{p(\cdot)} \|f\|_{p(\cdot)} \|g\|_{q(\cdot)}.$$

Lemma 3.2. *If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ and $K \subset \mathbb{R}^d$, $|K| < \infty$, then $\chi_K \in L_{p(\cdot)}(\mathbb{R}^d)$ and*

$$\|\chi_K\|_{p(\cdot)} \le |K| + 1.$$

We need also the next statement.

Theorem 3.3. *If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, then $L_{p(\cdot)} \subset W(L_1, \ell_\infty)(\mathbb{R}^d)$.*

Proof. Let $f \in L_{p(\cdot)}(\mathbb{R}^d)$ and $q(\cdot)$ the conjugate function of $p(\cdot)$. Then by Theorem 3.1 and Lemma 3.2,

$$\int_{[n,n+1)} |f(x)|\,dx \le C_{p(\cdot)} \|f\|_{p(\cdot)} \|\chi_{[n,n+1)}\|_{q(\cdot)} \le 2C_{p(\cdot)} \|f\|_{p(\cdot)},$$

for $n = (n_1, \dots, n_d) \in \mathbb{Z}^d$, where $n + 1 = (n_1 + 1, \dots, n_d + 1)$. Hence

$$\|f\|_{W(L_1,\ell_\infty)} \le 2C_{p(\cdot)} \|f\|_{p(\cdot)} < \infty,$$

which implies the theorem. □

The following three theorems can be found in Cruz-Uribe and Fiorenza [4, p.56, p.44, p.42]

Theorem 3.4. *If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, $p_+ < \infty$, then the set of bounded functions with compact support is dense in $L_{p(\cdot)}(\mathbb{R}^d)$.*

Theorem 3.5. *If $p \in \mathcal{P}(\mathbb{R}^d)$ and $p_+(\mathbb{R}^d \setminus \Omega_\infty) < \infty$, then the following properties are equivalent:*

(i) *convergence in norm,*
(ii) *convergence in modular.*

Theorem 3.6. *If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, then*

$$L_{p(\cdot)}(\mathbb{R}^d) \subset L_{p_+}(\mathbb{R}^d) + L_{p_-}(\mathbb{R}^d),$$

i.e., for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$ there exist $f_1 \in L_{p_-}(\mathbb{R})$ and $f_2 \in L_{p_+}(\mathbb{R})$ such that $f = f_1 + f_2$.

The next theorem says that $\sigma_T^\theta f(x)$ converges at every Lebesgue point.

Theorem 3.7. *If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ and θ has a radial majorant, then*

(i) for all Lebesgue points of $f \in L_{p(\cdot)}(\mathbb{R}^d)$,

$$\lim_{T\to\infty} \sigma_T^\theta f(x) = \int_{\mathbb{R}^d} \theta(y)\,dy \cdot f(x).$$

(ii) If in addition $p_+ < \infty$, then

$$\lim_{T\to 0_+} \sigma_T^\theta f(x) = 0$$

for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$ and for all $x \in \mathbb{R}^d$.

Proof. To prove (i), let $f \in L_{p(\cdot)}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ be a Lebesgue point of f. Using Theorem 3.3, we have $f \in W(L_1, \ell_\infty)(\mathbb{R}^d)$. By Theorem 2.2 we get that

$$\lim_{T\to\infty} \sigma_T^\theta f(x) = \int_{\mathbb{R}^d} \theta(y)\, dy \cdot f(x).$$

Consider (ii). Let $f \in L_{p(\cdot)}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ arbitrary. Then by Theorem 3.6 there exist $f_1 \in L_{p_-}(\mathbb{R}^d)$ and $f_2 \in L_{p_+}(\mathbb{R}^d)$ such that $f = f_1 + f_2$. Since $p_+ < \infty$ we can use Theorem 2.2 to obtain

$$\lim_{T\to 0_+} \sigma_T^\theta f(x) = \lim_{T\to 0_+} \sigma_T^\theta f_1(x) + \lim_{T\to 0_+} \sigma_T^\theta f_2(x) = 0,$$

which proves the theorem. □

Of course, the convergence in (i) holds almost everywhere (see also Cruz-Uribe and Fiorenza [4, p.197]). The first and the second statement of the next theorem can be found in Cruz-Uribe and Fiorenza [4, p.199].

Theorem 3.8. *Let $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, $p_+ < \infty$, $1/p(x) + 1/q(x) = 1$. If θ has a radial majorant and the maximal operator is bounded on $L_{q(\cdot)}(\mathbb{R}^d)$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$*

(i)

$$\left\|\sigma_T^\theta f\right\|_{p(\cdot)} \le C \left\|f\right\|_{p(\cdot)} \qquad (T > 0). \tag{3.1}$$

(ii)

$$\lim_{T\to\infty} \sigma_T^\theta f = \int_{\mathbb{R}^d} \theta(x)\, dx \cdot f \quad \textit{in the } L_{p(\cdot)}(\mathbb{R}^d)\textit{-norm.}$$

(iii) *If in addition $p_- > 1$, then*

$$\lim_{T\to 0_+} \sigma_T^\theta f = 0 \quad \textit{in the } L_{p(\cdot)}(\mathbb{R}^d)\textit{-norm}$$

for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$.

Proof. To prove (iii), fix $\varepsilon > 0$. By Theorem 3.4 there exists a bounded function g with compact support, such that $\|f - g\|_{p(\cdot)} < \varepsilon$. Using (3.1) we have

$$\left\|\sigma_T^\theta f\right\|_{p(\cdot)} \le \left\|\sigma_T^\theta (f-g)\right\|_{p(\cdot)} + \left\|\sigma_T^\theta g\right\|_{p(\cdot)} < C\varepsilon + \left\|\sigma_T^\theta g\right\|_{p(\cdot)}.$$

So it is enough to show that

$$\lim_{T\to 0_+} \left\|\sigma_T^\theta g\right\|_{p(\cdot)} = 0.$$

Since $p_+ < \infty$, then by Theorem 3.5 we have to show that

$$\lim_{T\to 0_+} \int_{\mathbb{R}^d} |\sigma_T^\theta g(x)|^{p(x)}\, dx = 0.$$

Let

$$g_0(x) := \frac{g(x)}{\|\theta\|_1 \|g\|_\infty}.$$

Then $\|g_0\|_\infty \le 1/\|\theta\|_1$ and

$$|\sigma_T^\theta g_0(x)| = |(g_0 * \theta_T)(x)| \le \|g_0\|_\infty \|\theta_T\|_1 = \|g_0\|_\infty \|\theta\|_1 \le 1.$$

Therefore

$$\begin{aligned}\lim_{T\to 0_+}\int_{\mathbb{R}^d}|\sigma_T^\theta g(x)|^{p(x)}\,dx &= \lim_{T\to 0_+}\int_{\mathbb{R}^d}(\|\theta\|_1\|g\|_\infty)^{p(x)}|\sigma_T^\theta g_0(x)|^{p(x)}\,dx\\ &\le (\|\theta\|_1\|g\|_\infty+1)^{p_+}\lim_{T\to 0_+}\int_{\mathbb{R}^d}|\sigma_T^\theta g_0(x)|^{p_-}\,dx.\end{aligned}$$

Here $1<p_-<\infty$ and $g_0\in L_{p_-}(\mathbb{R}^d)$, therefore by Theorem 2.3 we get that

$$\lim_{T\to 0_+}\int_{\mathbb{R}^d}|\sigma_T^\theta g_0(x)|^{p_-}\,dx=0,$$

which proves the theorem. □

The next theorem about the boundedness of the classical Hardy-Littlewood maximal operator in variable Lebesgue spaces can be found in Cruz-Uribe and Fiorenza [4, p.89].

Theorem 3.9. *Let $p(\cdot)\in\mathcal{P}(\mathbb{R}^d)$ and $1/p(\cdot)\in LH(\mathbb{R}^d)$.*

(i) *Then for all $f\in L_{p(\cdot)}(\mathbb{R}^d)$ and $t>0$*

$$\left\|t\chi_{\{x\in\mathbb{R}^d:Mf(x)>t\}}\right\|_{p(\cdot)}\le C\left\|f\right\|_{p(\cdot)}.$$

(ii) *If in addition $p_->1$, then for all $f\in L_{p(\cdot)}(\mathbb{R}^d)$*

$$\left\|Mf\right\|_{p(\cdot)}\le C\left\|f\right\|_{p(\cdot)}.$$

Remark 3.10. If $1/p(\cdot)\in LH(\mathbb{R}^d)$ and $p_+<\infty$, then $1/q(\cdot)\in LH(\mathbb{R}^d)$ and $q_->1$ so the maximal operator is bounded on $L_{q(\cdot)}(\mathbb{R}^d)$. Therefore if $1/p(\cdot)\in LH(\mathbb{R}^d)$, $p_+<\infty$ and θ has a radial majorant, then the hypotheses of Theorem 3.8 remain true.

4. The continuous wavelet transform

The continuous wavelet transform of f with respect to a wavelet g is defined by

$$W_gf(x,s):=|s|^{-d/2}\int_{\mathbb{R}^d}f(t)\overline{g(s^{-1}(t-x))}\,dt=\langle f,T_xD_sg\rangle,$$

$(x\in\mathbb{R}^d, s\in\mathbb{R}, s\neq 0)$ when the integral does exist. We suppose that $g,\gamma\in L_2(\mathbb{R}^d)$ and

$$\int_0^\infty|\widehat{g}(s\omega)||\widehat{\gamma}(s\omega)|\,\frac{ds}{s}<\infty$$

for almost $\omega\in\mathbb{R}^d$ with $\|\omega\|_2=1$. If

$$C_{g,\gamma}:=\int_0^\infty\overline{\widehat{g}}(s\omega)\widehat{\gamma}(s\omega)\,\frac{ds}{s}$$

is independent of ω, then the inversion formula holds for all $f\in L_2(\mathbb{R}^d)$:

$$\int_0^\infty\int_{\mathbb{R}^d}W_gf(x,s)T_xD_s\gamma\,\frac{dxds}{s^{d+1}}=C_{g,\gamma}\cdot f,$$

where the equality is understood in a vector-valued weak sense. Consider the operators

$$\rho_S f := \int_S^\infty \int_{\mathbb{R}^d} W_g f(x,s) T_x D_s \gamma \, \frac{dxds}{s^{d+1}}$$

and

$$\rho_{S,T} f := \int_S^T \int_{\mathbb{R}^d} W_g f(x,s) T_x D_s \gamma \, \frac{dxds}{s^{d+1}},$$

where $0 < S < T < \infty$. Let

$$C'_{g,\gamma} := - \int_{\mathbb{R}^d} (g^* * \gamma)(x) \ln(|x|) \, dx,$$

where $g^*(x) = \overline{g(-x)}$ is the involution operator. If g and γ both have radial log-majorants, then $C'_{g,\gamma}$ is finite by Lemma 2.4.

Li and Sun [13] proved that if g and γ radial, $\int_{\mathbb{R}^d} (g^* * \gamma)(x)\, dx = 0$, and both have a radial log-majorant, then for any $f \in L_p(\mathbb{R}^d)$ $(1 \le p < \infty)$

$$\lim_{S \to 0_+, T \to \infty} \rho_{S,T} f(x) = \lim_{S \to 0_+} \rho_S f(x) = C'_{g,\gamma} f(x)$$

at every Lebesgue point of f. Moreover, if $1 < p < \infty$, then the convergence holds in the $L_p(\mathbb{R}^d)$-norm for all $f \in L_p(\mathbb{R}^d)$. If $p = 1$, then

$$\lim_{S \to 0_+} \rho_S f = C'_{g,\gamma} f \qquad \text{in the } L_1(\mathbb{R}^d)\text{-norm}$$

for all $f \in L_1(\mathbb{R}^d)$. Under some similar conditions Weisz [19] proved similar results. In this paper we investigate the same questions on the variable Lebesgue spaces and we will prove similar theorems. Of course, $C_{g,\gamma} = C'_{g,\gamma}$ under some conditions (see Li and Sun [13]).

5. Convergence of ϱ_S and $\varrho_{S,T}$

We will denote the surface area of the unit ball in $\mathbb{R}^d$ by ω_{d-1}. The next theorem plays central role in this chapter. Under some conditions we lead back $\varrho_S f$ to a θ-summation.

Theorem 5.1. *Suppose that g, γ have radial log-majorants and*

$$\int_{\mathbb{R}^d} (g^* * \gamma)(x)\, dx = 0.$$

If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ and $p_+ < \infty$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$

$$\varrho_S f = \sigma^\theta_{1/S} f \quad (S > 0),$$

where θ is defined later in the proof.

Proof. Let $y \in \mathbb{R}^d$ be arbitrary and decompose $\varrho_S f(y)$

$$\begin{aligned}
\varrho_S f(y) &= \int_S^\infty \int_{\mathbb{R}^d} \frac{1}{s^{2d+1}} \int_{\mathbb{R}^d} f(t) \overline{g\left(\frac{t-x}{s}\right)} \gamma\left(\frac{y-x}{s}\right) dt dx ds \\
&= \int_S^\infty \int_{\|y-t\|_2<S} \frac{1}{s^{2d+1}} \int_{\mathbb{R}^d} f(t) \overline{g\left(\frac{t-x}{s}\right)} \gamma\left(\frac{y-x}{s}\right) dx dt ds \\
&- \int_0^S \int_{\|y-t\|_2\geq S} \frac{1}{s^{2d+1}} \int_{\mathbb{R}^d} f(t) \overline{g\left(\frac{t-x}{s}\right)} \gamma\left(\frac{y-x}{s}\right) dx dt ds \\
&+ \int_0^\infty \int_{\|y-t\|_2\geq S} \frac{1}{s^{2d+1}} \int_{\mathbb{R}^d} f(t) \overline{g\left(\frac{t-x}{s}\right)} \gamma\left(\frac{y-x}{s}\right) dx dt ds \\
&=: I - II + III.
\end{aligned}$$

We can write I and II as a convolution, similarly as in Li and Sun [13]:

$$I = (f * \varphi_{1/S})(y),$$

where

$$\varphi(t) := \int_1^\infty H\left(\frac{t}{u}\right) \frac{1}{u^{d+1}} \chi_{B(0,1)}(t)\, du$$

and $H := g^* * \gamma$. Since γ has radial log-majorant, $\gamma \in L_\infty(\mathbb{R}^d)$ and since $g \in L_1(\mathbb{R}^d)$, $H \in L_\infty(\mathbb{R}^d)$. Therefore if $t \in B(0,1)$, then

$$|\varphi(t)| \leq \|H\|_\infty \int_1^\infty \frac{1}{u^{d+1}}\, du = C\|H\|_\infty < \infty.$$

If $t \notin B(0,1)$, then $\varphi(t) = 0$. Thus $C\|H\|_\infty \chi_{B(0,1)}$ is a radial majorant of φ.

$$II = (f * \psi_{1/S})(y),$$

where

$$\psi(t) := \int_0^1 H\left(\frac{t}{u}\right) \frac{1}{u^{d+1}} \chi_{\mathbb{R}^d \setminus B(0,1)}(t)\, du.$$

Let $G := |g| * |\gamma|$. Then $H \leq G$, and since g, γ have radial log-majorants, Lemma 2.4 implies that H, G have radial log-majorants, too. Since G is radial, there exists η such that $G(x) = \eta(\|x\|_2)$. If $t \in B(0,1)$, then $\psi(t) = 0$. If $t \in \mathbb{R}^d \setminus B(0,1)$, then

$$\begin{aligned}
|\psi(t)| &\leq \int_0^1 G\left(\frac{t}{u}\right) \frac{1}{u^{d+1}}\, du = \int_0^1 \eta\left(\frac{\|t\|_2}{u}\right) \frac{1}{u^{d+1}}\, du \\
&= \frac{1}{\|t\|_2^d} \int_{\|t\|_2}^\infty \eta(s) s^{d-1}\, ds =: \zeta(t)
\end{aligned}$$

and let

$$\zeta(t) := \int_1^\infty \eta(s) s^{d-1}\, ds \leq \frac{1}{\omega_{d-1}} \|G\|_1 < \infty \qquad (t \in B(0,1)).$$

It is easy to see that $|\psi| \leq \zeta$, ζ is radial, non-increasing on $(0,\infty)$ and bounded. Moreover,

$$\begin{aligned}
\int_{\mathbb{R}^d} \zeta(t)\,dt &= \int_{B(0,1)} \zeta(t)\,dt + \int_{\mathbb{R}^d\setminus B(0,1)} \zeta(t)\,dt \\
&= C + \int_{\mathbb{R}^d\setminus B(0,1)} \frac{1}{\|t\|_2^d} \int_{\|t\|_2}^{\infty} \eta(s)s^{d-1}\,dsdt \\
&\leq C + \omega_{d-1} \int_1^{\infty} \frac{1}{r} \int_r^{\infty} \eta(s)s^{d-1}\,dsdr \\
&= C + \omega_{d-1} \int_1^{\infty} \eta(s)s^{d-1} \int_1^s \frac{1}{r}\,drds \\
&= C + \int_{\mathbb{R}^d\setminus B(0,1)} G(t)\ln(|t|)\,dt < \infty,
\end{aligned}$$

i.e., ζ is integrable so ζ is a radial majorant of ψ.

We will show that $III = 0$. To apply Fubini's theorem we will verify that

$$\int_0^{\infty} \int_{\|y-t\|_2 \geq S} \frac{1}{s^{d+1}} |f(t)| G\left(\frac{y-t}{s}\right)\,dtds < \infty. \tag{5.1}$$

Since G is radial

$$\begin{aligned}
&\int_{\|y-t\|_2 \geq S} |f(t)| \int_0^{\infty} \frac{1}{s^{d+1}} \eta\left(\frac{\|y-t\|_2}{s}\right)\,dsdt \\
&= \int_{\|y-t\|_2 \geq S} |f(t)| \int_0^{\infty} \frac{1}{\|y-t\|_2^d} \eta(u)u^{d-1}\,dudt \\
&= \frac{1}{\omega_{d-1}} \|G\|_1 \int_{\|y-t\|_2 \geq S} |f(t)| \frac{1}{\|y-t\|_2^d}\,dt. \qquad (5.2)
\end{aligned}$$

By Theorem 3.1

$$\int_{\|y-t\|_2 \geq S} |f(t)| \frac{1}{\|y-t\|_2^d}\,dt \leq C_{p(\cdot)} \|f\|_{p(\cdot)} \left\| \frac{1}{\|y-\cdot\|_2^d} \chi_{\{\|y-\cdot\|_2 \geq S\}} \right\|_{q(\cdot)},$$

where $1/p(x) + 1/q(x) = 1$ $(x \in \mathbb{R}^d)$. Let $\lambda := 1/S^d$. Then

$$\frac{1}{\lambda\|y-t\|_2^d} \leq \frac{1}{\lambda S^d} = 1 \qquad \text{and} \qquad \left(\frac{1}{\lambda\|y-t\|_2^d}\right)^{q(t)} \leq \left(\frac{1}{\lambda\|y-t\|_2^d}\right)^{q_-}.$$

If $p_+ < \infty$, then $q_- > 1$ and

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\frac{\chi_{\{\|y-t\|_2 \geq S\}}}{\lambda\|y-t\|_2^d}\right)^{q(t)} dt &\leq \int_{\|y-t\|_2 \geq S} \left(\frac{1}{\lambda\|y-t\|_2^d}\right)^{q_-} dt \\
&= S^{dq_-} \int_{\|y-t\|_2 \geq S} \frac{1}{\|y-t\|_2^{dq_-}}\,dt \\
&= \omega_{d-1} S^{dq_-} \int_S^{\infty} u^{-dq_-+d-1}\,du < \infty.
\end{aligned}$$

Moreover,

$$\frac{\frac{1}{\|y-t\|_2^d}\chi_{\{\|y-t\|_2\geq S\}}(t)}{\lambda}\leq\frac{1}{\lambda S^d}=1,$$

in other words,

$$\left\|\frac{\frac{1}{\|y-\cdot\|_2^d}\chi_{\{\|y-\cdot\|_2\geq S\}}}{\lambda}\right\|_{L_\infty(\Omega_\infty)}\leq 1,$$

where

$$\Omega_\infty=\{x\in\mathbb{R}^d:q(x)=\infty\}.$$

So

$$\varrho_{q(\cdot)}\left(\frac{\frac{1}{\|y-\cdot\|_2^d}\chi_{\{\|y-\cdot\|_2\geq S\}}}{\lambda}\right)<\infty$$

and

$$\left\|\frac{1}{\|y-\cdot\|_2^d}\chi_{\{\|y-\cdot\|_2\geq S\}}\right\|_{q(\cdot)}<\infty.$$

We get that (5.2) is finite so we can apply Fubini's theorem. Since H is radial, there exists ν such that $H(x)=\nu(\|x\|_2)$ and

$$\begin{aligned} III &= \int_0^\infty\int_{\|y-t\|_2\geq S} f(t)\frac{1}{s^{d+1}}H\left(\frac{y-t}{s}\right)dtds \\ &= \int_{\|y-t\|_2\geq S} f(t)\int_0^\infty\frac{1}{s^{d+1}}\nu\left(\frac{\|y-t\|_2}{s}\right)dsdt \\ &= \int_{\|y-t\|_2\geq S} f(t)\frac{1}{\|y-t\|_2^d}\int_0^\infty \nu(u)u^{d-1}\,dudt \\ &= \frac{1}{\omega_{d-1}}\int_{\|y-t\|_2\geq S} f(t)\frac{1}{\|y-t\|_2^d}\int_{\mathbb{R}^d}H(u)\,dudt= \\ &= \frac{1}{\omega_{d-1}}\int_{\|y-t\|_2\geq S} f(t)\frac{1}{\|y-t\|_2^d}\int_{\mathbb{R}^d}(g^**\gamma)(u)\,dudt=0. \end{aligned}$$

We have that

$$\varrho_S f(y)=(f*\varphi_{1/S})(y)-(f*\psi_{1/S})(y)=\big(f*(\varphi_{1/S}-\psi_{1/S})\big)(y)=:\sigma_{1/S}^\theta f(y),$$

where

$$\theta(y):=\varphi(y)-\psi(y).$$

Since φ and ψ have radial majorants, θ has, too. □

Using the previous theorem we can prove the convergence of $\varrho_S f$ and $\varrho_{S,T} f$ at Lebesgue points, almost everywhere and in the $L_{p(\cdot)}(\mathbb{R}^d)$-norm.

Theorem 5.2. *Suppose that g, γ have radial log-majorants and*

$$\int_{\mathbb{R}^d}(g^**\gamma)(x)\,dx=0.$$

If $p(\cdot)\in\mathcal{P}(\mathbb{R}^d)$, $p_+<\infty$, then for all Lebesgue points of $f\in L_{p(\cdot)}(\mathbb{R}^d)$,

(i)

$$\lim_{S\to 0_+} \varrho_S f(x) = C'_{g,\gamma}\cdot f(x).$$

(ii)

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f(x) = C'_{g,\gamma}\cdot f(x).$$

Proof. By Theorem 5.1 and Theorem 3.7 we have

$$\lim_{S\to 0_+} \varrho_S f(x) = \lim_{S\to 0_+} \sigma^{\theta}_{1/S} f(x) = \int_{\mathbb{R}^d} \theta(y)\,dy\cdot f(x),$$

i.e., it is enough to prove that

$$\int_{\mathbb{R}^d} \theta(y)\,dy = C'_{g,\gamma}.$$

We have

$$\int_{\mathbb{R}^d} \theta(y)\,dy = \int_{\mathbb{R}^d} \varphi(y)\,dy - \int_{\mathbb{R}^d} \psi(y)\,dy.$$

Here

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(y)\,dy &= \int_{B(0,1)}\int_1^\infty H\left(\frac{y}{u}\right)\frac{1}{u^{d+1}}\,dudy\\
&= \omega_{d-1}\int_0^1 r^{d-1}\int_1^\infty \nu\left(\frac{r}{u}\right)\frac{1}{u^{d+1}}\,dudr\\
&= \omega_{d-1}\int_0^1 \frac{1}{r}\int_0^r \nu(t)t^{d-1}\,dtdr\\
&= \omega_{d-1}\int_0^1 t^{d-1}\nu(t)\int_t^1 \frac{1}{r}\,drdt\\
&= -\int_{B(0,1)} H(t)\ln(|t|)\,dt\\
&= -\int_{B(0,1)} (g^* * \gamma)(t)\ln(|t|)\,dt
\end{aligned}$$

and we have similarly that

$$\int_{\mathbb{R}^d} \psi(y)\,dy = \int_{\mathbb{R}^d\setminus B(0,1)} (g^* * \gamma)(t)\ln(|t|)\,dt,$$

i.e.

$$\begin{aligned}
\int_{\mathbb{R}^d} \theta(y)\,dy &= -\int_{B(0,1)} (g^* * \gamma)(t)\ln(|t|)\,dt - \int_{\mathbb{R}^d\setminus B(0,1)} (g^* * \gamma)(t)\ln(|t|)\,dt\\
&= -\int_{\mathbb{R}^d} (g^* * \gamma)(t)\ln(|t|)\,dt = C'_{g,\gamma}.
\end{aligned}$$

To prove (ii), observe that

$$\varrho_{S,T} f(x) = \varrho_S f(x) - \varrho_T f(x).$$

Then use Theorem 5.1 and Theorem 3.7 to get

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f(x) = \lim_{S\to 0_+} \varrho_S f(x) - \lim_{T\to\infty} \sigma^\theta_{1/T} f(x) = C'_{g,\gamma} \cdot f(x) - 0,$$

which proves the theorem. □

Corollary 5.3. *Suppose that g, γ have radial log-majorants and*

$$\int_{\mathbb{R}^d} (g^* * \gamma)(x)\, dx = 0.$$

If $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$ and $p_+ < \infty$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$

(i)

$$\lim_{S\to 0_+} \varrho_S f = C'_{g,\gamma} f \qquad a.e.$$

(ii)

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f = C'_{g,\gamma} f \qquad a.e.$$

Theorem 5.4. *Suppose that g, γ have radial log-majorants and*

$$\int_{\mathbb{R}^d} (g^* * \gamma)(x)\, dx = 0.$$

Let $p(\cdot) \in \mathcal{P}(\mathbb{R}^d)$, $p_+ < \infty$, $1/p(x) + 1/q(x) = 1$. If the maximal operator is bounded on $L_{q(\cdot)}(\mathbb{R}^d)$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$,

(i)

$$\lim_{S\to 0_+} \varrho_S f = C'_{g,\gamma} \cdot f \qquad \text{in the } L_{p(\cdot)}(\mathbb{R}^d)\text{-norm.}$$

(ii) *If in addition $p_- > 1$, then for all $f \in L_{p\cdot}(\mathbb{R}^d)$*

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f = C'_{g,\gamma} \cdot f \qquad \text{in the } L_{p(\cdot)}(\mathbb{R}^d)\text{-norm.}$$

Proof. To prove (i), use Theorem 5.1 and Theorem 3.8

$$\lim_{S\to 0_+} \varrho_S f = \lim_{S\to 0_+} \sigma^\theta_{1/S} f = \int_{\mathbb{R}^d} \theta(x)\, dx \cdot f \qquad \text{in the } L_{p(\cdot)}(\mathbb{R}^d)\text{-norm.}$$

We have seen, that $\int_{\mathbb{R}^d} \theta(x)\, dx = C'_{g,\gamma}$.

We can prove (ii) similarly. Use Theorem 5.1 and Theorem 3.8 to obtain

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f = \lim_{S\to 0_+} \varrho_S f - \lim_{T\to\infty} \varrho_T f = C'_{g,\gamma} \cdot f$$

in the $L_{p(\cdot)}(\mathbb{R}^d)$-norm. The proof of the theorem is complete. □

By Remark 3.10 if we suppose that $1/p(\cdot) \in LH(\mathbb{R}^d)$ and $p_+ < \infty$, then the maximal operator is bonded on $L_{q(\cdot)}(\mathbb{R}^d)$. Therefore we have

Corollary 5.5. *Suppose that g, γ have radial log-majorants and*

$$\int_{\mathbb{R}^d} (g^* * \gamma)(x)\, dx = 0.$$

If $1/p(\cdot) \in LH(\mathbb{R}^d)$, $p_+ < \infty$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$,

(i) $\lim_{S\to 0_+} \varrho_S f = C'_{g,\gamma} \cdot f$ *in the $L_{p(\cdot)}(\mathbb{R}^d)$-norm.*

(ii) If in addition $p_- > 1$, then for all $f \in L_{p(\cdot)}(\mathbb{R}^d)$

$$\lim_{S\to 0_+, T\to\infty} \varrho_{S,T} f = C'_{g,\gamma} \cdot f \qquad \text{in the } L_{p(\cdot)}(\mathbb{R}^d)\text{-norm.}$$

References

[1] Butzer, P.L., Nessel, R.J., *Fourier Analysis and Approximation*, Birkhäuser Verlag, Basel, 1971.

[2] Cruz-Uribe, D., Fiorenza, A., Martell, J., Pérez, C., *The boundedness of classical operators on variable L^p spaces*, Ann. Acad. Sci. Fenn., Math., **31**(1)(2006), 239–264.

[3] Cruz-Uribe, D., Firorenza, A., Neugebauer, C.J., *The maximal function on variable L^p spaces*, Ann. Acad. Sci. Fenn., Math., **28**(1)(2003), 223–238.

[4] Cruz-Uribe, D.V., Fiorenza, A., *Variable Lebesgue spaces. Foundations and harmonic analysis*, New York, NY: Birkhäuser/Springer, 2013.

[5] Daubechies, I., *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[6] Diening, L., Harjulehto, P., Hästö, P., Růžička, M., *Lebesgue and Sobolev spaces with variable exponents*, Springer, Berlin, 2011.

[7] Feichtinger, H.G., Weisz, F., *The Segal algebra $\mathbf{S}_0(\mathbb{R}^d)$ and norm summability of Fourier series and Fourier transforms*, Monatshefte Math., **148**(2006), 333–349.

[8] Feichtinger, H.G., Weisz, F., *Wiener amalgams and pointwise summability of Fourier transforms and Fourier series*, Math. Proc. Camb. Phil. Soc., **140**(2006), 509–536.

[9] Gát, G., *Pointwise convergence of cone-like restricted two-dimensional $(C,1)$ means of trigonometric Fourier series*, J. Appr. Theory., **149**(2007), 74–102.

[10] Goginava, U., *The maximal operator of the Marcinkiewicz-Fejér means of d-dimensional Walsh-Fourier series*, East J. Appr., **12**(2006), 295–302.

[11] Gröchenig, K., *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, 2001.

[12] Kováčik, O., Rákosník, J., *On spaces $L^{p(x)}$ and $W^{k,p(x)}$*, Czech. Math. J., **41**(4)(1991), 592–618.

[13] Li, K., Sun, W., *Pointwise convergence of the Calderon reproducing formula*, J. Fourier Anal. Appl., **18**(2012), 439–455.

[14] Simon, P., *(C,α) summability of Walsh-Kaczmarz-Fourier series*, J. Appr. Theory, **127**(2004), 39–60.

[15] Trigub, R.M., Belinsky, E.S., *Fourier Analysis and Approximation of Functions*, Kluwer Academic Publishers, Dordrecht, Boston, London, 2004.

[16] Weisz, F., *Pointwise convergence in Pringsheim's sense of the summability of Fourier transforms on Wiener amalgam spaces*, Monatshefte Math. (to appear).

[17] Weisz, F., *Summability of Multi-dimensional Fourier Series and Hardy Spaces*, Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, Boston, London, 2002.

[18] Weisz, F., *Summability of Multi-Dimensional Trigonometric Fourier Series*, vol. 7. Surveys in Approximation Theory, 2012.

[19] Weisz, F., *Inversion formulas for the continuous wavelet transform*, Acta Math. Hungar., **138**(2013), 237–258.

Kristóf Szarvas
Eötvös Lorand University
Department of Numerical Analysis
1/C, Pázmány P. sétány
1117 Budapest, Hungary
e-mail: szarvaskristof@gmail.com

Ferenc Weisz
Eötvös Lorand University
Department of Numerical Analysis
1/C, Pázmány P. sétány,
1117 Budapest, Hungary
e-mail: weisz@inf.elte.hu

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 513–521

The isotomic transformation in the hyperbolic plane

Paul A. Blaga

Abstract. In this note, we introduce a hyperbolic analogue of the isotomic transformation, originally defined for Euclidean triangle and we investigate some of its properties.

Mathematics Subject Classification (2010): 51M09, 51M10.

Keywords: Hyperbolic plane, transformations.

1. Introduction

The aim of this paper is to introduce the *isotomic transformation* in the hyperbolic geometry and investigate some of its basic properties. There are several approaches to the geometry of the hyperbolic plane. We found that the approach most suitable for our purposes is the so-called *Cayley-Klein approach* (see [4], [9] or [5]). In this approach, the hyperbolic plane is thought of as being a region of the projective plane, bounded by a real, nondegenerate projective conic. This conic is defined by a polarity of the real projective plane,

$$\begin{cases} x_\mu = c_{\mu\nu}\xi_\nu, \\ \xi_\mu = C_{\mu\nu}x_\nu, \end{cases} \tag{1.1}$$

where we sum after all the possible values of the indices ($\mu, \nu = 0, 1, 2$). Here (x_μ) are the point coordinates, while (ξ_μ) are the line coordinates. As we are given a hyperbolic polarity, the matrices $[c]$ and $[C]$ are both symmetric and inverse to each other. Moreover, the conic given by the point equation

$$c_{\mu\nu}x_\mu x_\nu = 0 \tag{1.2}$$

is a nondegenerate real conic, called the *Absolute*. The equation of the Absolute, written in line coordinates, is

$$C_{\mu\nu}\xi_\mu\xi_\nu = 0. \tag{1.3}$$

We notice that, usually, we prescribe the polarity, therefore the equations (1.1) give, implicitly, the definition of the coordinates (homogeneous coordinates, of course).

The system of coordinates we are going to use is the system of *barycentric* (or *areal*) coordinates, introduced by Sommerville, in 1932 ([10]) by another method. For these coordinates, the two matrices that define the polarity are

$$[c_{\mu\nu}] = \begin{pmatrix} 1 & \cosh c & \cosh b \\ \cosh c & 1 & \cosh a \\ \cosh b & \cosh a & 1 \end{pmatrix} \tag{1.4}$$

and

$$[C_{\mu\nu}] = \frac{1}{\gamma} \begin{pmatrix} -\sinh^2 a & \sinh a \sinh b \cos C & \sinh a \sinh c \cos B \\ \sinh a \sinh b \cos C & -\sinh^2 b & \sinh b \sinh b \cos A \\ \sinh a \sinh c \cos B & \sinh b \sinh b \cos A & -\sinh^2 c \end{pmatrix}, \tag{1.5}$$

where

$$\gamma = 1 + 2\cosh a \cosh b \cosh c - \cosh^2 a - \cosh^2 b - \cosh^2 c > 0$$

is the determinant of the matrix $[c_{\mu\nu}]$.

It should be clear that the coordinates defined by a polarity are defined for any point of the projective plane, not just for the ordinary points. In contrast, the barycentric coordinates, defined by Sommerville by using a triangle and a unit point, as it is standard in projective geometry, are valid only for ordinary points. Thus, for Sommerville, the barycentric (point) coordinates are defined by

$$\begin{cases} X_0 = \sinh a \sinh u, \\ X_1 = \sinh b \sinh v, \\ X_0 = \sinh c \sinh w, \end{cases} \tag{1.6}$$

where a, b, c are the lengths of the sides BC, CA and AB, respectively, of the reference triangle, while u, v, w are the distances from the current point to these sides. These definitions are equivalent to those given by polarisation for ordinary points in the hyperbolic plane, but they don't make sense for ideal or ultra-infinite points for the very simple reason that the lengths u, v, w are not defined.

The homogeneous coordinates in the hyperbolic plane have not been very popular, lately. Most of the works on analytic hyperbolic geometry prefer the use of Cartesian or polar coordinates. Nevertheless, if somebody wants to investigate problems related to a hyperbolic triangle, it is more convenient to use some coordinates related closely to the triangle itself. Recently, (see [11]) Ungar introduced a set of barycentric coordinates, in the framework of the so-called *Einstein velocity space* model of hyperbolic geometry. We feel, however, that the Cayley-Klein (projective) model is closer to the intuition and, therefore, we use the barycentric coordinates introduced by Sommerville ([10]) and, afterwards, reformulated by Coxeter ([4]).

We introduce the following notations, that we will use again and again. For more details, see [2]. First of all, we denote by H^2 the hyperbolic plane, as a subset of the real projective plane. If (x) and (y) are two points, while $[\xi]$ and $[\eta]$ are two lines (from the real projective plane!), then

1. $(x, y) = c_{\mu\nu} x_\mu y_\nu$;
2. $[\xi, \eta] = C_{\mu\nu} \xi_\mu \eta_\nu$;
3. $\{x, \eta\} = x_\mu \xi_\eta$;

4. $\{\xi, y\} = \xi_\mu y_\nu$,

where, as usually, we sum after all the possible values of the indices. We mention that, if the lines $[\xi]$ and $[\eta]$ are the polars of the points (x) and (y), the all the all four brackets defined above are equal.

The brackets we introduce are very convenient for describing different entities related to hyperbolic geometry. We mention only some of them, that will be used in the paper.

1. The equation of the Absolute is $(x, x) = 0$ (in point coordinates) or $[\xi, \xi] = 0$ (in line coordinates);
2. a point (x) is an ordinary point iff $(x, x) > 0$;
3. a point (x) is an ultra-infinite point iff $(x, x) < 0$;
4. a line $[\xi]$ is ultra-infinite (i.e. lies outside the hyperbolic plane) iff $[\xi, \xi] > 0$;
5. the polar of any ordinary point of the hyperbolic plane is ultra-infinite and the polar of ultra-infinite point is an ordinary line;
6. the lines $[\xi]$ and $[\eta]$ are perpendicular iff $[\xi, \eta] = 0$;
7. if α is the angle between two lines, $[\xi]$ and $[\eta]$, then
$$\cos^2 \alpha = \frac{[\xi, \eta]^2}{[\xi, \xi] \cdot [\eta, \eta]}.$$
Notice that this relations makes sense iff the two lines are either both ordinary, either both ultra-infinite. We cannot compute, for instance, the angle between an ordinar line and an ultra-ideal one.
8. If (x) and (y) are two points and d is the distance between them, then
$$\cosh d = \frac{|(x, y)|}{\sqrt{(x, x) \cot (y, y)}}.$$
Again, we can only compute distances between two ordinary points or two ultra-infinite points, but not between an ordinary point and an ultra-infinite point.
9. We can, also, compute the distance d between a point (x) and a line $[\xi]$, by using the formula
$$\sinh d = \frac{|\{x, \xi\}|}{\sqrt{(x, x)} \cdot \sqrt{-[\xi, \xi]}}.$$
This distance can be computed iff both (x) and (ξ) are ordinary.

From now on, we shall use exclusively barycentric coordinates and we shall denote them with capital letters, (X_0, X_1, X_2). In this coordinates, as we saw, we have
$$\begin{aligned}(X, X) = (X_0)^2 + (X_1)^2 + (X_2)^2 + 2 \cosh c \cdot X_0 X_1 + \\ + 2 \cosh b \cdot X_0 X_2 + 2 \cosh a \cdot X_1 X_2.\end{aligned} \tag{1.7}$$

2. The transformation

The isotomic transformation for Euclidean triangles has been introduced by G. de Longchamps in 1866 (see [6] and [7]). We shall give here a similar definition, using the hyperbolic barycentric coordinates.

Definition 2.1. *We define, by analogy to the Euclidean case, the* isotomic transformation *as being a map*

$$\mathrm{Isot} : \mathbb{P}^2(\mathbb{R}) \setminus \mathcal{T} \to \mathbb{P}^2(\mathbb{R}),$$

defined by

$$\mathrm{Isot}\,(X_0, X_1, X_2) = \left(\frac{1}{X_0}, \frac{1}{X_1}, \frac{1}{X_2}\right). \tag{2.1}$$

Here $\mathcal{T}$ is the union of the three sides of the triangle ABC (thought of as projective *lines). We shall say that the points M and M' form an* isotomic pair. *We shall also say that M' is the* isotomic conjugate *or the* isotomic inverse *of M. We may, as well, say, again inspired from the classical case, that the two points are* reciprocal *(with respect to the triangle ABC).*

As Isot is defined on $\mathbb{P}^2(\mathbb{R}) \setminus \mathcal{T}$, none of the coordinates X_i vanishes, hence Isot is well defined.

Remark 2.2.
1. We might have tried, as well, to define the isotomic transformation just on points of the hyperbolic plane. Nevertheless, as we shall see later, the image of an ordinary point through the isotomic transformation is not always an ordinary point, it might be ideal or ultra-infinite.
2. By looking at the formula (2.1), the reader may think that the definition of the isotomic transformation is identical to the definition from the Euclidean/projective case. This is not the case, however, because the barycentric coordinates from the hyperbolic case are not the same with the classical barycenter coordinates.

Definition 2.3. *We shall say that two points on the side BC of hyperbolic triangle ABC (ordinary, ideal or ultra-infinite) are* isotomically symmetric *with respect to the midpoint A' of the side BC if they coordinates are $A_1(0, \alpha_1, \alpha_2)$ and $A'(0, 1/\alpha_1, 1/\alpha_2)$ or $A_1'(0, \alpha_2, \alpha_1)$.*

The following theorem justifies the name of "isotomic transformation".

Theorem 2.4. *If A_1 is an* ordinary *point on BC, then its isotomic symmetric A_1' is, also, ordinary, and $A'A_1 = A'A_1'$ (as hyperbolic lengths). Moreover, is A_1 is either ideal or ultra-infinite, the same holds true for A_1'.*

Proof. We notice, first of all, that $(A_1, A_1) = (A_1', A_1')$, therefore the two numbers are simultaneously zero, positive or negative. As such, the points A_1 and A_1' have the same character(ordinary, ideal or ultra-infinite).

Thus, we will consider the particular case when A_1 is an ordinary point, and, of course, the same is true for A_1'. We know, already, that the barycentric coordinates of A' are $(0, 1, 1)$. We compute, first, the length of the segment $A'A_1$. We have

$$\cosh A'A_1 = \frac{|(A', A_1)|}{\sqrt{(A', A') \cdot (A_1, A_1)}}.$$

On the other hand,

$$\begin{aligned}(A', A') &= 2(1+\cosh a) = 4\cosh^2\frac{a}{2},\\ (A_1, A_1) &= \alpha_1^2+\alpha_2^2+2\alpha_1\alpha_2\cosh a,\\ (A', A_1) &= 2(\alpha_1+\alpha_2)\cosh^2\frac{a}{2}.\end{aligned}$$

We have, therefore

$$\cosh A'A_1 = \frac{\left|2(\alpha_1+\alpha_2)\cosh^2\frac{a}{2}\right|}{2\cosh\frac{a}{2}\sqrt{\alpha_1^2+\alpha_2^2+2\alpha_1\alpha_2\cosh a}} = \frac{|\alpha_1+\alpha_2|\cosh\frac{a}{2}}{\sqrt{\alpha_1^2+\alpha_2^2+2\alpha_1\alpha_2\cosh a}}.$$

Now, it is easy to check that $(A_1', A_1') = (A_1, A_1)$ and $(A', A_1') = (A', A_1)$, therefore $\cosh A'A_1' = \cosh A'A_1$, hence $A'A_1' = A'A_1$. □

The previous theorem justifies the following definition:

Definition 2.5. *Two cevians of a hyperbolic triangle ABC, starting from the same vertex, are called* isotomic *if they cut the opposite side at isotomically symmetric points. We shall, also, say that the cevians are* isotomically conjugated.

Theorem 2.6. *If three cevians (starting from different vertices) are concurrent at a point, then their isotomic conjugates are also concurrent and the intersection points are, as well, isotomically conjugated.*

Proof. Let $M\left(X_0^0, X_1^0, X_2^0\right)$ be the intersection point of the three given cevians. It is easy to see that the equations of these cevians are

$$\begin{aligned}AM &: X_2^0X_1 - X_1^0X_2 = 0,\\ BM &: X_2^0X_0 - X_0^0X_2 = 0,\\ CM &: X_1^0X_0 - X_0^0X_1 = 0.\end{aligned}$$

As such, their intersection points with the sides BC, CA and AB, respectively, will be $A_1\left(0, X_1^0, X_2^0\right)$, $B_1\left(X_0^0, 0, X_2^0\right)$ and $C_1\left(X_0^0, X_1^0, 0\right)$, respectively. Then, according to the previous lemma, their symmetrics with respect to the midpoints of the respective sides will be $A_1'\left(0, X_2^0, X_1^0\right)$, $B_1'\left(X_2^0, 0, X_0^0\right)$ and $C_1'\left(X_1^0, X_0^0, 0\right)$, respectively.

We are, thus, led to the equations of the isotomically conjugated of the cevians AM, BM and CM:

$$\begin{aligned}AA_1' &: X_1^0X_1 - X_2^0X_2 = 0,\\ BB_1' &: X_0^0X_0 - X_2^0X_2 = 0,\\ CC_1' &: X_0^0X_0 - X_1^0X_1 = 0.\end{aligned}$$

It turns out that the three cevians *do* intersect, at the point

$$M'\left(1/X_0^0, 1/X_1^0, 1/X_2^0\right),$$

as we expected. □

3. The Steiner quadratix

As we mentioned before, one of the advantages of the Cayley-Klein approach to the hyperbolic plane is that, in this model, the hyperbolic geometry is, in a certain sense, a "sub-geometry" of the real projective plane. As such, we have access to all the points of the projective plane, although they are not treated on the same footing. We can treat any pair of lines as being intersecting lines, but some of them intersect at ideal or ultra-infinite points. We can write the equation of the line passing through an arbitrary pair of points, but some of the lines are either ultra-infinite (they don't intersect the hyperbolic plane) or lines at infinity (they are tangents to the Absolute). The downside is that we cannot compute distances and lengths when ideal or ultra-infinite points and lengths are involved.

We turn, for a while, to the "classical" language of hyperbolic geometry. Then, for instance, the theorem 2.6 can be reformulated as

Theorem 3.1. *Consider three cevians of a given triangle, starting from different vertices. If the three cevians belong to the same pencil of lines (concurrent, ultra-parallel or parallel), then the their isotomic conjugates also belong to the same pencil.*

The point is that we don't know *what kind* of pencil.

We ask the following question: *When three concurrent cevians of a given hyperbolic triangle, starting from different vertices turn, through the isotomic transformation, into three parallel lines?*

We know the answer in the classical Euclidean (or, rather, projective case): when they intersect on the line at infinity. But the things are similar, here, only that the line at infinity gets replaced by the Absolute. Indeed, three lines belong to the same pencil of parallel lines iff they intersect (according to the Cayley-Klein view of the hyperbolic geometry) on the Absolute. Therefore, we have the following theorem:

Theorem 3.2. *Let us assume that the cevians AA_1, BB_1 and CC_1 of the triangle ABC intersect at a point $M(X_0^0, X_1^0, X_2^0)$ (ordinary, ideal or ultra-infinite). Then the isotomic conjugates of the cevians are parallel (i.e. they intersect at an ideal point) iff M belongs to the curve*

$$\begin{aligned} &X_0^2X_1^2 + X_0^2X_2^2 + X_1^2X_2^2 + 2\cosh c X_0X_1X_2^2 + \\ &+ 2\cosh b X_0X_1^2X_2 + 2\cosh a X_0^2X_1X_2 = 0. \end{aligned} \tag{3.1}$$

*We shall call the curve (*3.1*) the* Steiner quadratix*. It is the hyperbolic analogue of the first Steiner ellipse.*

Proof. The isotomic conjugates of the cevians are parallel to each other iff they intersect at a point of the Absolute. But, as we saw earlier, the conjugates intersect at the point $M'\left(1/X_0^0, 1/X_1^0, 1/X_2^0\right)$. M' belongs to the Absolute iff

$$\frac{1}{\left(X_0^0\right)^2} + \frac{1}{\left(X_1^0\right)^2} + \frac{1}{\left(X_2^0\right)^2} + \frac{2}{X_0^0\cdot X_1^0}\cosh c + \frac{2}{X_0^0\cdot X_2^0}\cosh b + \frac{2}{X_1^0\cdot X_2^0}\cosh a = 0$$

or

$$\begin{aligned}&\left(X_0^0\right)^2 \cdot \left(X_1^0\right)^2 + \left(X_0^0\right)^2 \cdot \left(X_2^0\right)^2 + \left(X_1^0\right)^2 \cdot \left(X_2^0\right)^2 + \\ &+ 2\cosh c \cdot X_0^0 \cdot X_1^0 \cdot \left(X_2^0\right)^2 + 2\cosh b \cdot X_0^0 \cdot \left(X_1^0\right)^2 \cdot X_2^0 + \\ &+ 2\cosh a \cdot \left(X_0^0\right)^2 \cdot X_1^0 \cdot X_2^0 = 0,\end{aligned}$$

which shows that the point M belongs to the Steiner quadratix. □

Remark 3.3. Clearly, the vertices of the triangle ABC belong to the Steiner quadratix, which, thus, is not empty.

4. Some remarkable pairs of isotomic points

As examples, we use the hyperbolic analogs of some classical remarkable points from the geometry of the Euclidean triangles, the Gergonne group of points and the Nagel group of points. For the Euclidean points, see [1] and [8].

4.1. The Gergonne and Nagel Points

In [3], we introduced the Gergonne and Nagel points associated to a hyperbolic triangle. Exactly as it happens for a Euclidean triangle, the Gergonne point is the point obtained by intersecting the lines connecting the vertices of a hyperbolic triangle ABC to the points of contact of the incircle with the opposite sides. The incircle is, for any hyperbolic triangle, a proper circle. For the Nagel point, the definition has to be a little bit adapted to work for an arbitrary hyperbolic triangle. Thus, the Nagel point is obtained as intersection of the lines connecting the vertices of the triangle to the points of contact of the opposite sides to the corresponding excycles. Unlike the Euclidean case, an arbitrary hyperbolic triangle doesn't always have excircles. In some cases, these circles become equidistants or horocycles. We use the term "cycle" to cover all the possible situations.

In [3], we prove that, for each situation, the three cevians really intersect and, moreover, the intersection points are always ordinary. More specifically, we were able to prove that they barycentric coordinates are identical to the barycentric coordinates of their Euclidean analogues, i.e.

- for the Gergonne point we obtain

$$\Gamma = \Gamma\left(\tan\frac{A}{2}, \tan\frac{B}{2}, \tan\frac{C}{2}\right);$$

- for the Nagel point, we obtain

$$N = N\left(\cot\frac{A}{2}, \cot\frac{B}{2}, \cot\frac{C}{2}\right).$$

Thus, the Gergonne and Nagel points are isotomic to each other.

We can see immediately, without any computation, that the points are ordinary (as all the coordinates are strictly positive, they are in the interior of the triangle) and they are isotomic to each other.

4.2. The adjoint Gergonne and Nagel points

We introduce these points in [3], by analogy to the classical case. Thus, the *adjoint Gergonne points* $\Gamma_a, \Gamma_b, \Gamma_c$ are the analogues of the Gergonne points, for the excycles. Thus, consider, for instance the excycles that is tangent to the side BC in an interior points. We connect the tangency points with the opposite vertices. We proved in [3] that they intersect at a point Γ_a (which is not necessarily ordinary) and the same happens with the other two vertices of the triangle ABC.

We get, thus, three points

$$\Gamma_a = \Gamma_a\left(-\tan\frac{A}{2}, \cot\frac{B}{2}, \cot\frac{C}{2}\right),$$
$$\Gamma_b = \Gamma_b\left(\cot\frac{A}{2}, -\tan\frac{B}{2}, \cot\frac{C}{2}\right),$$
$$\Gamma_c = \Gamma_c\left(\cot\frac{A}{2}, \cot\frac{B}{2}, -\tan\frac{C}{2}\right).$$

The lines connecting the extremities of a side of the triangle ABC to the contact points of the excycles lying within the angles adjacent to this side, situated on the extensions of the opposite sides of the one considered and the line that connects the third vertex to the contact point of the incircle to the opposite side are concurrent at a point (ordinary, ideal or ultra-infinite). We get, thus (see [3]), three points N_a, N_b, N_c, called the *adjoint Nagel points* of the triangle ABC:

$$N_a = N_a\left(-\cot\frac{A}{2}, \tan\frac{B}{2}, \tan\frac{C}{2}\right),$$
$$N_b = N_b\left(\tan\frac{A}{2}, -\cot\frac{B}{2}, \tan\frac{C}{2}\right),$$
$$N_c = N_c\left(\tan\frac{A}{2}, \tan\frac{B}{2}, -\cot\frac{C}{2}\right).$$

It can be seen that each adjoint Nagel point is the isotomic conjugate of the corresponding adjoint Gergonne point.

Acknowledgments. The author would like to thank Professor Dorin Andrica for the suggestions that led to the improvement of the paper.

References

[1] Andrica, D., Nguyen, K.L., *A note on the Nagel and Gergonne points*, Creative Math. & Inf., **17**(2008), 127-136.

[2] Blaga, P.A., *Barycentric and trilinear coordinates in the hyperbolic plane*, Automat. Comput. Appl. Math., **22**(2013), no. 1, 49-58.

[3] Blaga, P.A., *On the Gergonne and Nagel points for a hyperbolic triangle*, in preparation.

[4] Coxeter, H., *Non-Euclidean Geometry*, 5th Edition, University of Toronto Press, 1965.

[5] Kowol, G., *Projektive Geometrie und Cayley-Klein Geometrien der Ebene*, Birkhäuser, 2009.

[6] de Longchamps, G., *Mémoire sur une nouvelle méthode de transformation en géométrie*, Annales Scientifiques de l'École Normale Supérieure, Sér. 1, **3**(1866), 321-341.

[7] de Longchamps, G., *Étude de géométrie comparée, Avec applications aux sections coniques et aux courbes d'ordre supérieure, particulièrement à une famille de courbes du sixiéme ordre et de la quatriéme class*, Nouvelle Annales de Mathématiques, **12**(1866), 118-128.

[8] Mihalescu, C., *The Geometry of Remarkable Elements* (Romanian), Editura Tehnică, 1957.

[9] Onishchik, A., Sulanke, R., *Projective and Cayley-Klein Geometries*, Springer, 2006.

[10] Sommerville, D., *Metrical coordinates in non-Euclidean geometry*, Proceedings of the Edinburgh Mathematical Society, **3**(2)(1932), 16-25.

[11] Ungar, A., *Barycentric Calculus in Euclidean and Hyperbolic Geometry*, World Scientific, 2010.

Paul A. Blaga
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: `pablaga@cs.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 523–531

Circular mappings with minimal critical sets

Dorin Andrica, Adela Lupescu and Cornel Pintea

Abstract. We provide classes of manifolds M satisfying the relation $\varphi_{S^1}(M) = \varphi(M)$, we discuss the situation $\varphi_{S^1}(M) = 1$, and we formulate a circular version of the Ganea conjecture.

Mathematics Subject Classification (2010): 58E05, 57R70, 57R35, 57M10, 55M30.

Keywords: Circular function, φ-category, circular φ-category, covering map, fundamental group, compact surface, projective space, Morse function, Ganea's conjecture.

1. Introduction

The systematic study of the smooth circular functions defined on a manifold was initiated by E.Pitcher in the articles [23],[24]. His goal was to extend in this context the classical Morse theory for real-valued functions. The importance of this study was pointed out by Novikov in the early 1980s. The Morse - Novikov theory is now a large and actively developing domain of Differential Topology, with applications and connections to many geometrical problems (see the monographs [11] and [21]).

The *φ-category* of a manifold M is $\varphi(M) = \min\{\mu(f) : f \in \mathcal{C}^\infty(M, \mathbb{R})\}$, and it represents the φ-category of the pair $(M, \mathbb{R})$.

The *circular φ-category* of a manifold M was introduced in the paper [4]. It is defined as the φ-category of the pair (M, S^1) corresponding to the family $C^\infty(M, S^1)$, where S^1 is the unit circle. That is

$$\varphi_{S^1}(M) = \min\{\mu(f) : f \in \mathcal{C}^\infty(M, S^1)\},$$

where $\mu(f)$ denotes the cardinality of the critical set of mapping $f : M \to S^1$.

If we restrict the class of smooth functions to its subclass of Morse functions, then we obtain, in the real case, the *Morse-Smale characteristic*

$$\gamma(M) = \min\{\mu(f) : f \in \mathcal{C}^\infty(M, \mathbb{R}),\ f - \text{ Morse}\},$$

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

and the *circular Morse-Smale characteristic*

$$\gamma_{S^1}(M) = \min\{\mu(f) : f \in \mathcal{C}^\infty(M, S^1),\ f - \text{ circular Morse function}\}$$

in the circular case. For the Morse-Smale characteristic of the closed surfaces we refere the reader to [5]. The inequalities

$$\varphi_{S^1}(M) \leq \varphi(M), \quad \gamma_{S^1}(M) \leq \gamma(M) \tag{1.1}$$

rely on the property $C(\exp \circ g) = C(g)$ which is quite obvious due to the property of the exponential map to be a local diffeomorphism. Thus, the quality of a real valued function $g : M \longrightarrow \mathbb{R}$ to be Morse is transmited to the function $\exp \circ g$ and the second inequality of (1.1) is also justified. On the other hand, the inequalities

$$\varphi(M) \leq \gamma(M), \quad \varphi_{S^1}(M) \leq \gamma_{S^1}(M) \tag{1.2}$$

are obvious.

One of the main goals of this paper is to provide classes of manifolds M satisfying (1.1) with equality, i.e. $\varphi_{S^1}(M) = \varphi(M)$ and $\gamma_{S^1}(M) = \gamma(M)$. In the last section we discuss the situation $\varphi_{S^1}(M) = 1$ and we formulate a circular version of the Ganea conjecture.

2. Manifolds with $\varphi_{S^1}(M) = \varphi(M)$ and $\gamma_{S^1}(M) = \gamma(M)$

Let us first observe that the inequality $\varphi_{S^1}(M) \leq \varphi(M)$ ensured by (1.1) can be strict. Indeed, the m-dimensional torus $T^m = S^1 \times \cdots \times S^1$ (m times) has, according to [1, Example 3.6.16], the φ-category $\varphi(T^m) = m + 1$. On the other hand, every projection $T^m \to S^1$ is a trivial differentiable fibration, hence it has no critical points, implying $\varphi_{S^1}(T^m) = 0$. This example is part of the following more general remark. For a closed manifold M we have $\varphi_{S^1}(M) = 0$ if and only if there is a differentiable fibration $M \to S^1$. Indeed, the existence of a differentiable fibration $M \to S^1$ ensures the equality $\varphi_{S^1}(M) = 0$, as the fibration itself has no critical points at all. Conversely, the equality $\varphi_{S^1}(M) = 0$ ensures the existence of a submersion $M \to S^1$, which is also proper, as its inverse images of the compact sets in S^1 are obviously compact. Thus, by the well-known Ehresmann's fibration theorem (see for instance the reference [10, p. 15]) one can conclude that our submersion is actually a locally trivial fibration. Note that this property works for arbitrary closed target manifolds, not just for the circle S^1.

Assume that every smooth (Morse) circle valued function $f : M \longrightarrow S^1$ can be lifted to a smooth (Morse) real valued function $\tilde{f} : M \longrightarrow \mathbb{R}$, i.e. we have $\exp \circ \tilde{f} = f$. Since the universal cover $\exp : \mathbb{R} \longrightarrow S^1$ is a local diffeomorphism, it follows that $\mu(f) = \mu(\tilde{f}) \geq \varphi(M)$, for every smooth function $f : M \longrightarrow S^1$. This shows that the inequalities $\varphi_{S^1}(M) \geq \varphi(M)$, $\gamma_{S^1}(M) \geq \gamma(M)$ hold, which combined to the general inequalities (1.1), leads to the following result.

Proposition 2.1. ([6]) *Let M be a connected smooth manifold. If M satisfies the lifting property* $\mathrm{Hom}\,(\pi(M), \mathbb{Z}) = 0$, *then $\varphi_{S^1}(M) = \varphi(M)$ and $\gamma_{S^1}(M) = \gamma(M)$. In particular $\varphi_{S^1}(M) = \varphi(M)$ and $\gamma_{S^1}(M) = \gamma(M)$ whenever the fundamental group of M is a torsion group.*

2.1. On the categories of some Grassmann manifolds

Proposition 2.2. *If $n \geq 2$ is an integer, then $\varphi_{S^1}(S^n) = \varphi(S^n) = \gamma_{S^1}(S^n) = \gamma(S^n) = 2$ and*

$$\begin{aligned} \varphi_{S^1}(\mathbb{RP}^n) &= \varphi(\mathbb{RP}^n) = \gamma_{S^1}(\mathbb{RP}^n) = \gamma(\mathbb{RP}^n) = \mathrm{cat}(\mathbb{RP}^n) = \\ \varphi_{S^1}(\mathbb{CP}^n) &= \varphi(\mathbb{CP}^n) = \gamma_{S^1}(\mathbb{CP}^n) = \gamma(\mathbb{CP}^n) = \mathrm{cat}(\mathbb{CP}^n) = n+1, \end{aligned}$$

where $\mathrm{cat}(\mathbb{CP}^n)$ *stands for the Lusternik-Schnirelmann category of the complex projective space $\mathbb{CP}^n$.*

Proof. We shall only justify the equalities

$$\varphi_{S^1}(\mathbb{CP}^n) = \varphi(\mathbb{CP}^n) = \gamma(\mathbb{CP}^n) = \gamma_{S^1}(\mathbb{CP}^n) = \mathrm{cat}(\mathbb{CP}^n) = n+1,$$

as the other equalities have been already proved in [6]. The equalities $\varphi_{S^1}(\mathbb{CP}^n) = \varphi(\mathbb{CP}^n)$ and $\gamma_{S^1}(\mathbb{CP}^n) = \gamma(\mathbb{CP}^n)$ follow from Proposition 2.1 taking into account the simply-connectedness of the complex projective space $\mathbb{CP}^n$. On the other hand the inequality $\varphi(\mathbb{CP}^n) \leq \gamma(\mathbb{CP}^n)$ follow from the general inequality (1.2). Therefore $\varphi_{S^1}(\mathbb{CP}^n) = \varphi(\mathbb{CP}^n) \leq \gamma(\mathbb{CP}^n) = \gamma_{S^1}(\mathbb{CP}^n)$. In order to prove the equalities $\gamma(\mathbb{CP}^n) = \mathrm{cat}(\mathbb{CP}^n) = n+1$ we observe that

$$\gamma(\mathbb{CP}^n) \leq \mu(f) = \mathrm{card}(C(f)) = n+1,$$

as the function

$$f : \mathbb{CP}^n \longrightarrow \mathbb{R}, \quad f([z_1, \ldots, z_{n+1}]) = \frac{|z_1|^2 + 2|z_2|^2 + \cdots + n|z_n|^2 + (n+1)|z_{n+1}|^2}{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 + |z_{n+1}|^2}.$$

is a Morse function with the $n+1$ critical points

$$[1, 0, \ldots, 0], [0, 1, \ldots, 0], \ldots, [0, 0, \ldots, 1] \in \mathbb{CP}^n \text{ [19, p. 89].}$$

Thus $\varphi(\mathbb{CP}^n) \leq \gamma(\mathbb{CP}^n) \leq n+1$. Finally, we use the well-known inequality $\varphi(\mathbb{CP}^n) \geq \mathrm{cat}(\mathbb{CP}^n)$ and the relation $\mathrm{cat}(\mathbb{CP}^n) = n+1$ [9, p. 3, pp. 7-13]. □

Note that the equalities $\varphi_{S^1}(\mathbb{RP}^n) = \varphi(\mathbb{RP}^n) = \mathrm{cat}(\mathbb{RP}^n) = n+1$ are being similarly proved in [6] by using the $\mathbb{Z}_2$ structure of the fundamental group of $\mathbb{RP}^n$, the Morse function

$$F_n : \mathbb{RP}^n \longrightarrow \mathbb{R}, \; F_n([x_1, \ldots, x_{n+1}]) = \frac{x_1^2 + 2x_2^2 + \cdots + nx_n^2 + (n+1)x_{n+1}^2}{x_1^2 + x_2^2 + \cdots + x_n^2 + x_{n+1}^2},$$

whose critical set is $C(F_n) = \{[1, 0, \ldots, 0], [0, 1, \ldots, 0], \ldots, [0, 0, \ldots, 1]\}$, and the well-known relations $\varphi(\mathbb{RP}^n) \geq \mathrm{cat}(\mathbb{RP}^n) = n+1$ [22, pp. 190-192].

Proposition 2.3. *If $n \geq 3$ and $1 \leq k \leq n-1$, then*

$$\varphi_{S^1}(G_{k,n}) = \varphi(G_{k,n}) \leq \gamma(G_{k,n}) = \gamma_{S^1}(G_{k,n}) \leq \binom{n+k}{k},$$

where $G_{k,n}$ stands for the Grassmann manifold of all k-dimensional subspaces of the space $\mathbb{R}^{n+k}$.

Proof. The equalities $\varphi_{S^1}(G_{k,n}) = \varphi(G_{k,n})$ and $\gamma_{S^1}(G_{k,n}) = \gamma(G_{k,n})$ follow due to Proposition 2.1 and the $\mathbb{Z}_2$ structure of the fundamental group of $G_{k,n}$. Thus $\varphi_{S^1}(G_{k,n}) = \varphi(G_{k,n}) \leq \gamma(G_{k,n}) = \gamma_{S^1}(G_{k,n})$. Recall that $G_{k,n}$ can be embedded into the projective space $\mathbb{RP}^{n+k-1}$ via the Plücker embedding

$$p : G_{k,n} \hookrightarrow P\left(\Lambda^k(\mathbb{R}^{n+k})\right) = \mathbb{RP}^{d(n,k)-1},\ p(W) = [w_1 \wedge \cdots \wedge w_k],$$

where $\{w_1, \ldots, w_k\}$ is an arbitrary basis of W and $d(n,k)$ stands for the dimension of $\Lambda^k(\mathbb{R}^{n+k})$, i.e.

$$d(k,n) = \binom{n+k}{k}.$$

The composed function $F_{d(k,n)-1} \circ p : G_{k,n} \longrightarrow \mathbb{R}$ is, according to Hangan [15], a Morse function with $d(k,n)$ critical points and show that $\gamma(G_{k,n}) \leq \mu\left(F_{d(k,n)-1} \circ p\right) = d(k,n)$. □

Corollary 2.4. *If $n = 1$ or $k = 1$ or ($n = 2$ and $k = 2p-1$ for some p) or ($n = 2p-1$ and $k = 2$), then $nk \leq \varphi_{S^1}(G_{k,n}) = \varphi(G_{k,n}) \leq \gamma_{S^1}(G_{k,n}) = \gamma(G_{k,n}) \leq \binom{n+k}{k}$.*

Proof. We only need to use the inequality $\varphi(G_{k,n}) \geq \operatorname{cat}(G_{k,n})$ and the equalities $\operatorname{cat}(G_{k,n}) = nk$, proved by Berstein [8], whenever $n = 1$ or $k = 1$ or ($n = 2$ and $k = 2p-1$ for some p) or ($n = 2p-1$ and $k = 2$). □

2.2. On the categories of some classical Lie groups

Proposition 2.5. *If $n \geq 3$, then the following relations hold*

$$\varphi_{S^1}(SO(n)) = \varphi(SO(n)) \leq \gamma(SO(n)) = \gamma_{S^1}(SO(n)) \leq 2^{n-1}.$$

Proof. The equalities $\varphi_{S^1}(SO(n)) = \varphi(SO(n))$ and $\gamma(SO(n)) = \gamma_{S^1}(SO(n))$ follow from Proposition 2.1 by using the fundamental group of $SO(n)$ which is $\mathbb{Z}_2$. Thus $\varphi_{S^1}(SO(n)) = \varphi(SO(n)) \leq \gamma(SO(n)) = \gamma_{S^1}(SO(n))$. In order to prove the inequality $\gamma(SO(n)) \leq 2^{n-1}$ we observe that

$$\gamma(SO(n)) \leq \mu(f) = \operatorname{card}(C(f)) = 2^{n-1},$$

where $f : SO(n) \longrightarrow \mathbb{R}$, $f([a_{ij}]_{n\times n}) = a_{11} + 2a_{22} + \cdots + na_{nn}$ is a Morse function. The critical set of f consists in all diagonal matrices D with ± 1 as diagonal entries and $\det(D) = 1$ [19, p. 92]. In other words, $C(f)$ is the collection of all diagonal matrices D with an even number of -1 on the main diagonal. The number of such diagonal matrices is $\binom{n}{0} + \binom{n}{2} + \cdots = 2^{n-1}$, i.e. $\mu(f) = 2^{n-1}$. □

Remark 2.6. *If $n \geq 3$, then the following relations hold*

$$\varphi_{S^1}(Spin(n)) = \varphi(Spin(n)) \leq \gamma(Spin(n)) = \gamma_{S^1}(Spin(n)) \leq 2^n.$$

Moreover, $\varphi(Spin(9)) \geq cat(Spin(9)) = 9$ [17]. We only need to justify the inequality $\gamma_{S^1}(Spin(n)) \leq 2^n$, as the other ones rely on the general inequalities (1.2) and the simply connectedness of $Spin(n)$. The inequlity $\gamma_{S^1}(Spin(n)) \leq 2^n$ follows from the *general inequality* $\gamma_{S^1}(\tilde{M}) \leq k \cdot \gamma_{S^1}(M)$, *where* $\tilde{M}$ *is a* k*-fold cover of* M [5, Proposition 1.5], *taking into account that the universal cover* $Spin(n) \longrightarrow SO(n)$ *is a* 2*-fold cover.*

Corollary 2.7. $9 \leq \varphi(SO(5)) = \varphi_{S^1}(SO(5)) \leq \gamma_{S^1}(SO(5)) = \gamma(SO(5)) \leq 16.$

Proof. The relations $\varphi(SO(5)) = \varphi_{S^1}(SO(5)) \leq \gamma_{S^1}(SO(5)) = \gamma(SO(5)) \leq 16$ follow from Proposition 2.5 and the left hand side inequality follows by means of the following well-known relations $\varphi(SO(5)) \geq \operatorname{cat}(SO(5))$ and $\operatorname{cat}(SO(5)) = 9$ [9, p. 279], [18]. □

Unfortunately, we do not know at this moment the precise values of these categories among the values $9, 10, \dots, 16$.

Proposition 2.8. *The following relations hold:*

1. $n \leq \varphi(U(n)) \leq \gamma(U(n)) \leq 2^n$.
2. $n - 1 \leq \varphi_{S^1}(SU(n)) = \varphi(SU(n)) \leq \gamma(SU(n)) = \gamma_{S^1}(SU(n)) \leq 2^{n-1}$.

Proof. (1) In order to prove the inequality $\gamma(U(n)) \leq 2^n$ we recall that

$$\gamma(U(n)) \leq \mu(f) = \operatorname{card}(C(f)) = 2^n,$$

where $f : U(n) \longrightarrow \mathbb{R}$, $f([z_{ij}]_{n\times n}) = \operatorname{Re}(z_{11} + 2z_{22} + \cdots + nz_{nn})$, which is a Morse function and its critical set consists in all diagonal matrices D with ± 1 as diagonal entries [19, p. 98]. The number of such diagonal matrices is obviously 2^n. For the left-hand-side inequality we have $\varphi(U(n)) \geq \operatorname{cat}(U(n))$ and $\operatorname{cat}(U(n)) = n$ [25].

(2) The equalities $\varphi_{S^1}(SU(n)) = \varphi(SU(n))$ and $\gamma_{S^1}(SU(n)) = \gamma(SU(n))$ follows from Proposition 2.1 by using the simply conectedness of $SU(n)$. Consequently $\varphi_{S^1}(SU(n)) = \varphi(SU(n)) \leq \gamma(SU(n)) = \gamma_{S^1}(SU(n))$. In order to prove the inequality $\gamma(SU(n)) \leq 2^{n-1}$ we observe that

$$\gamma(SU(n)) \leq \mu\left(f\big|_{SU(n)}\right) = \operatorname{card}\left(C\left(f\big|_{SU(n)}\right)\right) = 2^{n-1},$$

as the restricted function $f\big|_{SU(n)}$ is also a Morse function and its critical set consists in all diagonal matrices D with ± 1 as diagonal entries and $\det(D) = 1$ [19, p. 99]. In other words, $C\left(f\big|_{SU(n)}\right)$ is the collection of all diagonal matrices D with an even number of -1 on the main diagonal. The number of such diagonal matrices is $\binom{n}{0} + \binom{n}{2} + \cdots = 2^{n-1}$, i.e. $\mu(f) = 2^{n-1}$. The left-hand-side inequality follows by means of the relations $\varphi(SU(n)) \geq \operatorname{cat}(SU(n))$ and $\operatorname{cat}(SU(n)) = n - 1$ [25]. □

Remark 2.9. *The inequality* $\varphi(U(n)) \leq \varphi_{S^1}(U(n))$ *might be strict as the unitary group is diffeomorphic (but not isomorphic) to the product* $SU(n) \times S^1$ [19, p. 103] *and Proposition* 2.1 *does not apply, since the fundamental group of* $U(n)$ *is therefore* $\mathbb{Z}$.

2.3. On the categories of some products and connected sums

In this subsection we shall rehearse several computations of (circular) φ-category proved in the previous work [6].

If $k, l, m_1, \ldots, m_k \geq 2$, are integers, then the following relations hold:

1. $\varphi_{S^1}(S^{m_1} \times \cdots \times S^{m_k}) = \varphi(S^{m_1} \times \cdots \times S^{m_k}) = k+1$.
2. $\varphi_{S^1}(\mathbb{RP}^{m_1} \times \cdots \times \mathbb{RP}^{m_k}) = \varphi(\mathbb{RP}^{m_1} \times \cdots \times \mathbb{RP}^{m_k}) \leq m_1 + m_2 + \cdots + m_k + 1$.
3. $\varphi_{S^1}(L(7,1) \times S^4) = \varphi(L(7,1) \times S^4) = \varphi_{S^1}(L(7,1) \times S^4) = \varphi(L(7,1) \times S^4) = 5$, where $L(r,s)$ is the lens space of dimension 3 of type (r,s).
4. $\varphi_{S^1}(\mathbb{RP}^k \times S^l) = \varphi(\mathbb{RP}^k \times S^l) \leq k+2$.

The proofs of the equalities

$$\varphi(S^{m_1} \times \cdots \times S^{m_k}) = k+1$$
$$\varphi(L(7,1) \times S^4) = \varphi(L(7,1) \times S^4) = 5$$

have been done by C. Gavrilă [14, Proposition 4.6, Example 4.7] and the estimate $\varphi(\mathbb{RP}^k \times S^l) \leq k+2$ relies on [14, Proposition 4.19].

An immediate consequence of Proposition 2.1 is the following

Corollary 2.10. *If $M_1^n, \ldots, M_r^n, n \geq 3$, are connected manifolds with torsion fundamental groups, then $\varphi_{S^1}(M_1 \# \cdots \# M_r) = \varphi(M_1 \# \cdots \# M_r)$. In particular the following equality $\varphi_{S^1}(r\mathbb{RP}^n) = \varphi(r\mathbb{RP}^n)$ holds, where $r\mathbb{RP}^n$ stands for the connected sum $\mathbb{RP}^n \# \cdots \# \mathbb{RP}^n$ of r copies of $\mathbb{RP}^n$.*

The following result is mentioned in the monograph [9, p. 221].

Lemma 2.11. *If M and N are closed manifolds, then the following inequality holds $\varphi(M \# N) \leq \max\{\varphi(M), \varphi(N)\}$. In particular $\varphi(X \# X) \leq \varphi(X)$ for every closed manifold X.*

Recall that P_g denotes the closed connected non-orientable surface $\mathbb{RP}^2 \# \cdots \# \mathbb{RP}^2$ of genus g, and Σ_g stands for the closed connected orientable surface $T^2 \# \cdots \# T^2$ of genus g.

Based on Corollary 2.10 and Lemma 2.11 we were able to prove in [6] the following relations

- $\varphi(\Sigma_g) = \varphi(P_g) = 3, g \geq 1$;
- $2 \leq \varphi(r\mathbb{RP}^n) = \varphi_{S^1}(r\mathbb{RP}^n) \leq n+1, \ r \geq 1, \ n \geq 3$.
- If $k, l \geq 2$ are positive integers, then

$$\varphi_{S^1}\left((S^k \times S^l) \# \cdots \# (S^k \times S^l)\right) = \varphi\left((S^k \times S^l) \# \cdots \# (S^k \times S^l)\right) = 3. \quad (2.1)$$

3. Manifolds with $\varphi_{S^1}(M) = 1$ and the circular version of the Ganea conjecture

We do not have any example of a closed manifold M such that $\operatorname{cat}(M) < \varphi(M)$, and also the equality $\operatorname{cat}(M) = \varphi(M)$ is proved only for some isolated classes of manifolds. An example in this respect is given by the connected sum

$(S^k \times S^l)\#\cdots\#(S^k \times S^l)$, $k,l \geq 2$, justified by equality in (2.1). In order to emphasize the difficulty of the above mentioned problem, assume that the equality $\mathrm{cat}(M) = \varphi(M)$ holds for every closed manifold. Let us only look to the following particular situation: $\mathrm{cat}(M) = \varphi(M) = 2$. From $\mathrm{cat}(M) = 2$ one obtains that M is a homotopy sphere. Taking into account the well-known Reeb's result, from the equality $\varphi(M) = 2$ it follows that M is a topological sphere. Therefore, the equalities $\mathrm{cat}(M) = \varphi(M) = 2$ are related to the Poincaré conjecture, proved by Perelmann, it follows for instance that for any closed manifold with $\mathrm{cat}(M) = 2$ we have $\varphi(M) = 2$ and therefore $\mathrm{cat}(M) = \varphi(M) = 2$.

Taking into account these comments, in the article [6] we have formulated the following Reeb type problem for circular functions : *Characterize the closed manifolds M^m with the property $\varphi_{S^1}(M)=1$.*
When $m = 2$, one example of such a manifold, suggested to us by L. Funar, is given by the closed orientable surface Σ_g of genus $g \geq 2$, i.e. we have the following result :

Proposition 3.1. *The following relation holds : $\varphi_{S^1}(\Sigma_g) = 1, g \geq 2$.*

Proof. We will construct a function with one critical point from Σ_g to S^1 by composing the projection $p : T^2 = S^1 \times S^1 \to S^1$, $p(x,y) = x$, with a map $f : \Sigma_g \to T^2$ having precisely one critical point. The existence of the map f is assured by [2] (see also [3] and [12]) as $\varphi(\Sigma_g, T^2) = 1$, and the projection p is a fibration, i.e. the critical set $C(p)$ is empty. Therefore, the composed function $p \circ f$ has at most one critical point as $C(p \circ f) \subseteq C(f)$ and $\mathrm{card}(C(f)) = 1$. This shows that $\varphi_{S^1}(\Sigma_g) \leq 1$. For the opposite inequality, assume that $\varphi_{S^1}(\Sigma_g) = 0$ and consider a fibration $g : \Sigma_g \to S^1$, whose fiber F is a compact one dimensional manifold without boundary, i.e. a circle or a disjoint union of circles. By applying the product property of the Euler-Poincaré characteristic associated to the fibration $F \hookrightarrow \Sigma_g \xrightarrow{g} S^1$, one obtains $2-2g = \chi(\Sigma_g) = \chi(F)\chi(S^1) = 0$ as $\chi(S^1) = 0$, a contradiction with the initial assumption $g \geq 2$. □

In what follows we rely on the following relation

$$\varphi_{S^1}(M \times N) \leq \varphi_{S^1}(M) \cdot \varphi_{S^1}(N). \tag{3.1}$$

(see [6]) in order to produce other examples of closed manifolds X with $\varphi_{S^1}(X) = 1$. In fact, we will prove that the following class of closed manifolds

$$\mathcal{M}_1 := \{X - \text{closed manifold } : \varphi_{S^1}(X) = 1 \text{ and } \chi(X) \neq 0\}$$

is closed with respect to the cross product. More precisely, we have:

Proposition 3.2. *If $M, N \in \mathcal{M}_1$, then $M \times N \in \mathcal{M}_1$.*

Proof. If $M, N \in \mathcal{M}_1$, then, due to inequality 3.1, we conclude that $\varphi_{S^1}(M \times N) \leq \varphi_{S^1}(M) \cdot \varphi_{S^1}(N) = 1$. We now assume that $\varphi_{S^1}(M \times N) = 0$, i.e. there exists a fibration $F \hookrightarrow M \times N \longrightarrow S^1$. Since the Euler-Poincaré characteristic is multiplicative with respect to fibrations and vanishes on Lie groups, we deduce that $\chi(M \times N) = \chi(F) \cdot \chi(S^1)$, i.e. $\chi(M)\chi(N) = 0$, a contradiction with the initial assumption $\chi(M)$, $\chi(N) \neq 0$. □

The following example shows the existence of even dimensional manifolds X^{2k} with $\varphi_{S^1}(X) = 1, k = 1, 2, \dots$.

Example 3.3. *If $g_1, \dots, g_k \geq 2$, then $\varphi_{S^1}(\Sigma_{g_1} \times \cdots \times \Sigma_{g_k}) = 1$, where Σ_g stands for the closed oriented surface of genus g. Moreover, if M is a closed manifold, then*

$$\varphi_{S^1}(M \times \Sigma_{g_1} \times \cdots \times \Sigma_{g_k}) \leq \varphi_{S^1}(M).$$

Ganea's conjecture is a claim in Algebraic Topology, now disproved. It states that

$$\mathrm{cat}(X \times S^n) = \mathrm{cat}(X) + 1, n > 0,$$

where $\mathrm{cat}(X)$ is the Lusternik-Schnirelmann category of the topological space X, and S^n is the n-dimensional sphere. The conjecture was formulated by T. Ganea in 1971 (see the original reference [13]). Many particular cases of this conjecture were proved, till finally N. Iwase [16] gave a counterexample in 1998. The φ-category version of Ganea's conjecture has been studied by C. Gavrilă [14]. Now we formulate the φ_{S^1}-version of this conjecture :

Conjecture. *For every closed manifold N with $\varphi_{S^1}(N) = 1$, and for every closed manifold M, the following relation holds :*

$$\varphi_{S^1}(M \times N) = \varphi_{S^1}(M).$$

References

[1] Andrica, D., *Critical Point Theory and Some Applications*, Cluj University Press, 2005.

[2] Andrica, D., Funar, L., *On smooth maps with finitely many critical points.*, J. London Math. Soc., **69**(2)(2004), 783-800.

[3] Andrica, D., Funar L., *On smooth maps with finitely many critical points. Addendum*, J. London Math. Soc., **73**(2)(2006), 231–236.

[4] Andrica, D., Mangra, D., *Morse-Smale characteristic in circle-valued Morse theory*, Acta Universitatis Apulensis, **22**(2010), 215-220.

[5] Andrica, D., Mangra, D., Pintea, C., *The circular Morse-Smale characteristic of closed surfaces*, Bull. Math. Soc. Sci. Math. Roumanie, **57**(105)(2014), no. 3, 235-242.

[6] Andrica, D., Mangra, D., Pintea, C., *Aspects of Global Analysis of Circle-Valued Mappings*, in "Topics in Mathematical Analysis and Applications", L. Tóth and Th.M. Rassias, Eds., 79-109, Springer, 2014.

[7] Andrica, D., Pintea, C., *Recent results on the size of critical sets*, in "Essays in Mathematics and Applications" (In honor of Stephen Smale's 80th Birthday), P. Pardalos and Th.M. Rassias, Eds., 17-35, Springer, 2012.

[8] Berstein, I., *On the Lusternik-Schnirelmann category of Grassmannians*, Proc. Camb. Phil. Soc., **79**(1976), 129-134.

[9] Cornea, O., Lupton, L., Oprea, J., Tanré, T., *Lusternik-Schnirelmann category*, Mathematical Surveys and Monographs, vol. 103, American Mathematical Society, Providence, RI, 2003.

[10] Dimca, A., *Singularities and Topology of Hypersurfaces*, Springer-Verlag, Berlin, 1992.

[11] Farber, M., *Topology of Closed One-Forms*, Mathematical Surveys and Monographs, Vol. 108 AMS, 2003.

[12] Funar, L., *Global classification of isolated singularities in dimensions* $(4,3)$ *and* $(8,5)$, Ann. Scuola Norm. Sup. Pisa Cl. Sci., **10**(2011), 819-861.

[13] Ganea, T., *Some problems on numerical homotopy invariants*, Lecture Notes in Mathematics (Berlin, Springer-Verlag) **249**(1971), 13-22.

[14] Gavrilă, C., *Functions with minimal number of critical points*, Ph.D. Thesis, Heidelberg, 2001.

[15] Hangan, Th., *A Morse function on Grassmann manifolds*, Journal of Differential Geometry, **2**(1968), no. 4, 363-367.

[16] Iwase, N., *Ganea's conjecture on Lusternik-Schnirelmann category*, Bulletin of the London Mathematical Society, **30**(6)(1998), 623-634.

[17] Iwase, N., Kono, A., *Lusternick-Schnirelmann category of Spin(9)*, Trans. Amer. Math. Soc., **359**(2007), 1517-1526.

[18] James, I.M., Singhof, W., *On the category of bre bundles, Lie groups, and Frobenius maps*, Higher Homotopy Structures in Topology and Mathematical Physics (Poughkeepsie, NY, 1996), Contemp. Math., vol. 227, Amer. Math. Soc., Providence, RI, 1999, 177-189.

[19] Matsumoto, Y., *An introduction to Morse Theory*, Iwanami Series in Modern Mathematics, 1997 (Translations of Mathematical Monographs, Vol. 208 AMS, 2002).

[20] Nicolaescu, L., *An Invitation to Morse Theory* (Universitext), Springer, 2011.

[21] Pajitnov, A., *Circle-Valued Morse Theory*, Walter de Gruyter, Berlin, New York, 2006.

[22] Palais, R.S., Terng, C.L., *Critical Point Theory and Submanifold Geometry*, Lecture Notes in Mathematics 1353, Springer-Verlag, Berlin, 1988.

[23] Pitcher, E., *Critical points of a map to a circle*, Proc. Nat. Acad. Sci. USA, **25**(1939), 428-431.

[24] Pitcher, E., *Inequalities of critical point theory*, Bull. Amer. Math. Soc., **64**(1958), no. 1, 1-30.

[25] Singhof, W., *On the Lusternik-Schnirelmann category of Lie groups*, Math. Zeit., **145**(1975), 111-116.

Dorin Andrica
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: `dandrica@math.ubbcluj.ro`

Adela Lupescu
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: `adela@yahoo.com`

Cornel Pintea
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street, 400084 Cluj-Napoca, Romania
e-mail: `cpintea@math.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 533–542

Enclosing the solution set of overdetermined systems of interval linear equations

Szilvia Huszárszky and Lajos Gergó

Abstract. We describe two methods to bound the solution set of full rank interval linear equation systems $\mathbf{A}x = \mathbf{b}$ where $\mathbf{A} \in \mathbb{IR}^{m\times n}$, $m \geq n$ is a full rank interval matrix and $\mathbf{b} \in \mathbb{IR}^m$ is an interval vector. The methods are based on the concept of generalized solution of overdetermined systems of linear equations. We use two type of preconditioning the $m \times n$ system: multiplying the system with the generalized inverse of the midpoint matrix or with the transpose of the midpoint matrix. It results an $n \times n$ system which we solve using Gaussian elimination or the method provided by J. Rohn in [8]. We give some examples in which we compare the efficiency of our methods and compare the results with the interval Householder method [11].

Mathematics Subject Classification (2010): 65G06.

Keywords: Interval linear equation, overdetermined, preconditioning.

1. Introduction

An interval matrix, $\mathbf{A}$, is a matrix whose elements are intervals, an interval vector, $\mathbf{b}$, is a vector whose components are intervals. Let $\mathbf{A} = [\underline{A}, \overline{A}]$ be an $m \times n$ interval matrix and $\mathbf{b} = [\underline{b}, \overline{b}]$ an m-dimensional interval vector. We suppose that $m \geq n$ and the interval matrix $\mathbf{A}$ has full rank, i.e., all real matrices $A \in \mathbf{A}$ have full rank. Consider the set of linear equations

$$\mathbf{A}x = \mathbf{b}. \tag{1.1}$$

The set of solutions of such problem is given by

$$\sum(\mathbf{A}, \mathbf{b}) = \left\{ \widetilde{x} \in \mathbb{R}^n \mid \exists A \in \mathbf{A}, \exists b \in \mathbf{b} : \ \|A\widetilde{x} - b\| = \min_{x\in\mathbb{R}^n} \|Ax - b\| \right\},$$

i.e., the minimalization of $\|Ax - b\|$ for any $A \in \mathbf{A}$ and any $b \in \mathbf{b}$.

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

In recent years, much attention has been paid to systems of interval linear equations (1.1) with square interval matrices (see, for example, [1], [3], [4] [5]). Lot of works were performed to compute an enclosure interval vector of the set $\sum(\mathbf{A},\mathbf{b})$ which becomes as follows,

$$\sum(\mathbf{A},\mathbf{b}) = \{x \in \mathbb{R}^n \mid \exists A \in \mathbf{A}, \exists b \in \mathbf{b} : \; Ax = b\}.$$

The solution set is generally of a complicated non-convex structure. In practical computations, therefore, we look for an enclosure of it, i.e., for an interval vector $\mathbf{x}$ satisfying

$$\sum(\mathbf{A},\mathbf{b}) \subseteq \mathbf{x}.$$

If $\mathbf{A}$ is regular, then the intersection of all enclosures of $\sum(\mathbf{A},\mathbf{b})$ forms an interval vector which is called the interval hull of $\sum(\mathbf{A},\mathbf{b})$. If $\mathbf{A}$ is singular, then $\sum(\mathbf{A},\mathbf{b})$ is either empty, or unbounded and the interval hull is not defined in this case.

We note that it is especially interest if the matrix of the interval linear equation system is non-squared. Several papers have been published in this topic. Further details can be found for example in [4], [11]. In the present paper, we propose additional methods for the full rank case. The purpose of this work is to give an enclosure interval vector of the solution set of the overdetermined systems of interval linear equations. These methods are based on Hansen's preconditioning and the concept of generalized solution of full rank overdetermined systems of linear equations.

In the next section, we introduce some notation. In subsection 3.1 we describe variations of the preconditioning. This preconditioning results an $n \times n$ interval linear equation system. In subsection 3.2 we describe which methods have been used to solve the square interval linear equation system. In section 4 we give some examples in which we compare the efficiency of our methods and compare the results with the preconditioning interval Householder method [11] and with the interval Cholesky method [12] applied to the symmetric $n \times n$ interval linear system $\mathbf{A}^T\mathbf{A}x = \mathbf{A}^T\mathbf{b}$.

2. Notations and operations

We denote the set of real compact intervals by $\mathbb{IR}$ whose elements are $[a] = [\underline{a}, \overline{a}] = \{x \in \mathbb{R} \mid \underline{a} \le x \le \overline{a}\}$, for $\underline{a} \le \overline{a}$ and $\underline{a}, \overline{a} \in \mathbb{R}$. The set of $m \times n$ matrices over the real compact intervals is denoted by $\mathbb{IR}^{m \times n}$.

Let $[a] = [\underline{a}, \overline{a}]$ and $[b] = [\underline{b}, \overline{b}]$ are real compact intervals and let $* \in \{+, -, \cdot, :\}$. Then arithmetic operations on intervals are defined by [1]

$$[a] * [b] = \{x * y \mid \underline{a} \le x \le \overline{a}, \; \underline{b} \le y \le \overline{b}\}.$$

It is assumed that $0 \notin [b]$ in the case of division. We note that $[a]*[b]$ is a real compact interval and

$$[a] * [b] = [\min\{\underline{a} * \underline{b}, \; \underline{a} * \overline{b}, \; \overline{a} * \underline{b}, \; \overline{a} * \overline{b}\}, \; \max\{\underline{a} * \underline{b}, \; \underline{a} * \overline{b}, \; \overline{a} * \underline{b}, \; \overline{a} * \overline{b}\}].$$

For $\mathbf{A}, \mathbf{B} \in \mathbb{IR}^{m \times n}$, $\mathbf{A} \pm \mathbf{B}$ is the $m \times n$ interval matrix whose elements are $\mathbf{A}_{ij} \pm \mathbf{B}_{ij}$. If $\mathbf{A} \in \mathbb{IR}^{m \times n}$ and $\mathbf{B} \in \mathbb{IR}^{n \times r}$ than $\mathbf{A} \cdot \mathbf{B}$ is the $m \times r$ interval matrix whose elements are given by

$$(\mathbf{A} \cdot \mathbf{B})_{ij} = \sum_{k=1}^{n} \mathbf{A}_{ik} \cdot \mathbf{B}_{kj}.$$

For $\mathbf{A} \in \mathbb{IR}^{m\times n}$ and $\mathbf{b} \in \mathbb{IR}^n$, $\mathbf{A} \cdot \mathbf{b}$ is an m-dimensional interval vector whose components are defined by

$$(\mathbf{A} \cdot \mathbf{b})_i = \sum_{k=1}^{n} \mathbf{A}_{ik} \cdot \mathbf{b}_k.$$

If $[a] \in \mathbb{IR}$ and $\mathbf{b} \in \mathbb{IR}^n$, $[a] \cdot \mathbf{b}$ is an interval vector whose components are given by

$$([a] \cdot \mathbf{b})_i = [a] \cdot \mathbf{b}_i.$$

For an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ we define

$$A_c := \frac{1}{2}(\underline{A} + \overline{A})$$

the midpoint matrix whose each element $(A_c)_{ij}$ corresponds to the midpoint of the element $\mathbf{A}_{ij}$ of $\mathbf{A}$. Let

$$\Delta := \frac{1}{2}(\overline{A} - \underline{A})$$

denote the radius matrix whose each element Δ_{ij} corresponds to the radius of the element $\mathbf{A}_{ij}$ of $\mathbf{A}$. Then $\underline{A} = A_c - \Delta$ and $\overline{A} = A_c + \Delta$, so that we also can write $\mathbf{A} = [A_c - \Delta, A_c + \Delta]$. Similarly, for an interval vector $\mathbf{b} = [\underline{b}, \overline{b}]$

$$b_c := \frac{1}{2}(\underline{b} + \overline{b})$$

the midpoint vector and

$$\delta := \frac{1}{2}(\overline{b} - \underline{b})$$

the radius vector thus $\mathbf{b} = [b_c - \delta, b_c + \delta]$.

3. Solving overdetermined linear interval systems

3.1. Preconditioning

We now describe two ways to obtain the preconditioning matrix. As we have seen in the square case, when solving the system $\mathbf{A}x = \mathbf{b}$ of linear equations using interval version of methods such as Gaussian elimination, it is generally advisable to precondition the system. The most commonly used method of preconditioning is to multiply by an approximate inverse of A_c. The products $A_c^{-1}\mathbf{A}$ and $A_c^{-1}\mathbf{b}$ are computed using interval arithmetic. The solution set of the preconditioned equation

$$A_c^{-1}\mathbf{A}x = A_c^{-1}\mathbf{b}$$

contains the solution set of the original equation (see [1]).

When the interval elements of $\mathbf{A}$ and $\mathbf{b}$ are narrow, preconditioning increases the size of the solution set only slightly. When the intervals are wide, preconditioning can substantially increase the size of the solution set. If preconditioning is not used, interval widths generally grow so fast during the solution process that the final results are of little use. This preconditioning was introduced by E.R. Hansen in [2].

In our case, when the matrix of the interval linear equation is not square, $m > n$, two way due to the preconditioning. Our first method of preconditioning is to multiply by the generalized inverse of the midpoint matrix of $\mathbf{A}$. Since the system we consider is overdetermined and all $A \in \mathbf{A}$ has full rank, the generalized inverse of A_c is given by

$$A_c^+ = (A_c^T A_c)^{-1} A_c^T.$$

The products $\widetilde{\mathbf{A}}_\mathbf{1} := A_c^+ \mathbf{A}$ and $\widetilde{\mathbf{b}}_\mathbf{1} := A_c^+ \mathbf{b}$ are computed using interval arithmetic. After the preconditioning, we get the following $n \times n$ interval linear equation system

$$\widetilde{\mathbf{A}}_\mathbf{1} x = \widetilde{\mathbf{b}}_\mathbf{1} \tag{3.1}$$

which we can solve by one of the several existing method.

The second way of preconditioning is comes from the idea of the point (non-interval) case. Our second method of preconditioning is to multiply by the transpose of A_c. Just like in the previous case, the products $\widetilde{\mathbf{A}}_\mathbf{2} := A_c^T \mathbf{A}$ and $\widetilde{\mathbf{b}}_\mathbf{2} := A_c^T \mathbf{b}$ are computed using interval arithmetic. After the preconditioning we get the following $n \times n$ interval linear equation system

$$\widetilde{\mathbf{A}}_\mathbf{2} x = \widetilde{\mathbf{b}}_\mathbf{2} \tag{3.2}$$

which we can solve by one of the several existing method.

As we will see in section 4, the bounds on the solution set of (3.1) is usually narrower then the bounds on the solution set of (3.2). On the other hand, the calculation of the transpose of the midpoint matrix requires fewer arithmetic operations than the calculation of the generalized inverse of A_c.

3.2. Bounding the solution of interval linear equations

Preconditioning described above results an $n \times n$ interval linear equation system. Several methods were developed to compute an enclosure interval vector of the set $\sum(\widetilde{\mathbf{A}}, \widetilde{\mathbf{b}})$. For detailed we refer to [6], [7], [8].

First algorithm was used to bound the solution set of (3.1) and (3.2) was the method provided by J. Rohn in [8]. This algorithm either computes the interval hull of the solution set of the system of interval linear equations $\widetilde{\mathbf{A}} x = \widetilde{\mathbf{b}}$, or finds a singular matrix $S \in \widetilde{\mathbf{A}}$. It has been proved the algorithm terminates in finite number of steps for each $n \times n$ interval matrix $\widetilde{\mathbf{A}}$ and for each n-dimensional interval vector $\widetilde{\mathbf{b}}$. In this algorithm we have to solve an equation of the form

$$Ax + B|x| = b \tag{3.3}$$

where $A, B \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$, which is called an absolute value equation. A very efficient algorithm for the solution of equation (3.3) was described by J. Rohn in [9], [10].

Since the interval matrix $\widetilde{\mathbf{A}}_\mathbf{i}$ ($i \in \{1, 2\}$) is non-singular, Gaussian elimination also can be used to bound the solution set of (3.1) and (3.2). The direct generalization of the Gaussian algorithm was described in [1]. As we will see in example 4.1 sometimes

Gaussian elimination gives the same result as Rohn's algorithm. Generally Rohn's method provide better enclosure of the solution set.

4. Numerical examples

Let $\mathbf{A} \in \mathbb{IR}^{m\times n}$ be a full rank interval matrix where $m > n$ and $\mathbf{b} \in \mathbb{IR}^m$. Let $A_c = Q \cdot R$ be the QR factorization of A_c. The $n \times n$ triangular real matrix R_1 is obtained by dropping from R the last $m - n$ rows. Let $\mathbf{B} := (Q^T\mathbf{A}) \cdot R_1^{-1}$ and $\mathbf{c} := Q^T\mathbf{b}$. The subvector of c_c obtained by dropping the last $m - n$ components is denoted by x_0. Let the interval vector $\mathbf{d}$ is given by d_c is the subvector of c_c obtained by replacing the first n components by zeros and $\delta_d := \Delta_B \cdot |x_0| + \delta_c$. Let the interval vector $\mathbf{h}$ is given by $h_c := d_c$ and $\delta_h := \Delta_B \cdot x_0$. The following results was showed by A.H. Bentbib in [11].

$$\sum(\mathbf{A},\mathbf{b}) \subseteq R_1^{-1} \cdot \sum(\mathbf{B},\mathbf{c}) \subseteq R_1^{-1} \cdot \Big(x_0 + \sum(\mathbf{B},\mathbf{d})\Big)$$

and

$$\sum(\mathbf{A},\mathbf{b}) \subseteq R_1^{-1} \cdot \Big(\sum(\mathbf{B},c_c) + \sum(\mathbf{B},\widetilde{\mathbf{c}})\Big) \subseteq$$
$$\subseteq R_1^{-1} \cdot \Big(x_0 + \sum(\mathbf{B},\mathbf{h}) + \sum(\mathbf{B},\widetilde{\mathbf{c}})\Big)$$

where $\widetilde{\mathbf{c}}$ denote the interval vector whose component $\widetilde{\mathbf{c}}_i$ corresponds to the centered interval $[-(\delta_c)_i, (\delta_c)_i]$.

Let the interval vector given by the Householder method applied to the overdetermined full rank interval linear equation system $\mathbf{A}x = \mathbf{b}$ is denoted by $\mathbf{x}_H(\mathbf{A},\mathbf{b})$. We denote by $\mathbf{x}_{Ch}(\mathbf{A},\mathbf{b})$ the interval vector given by the interval Cholesky method [12] applied to symmetric $n \times n$ interval linear equation system $\mathbf{A}^T\mathbf{A}x = \mathbf{A}^T\mathbf{b}$. A.H. Bentbib compared the interval vectors $\mathbf{v}_1 = \mathbf{x}_H(\mathbf{A},\mathbf{b})$, $\mathbf{v}_2 = R_1^{-1}\mathbf{x}_H(\mathbf{B},\mathbf{c})$, $\mathbf{v}_3 = R_1^{-1}(x_0 + \mathbf{x}_H(\mathbf{B},\mathbf{d}))$, $\mathbf{v}_4 = R_1^{-1}(\mathbf{x}_H(\mathbf{B},\mathbf{c_c}) + \mathbf{x}_H(\mathbf{B},\widetilde{\mathbf{c}}))$ and $\mathbf{v}_5 = R_1^{-1}(x_0 + \mathbf{x}_H(\mathbf{B},\mathbf{h}) + \mathbf{x}_H(\mathbf{B},\widetilde{\mathbf{c}}))$ which all contains $\sum(\mathbf{A},\mathbf{b})$.

Let us denote by $\mathbf{x}_1$ the interval vector given by J. Rohn's method [8] applied to the square interval linear equation system $A_c^+\mathbf{A}x = A_c^+\mathbf{b}$, by $\mathbf{x}_2$ the interval vector given by J. Rohn's method applied to the square interval linear equation system $A_c^T\mathbf{A}x = A_c^T\mathbf{b}$. Let $\mathbf{x}_3$ denote the enclosure interval vector of the solution set of the square interval linear equation system $A_c^+\mathbf{A}x = A_c^+\mathbf{b}$ and $\mathbf{x}_4$ denote the enclosure interval vector of the solution set of the square interval linear equation system $A_c^T\mathbf{A}x = A_c^T\mathbf{b}$ by using Gaussian elimination.

Example 4.1. Let us consider the following interval linear system

$$\begin{pmatrix} [0.1,\ 0.3] & [0.9,\ 1.1] \\ [8.9,\ 9.1] & [0.4,\ 0.6] \\ [0.9,\ 1.1] & [6.9,\ 7.1] \end{pmatrix} \cdot x = \begin{pmatrix} [0.8,\ 1.2] \\ [-0.2,\ 0.2] \\ [1.8,\ 2.2] \end{pmatrix}.$$

The following results was published by A.H. Bentbib:

$$\mathbf{x}_{Ch} = \begin{pmatrix} [-0.0642,\ 0.0285] \\ [0.2408,\ 0.3692] \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} [-0.0761,\ 0.0362] \\ [0.2199,\ 0.4024] \end{pmatrix},$$

$$\mathbf{v}_2 = \begin{pmatrix} [-0.0616,\ 0.0294] \\ [0.2579,\ 0.3485] \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} [-0.0558,\ 0.0232] \\ [0.2560,\ 0.3486] \end{pmatrix},$$

$$\mathbf{v}_4 = \begin{pmatrix} [-0.0620,\ 0.0296] \\ [0.2564,\ 0.3485] \end{pmatrix}, \quad \mathbf{v}_5 = \begin{pmatrix} [-0.0558,\ 0.0232] \\ [0.2560,\ 0.3486] \end{pmatrix}.$$

We have the following results:

$$\mathbf{x}_1 = \begin{pmatrix} [-0.0451,\ 0.0121] \\ [0.2614,\ 0.3447] \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} [-0.0532,\ 0.0186] \\ [0.2557,\ 0.3516] \end{pmatrix},$$

$$\mathbf{x}_3 = \begin{pmatrix} [-0.0451,\ 0.0121] \\ [0.2614,\ 0.3447] \end{pmatrix}, \quad \mathbf{x}_4 = \begin{pmatrix} [-0.0532,\ 0.0188] \\ [0.2544,\ 0.3517] \end{pmatrix}.$$

Let e denote the real vector whose all components are equal to 1 and by E we denote the real $m \times n$ matrix whose elements are all equal to 1. We illustrate the real vectors $\underline{x}$ and $\overline{x}$ according to the index of component i which is varying from 1 to n.

Example 4.2. (See Figures 1-4.) Let $\mathbf{A} \in \mathbb{IR}^{m \times n}$ is given by

$$A_c = rand(m,n) + 4E - 2I, \qquad \Delta = \varepsilon_1 \cdot E.$$

The interval vector $\mathbf{b} \in \mathbb{IR}^m$ is given by

$$b_c = A_c \cdot e, \qquad \delta = \varepsilon_2 \cdot e.$$

Example 4.3. (See Figures 5-7.) Let $\mathbf{A} \in \mathbb{IR}^{m \times n}$ is given by

$$A_c = rand(m,n) + 3I, \qquad \Delta = \varepsilon_1 \cdot E.$$

The interval vector $\mathbf{b} \in \mathbb{IR}^m$ is given by

$$b_c = A_c \cdot e, \qquad \delta = \varepsilon_2 \cdot e.$$

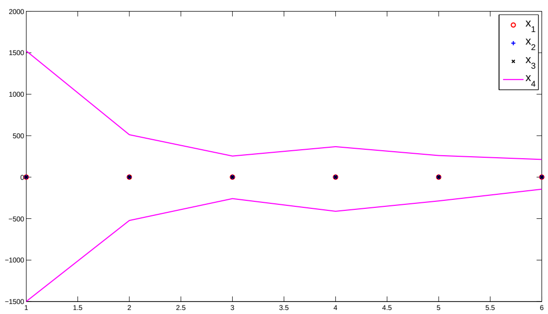


Figure 1. For $m = 10$, $n = 6$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-3}$.

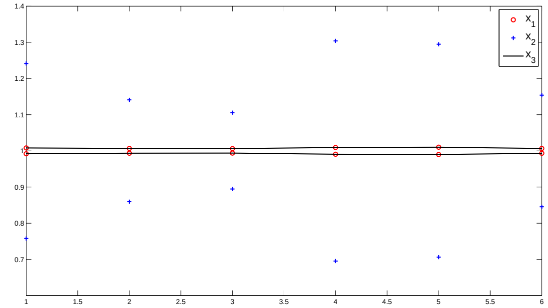


Figure 2. For $m = 10$, $n = 6$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-3}$.

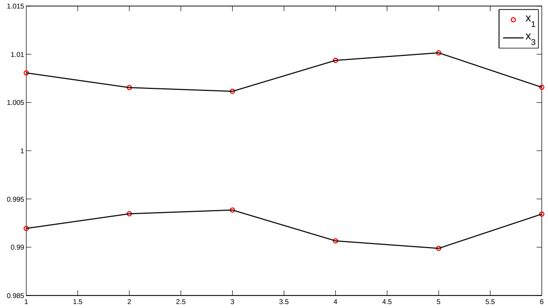


Figure 3. For $m = 10$, $n = 6$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-3}$.

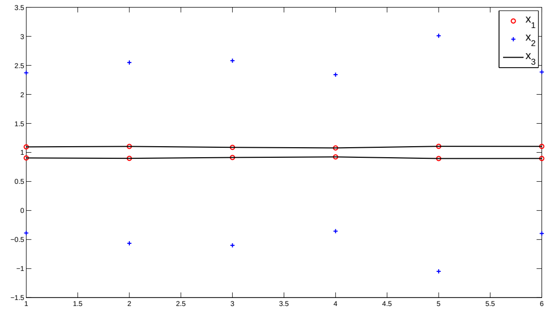


Figure 4. For $m = 10$, $n = 6$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-1}$.

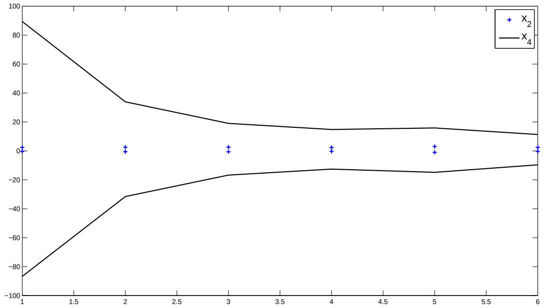


Figure 5. For $m = 10$, $n = 6$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-1}$.

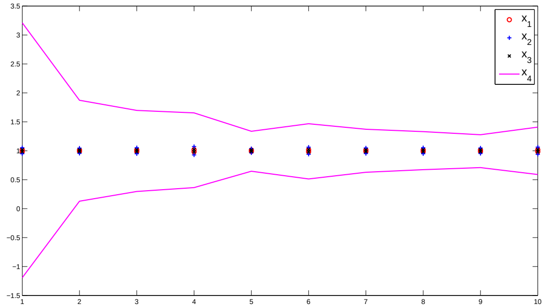


Figure 6. For $m = 20$, $n = 10$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-2}$.

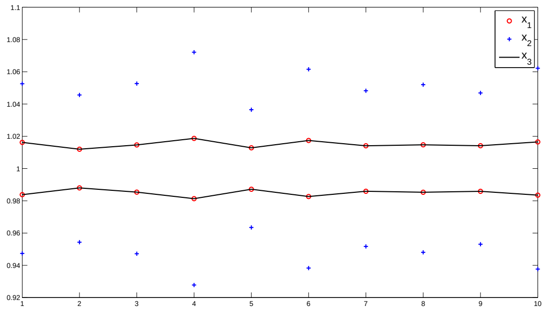


Figure 7. For $m = 20$, $n = 10$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-2}$.

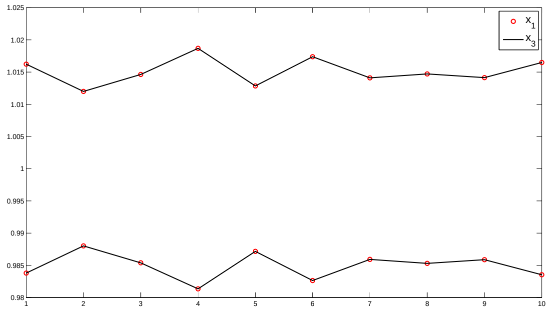


Figure 8. For $m = 20$, $n = 10$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-2}$.

We note that in our experiments (see Example 4.2, 4.3) we got the same results for $\mathbf{x}_1$ and $\mathbf{x}_3$ using two different methods to solve equation (3.1). Namely, we applied Rohn's method and Gaussian elimination. It would be a very interesting question how we can characterize those systems where the equality holds.

References

[1] Alefeld, G., Herzberger, J., *Introduction to Interval Computations*, Academic Press, New York, 1983.

[2] Hansen, E.R., Smith, R.R., *Interval arithmetic in matrix computations, Part 2*, SIAM J. Numerical Analysis, **4**(1967), 1-9.

[3] Hansen, E.R., *Bounding the solution of interval linear equations*, SIAM J. Numerical Analysis, **29**(1992), 1493-1503.

[4] Ning, S., Kearfott, R.B., *A comparison of some methods for solving linear interval equations*, SIAM J. Numerical Analysis, **34**(1997), 1289-1305.

[5] Rohn, J., *Cheap and tight bounds: The recent result by E. Hansen can be made more efficient*, Interval Comput., **4**(1993), 13-21.

[6] Rohn, J., *Systems of linear interval equations*, Linear Algebra and Its Applications, **126**(1989), 39-78.

[7] Rohn, J., *Solvability of systems of linear interval equations*, SIAM J. Matrix Analysis Appl., **25**(2003), 237-245.

[8] Rohn, J., *An algorithm for computing the hull of the solution set of interval linear equations*, Linear Algebra and Its Applications, **435**(2011), 193-201.

[9] Rohn, J., *An algorithm for solving the absolute value equation*, Electronic Journal of Linear Algebra, **18**(2009), 589-599.

[10] Rohn, J., *An algorithm for solving the absolute value equation: An improvement*, Technical Report 1063, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, January 2010.

[11] Bentbib, A.H., *Solving the full rank interval least squares problem*, Applied Numerical Mathematics, **41**(2002), 283-294.

[12] Alefeld, G., Mayer G., *The Cholesky method for interval data*, Linear Algebra and Its Applications, **194**(1993), 161-182.

Szilvia Huszárszky
Eötvös Loránd University
Faculty of Informatics, Department of Numerical Analysis
1/C, Pázmány Péter sétány, 1117 Budapest, Hungary
e-mail: szilvia.huszarszky@gmail.com

Lajos Gergó
Eötvös Loránd University
Faculty of Informatics, Department of Numerical Analysis
1/C, Pázmány Péter sétány, 1117 Budapest, Hungary
e-mail: gergo@inf.elte.hu

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 543–556

Comparison of Riemann solvers in fluid dynamics by weighted error number

Csaba Müller and Lajos Gergó

Abstract. After using a numerical method our eyes are good witnesses whether that method is good or not. We aim to provide, for first order hyperbolic systems, a number that measures, determines the quality of a method instead of deciding by figures. This number is based on the ℓ_1 vector norm of the error vector, combined with weighting. This weight vector has bigger values near discontinuities and kinks because most of the Riemann-solvers have difficulties (including numerical diffusion and oscillations) in solving the equations near these states.

Mathematics Subject Classification (2010): 65M06.

Keywords: Riemann-solver, hyperbolic equation.

1. Introduction

Our primary research field is numerical methods for first order hyperbolic equations. This type of equations, systems are used in several places, for example in fluid dynamics, shallow water calculations, and many other places. For example if we want to model blood flow in human vascular system then the obtained equation will be hyperbolic too but a much more complicated one. For a detailed biomechanical view for this subject see [10].

In this case wall of veins can't be considered as a rigid tube, it is flexible, it can narrow and broaden. Thereby a new source term appears in the system which will depend on the solution itself. There is another big difficulty because in this case junctions have to be studied. In this paper we work with a simpler system of equations, namely the Euler system in fluid dynamics. For more detailed theory of hyperbolic equations see the Godlewski-Raviart book [3].

Our objective is to assign a number to a given numerical solution. This number should show us how that method can perform near critical regions. How close the numerical solution to the exact solution is; if the given method produces a solution at

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

all. As we will see there are test cases where certain methods are unable to produce numerical results because of their properties. This is because most of the arising physical properties in modeling gas flows (pressure, density and energy) can't be negative. However in certain cases oscillations could occur with a portion of numerical methods. If these oscillations are big "enough" then they can reach negative region and a negative value in density for example ruins all the calculations.

These problems could be avoided by minor modifications of the methods but our goal is to use and study them in their "original" forms. We would do a note here. These methods usually lose slightly from their good properties due to the modifications and the running time could also increase by these extra checks.

1.1. Computer configuration used for tests

Hardware configuration.

- CPU: AMD FX-8350, 4.0-4.2 GHz
- RAM: 12 GB, DDR3-1600 MHz

Software configuration.

- Operating system: Microsoft Windows 7 (64-bit, Professional version)
- MATLAB version 2010b

2. Euler system

Our test equation is the Euler system in fluid dynamics. It is given as

$$\frac{\partial \vec{u}}{\partial t} + \frac{\partial \vec{f}\,(\vec{u})}{\partial x} = 0,$$

that is in the so-called conservative form. The solution vector $\vec{u}$ has three components, namely $\vec{u} = [\rho, m, e]$ where ρ is the density, $m = \rho u$ is the mass flow component and e is the total energy.

The function $\vec{f}$ is known as flux. It contains three components as

$$\vec{f}\,(\vec{u}) = \left[m, \quad \frac{m^2}{\rho} + p, \quad \frac{m}{\rho}\,(e + p) \right],$$

where p is the pressure. It can be calculated by the equation

$$e = \frac{p}{\gamma - 1} + \frac{m^2}{2\rho},$$

where γ is ratio of specific heats, a constant depending on the gas. In our tests we used $\gamma = 1.4$ which is the case of air.

Physically this system describes gas flow and state changes over time in a rigid one-dimensional tube with given initial values. It is easy to see that Euler system is nonlinear. Nonlinearity always brings additional complexity compared to the simpler linear cases. This is even more true in solving nonlinear partial differential equations numerically. In our case, complexity of the problem is caused by the nonlinearity of the flux and the discontinuity of initial values.

Characteristics are important in the case of hyperbolic equations (see [1], [12], [18]). In linear systems these characteristics are parallel lines. The components of the solution are constant along these lines, only depending on initial values and the source terms if the system is not homogeneous. Some methods were discussed by Roe [13] and LeVeque, Yee [9] for systems with source terms.

The aforementioned property also provides a simple method for calculating the solution, this is called the method of characteristics (see in [14]). The initial values should be moved along characteristic lines. Additionally if the system is inhomogeneous the source term should be integrated above a specific section of the characteristic line. In numerical mathematics there are many known methods to integrate a function above a finite line segment.

Characteristics are not parallel in the case of nonlinear equations. In addition they could intersect each other so we can not use this simple method.

But characteristics are still important in solving this system numerically. They describe the propagation speed of waves, S. If we want to guarantee stability for a numerical method then we should use a time step τ such that the following inequality holds in all grid points and for all coordinates

$$CFL := \frac{\tau\,|S|}{h} \leq c,$$

where h is the spatial step. The value of c is 1 for Lax-Friedrichs and Lax-Wendroff methods; it is $\frac{1}{2}$ for Godunov-type methods, so that the waves do not cross the cell borders (see [18]). CFL (Courant-Friedrichs-Levy) is called Courant number.

So the stability of a method depends on the Courant number, therefore we need to determine this number during calculations. Because τ and h depends only on discretization, we need to calculate S in all grid points.

This value S depends on none other than the eigenvalues of Jacobian matrix of the flux which is nothing other than the slope of characteristics. In linear case, these eigenvalues are constant, but in our case they depend on the solution as well.

Analytically the eigenvalues of the Euler system are as follows

$$\lambda_1 = u - c, \lambda_2 = u, \lambda_3 = u + c,$$

where u is the velocity of the gas and c is the speed of sound which can be calculated by

$$c = \sqrt{\gamma \frac{p}{\rho}}\,.$$

We determine approximation to this number in all grid points during the entire calculation and examine whether the maximal absolute value from these numbers meets the Courant condition or not. In practice this is the easiest way to guarantee the stability.

3. Test cases

In our tests we solved the Euler equation with different initial values (see Table 1). These initial values are from the book of Toro [18]. We assume that the initial values consist of two different constant states with a discontinuity in x_0 as follows

$$\vec{u}_0(x) := \vec{u}(x,t)\,|_{t=0} = \begin{cases} [\,\rho_L, m_L, e_L]^T\,(=: \vec{u}_L) & \text{if } x < x_0 \\ [\,\rho_R, m_R, e_R]^T\,(=: \vec{u}_R) & \text{if } x_0 < x \end{cases}$$

where the indices L and R refer to the left and right constant states.

For this special type of initial conditions the initial value problem is called Riemann problem. The spatial domain is taken as interval $[-3,3]$ and our time domain is $[0,T]$ where T is a parameter of the test case. The calculations were done on a finite interval so we need boundary conditions. In our cases we used transmissive boundary conditions (see [2]). The spatial domain was divided to 600 subintervals, so h is fixed at 0.01 in all cases. Time steps number M could be changed in order to guarantee stability. Then

$$\tau = \frac{T}{M}$$

can be applied to determine time step size.

The spatial-time graph can be divided into 4 parts in the case of Riemann problem by 3 lines. In all subparts of the graph, gas states (velocity, pressure, density and energy) are constant.

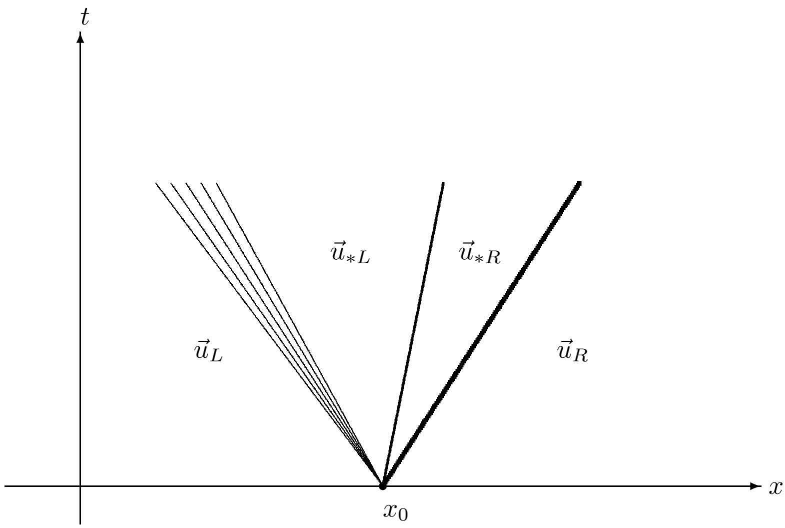


FIGURE 1. Characteristic types

These three lines are characteristic lines starting from the discontinuity of initial values (see Figure 1). There are three different types of characteristics. There is a conventional way of marking these waves by type. Thick lines mark the so-called shock waves, thinner lines mark the contact discontinuity in the middle and the fan-like marking is for rarefaction waves. For more detailed descriptions of waves, see the book of Whitham [20].

In our case the middle wave will be a contact wave. Only density and total energy change along a contact wave, velocity and pressure are equal on both sides of this wave. The region between the two outer waves is called star region.

It is important to note that while contact and shock waves produce a discontinuity in solution, rarefaction waves do not. They blur and link the values from the adjacent regions continuously. This is the reason of the fan-like marking in figures.

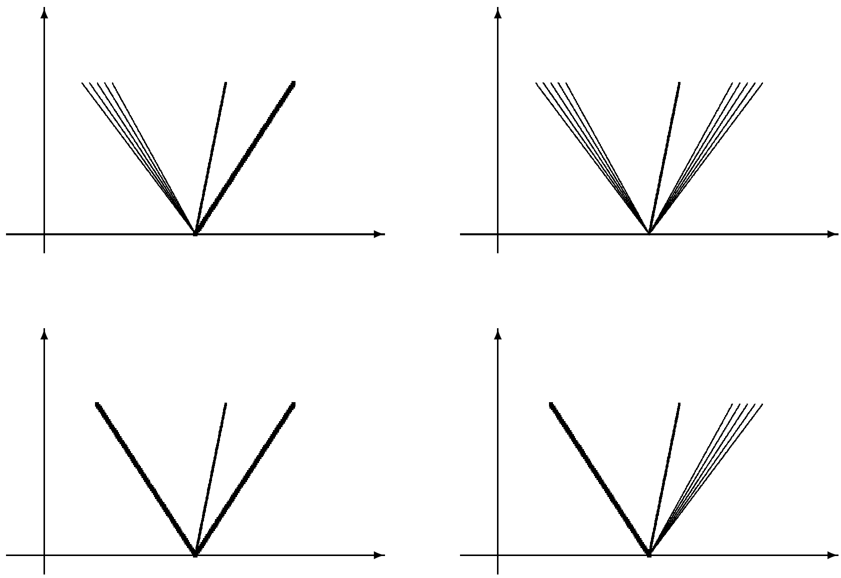


Figure 2. Possible characteristic layouts

On our example graph there is a rarefaction wave on the left side and a shock wave on the right. There are 4 possible layout as you can see in figure 2. On a given side the wave type depends on how the pressure on that particular side compares to the pressure in star region. If pressure is higher in the star region then there will be a shock wave at a given side , if lower (or equal) then there will be a rarefaction wave.

After presenting the wave types and possible layouts we describe each test cases in a few words.

Test	x_0	T	ρ_L	u_L	p_L	ρ_R	u_R	p_R
1	0	1	1	0	1	0.125	0	0.1
2	−1.2	1.2	1	0.75	1	0.125	0	0.1
3	0	0.9	1	−2	0.4	1	2	0.4
4	0	0.072	1	0	1000	1	0	0.01
5	−0.6	0.21	5.99924	19.5975	460.894	5.99242	−6.19633	46.095
6	1.8	0.072	1	−19.59745	1000	1	−19.59745	0.01
7	0	12	1.4	0	1	1	0	1
8	0	12	1.4	0.1	1	1	0.1	1

TABLE 1. Initial values and parameters

Test case 1 is a very popular test case for this equation, called SOD test case. Its characteristics layout is the same as in figure 1, so there is a rarefaction wave on the left, middle wave is contact discontinuity. This is moving to the right slowly by the time. The right wave is a shock wave in this case.

Test case 2 is very similar to test 1, but the initial velocity isn't 0 on the entire interval, only on the right side, it is 0.75 on the left. Wave structure is the same as in test 1 but wave slopes, speeds differ from those. Numerical results are also very similar.

In test case 3 there are two rarefaction waves symmetrically to 0. These rarefactions cover two long intervals. Generally, numerical methods do not handle these long rarefaction waves well. Furthermore close to vacuum state appears in this test, which causes the Lax-Wendroff method to fail.

The wave layout in test 4 is the same as in test 1 and 2 but in this case contact and shock waves are extremely close to each other. Robustness of the method can be measured by this test case. Initial values differ several orders of magnitude on two sides which causes another difficulty to numerical methods.

Test 5 is very similar to test 4 but in this case left wave will also be a shock wave and there will be bigger distance between contact discontinuity and right shock wave. Numerical results are also similar to those seen in the latest test.

Test 6 is almost exactly the same as test 4 except the initial velocity is not 0. Perhaps the biggest differences are visible among methods in this test case. Because of this we make a comparative figure (see Figure 3) for the obtained numerical approximations by different methods. We only represent the density plots since density graph is always the most interesting one.

Test 7 is a trivial test case, because the gas is in steady state. There is one trivial rarefaction wave on both sides. The contact wave stays at $x_0 = 0$ and does not move, however bunch of numerical methods blur this discontinuity along left and right states because of numerical diffusion. Some methods can produce exact solution in this case because of the simplicity of initial values and trivial wave structure.

Test 8 is almost the same as case 7, except the contact wave will move to the right slowly by the time. The results are also very similar to those produced in the latest test case, but there is no method producing exact solution.

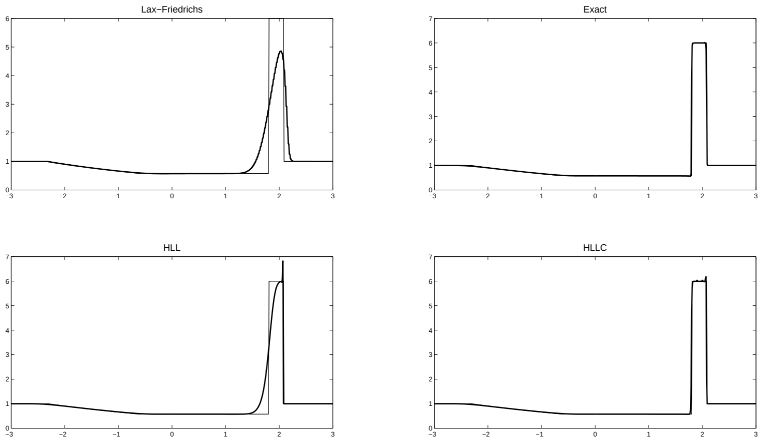


FIGURE 3. Test 6: obtained density plots by different methods

4. Approximate Riemann-solvers

We study conservative numerical methods in the following form

$$\vec{u}_i^{j+1} = \vec{u}_i^j - \frac{\tau}{h}\left(\vec{f}_{i+\frac{1}{2}}^{j+\frac{1}{2}} - \vec{f}_{i-\frac{1}{2}}^{j+\frac{1}{2}}\right),$$

where $\vec{u}_i^j$ means the numerical approximation at time level j in the ith spatial point and $\vec{f}_{i\pm\frac{1}{2}}^{j+\frac{1}{2}}$ is the left and right intercell numerical flux. The studied methods differ only in the calculation, definition of this numerical flux.

The basic idea comes from Godunov [4]. The initial value problem could be solved by calculating exact or approximate solution of a Riemann problem in each cell of the grid. We get the intercell numerical fluxes from these solutions. Then we can do a time-step.

Five different Riemann-solvers were tested, the Lax-Friedrichs, Lax-Wendroff, HLL solver, HLLC solver, and the exact one. Intercell flux can be expressed without solving a Riemann problem in the case of Lax-Friedrichs and Lax-Wendroff solvers.

The Lax-Friedrichs [7] solver has an important advantage over the other (except the exact) solvers, namely this is a monotone solver. Therefore it will not produce oscillations near the regions with difficulties, according to Godunov's theorem (see [4]). On the other hand it has a disadvantage, it generates a high numerical diffusion near the contact discontinuity.

The Lax-Wendroff [8] solver is second order for linear problems, and therefore it can not be monotone method, so it will produce oscillations near the discontinuities. It is a disadvantage, but on the other hand it limits the discontinuities to a smaller interval. It means that the numerical diffusion will be much smaller using Lax-Wendroff scheme than in the case of using a first order methods.

The other three tested methods are really based on Riemann problem. The HLL solver [6] uses approximations to the two outer waves' speed only. It has problems with middle wave, in the form of high numerical diffusion. The HLLC solver is an improved version of the HLL solver introduced by Toro [17]. It tries to reduce numerical diffusion in the way of using estimates for all three wave speeds, C refers to the contact.

Finally the exact Riemann solver. It computes the exact solution of the Riemann subproblems. There are many works related to the exact solver, for example [4], [5], [11], [15], [16] and [19]. We used the version from [18, Chap. 4].

5. Numerical results

All test cases (see Table 1) were computed by all the mentioned numerical methods and all of the obtained approximations were studied.

All of our test cases are Riemann problems. There exists exact solver for these type of problems, as mentioned above. This can be used to compare the obtained numerical approximations to the exact solution.

Our primary goal is to specify the error at time T as follows

$$\|\vec{u}(\cdot,T)-\vec{u}_{num}(\cdot,T)\|_{L_1}=\int\limits_{-3}^{3}\left|\vec{u}\left(x,T\right)-\vec{u}_{num}\left(x,T\right)\right|\,\mathrm{d}x, \tag{5.1}$$

this is the L_1 norm of the difference between the exact and the approximate solution, where $\vec{u}_{num}$ is the numerical solution. But numerical solution gives values only in grid points, therefore we can not integrate this difference along the specified interval. We could interpolate the numerical values and then integrate using this interpolation, but there is a more simple way.

We evaluate the exact solution only in grid points at time level T, denote this vector of values by $\vec{U}^{(i)}_{exact}$, $(i=0,1,\dots 600)$. Then we can get the discretization of L_1 norm by calculating the

$$\left\|\vec{U}_{exact}-\vec{U}_{num}\right\|_{\ell_1}=\sum_{i=0}^{600}\left|\vec{U}^{(i)}_{exact}-\vec{U}^{(i)}_{num}\right|$$

ℓ_1 vector norm, where $\vec{U}^{(i)}_{num}$ is the numerical solution at time level T in the ith spatial grid point. The $\vec{U}^{(i)}_{exact}$ and $\vec{U}^{(i)}_{num}$ contains multiple values, so these calculations could be made coordinate-by-coordinate.

We remark here that if we use interpolation to calculate the error formula (5.1) then the results would be almost the same. For example if we use the trapezoidal rule then the result differs only by a multiplication factor of h, because at the boundaries the error is 0, at the inner points the trapezoidal rule multiplies by h.

It is not our goal. We want to calculate the error with higher weights near critical regions. For this reason our exact solver returns the types and places of waves at time $t=T$. The mentioned critical regions are the locations of contact, shock waves and head and tail of rarefaction waves at time $t=T$.

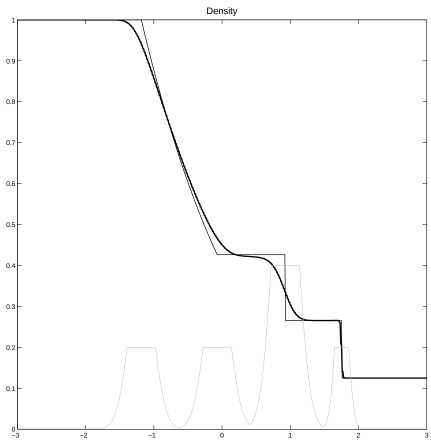


FIGURE 4. Weight vector on the density plot

Weights are developed as follows. We use

$$w(x) = min\left\{c_1 e^{-c_3\,(x-c_2)^2};\ c_4\right\}, \quad x \in [-3,3]$$

functions. 3 to 5 functions of this type and the constant 1 function were taken. Then the weight in the ith point of the grid (w_i) is the maximal value of the previous functions' value in that given point.

It should be noted that we have actually two different weight vectors because contact discontinuity only appears in the density and total energy graphs. One of the weight vectors is used to these coordinates. The other weight vector ignores the contact wave. We use this to calculate error of the pressure and velocity components.

We tuned the constants ($c_1, \ldots, c_4$) to focus weights to critical regions. Based on our experience we use 3 truncated exponential functions. The function

$$min\left\{200\,e^{-64\,(x-c)^2}, 100\right\}$$

is used in case of shock waves where c is the place of shock at time $t = T$;

$$min\left\{200\,e^{-16\,(x-c)^2}, 100\right\}$$

is used to the head and tail of a rarefaction wave, c is the place of the head/tail of the given rarefaction;

$$min\left\{400\,e^{-64\,(x-c)^2}, 200\right\}$$

is used to contact discontinuities, c denotes the place of the contact wave. In all function x takes value from the $[-3,3]$ interval.

Test	Lax-Friedrichs	Lax-Wendroff	HLL	HLLC	Exact
1	0.0294 0.0163 s	0.0075 0.0401 s	0.0156 2.8603 s	0.0151 3.0637 s	0.0144 91.5954 s
2	0.0413 0.0373 s	0.0096 0.0598 s	0.0160 4.2828 s	0.0162 4.5213 s	0.0163 126.875 s
3	0.0193 0.0303 s	– –	0.0237 3.0494 s	0.0237 3.1376 s	0.0214 98.2901 s
4	47.2754 0.0409 s	– –	28.2834 5.0140 s	27.9647 5.3397 s	27.8372 129.833 s
5	86.1165 0.0627 s	– –	30.2774 7.7901 s	31.0984 8.0085 s	30.307 276.242 s
6	14.0664 14.0664 s	– –	11.1666 5.3077 s	10.2964 5.5960 s	9.5426 130.901 s
7	0.0323 0.0934 s	$\approx 10^{-16}$ 0.1540 s	0.0316 18.4428 s	0 19.5587 s	0 819.263 s
8	0.0329 0.1132 s	0.0088 0.1779 s	0.0318 20.091 s	0.0131 21.174 s	0.0131 657.2296 s

TABLE 2. Summary of results

Figure 4 illustrates the resulting weight vector with the exact solution and a numerical approximation by Lax-Friedrichs solver of the density plot for SOD test case. In the figure thinner line marks the exact solution, the thicker one marks the numerical approximation while the lighter line marks the weight vector scaled down by 500.

Then the weighted error is defined in the following steps. First we calculate the

$$error_w = \frac{1}{W} \sum_{i=0}^{600} w_i \left|numerical_i - exact_i\right|$$

numbers for all coordinates where $W = \sum_{i=0}^{600} w_i$ with the corresponding weight vector's values. To be totally clear $numerical_i$ marks the given coordinate of the numerical approximation in ith grid point, $exact_i$ stays for the exact solution in that point. After calculating this number for all four coordinates we average these to get the final weighted error of the approximation.

We summarize these results in Table 2. There are two numbers in each cell of this table. Weighted error numbers are at top and the runtimes are at bottom of each cell.

Each test was calculated with the highest possible Courant number that holds stability. We get the slightest numerical diffusion this way.

We examined how methods work if using not 1 but 0.75, 0.5 and 0.25 as Courant numbers. You can find these results in Table 3 using Lax-Friedrichs method.

Results get worse in all test cases, because more time-steps should be done to reach the desired $t = T$ level as we have smaller time step size.

We make another figure (Figure 5) to illustrate this behavior. On all of these figures thinner line marks the exact solution, thicker one marks the numerical solution using the Lax-Friedrichs method with the corresponding CFL number can be found

Test	$CFL \leq 1$	$CFL \leq 0.75$	$CFL \leq 0.5$	$CFL \leq 0.25$
1	0.0294	0.0394	0.0546	0.0836
2	0.0413	0.0523	0.0702	0.1049
3	0.0193	0.0364	0.0566	0.0897
4	47.2754	58.7429	74.6345	99.69
5	86.1165	112.4286	157.1259	239.5929
6	14.0664	23.4655	34.5278	54.0884
7	0.0323	0.0342	0.0366	0.0399
8	0.0329	0.0348	0.0372	0.0404

TABLE 3. Lax-Friedrichs method with different Courant numbers

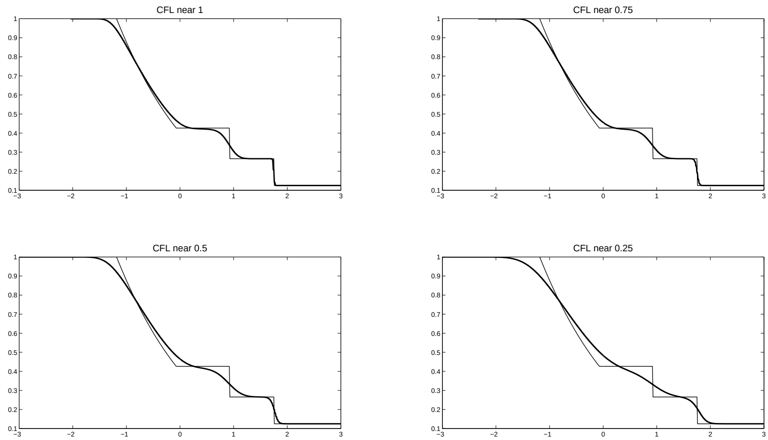


FIGURE 5. Test 1 with different Courant numbers using Lax-Friedrichs method

in the title of the subfigure. We mentioned that doing more time steps increases numerical diffusion. For example when CFL is near 0.25 then the obtained solution is almost a straight line between side states. It is impossible to recognize discontinuities based on this approximation.

These modified CFL tests were done with Lax-Wendroff method. In this case numerical diffusion grow only slightly but oscillations increasing and widening (see Figure 6). In this figure thinner line marks the exact solution, thicker line marks the numerical approximation using the Lax-Wendroff method with CFL number as in the title of the subfigure.

The other three methods do not produce worse solutions (maybe a little bit worse, barely visible differences) calculating with lower Courant number. The reason of this could be that they divide the whole problem to many Riemann problems and they use some approximation of wave speeds. They can keep numerical diffusion under control this way.

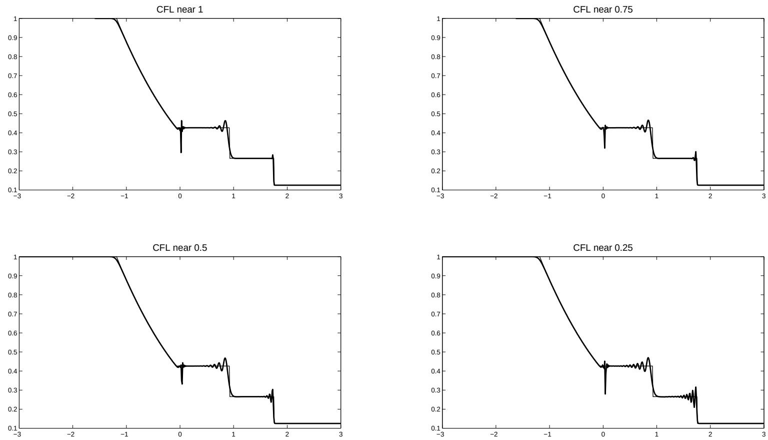


FIGURE 6. Test 1 with different Courant numbers using Lax-Wendroff method

6. Remarks and conclusion

The space step size, h was fixed at 0.01 in all of our tests. We remark that in the case we use a smaller space step our results will keep the same pattern. Of course the approximations would be more accurate, numerical diffusion and oscillations would take a narrower interval. But we should take more time steps in order to keep the CFL number below 1 and therefore computational time would increase significantly. For example if we use half of the original spatial step size ($h = 0.005$), it means twice as much intervals as in the original case, but we should take about twice as much time steps because of the CFL condition. So the calculation time increases with a multiplier of 4, but the accuracy of the approximation would not be 4 times better. All in all taking a fixed h was just a simplification to our test procedure.

The runtime of the exact solver is incredibly high in comparison with other solvers. That is because it has to calculate the exact solution of a Riemann problem in each cell. Which is a very time-consuming task. In addition we can see that its weighted error values are not much better compared for example to the $HLLC$ solver's values. So using the exact solver is only recommended for very sharpened cases.

Lax methods are much quicker than their counterparts in all cases. The explanation is simple, these two solvers could be written in a closed formula, as mentioned above, so they actually do not need to divide the problem into subproblems thus considerably simplifying the process of calculation.

Furthermore, we remark that in general cases we do not have an exact solver, so we can not produce exact solution. We could not compare our results to the exact solution. In this case we should make a quasi-exact solution using some monotone (eg. Lax-Friedrichs) method with very fine spatial discretization. Numerical diffusion should be corrected after the calculation in order to use this as a quasi-exact solution and compare other methods to this result.

References

[1] Dafermos, C.M., *Generalized characteristics in hyperbolic systems of conservation laws*, Archive for Rational Mechanics and Analysis, **107**(1989), 127–155.

[2] Giles, M.B., *Non-reflective boundary conditions for euler equation calculations*, AIAA Journal, **28**(1990), 2050–2058.

[3] Godlewski, H., Raviart, P.A., *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer, 1996.

[4] Godunov, S K., *A finite difference method for the computation of discontinuous solutions of the equations of fluid dynamics*, Mat. Sb., **47**(1959), 357–393.

[5] Gottlieb, J.J., Groth, C.P.T., *Assessement of riemann solvers for unsteady one-dimensional inviscid flows of perfect gases*, J. Comput. Phys., **78**(1988), 437–458.

[6] Harten, A., Lax, P.D., van Leer, B., *On upstream differencing and godunov type methods for hyperbolic conservation laws*, SIAM Rev., **25**(1983), 35–61.

[7] Lax, P.D., *Weak solutions of nonlinear hyperbolic equations and their numerical computation*, Comm. Pure Appl. Math., **7**(1954), 159–193.

[8] Lax, P.D., Wendroff, B., *Systems of conservation laws*, Comm. Pure Appl. Math., **13**(1960), 217–237.

[9] LeVeque, R., Yee H.C., *A study of numerical methods for hyperbolic conservation laws with source terms*, J. Comp. Phys., **86**(1990), 187.

[10] Parker, K.H., Jones, C.J.H., *Forward and backward running waves in the arteries: Analysis using the method of characteristics*, Journal of Biomechanical Engineering, **112**(1990), 322–326.

[11] Pike, J., *Riemann solvers for perfect and near-perfect gases*, AIAA Journal, **31**(10)(1993), 1801–1808.

[12] Roe, P.L., *Characteristic-based schemes for the euler equations*, Ann. Rev. Fluid Mech., **18**(1986), 337–365.

[13] Roe, P.L., *Upwind differenced schemes for hyperbolic conservation laws with source terms*, In Proc. Conf. Hyperbolic Problems, (Carasso, Raviart, Serre editors), Springer, 1986, 41–51.

[14] Sarra, S.A., *The method of characteristics with applications to conservation laws*, JOMA, **3**(2003).

[15] Schleicher, M., *Ein einfaches und effizientes verfahren zur loesung des riemann-problems*, Z. Flugwiss. Weltraumforsch., **17**(1979), 265–269.

[16] Toro, E.F., *A fast riemann solver with constant covolume applied to the random choice method*, Int. J. Meth. Fluids, **9**(1989), 1145–1164.

[17] Toro, E.F., Spruce, M., Speare, W., *Restorian of the courant surface in the hll-riemann solver*, Shock Waves, **4**(1994), 25–34.

[18] Toro, E.F., *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, 3rd edition, 1999.

[19] van Leer, B., *Towards the ultimate conservative difference scheme v. a second order sequel to godunov's method*, J. Comput. Phys., **32**(1979), 101–136.

[20] Whitham, G.B., *Linear and Nonlinear Waves*, John Wiley & Sons, 1999.

Csaba Müller
Eötvös Loránd University
Faculty of Informatics
1/C, Pázmány Péter sétány
1117 Budapest, Hungary
e-mail: `csaba108@gmail.com`

Lajos Gergó
Eötvös Loránd University
Faculty of Informatics
1/C, Pázmány Péter sétány
1117 Budapest, Hungary
e-mail: `gergo@cs.elte.hu`

Stud. Univ. Babeş-Bolyai Math. 59(2014), No. 4, 557–566

Distributed computing of simultaneous Diophantine approximation problems

Norbert Tihanyi, Attila Kovács and Ádám Szűcs

Abstract. In this paper we present the *Multithreaded Advanced Fast Rational Approximation* algorithm – `MAFRA` – for solving n-dimensional simultaneous Diophantine approximation problems. We show that in some particular applications the Lenstra-Lenstra-Lovász (L^3) algorithm can be substituted by the presented one in order to reduce their practical running time. `MAFRA` was implemented in the following architectures: an Intel Core i5-2450M CPU, an AMD Radeon 7970 GPU card and an Intel cluster with 88 computing nodes.

Mathematics Subject Classification (2010): 68R01, 11J68.

Keywords: Diophantine approximation, rational approximation.

1. Introduction

1.1. Diophantine approximations

Approximating an irrational α with rationals is called Diophantine approximation or rational approximation. The theory of continued fractions provides one of the most effective methods of rational approximation of a real number [1]. *Simple continued fractions* are expressions of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}$$

where a_i-s are integers with $a_1, a_2 \ldots > 0$. The sequence $C_0 = a_0,\ C_1 = a_0 + \frac{1}{a_1},\ \ldots$ are called *convergents*. Every convergent $C_m = p_m/q_m$ represents a rational number. An infinite continued fraction $[a_0; a_1, \ldots, a_m]$ is called *convergent* if the limit

$$\alpha = \lim_{m\to\infty} C_m$$

This paper was presented at the 10th Joint Conference on Mathematics and Computer Science (MaCS 2014), May 21-25, 2014, Cluj-Napoca, Romania.

exists. It is known that no better rational approximation exists to the irrational number α with smaller denominator than the convergents (see e.g: [2]). Fractions of the form

$$\frac{p_{m-1}+jp_m}{q_{m-1}+jq_m}\ (1\le j\le a_{m+2}-1)$$

are called *intermediate (or semi-) convergents.* Calculating intermediate convergents can be used to get *every* rational approximation between two consecutive convergents p_m/q_m and p_{m+1}/q_{m+1}. Adolf Hurwitz (1859-1919) proved in 1891 that for each irrational α there are infinitely many pairs (p,q) of integers which satisfy

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^2\sqrt{5}}\,.$$

Approximating more than one irrationals at the same time is called *simultaneous Diophantine approximation.* The challenge in this case is that for given real numbers $\alpha_1,\alpha_2,\dots,\alpha_n$ and $\varepsilon>0$ find $p_1,p_2,\dots,p_n,q\in\mathbb{Z}$ such that

$$\left|\alpha_i-\frac{p_i}{q}\right|<\varepsilon \tag{1.1}$$

for all $1\le i\le n$. The continued fraction approximation method can be used efficiently for constructing solutions in one or two dimensions. In higher dimensions the situation is more challenging. In 1982 Arjen Lenstra, Hendrik Lenstra and László Lovász invented a polynomial time lattice basis reduction algorithm (L^3) that can be used for solving simultaneous Diophantine approximations [3]. If $\alpha_1,\alpha_2,\dots,\alpha_n$ are irrationals and $0<\varepsilon<1$ then there is a polynomial time algorithm to compute integers $p_1,p_2,\dots,p_n,q\in\mathbb{Z}$ such that

$$1\le q\le 2^{n(n+1)/4}\varepsilon^{-n}\ \text{and}\ |q\cdot\alpha_i-p_i|<\varepsilon$$

for all $1\le i\le n$. The algorithm L^3 can be used effectively for solving Diophantine approximations in higher dimensions, however, it can not be used to generate thousands or millions of $q\in\mathbb{Z}$ that satisfy (1.1) even with varying reduction parameters. Consider the set of irrationals $\Upsilon=\{\alpha_1,\alpha_2,\dots,\alpha_n\}$. Let $\varepsilon>0$ and let us define the set

$$\Lambda(\Upsilon,\varepsilon)=\{k\in\mathbb{N}:\|k\alpha_i\|<\varepsilon \text{ for all } \alpha_i\in\Upsilon\} \tag{1.2}$$

where $\|\cdot\|$ denotes the nearest integer distance function, i.e.

$$\|z\|=\min\{|z-j|,j\in\mathbb{Z}\}\,.$$

In general, the following computational challenges can be stated: (1) generate as many elements of $\Lambda=\Lambda(\Upsilon,\varepsilon)$ as possible in a given time frame, and (2) generate a predefined (huge) number of solutions as fast as possible. In this paper we consider the following number-theoretic challenge:

Challenge: Determine 1 billion elements of the set

$$\Lambda\left(\left\{\frac{\log(p)}{\log(2)}\,,p \text{ prime}\,,3\le p\le 31\right\},0.01\right) \tag{1.3}$$

as fast as possible. This challenge is a 10-dimensional simultaneous Diophantine approximation problem. Generating such a huge amount of integers with L^3 would be very time-consuming on an average desktop PC. The first two authors of this paper recently presented a method for solving n-dimensional Diophantine approximation problems efficiently [4]. The main idea is the following:

Theorem 1.1. *Let* $\Upsilon = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ *be a set of irrationals and* $\varepsilon > 0$ *real. Then there is a set* Γ_n *with* 2^n *elements with the following property: if* $k \in \Lambda$ *then* $(k+\gamma) \in \Lambda$ *for some* $\gamma \in \Gamma_n$.

It was also presented that the generation of Γ_n can be done very efficiently for small dimension (e.g: $n < 20$). In our particular case using the set Γ_{10} it is possible to generate arbitrarily many integers $k \in \Lambda$.

The main goal of this paper is to improve the implementation of the existing algorithms and develop an even faster method than the one presented in [4]. We refer to this new algorithm as `MAFRA` – *Multithreaded Advanced Fast Rational Approximation.*

1.2. Practical usage of MAFRA

Fast algorithms for solving Diophantine approximations can be used in many fields of computer science. In some particular applications the L^3 algorithm can be substituted by `MAFRA` in order to reduce their practical running time. We used `MAFRA` for locating large values of the Riemann zeta function on the critical line. The Riemann-Siegel formula can be calculated by

$$Z(t) = 2 \sum_{n=1}^{\lfloor \sqrt{t/2\pi} \rfloor} \frac{1}{\sqrt{n}} \cos(\theta(t) - t \cdot \ln n) + O(t^{-1/4}), \tag{1.4}$$

where $\theta(t) = \arg(\Gamma(1/4 + \frac{it}{2})) - \frac{1}{2} t \ln \pi$. In 1989 Andrew M. Odlyzko presented a method for predicting large values of $Z(t)$. "We need to find a t for which there exist integers $m_1, \ldots, m_n$ such that each of $t \ln p_k - 2\pi m_k$ is small $(1 \leq k \leq n)$" [5]. This is a simultaneous Diophantine approximation problem like (1.3). By applying `MAFRA` one can solve this kind of approximation problem much faster than with L^3 for small dimensions ($n < 20$). We implemented `MAFRA` in order to be able to measure the practical running time in different architectures.

2. Algorithms for solving Diophantine approximation problems

It is known that Algorithm 2.1 solves our challenge efficiently [4].

Algorithm 2.1. – **(FRA)** – **Fast Rational Approximation**

Require: `bound` ▷ default is one billion
Require: k ▷ starting point, the default is zero

$\Gamma \leftarrow$ Apply Algorithm PRECALC from [4]
$\Upsilon \leftarrow \frac{\log(p)}{\log(2)}$, p prime, $3 \leq p \leq 31$
`counter` $\leftarrow 0$, $\varepsilon \leftarrow 0.01$
while `counter` $<$ `bound` **do**
  for $i = 1 \to 1024$ **do**
    `find` $\leftarrow$ TRUE
    for $j = 1 \to 10$ **do**

```
            a ← Frac((k + Γ[i]) · ϒ[j])
            if (a > ε) and (a < 1 − ε) then
                find ← FALSE
                break                                   ▷ Leave the for loop
            end if
        end for
        if find = TRUE then
            k ← k + Γ[i]
            counter ← counter + 1
            break
        end if
    end for
end while
```

The test system was an Intel®Core i5-2450M CPU with Sandy Bridge architecture and the development environment was the PARI/GP computer algebra system. Using this setup it was possible to produce 100 000 appropriate integers within 22.16 seconds. In order to achieve better performance the development environment had been changed to native C using the GNU MP 5.1.3 multi precision library. In Algorithm 2.1 the Precalc function calculates Γ_{10} in few minutes. After the calculation of the 1024 elements of Γ_{10} one can generate arbitrarily many $k \in \Lambda$ very efficiently. With the improved C code it was possible to produce 100 000 integers within 2.65 seconds. This is approximately 10 times faster than the PARI/GP implementation.

It is important to note a significant difference between our 10-dimensional Challenge (1.3) and the 7-dimensional Challenge presented in [4]. In that paper the solution set was defined in the following way:

$$\Omega(\Upsilon, \varepsilon, a, b) = \{k \in \mathbb{N} : a \le k \le b, \|k\alpha_i\| < \varepsilon \text{ for all } \alpha_i \in \Upsilon\}.$$

As it can be seen, the elements of the Ω are bounded. In (1.2) we redefined Ω without boundaries. This "small" change of the definition allows us to design and develop an even faster algorithm.

Algorithm 2.2. – (AFRA) – Advanced Fast Rational Approximation

```
Require: bound                                  ▷ default is one billion
Require: k                          ▷ starting point, the default is zero
Γ ← Apply Algorithm Precalc from [4]
ϒ ← log(p)/log(2), p prime, 3 ≤ p ≤ 31
ε ← 0.01
counter ← 0
while counter < bound do
    sum ← 0
    for i = 1 → 10 do
        a ← Frac(k · ϒ[i])
        if (a < ε) then
            sum ← sum + 2^i
        end if
```

```
    end for
    sum ← abs(sum − 1024)                              ▷ binary complementer
    counter ← counter + 1
    k ← k + Γ[sum]
end while
```

Algorithm 2.2 (AFRA) is substantially different from Algorithm 2.1 (FRA). Let $k \in \Lambda$. FRA always finds the smallest $\gamma \in \Gamma_{10}$ where $(k+\gamma) \in \Lambda$. It is easy to see that in the worst case this algorithm goes through all the 1024 elements of Γ_{10} (see Algorithm 2.1, line 5). In each step the algorithm has to check whether $(k+\gamma) \in \Lambda$ or not (see line 9). AFRA finds one element from Γ_{10} — not necessary the smallest one[1] — that satisfies $(k+\gamma) \in \Lambda$ (Lemma 8 in [4] ensures finding the appropriate $\gamma \in \Gamma_{10}$ efficiently). Algorithm 2.2 is therefore faster, however, adding some γ to k produces larger values in Λ. It can be concluded that FRA is a better choice for solving bounded challenges like $\Omega(\Upsilon, \varepsilon, a, b)$. For solving unbounded challenges, like our particular 10-dimensional case, AFRA is much better. We implemented Algorithm 2.2 in native C. In our test system we were able to generate 100 000 integers $\in \Lambda$ in 0.434 seconds[2]. This is almost ten times faster than the Algorithm 2.1 implementation.

Let us compare the algorithms FRA and AFRA with exact numbers. Consider the following challenge: generate as many integers as possible in the set

$$\Omega\left(\left\{\frac{\log(p)}{\log(2)}\,, p \text{ prime}\,, 3 \le p \le 31\right\}, 0.01, 0, 2 \times 10^{19}\right) \tag{2.1}$$

This challenge differs from (1.3) since the elements of Ω are bounded. As we mentioned FRA is a better choice for a bounded challenge. Solving (2.1) by FRA one can produce 13 different integers between 0 and 2×10^{19}. These integers are presented in Table 1. It is easy to verify that every integer k in TABLE 2 satisfies the following:

$$\left\| k \frac{\log(p)}{\log(2)} \right\| < 0.01$$

for all $3 \le p \le 31$.

Table 1. FRA output between 0 and 2×10^{19}

1023317259883927888	4791256480457711184	7100801081230345000
17119933792261461700	20887873012835245666	34231068906304666630
54413427995085417300	75400638401263513339	84067970173856116672
101187903966117578842	105039984658753315688	110219518481847742122
190360506577505848788		

These integers were generated in 0.015 seconds. AFRA is almost 10 times faster than FRA, however, inappropriate for solving this particular “bounded” challenge.

[1]The set of integers in Γ_{10} are ordered in the following way: every integer in Γ_n is represented by an n-dimensional binary vector (see Lemma 8 in [4]). Γ_{10} contains integers ordered by the values of this binary vector (e.g: 0000000000, 0000000001, 0000000010, 0000000011 etc.)

[2]During the measurements Input/Output costs are not cummulated. Displaying the 100 000 integers from the memory would take approximately 5-6 seconds.

With AFRA we can produce only one integer solution, which is 2298677471355273619. The next integer would be 18396312148683633196 which is already out of the upper bound 2×10^{19}.

We conclude that challenge (1.3) is unbounded, so when the size of the integers is unimportant then Algorithm AFRA is the right solution.

3. Computing methods and results

To make the generation even faster we modified our C code in order to be able to run in parallel using pthreads (IEEE Std. 1003.1c-1995.). We refer to the multithreaded version of AFRA as MAFRA– *Multithreaded Advanced Fast Rational Approximation* algorithm.

In this section we present the measured running time of MAFRA for different architectures. The first test environment was a simple Sandy Bridge Intel Core i5-2450M with 4 GB RAM. The second hardware was a Super Computing Cluster called ATLAS with 90x Intel Xeon E5520 Nehalem Quad Core 2.26 GHz Processors and 0.6TB RAM. The third hardware was an ATI Radeon 7970 GPU card.

3.1. Test – Core i5-2450M Laptop

Our first test environment was a simple home desktop PC. It was an Intel Core i5-2450M Sandy Bridge CPU with 4 GB RAM having 2 cores. Generating 100 000 integers $\in \Lambda$ for solving the 10 dimensional challenge with the algorithm MAFRA took 0.234 sec. Our newly implemented, optimized and multithreaded C code is effective, however, generating 1 billion elements of (1.3) with this architecture would take approximately 39 minutes.

3.2. Test – ATLAS Computing Cluster

Our second test environment was the ATLAS Supercomputing Cluster that is operating in the Eötvös Loránd University, Budapest. The most important characteristics of ATLAS are the following: the architecture consists of one dedicated Headnode and 44 Computing nodes.

1x Headnode:

1. 2x Intel Xeon E5520 Nehalem Quad Core 2.26 GHz Processor with 8 MB cache (HyperThreading OFF)
2. 72 Gbyte RAM
3. 10 Gbit eth interface to the 44 computing nodes

44x Computing Nodes:

1. 2x Intel Xeon E5520 Nehalem Quad Core 2.26 GHz Processor with 8 MB cache (HyperThreading ON)
2. 12 Gbyte RAM

Each Nehalem Quad core CPU has 4 physical cores with SSE extension. Each node has a 2×36.256 GFLOP/sec peak performance (see [6]) calculated by the following formula:

$FLOPS = 4\,\text{cores} \times 2.266\text{GHz} \times 2\,(\text{SIMD double prec.}) \times 2\,(\text{MUL, ADD})$
$= 36.256$ GFLOP/sec.

There are 44 computing nodes which contain 88 physical CPU. The total number of physical cores are 352 (4×88). With hyper-threading the number of cores can be doubled to 704 virtual core. The peak performance of the ATLAS Computing Cluster is $72.512 \times 44 = 3190.528$ GFLOP/sec. With full performance ATLAS takes 12.6 kW, 34.2 A, and cosFI= 0.95.

Generating 100 000 integers in one computing node took approximately 0.175 sec. Remember that in the previous test the Intel Core i5-2450 had 2 cores with 4 threads. If the number of threads is less than the number of dimensions then the multithreaded running is obvious; every thread checks whether $(k+\gamma)\cdot\Upsilon[i] < \varepsilon$ for all $i < n$ where n denotes the dimension. ATLAS has 44 different nodes which are much more than the number of dimensions in our particular case. If one wanted to use all of the cores then the best way would be to run 44 copies of AFRA in each node. In this case each node should start from different starting points. Generating 44 different appropriate starting points for each copies of AFRA can be done very effectively with the L^3 algorithm. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be irrational numbers and let us approximate them with rationals admitting an $\varepsilon > 0$ error. Let $X = \beta^{n(n+1)/4}\varepsilon^{-n}$ and let the matrix A be the following:

$$A = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \alpha_1 X & X & 0 & \ldots & 0 \\ \alpha_2 X & 0 & X & \ldots & 0 \\ \vdots & & & & \vdots \\ \\ \alpha_n X & 0 & 0 & \ldots & X \end{bmatrix}.$$

Applying the L^3 algorithm for A the first column of the resulting matrix contains the vector $[q, p_1, p_2, p_3, \ldots, p_n]^T$ which satisfies

$$\left|\alpha_i - \frac{p_i}{q}\right| < \frac{\varepsilon}{q} \text{ and } 0 < q \leq \beta^{n(n+1)/4}\varepsilon^{-n}$$

for all $1 \leq i \leq n$, where β is an appropriate reduction parameter. Using `MAFRA` in accordance with L^3 it is possible to generate 4.4 millions of integers within $0.175 + \delta$ seconds where δ is the generating time of the 44 starting points not exceeding 5000 ms. With the ATLAS Computing Cluster calculating exactly one billion integers that satisfy (1.3) took approximately 39.7 seconds.

Generating the 44 integers as starting points with L^3 can be done very effectively, however, we would like to emphasize again that the L^3 algorithm is ineffective in generating many solutions (e.g. one billion).

3.3. Test – ATI Radeon 7970 GPU

The third test environment was a Sapphire Vapor-X ATI Radeon 7970 6GB GDDR5 GHz Edition GPU card. Modern graphic cards can be other promising solutions for solving high performance computations. Clearly, in order to implement another fast method for our Diophantine approximation problem one has to take into consideration the usage of GPU cards. In our case the multithreaded version of `FRA` and `AFRA` were implemented for the GPU. In the first step, however, we faced with the following problem: there were not any fast *quadruple* precision packages on the GPU. Although some similar packages for the older GPU cards were found written by Andrew Thall [7] and Eric Bainville [8], these packages found to be inappropriate to solve our particular challenge. The problem with the package written by Andrew Thall is that it uses too much branching and function calling in the program which costs a lot clock cycles. It comes from the behaviour of the graphical processing unit which evaluates both the `if` and the `else` part of the conditional, and after the computation it uses that data where the logical value was TRUE. `FRA` and `AFRA` contain a lot of logical evaluation, so the usage of this package was not convenient for our purposes. The other package, which was written by Eric Bainville, is faster, but it is for fixed point numbers which was inappropriate, as well.

In conclusion, we developed our own multiplication, addition, subtraction and truncation methods. We applied the Karatsuba multiplication algorithm and some bitwise tricks for the addition and truncation methods. After measuring the running speeds on this architecture it turned out that using the `AFRA` algorithm on the GPU approximately 50 times performance drop-down could be measured without using the L^3 algorithm for the generation of the starting points[3].

`bound`	AFRA Running speed in seconds
1	0.0254221
10	0.183748
100	1.35351
1000	12.7255
10000	127.038
100000	1200.7

After examining `AFRA` we can state that the main problem with this "linear" algorithm is that it was not possible to distribute enough threads on the GPU. Consider for example our 10-dimensional case. One had to add the 1024 integers to the partial results and then multiply them with the irrationals. The problem with this solution is that in the quadruple–adder kernel it was not possible to send in enough threads lowering or hiding the latency. In our case the global work size was twice as big as the local work size, which leaded to performance drop-down. In order to avoid the big performance drop-down we utilized every threads on the GPU just like in the ATLAS Super Cluster. For example, if we want to use 2048 threads on the GPU, then we would have to generate 2048 different starting points with the L^3 algorithm to feed all the threads on the GPU. We also modified a bit the number representation in order to achieve higher speed on this architecture. In that particular case our measurements show that generating 100 000 different integers on the 7970 GPU it is 4 times faster than on the CPU.

[3]Measuring speeds on the GPU is only an approximation.

Combination of the CPU version of L^3 and the GPU version of `MAFRA` turned out to be a very effective way to solve simultaneous Diophantine approximation problems. A supercomputer with GPU accelerators would be a nice solution for this problem.

4. Further Researches

As we stated in the introduction we used `MAFRA` for locating large values of the Riemann zeta function on the critical line. It was possible to substitute L^3 with `MAFRA` in order to achieve a much better performance of finding large values. We have implemented `MAFRA` algorithm to the GRID system of the Hungarian Academy of Sciences and solving simultaneous Diophantine approximation problems very effectively. We plan to continue our research in this direction.

Acknowledgment. The authors gratefully acknowledge the constructive comments of the referee. The authors would like to thank the opportunity to grant access to the ATLAS Super Cluster operating at Eötvös Loránd University. The research of the second author was partially supported by the European Union and co-financed by the European Social Fund (ELTE Támop-4.2.2/B-10/1-2010-0030).

References

[1] Khinchin, A.Y., *Continued fractions*, Translated from the third (1961) Russian edition, Reprint of the 1964 translation, Dover, Mineola, NY, 1997.

[2] Niven, I., Zuckerman, H.S., Montgomery, H.L., *An Introduction to the Theory of Numbers*, Fifth Edition, 1991, Chapter 7, pp. 338.

[3] Lenstra, A.K., Lenstra Jr., H.W., Lovász, L., *Factoring polynomials with rational coefficients*, Mathematische Annalen, **261**(1982), no. 4, 515–534.

[4] Kovács, A., Tihanyi, N., *Efficient computing of n-dimensional siultaneous Diophantine approximation problems*, Acta Univ. Sapientia, Informatica, **5**(2013), no. 1, 16–34.

[5] Odlyzko, A.M., *The* 10^{20}*-th zero of the Riemann zeta function and 175 million of its neighbors*, 1992
http://www.dtc.umn.edu/∼odlyzko/unpublished/index.html

[6] *Peak performance of Intel CPU's*,
http://download.intel.com/support/processors/xeon/sb/xeon_5500.pdf

[7] Thall, A., *Extended-Precision Floating-Point Numbers for GPU Computation*,
http://www.caesar.elte.hu/hpc/atlasz-hw.html

[8] Web page of Eric Bainville, http://www.bealto.com/cv.html

[9] PARI/GP Computeralgebra system, http://pari.math.u-bordeaux.fr/

[10] Web page of the ATLAS Computing Cluster, Eötvös Loránd University,
http://www.caesar.elte.hu/hpc/atlasz-hw.html (in hungarian).

[11] Thomadakis, M.E., *The Architecture of the Nehalem Processor and Nehalem-EP SMP Platforms http://sc.tamu.edu/systems/eos/nehalem.pdf*, Texas *A&M* University, March 17, 2011.

[12] Sapphire Vapor-X HD 7970 GHz Edition 6GB DDR5,
http://www.sapphiretech.com/presentation/product/

Norbert Tihanyi
Eötvös Loránd University, Faculty of Informatics
Department of Computer Algebra
Budapest, Hungary
e-mail: ntihanyi@compalg.inf.elte.hu

Attila Kovács
Eötvös Loránd University, Faculty of Informatics
Department of Computer Algebra
Budapest, Hungary
e-mail: attila.kovacs@compalg.inf.elte.hu

Ádám Szűcs
Eötvös Loránd University, Faculty of Informatics
Department of Computer Algebra
Budapest, Hungary
e-mail: adam.szucs@compalg.inf.elte.hu