ISSN: 0025-5742

THE MATHEMATICS STUDENT

Volume 86, Numbers 1-2, January-June (2017)
(Issued: May, 2017)

PUBLISHED BY
THE INDIAN MATHEMATICAL SOCIETY
www.indianmathsociety.org.in

THE MATHEMATICS STUDENT

Edited by J. R. PATADIA

In keeping with the current periodical policy, THE MATHEMATICS STUDENT will seek to publish material of interest not just to mathematicians with specialized interest but to the postgraduate students and teachers of mathematics in India. With this in view, it will ordinarily publish material of the following type:

1. the texts (written in a way accessible to students) of the Presidential Addresses, the Plenary talks and the Award Lectures delivered at the Annual Conferences.
2. general survey articles, popular articles, expository papers, Book-Reviews.
3. problems and solutions of the problems,
4. new, clever proofs of theorems that graduate / undergraduate students might see in their course work,
5. research papers of interest to larger readership, and
6. articles that arouse curiosity and interest for learning mathematics among readers and motivate them for doing mathematics.

Articles of the above type are invited for publication in THE MATHEMATICS STUDENT. Manuscripts intended for publication should be submitted online in the LaTeX and .pdf file including figures and tables to the Editor J. R. Patadia on E-mail: *msindianmathsociety@gmail.com*

Manuscripts (including bibliographies, tables, *etc.*) should be typed double spaced on A4 size paper with 1 inch (2.5 cm.) margins on all sides with font size 10 pt. in LaTeX. Sections should appear in the following order: Title Page, Abstract, Text, Notes and References. Comments or replies to previously published articles should also follow this format with the exception of abstracts. In LaTeX the following preamble be used as is required by the Press:

```
\ documentclass[10 pt,a4paper,twoside,reqno]{amsart}
\ usepackage {amsfonts, amssymb, amscd, amsmath, enumerate, verbatim, calc}
\ renewcommand{\ baselinestretch}{1.2}
\ textwidth=12.5 cm
\ textheight=20 cm
\ topmargin=0.5 cm
\ oddsidemargin=1 cm
\ evensidemargin=1 cm
\ pagestyle{plain}
```

The details are available on Indian Mathematical Society website: *www.indianmathsociety.org.in*

Authors of articles / research papers printed in the the Mathematics Student as well as in the Journal shall be entitled to receive a *soft copy (PDF file with watermarked "Author's copy")* of the paper published. There are no page charges. However, if author(s) (whose paper is accepted for publication in any of the IMS periodicals) *is (are) unable to send the LaTeX file of the accepted paper, then a charge Rs. 100 (US $ 10) per page will be levied for LaTeX typesetting charges.*

All business correspondence should be addressed to S. K. Nimbhorkar, Treasurer, Indian Mathematical Society, Dept. of Mathematics, Dr. B. A. M. University, Aurangabad - 431 004 (Maharashtra), India. E-mail: *sknimbhorkar@gmail.com*

In case of any query, the Editor may be contacted.

ISSN: 0025-5742

THE MATHEMATICS STUDENT

Volume 86, Numbers 1-2, January-June, (2017)
(Issued: May, 2017)

PUBLISHED BY
THE INDIAN MATHEMATICAL SOCIETY
www.indianmathsociety.org.in

ISSN: 0025-5742

Published by Prof. N. K. Thakare for the Indian Mathematical Society, type set by J. R. Patadia at 5, Arjun Park, Near Patel Colony, Behind Dinesh Mill, Shivanand Marg, Vadodara - 390 007 and printed by Dinesh Barve at Parashuram Process, Shed No. 1246/3, S. No. 129/5/2, Dalviwadi Road, Barangani Mala, Wadgaon Dhayari, Pune 411 041 (India). Printed in India.

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June, (2017)

CONTENTS

****

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 01-10

ROAD TO MATHEMATICAL SCIENCES IN INDIA - A RELOOK*

D. V. PAI

Abstract. By mathematical sciences, we would understand here a broad spectrum of knowledge that encompasses pure mathematics, applied mathematics, statistics, mathematics of OR, computational mathematics, mathematical physics, mathematical biology, mathematical economics, etc. One aims in this talk to take a relook at the road to mathematical sciences traversed in India since antiquity, with a view to try to gain an understanding about the current status of this domain in modern times.

1. Early Mathematics in India-Some Historical Observations

The road to mathematical sciences in India began from antiquity. India has every reason to feel proud of its rich heritage in mathematics and astronomy. The roots of mathematics in India are visible in the vedic literature which is nearly 3500-4000 years old. The *sulba sutras* - the vedic texts for consruction of ritual altars contain a lot of geometric results and constructions which include, among others, a statement of the Pythagoras theorem, approximation of the number 'pi', approximation of square root of number two upto five decimal places, etc. Undoubtedly, two of the most striking contributions of ancient Indian mathematics are the decimal place value system, which seems to have been discovered as early as in the Harappan period, and the number zero, *'Sunya'*, which could be considered as a profound gift of India to the mankind in the domains of mathematics and philosophy. To put this in the words attributed to Laplace:

"The ingenious method of expressing every possible number using a set of ten symbols (each having a place value and an absolute value) emerged in India. Its simplicity lies in the way it facilitated calculation and placed arithmetic foremost amongst useful inventions. The importance of this invention is more readily appreciated when one considers that it was beyond two of the greatest men of antiquity-Archimedes and Apollonius"; and the next quote from Bourbaki [7], p. 46:

* The text of the Presidential Address (general) delivered at the 82^{nd} Annual Conference of the Indian Mathematical Society held at the University of Kalyani, Kalyani-741 235, Nadia, West Bengal, India during December 27 - 30, 2016.

".....It must be noted moreover that the conception of zero as a number (and not as a simple symbol of separation) and its introduction into calculations also count amongst the original contributions of the Hindus."

Indeed, it is often said that the early seeds, sown in India, led through the efforts of the Arabs, to the revival of mathematics in Europe, the Middle East and even China. Chronologically, next comes mathematics of Jains, the so-called *Jaina mathematics*, whose characteristic was fascination for large numbers and attempts to understand various types of infinities and infinitecimals. Buddhists also understood infinite and indeterminate numbers. The apex of mathematical achievements of ancient India occured during the so-called *classical period* of Indian mathematics, which saw the legendary contributions of the mathematician-astronomers: Aryabhatta (476-550), Varahamira (505-587), Brahmagupta (598-670), Bhaskara (600-680) and Bhaskaracharya (1114-1185). The schools established by some of them are well acknowledged universally. Much later, during the 16th century came the flourishing school of Kerala, whose prominant mathematicians were Madhava, Parameshvara, Nilkantha, Jyestadevan and Achyutan. It has been recognized since the early forties that this school has anticipated by more than 200 years (albeit, with less of rigour than their western counterparts), a number of mathematical results in *analysis* involving infinite series (such as the *Gregory-Newton series* for the inverse tangent, etc.) and calculus, which were later invented by Newton and Leibniz in the 18th century. Madhava (1340-1425) appears to be the frst one to take the decisive step forward from finite procedures of ancient indian mathematics to treat their limit passage to *sin* and *cos* functions.

2. Mathematical Sciences in India During the 20th Century

Among the Indian mathematicians of the early part of the 20th century, the name which undoubtedly comes most prominantly to the mind is that of Srinivasa Ramanujan (1887-1920). He indeed belonged to the celebrated world class of mathematicians comprising of names such as Descartes, Euler, Gauss, Riemann, Hilbert, Poincare. Ramanujan was the first Indian mathematician, a self-taught genious, to have gained recognition from the west in his life time. It can perhaps be undisputably said that his innovative genius is yet to be surpassed in India, even nearly 100 years after he physically left the scene.

In order to analyse the status of mathematical sciences in India of the 20th century using the yardstick of the Ph.D. thesis produced, it seems interesting to review the statistics of Ph.D. thesis from India during this period as cited in Kapur [1]. As observed there, in the first two decades of the 20th century, The Indian share of world research in mathematical sciences was negligible, nearly of the order of 0.1 per cent. It reached 3-4 per cent towards the last decade of the 20th century, and in some fields like probability and mathematical statistics, it was found as high as 12-15 per cent. Keeping in view the population of our country as well

as its rich heritage in mathematics, it may not be too ambitious to expect that our contribution to world research in quantitative terms could reach a reasonable level, 10-15 per cent, if not 20 per cent, in the *next decade* of the 21st century. There are good reasons for this expectation, as we shall soon see. However, while the contribution in quantitative terms may seem going in the right direction, its impact on world mathematics hardly seems to indicate that it measures upto the expected attainment of quality. It is worth emphasizing that one of the visible reasons for this lack of quality in the research output is the fact that most universities do not insist on course-work for their Ph.D. programmes. This has a visible adverse effect on the quality expectations from a Ph.D. mathematician. The situation in this regard needs to be rectified for any intended *qualitative thrust* in research. It must be emphasized here, however, that the programmes such as *Advanced Training in Mathematics (ATM) Programme* sponsored by National Board of Higher Mathematics (NBHM), *National Programme on Differential Equations (NPDE)* sponsored by SERB, etc., are playing their due share in the intended goals of training manpower in these fields, as well as for creating the necessary quality awareness in research.

2.1. **Contributions of Professional Societies.** It is interesting to record below some of the historic facts which were already mentioned earlier by others. See, e.g., Kapur [1].

- Higher education in the modern sense really started in India in the historic year 1857 with the establishment of three universities in Calcutta, Bombay and Madras. Mathematics was taught in all these universities.
- Professional societies have often helped in fostering research culture in the universities. It may be worth recalling that the first set of Ph.D. theses in India in the 20th century came from Calcutta University. As obsered in Kapur [1], this was not a mere coincidence. The founding of the *Calcutta Mathematical Society* in 1908 and its the then dynamic President, Sir Asutosh Mukherjee were two important factors contributing to this developement. The next set of Ph.D. theses came from Madras University. It is a known fact that the founding of the *Indian Mathematical Society* in 1907 has had a positive effect in this developement. However, due to a lack of availability of research mathematicians, the initiation of Ph.D. programme there had to wait upto 1927, when Prof. Vaidyanathswamy joined the mathematics department.
- It may be worth recalling that although Bombay University started contemporarily with Calcutta and Madras Universities, its first Ph.D. in mathematics came as late as in 1942. A reason for this, which is often stated, is that talented mathematics students from the western region

mostly preferred to go to Cambridge and got more fascinated by Cambridge Mathematical Tripos. Some of them became senior 'wrangler's and became good text book writers; however, they did not lay sufficient emphasis on developing research inclinations. Lack of a professional society based in Mumbai was yet another reason for not being able to foster initial research culture in the region.

- Foundations of initial research culture in mathematics in north India can be really attributed to Prof. Ganesh Prasad. He seems to have been inspired by his stint during 1914-17 as Professor of Applied Mathematics at Calcutta University. While working as Professor at Benaras Hindu University during 1917-23, he founded the *Benaras Mathematical Society*, whose name was later changed to *Bharat Ganit Parishad*. This society does seem to have contributed quite well to the development of Mathematics in North India.
- As far as statistics is concerned, it is well known that its development and growth in India can be traced back to the founding of the *Indian Statistical Institute (ISI)* at Kolkata in 1931 by Prof. P. C. Mahalanobis and the starting of the journal *Sankhya* by him in 1933. The ISI at Kolkata also contributed to the development of Operations Research (OR) in India. The joint efforts of ISI, Calcutta, and Professors D. S. Kothari and R. S. Verma from Delhi University led to the founding of the *Operations Research Society* of India, which seems to have contributed quite well to the theory and practice of OR in India.
- The most prominant names of mathematicians which come to the mind for contributing to the growth of applied mathematics in India during the early part of the 20th century, include Professors N. R. Sen, S. N. Bose and B. B. Sen at Kolkata, Prof. A. C. Banerjee at Allahabad, Professors B. R. Seth and P. L. Bhatnagar at Delhi. The founding of the *Indian Society of Theoretical and Applied Mathematics* by Prof. B. R. Seth in 1956 at IIT Kharagpur is known to have contributed substantially to the growth of research in applied mathematics in India during the 20th century.

3. Mathematical Sciences in Modern India

After independence, many new universities came up. Their number has grown to more than 700 today (Central Universities 46, State Universities 343, Deemed Universities /Deemed to be Universities 128, Private Universities 225: Total 742; These do not include 82 central Institutions like IITs, IISc, IISERs, etc.). With Pandit Jawaharlal Nehru's vision for creating a scientific and technological base for the country, to the existing universities, were added in the beginning, National laboratories and five Indian Institutes of Technology (IIT) at Kharagpur, Mumbai, Kanpur, Chennai and Delhi. Subsequently, two more IITs came up one at

Guwahati and the other at Roorkee. During 2008-09 eight new IITs sprang up at Hyderabad, Gandhinagar, Indore, Ropar, Jodhpur, Patna, Mandi and Bhubhaneshwar. Also, IT BHU and ISM Dhanbad were given the status of IITs. More recently, six new IITs are on the anvil; in various stages of being set up. These are at Tirupathi, Palakkad, Jammu, Chhattisgarh, Goa, and Dharwar which makes it a family of 17 IITs which will soon reach the number 23. It is the right time for the country to think of sowing the right seeds of mathematical sciences in the ambiance of new IITs so that the subject eventually flowers in the expected directions.

4. Role of Special Institutes and Centres of Excellence

A number of special institutes in India where mathematics has a central role to play are listed below along with some of the various mathematical interests that are nurtured and emphasised there. Undoubtedly, each of these institutes has striven hard to play its role in building up the necessary infrastructure of research in the stated areas, as well as, has striven hard to reach the necessary international stature in these fields.

- **Tata Institute of Fundamental Research (TIFR), Mumbai: School of Mathematics**

  Algebra, Number Theory, Topology, Harmonic Analysis, Ergodic Theory, Combinatorics.

  School of Technology and Computer Science:
  Stochastic Processes.
- **Indian Institute of Sciences (IIsc), Bangalore, Division of Mathematical and Physical sciences: Mathematics Department**

  Algebra, Algebraic Geometry, Topology & Geometry, Functional Analysis & Operator Theory, Harmonic Analysis, Nonlinear waves, Hyperbolic Equations, Nonlinear dynamics, Probability & Stochastic processes, Time series analysis, Several Complex Variables.
- **Tata Institute of Fundamental Research-Centre for Applicable Mathematics (TIFR-CAM), Bangalore**

  PDEs, Numerical Analysis, Homogenization, Nonlinear Functional Analysis, Optimal Controls, Variational Analysis, Stochastic Analysis.
- **Chennai Mathematical Institute (CMI), Chennai**

  Algebra, Analysis, Differential Equations, Geometry and Topology.
- **Institute of Mathematical Sciences (IMSC), Chennai**

  Algebra, Algebraic Geometry, Number Theory, PDEs, Representation Theory and Topology, Ergodic Theory, Non-commutative Geometry.
- **Harish Chandra Research Institute (HRI), Allahabad** *Algebra, Lie Algebra, Geometry-Discontinuous Groups, Riemann surfaces, Algebraic Topology, Number Theory-Algebraic, Analytic, and Combinatorial.*

Traditionally, Statistics, Probability, and Stochastic Processes have been the main themes pursued at the Indian Statistical Institutes (ISIs) listed below.

- **Indian Statistical Institute Kolkata (ISI Kolkata)**
  Statistics, Mathematics, Computer Science, Quantitative Economics, OR and Information Science, Probability and Stochastic Processes, Statistics on Non-Euclidean Manifolds, Robust & Nonparametric Techniques, Geometry of Banach Spaces, Commutative Algebra, Noncommutative Geometry.
- **Indian Statistical Institute Bangalore (ISI Bangalore)**
  Statistics, Probability, Ergodic Theory & Dynamic Systems, Operator Algebras and Quantitative Probability, Algebra and Number Theory, Algebraic Geometry and Topology, Functional Analysis and Operator Theory, Harmonic Analysis.
- **Indian Statistical Institute Delhi (ISI Delhi)**
  Statistical Computing, Probability Theory, Stochastic Processes, Combinatorial Matrix Theory, Linear Algebra, Markov Processes, Heavy Tailored Distributions, Time Series, Reliability, Nonlinear Regression, Number Theory.

In addition, there are two more centres of ISI added recently, one at Chennai: **ISI, Chennai Centre,** and the other at Tezpur: **ISI, North-East Centre.**

4.1. **IISERS: a New Initiative.** The Government of India, based on the recommendation of Scientific Advisory Council to the Prime Minister **(SAC-PM)**, through the Ministry of Human Resource Development **(MHRD)**, took a bold initiative of establishing five Indian Institutes of Science Education and Research (IISER) since 2006, patterned broadly on the lines of I.I.Sc., Bangalore. These institutes are currently located in Kolkata, Pune, Mohali, Bhopal and Tiruvananthapuram. To these one may add one more IISER which has started functioning from Tirupathi since August 2015 while two more IISERs are planned to start functioning one from Berhampur (Odisha) and another from Nagaland. These steps appear quite timely, since it was increasingly realized during the first decade of the 21st century that due to the prevailing IT boom as well as the visible rat race of coaching classes specialized in training students for admission to the prestigious B.Tech. programmes of IITs, science education in the country, in general, was taking a back seat. IISERs were conceived as a unique initiative in the direction of uplifting science education in India in which teaching and education are to be totally integrated with the state-of-art research, nurturing both curiosity and creativity in an intellectually vibrant atmosphere of research. It is hoped that IISERs are likely to become Science Institutes of the highest caliber and reach the prestigious position in the global setting that IISc, IIMs and IITs presently enjoy.

For the sake of completeness, we list below some of the areas of interest in mathematical sciences pursued at each of the IISERs and also the current academic programmes (as available on their web-sites).

- **Indian Institute of Science Education and Research Pune (IISER Pune)**

  Algebra, Number Theory, Mathematical Biology, Cryptography, Algebraic Geometry, Combinatorics, Knot Theory, Lagland's Program, Linear Algebraic Groups, Representation Theory, Several Complex Variables.

  Academic Programmes: Integrated Master's level (MS) programme involving Biological, Chemical, Mathematical & Physical Sciences, Doctoral Programme (Ph.D.).

- **Indian Institute of Science Education and Research Mohali (ISER MOHALI)**

  Algebraic Geometry, Valuation Theory, Algebra, Algebraic Geometry, Functional Analysis, Groups and Geometry, Differential Algebra.

  Academic Programmes: Integrated Master's level (MS) Programme, Doctoral Programme (Ph.D.), Integrated Doctoral Programme (Int. Ph.D.).

- **Indian Institute of Science Education and Research Bhopal (ISER Bhopal)**

  Complexity and Computational Number Theory, Representation Theory, Algebra, Analysis, Algebraic Geometry, Differential Geometry, Low Dimensional Geometric Topology, Operator Algebras.

  Academic Programmes: BS-MS Dual Degree Programme, Doctoral Programme (Ph.D.).

- **National Institute of Science Education and Research Bhubaneswar (NISER Bhubaneswar)**

  Harmonic Analysis, Theory of Operator Spaces, Theoretical Computer Science, Representation of Geometries, Number Theory, Algebraic Graph Theory.

- **Indian Institute of Science Education and Research Thiruvananthapuram (IISER Thiruvananthapuram)**

  Numerical Functional Analysis, Mathematical Finance, Combinatorial Number Theory, PDEs, Control Theory, Hyperbolic System of Conservation Laws, Functional Analysis, Probability Theory, Commutative Algebra, Differential Geometry.

  Academic Programmes: Five year integrated Master's MS programme and Doctoral (Ph.D.) programme.

- **Indian Institute of Science Education and Research Kolkata (IISER Kolkata)**

Commutative Algebra, Algebraic Geometry, Geometric Group Theory, Spectral Graph Theory, Mathematical Biology, Reliability and Statistics.
Academic Programmes: Integrated BS-MS Dual Degree Programme, MS Programme, Integrated Doctoral (Ph.D.)Programme, Post-Doctoral Programme.

As mentioned in Sathyamurthy [3], the IISERS are presently targeting to admit nearly 200 students at the BS-MS level and a similar number at the doctoral level in each of these institutes, thus expecting to add to an annual steady output of 1000-1500 Master's degree-holders and 1000-1500 Ph.D.'s from the IISER system. This is expected to contribute to a steady increase in the pool of scientists in the country. Each IISER is expected to reach eventually a faculty strength of 200 (presently, it ranges between 60-100) and 200 postdoctoral fellows. Based on a recently conducted peer review of the departments, as reported in Satyamurthy [3], despite the fact that the IISERs are only 8-10 years old, the total number of publications contributed, 3346 during the period 2007-15, puts them almost on par with the established IITs in terms of the per capita output of the faculty. It is also interesting to observe that a number of publications from IISERs have been contributed in peer-reviewed journals with *BS-MS students as lead authors.*

4.2. **UGC Centres of Excellence.** In the initial phase, the University Grants Commission (UGC)supported research in mathematics through the advanced centres set by it at Punjab University, Chandigarh, Bombay University, Mumbai, Ramanujan Institute of Mathematics at Chennai, and the Department of Applied Mathematics at Calcutta University, Kolkata, the last one being the only centre of applied mathematics. Later in the 1980's arising from its Special Assistance Programme (SAP Programme) to the promising University Departments, UGC also created centres of excellence in mathematics at Pune, Bangalore, Coimbatore and also some other places. As good models, one may cite the *UGC Advanced Centre in Fluid Mechanics* in the Central College, Bangalore built around the mentor Prof. N. Rudraiah and the *Centre for Nonlinear Dynamics* in the Bharathidasan University, Tiruchirapalli built around the mentor Prof. M. Lakshamanan. Despite many odds, these centres seem to have functioned quite well under the dynamic leadership and vision of their mentors.

5. DST support to a few exceptional initiatives in Mathematical sciences

The Programme Advisory Committee for Mathematical Sciences (PACMS), SERC created a document entitled *Vision for R&D in Mathematical Sciences.* In this document, ten broad thrust areas in Mathematical Sciences were identified for support. Among other things, this led to:

(1) Support for establishment of the ***National Centre for Advanced Research in Discrete Mathematics (CARDMATH)***;

(2) Support for establishment of the ***Centre for Mathematical Sciences (CMS)*** at Pala, Kerala;
(3) Support for establishment of the ***Centre for Interdisciplinary Mathematics (CIMS)*** at BHU, Varanasi;
(4) Support for establishment of the ***Centre for Mathematical Biology*** at IISc, Bangalore;
(5) Support to ***IISc Mathematics Initiative (IMI)***;
(6) Support to ***National Mathematics Initiative (NMI)***;
(7) Support for establishment of the ***Centre for Research in Mathematical Sciences (CMS)*** at Banasthali University, Rajasthan;
(8) Support to the ***National Programme on Differential Equations (NPDE)***;
(9) Support to the ***National Network on Computational and Mathematical Biology***.

6. Status of Applied Mathematics in India

In the 60's and the early part of 70's, most faculty members at IISc and at all the IIT's worked in traditional domains of applied mathematics. As noted in [5], presently, at IIsc as well as at most of IIT's, interesting work in traditional areas of applied mathematics for the most part, gets done in engineering departments. The areas of interest in applied mathematics(the so-called *modern applied mathematics*) which today are found increasingly of interest are: *coding theory, cryptography, inverse problems, image recovery, communication networks and neuroscience, financial mathematics, computational and mathematical biology, dynamical systems*, etc. Mention must also be made about the dedicated groups of mathematicians working on PDE's and the numerics of PDE's at the TIFR (CAM) Centre, IIT Bombay, IIT Kanpur, IIT Kharagpur, IIT Gandhinagar, IIT MAdras, IIT Guwahati, among others.

From an elevated perspective, as observed in a recent peer review of applied mathematics in India,"it appears that India has a strong pitch and legacy of pure mathematics, and that in the last two decades or so, the number of research centres and quality teaching institutions devoted to pure mathematics have grown". However, there is a general feeling that "applied mathematics is lagging". On the other hand, even in the areas of pure mathematics, the available number of well-trained mathematics students and researchers is still a matter of concern, particularly for those involved in the task of hiring of quality faculty and research staff for their institutions. It must be emphasized, nonetheless, that applied mathematics plays a key role for high class innovations and solutions of complex technical problems faced in industry, and this requires ingenious skills coming from a good training in mathematics both pure and applied.

Finally, keeping in view that many new national institutions (new IITs and IISERs) have entered the scene in the last few years, it would appear desirable for the organic growth of mathematical sciences in the country to recommend that these institutes may possibly be entrusted the task of systematically nurturing and synergizing *applied and applicable mathematics*.

References

[1] Kapur, J. N., Development of Mathematical Sciences in India during the Twentieth Century, *Indian Journal of History of Science*, **27**(4), 1992, 389–407.

[2] Narasimhan, R., *The coming age of mathematics in India*, in: Hilton P., Hirzebruch F., Remmert R., eds., Miscellanea Mathematica, Berlin, Springer-Verlag, 1991, 235–258.3.

[3] Sathyamurthy, N., IISERs: Emerging science universities of India, Guest Editorial, *Current Science*, **110**(5), 2016, 747–748.

[4] Seshadri, C. S., *Mathematics in India during the last fifty years*, text of the talk delovered at the Indo-French Seminar on History of Development of Science in India and in France, Madras, October, 1992.

[5] Sreenivasan, K. R., A perspective on the status of mathematics in India, *Current Science*, **93**(8), 2007, 1080–1087.

[6] Varadarajan, V. S., Mathematics in and out of Indian universities, *The Mathematical Intelligencer*, **5**, 1983, 38–42.

[7] Bourbaki, Nikolas, *Elements of the History of Mathematics*, Springer Verlag, Berlin, Heidelberg, and New York, (1998).

D. V. Pai
Visiting Professor, Mathematics
Indian Institute of Technology Gandhinagar
Palaj, Gandhinagar, Gujarat-382355, India
E-mail: *dvp@iitgn.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 11-29

SOME HIGHLIGHTS OF OUR RESEARCH CONTRIBUTIONS*

D. V. PAI

Abstract. In the first part of this exposition, one will review certain early contributions of the author in the area of Optimization and Approximation, which are of some *nostalgic value* to him. In the second part, one will be mostly concerned with *stability and well-posedness* considerations of problems in approximation theory. Specifically, the required hyperspace topologies on certain subfamilies of nonempty closed sets will be reviewed here in the context of continuity and well-posedness of the *prox multifunction* and the *restricted center multifunction*.

1. Resume of Some Highlights of Major Contributions

I feel greatly honored that the Indian Mathematical Society, which is the oldest and the biggest of the mathematical societies in the country, has elected me to this august office of its President for the year 2016-17. It also entrusted me with the task of presiding over its 82nd Annual Conference which took place at the University of Kalyani, Kalyani, West Bengal in December, 2016. At the outset, I must express my humble and deep sense of indebtedness to the Council and the General Body of the Society for giving me this valuable opportunity.

Before giving a brief resume of some of the highlights of our work, since most of it involves global approximation, it may be appropriate to give below some of the pertinent quotes:

"Because the shape of the whole universe is most perfect, and, in fact designed by the wisest creator, nothing in all of the world will occur in which no maximum or minimum rule is shining forth".

Leonard Euler

"The profound significance of well-posed problems for advancement of mathematical science is undeniable".

David Hilbert

* The text of the Presidential Address (technical) delivered at the 82^{nd} Annual Conference of the Indian Mathematical Society held at the University of Kalyani, Kalyani-741 235, Nadia, West Bengal, India during December 27 - 30, 2016.

Mathematics Subject Classifications: 41A28, 41A52, 41A65.

Key words and Phrases: smooth norm, multioptimum, prox multifunction, restricted center multifunction, hypertopologies.

Our initial contributions arose from a study of the notion of *prox points* of a pair of convex sets in a normed linear space in Pai [29]. Existence, uniqueness, characterization and computability of prox points were studied there. A new characterization of smooth normed spaces was obtained in Pai [30] using this notion. Subsequently, in Pai [31], this result was considerably generalized by providing an answer to the following question in *Convex Analysis*: When is a *multioptimum* of a convex functional F defined on a catesian product of Hausdorff locally convex spaces for a cartesian product of convex subsets of these spaces an *optimum* for the same? This result was further generalized in Pai [32] to a cartesian product of certain regular subsets of these spaces. Connection of this result with *Nash equilibrium point* in Game Theory was considered in Pai and Veeramani [34]. Later, an interesting generic theorem was contributed in Beer and Pai [9] for points of single-valuedness of the *prox multifunction* for pairs of convex sets, along with certain applications to some results for best approximation and fixed points of convex-valued multifunctions. Hyperspace topologies related to the stability of the prox map were also investigated in Pai and Deshpande [39]. These investigations also seem to have drawn the attention of some other workers in this domain. For example, mention must be made here of the interesting contributions by De Blasi, Myjak and Papini [17, 18]. Also, the papers of Li and Ni [24], and Li and Xu [25] may be mentioned in this connection. All these articles refer to our early contributions, particularly to those in Beer and Pai [9]. Reference to some of these early works may also be found in many articles referred to in the recent survey articles of Veeramani and Rajesh [55], Singh and Singh [52].

Our more recent contributions are to the topic *Stability and Well-posedness in Optimization and Approximation*. This work began with our joint efforts with G. Beer in the early part of 90's, and it resulted in our joint papers [8, 9, 10]. Various notions of convergence of convex sets and their relation to the stability of the restricted Chebyshev centers were analysed in Beer and Pai [8], and this led to a subtle generic theorem for points of single-valuedness of the restricted center multifunction. Topologies related to stability of restricted center multifunction have also been discussed in Pai and Deshpande [39]. This work seems to have been taken note of, by some other contemporary co-workers in set-valued analysis. In this connection, among others, the articles of Attouch and Beer [1, 2], Beer and Borwein [7], Attouch, Moudafi and Riahi [3], and the monograph of Beer [15] must be mentioned. This investigation was continued further in the Ph.D. thesis of Shunmugaraj [49], and it resulted in our joint papers [50, 51]. A notable contribution in Shunmugaraj and Pai [50] was a new notion of convergence of sets which was initially called *Shunmugaraj-Pai convergence* in the paper by Sonntag and Zalinescu [54]. This notion of convergence was subsequently called the *bounded*

proximal convergence by other co-workers in this domain ([13, 14]), and they studied the corresponding hyperspace topology called the *bounded proximal topology*. The paper Shunmugaraj and Pai [50] has been referred to in many articles and monographs. A partial list of these would include [53, 13, 14], and also the two monographs [15, 48].

In Shunmugaraj and Pai [51], an interesting generic uniqueness theorem for solution sets for convex optimization problems was established. This work was further continued in the Ph.D. thesis of Deshpande [19] which also resulted in our joint publications [39, 40, 41]. In Pai and Deshpande [41], we contributed a unified approach to *hypertopologies* on collections of certain subsets of a Hausdorff uniform space, and more importantly, identified a suitable topology on the family of proper convex and lower semicontinuous functions defined on a Hausdorff locally convex space for which the *Young Fenchel transform* of convex analysis is bicontinuous. This result improved a previously known result due to Mosco, Joly and Beer. In effect, this paper implemented the *desiratarium* expressed at the end of the second para of Notes and References of the monograph [15].

More recently, our main contributions have been to the study of *strong uniqueness of simultaneous approximation*, a topic of continuing interst in the literature in *Approximation Theory*. This began with our joint work with P.J. Laurent [23]. In this article, we have contributed an important formula for the subdifferential of restricted Chebyshev radius of a bounded set (such a formula is of continuing interest in Convex Analysis after the seminal work of M. Valadier in this direction). This led us to establishing strong uniqueness of restricted Chebyshev centers for certain Haar-like subspaces. The papers [23, 42] have been referred to in many articles of C. Li and his co-workers, cf., e.g. [26, 27]. This work was subsequently continued in our investigations reported in [35, 36, 38, 43, 21, 44], and also in the Ph.D. thesis of K.Indira [22]. An interesting study of various properties of certain triplets leading to the existence and stability of restricted centers of sets was also contributed in [42]. Updated overviews of this topic related to well-posedness of the underlying problems have been contributed in [37, 43, 46]. In Indira and Pai [21], we have presented some important results for lower semicontinuity of the restricted center multifunction and Hausdorff strong uniqueness of best simultaneous approximation. These results are further extended to spaces of vector-valued functions in [45]. In this direction, mention must also be made of our book [28] contributed jointly with H. N. Mhaskar.

2. On Stability of the Prox Multifunction

2.1. **A characterization of smooth normed linear spaces.**

Definition 2.1. A normed linear space X is called *smooth* if each point of the unit sphere $S(X) = \{x \in X : \|x\| = 1\}$ has a unique support hyperplane

to the closed unit ball $U(X) = \{x \in X : \|x\| \leq 1\}$, or equivalently, if for each $x \in X, x \neq 0$, there corresponds a unique Hahn-Banach functional $x^* \in X^*$ such that $\|x^*\| = 1$ and $x^*(x) = \|x\|$.

Our initial contributions arose from a study of the notion of *prox points* of a pair of convex sets in a normed linear space.

Definition 2.2. Given a pair U, V of convex sets in a normed linear space X, the points $\overline{u}, \overline{v}$ in U, V respectively are called *prox points* of the pair of sets U, V if

$$\|\overline{u} - \overline{v}\| = D(U, V) = \inf_{u \in U, v \in V} \|u - v\|.$$

Existence, uniqueness, characterization and computability of prox points of a pair of convex sets were studied in Pai [29]. If the points $\overline{u} \in U, \overline{v} \in V$ are prox points then they are clearly the points that are mutually nearest to each other from the respective set. However, it is seen from examples in Pai [30] that, in general, the converse of this statement is false even for convex Chebyshev sets. Motivated by these examples, the following geometric property (P) was introduced in [30] for normed linear spaces.

Property (P)

- (P) For each pair U, V of convex subsets of X and points $\overline{u} \in U, \overline{v} \in V, \overline{u}$ being a nearest point of $\overline{v}$ in U and $\overline{v}$ being a nearest point of $\overline{u}$ in V, imply that $\overline{u}, \overline{v}$ are prox points of U, V.
- In Cheney and Goldstein [16], it was shown that (P) holds for a Hilbert space when U, V are closed and convex.
- In Pai [29] it was observed that (P) holds for X if its dual X^* is strictly convex.

The following result of Pai [30] gives a geometric characterization of smooth normed spaces.

Theorem 2.3. *For a normed linear space X, the following statements are equivalent.*

(1) X is smooth.
(2) X satisfies property (P).
(3) The norm $\|.\|$ in X is Gâteaux-differentiable at each nonzero point of X.

2.2. **Multioptimum of a convex functional.** Subsequently, in Pai [31], the above theorem was further generalized by providing an answer to the following question in *Convex Analysis*:

- When is a *multioptimum* of a convex functional F defined on a cartesian product of Hausdorff locally convex spaces for a cartesian product of certain convex subsets of these spaces an *optimum* for the same?
- This result was also further generalized to a cartesian product of certain regular subsets of these spaces in [32].

- In Pai and Veeramani [34] connection of this result to *Nash equilibrium point* in Game Theory was explored.

More precisely, the two questions raised in [31] were as follows:

- Question 1. Let f be a convex functional defined on a Hausdorff locally convex linear topological space X. Let U, V be a pair of convex sets in X. A pair $(\overline{u}, \overline{v}) \in U \times V$ is called a *multioptimum* for f if
$$f(\overline{u} - \overline{v}) = \inf_{v \in V} f(\overline{u} - v) = \inf_{u \in U} f(u - \overline{v}),$$
and it is simply called an *optimum* for f if $\overline{u} - \overline{v}$ is an optmum for f on $U - V$,:
$$f(\overline{u} - \overline{v}) = inf_{u \in U, v \in V} f(u - v).$$

The question that we asked was: Under what conditions is a multioptimum $(\overline{u}, \overline{v})$ an optimum for f?

Remark 2.4. Put differently, we were asking: When is a pair $(\overline{u}, \overline{v}) \in U \times V$ of elements that are mutually f-nearest to each other from the respective set, f-prox?

More generally, in [31], we were concerned with:

- Question 2: Let $X_i, i = 1, 2, \ldots, n$, be Hausdorff locally convex linear topological spaces, and let $K_i \subset X_i, i = 1, 2, \ldots, n$, be convex sets. Let F be a convex functional defined on $\prod_{i=1}^{n} X_i$. Let $\overline{x_i} \in K_i,\ i = 1, 2, \ldots, n$ be given points. Denote by $\psi_i,\ i = 1, 2, \ldots, n$, the convex functionals defined on X_i by
$$\psi_i(x_i) = F(\overline{x_1}, \overline{x_2}, \ldots, \overline{x_{i-1}}, x_i, \overline{x_{i+1}}, \ldots, \overline{x_n}).$$
One calls $(\overline{x_1}, \ldots, \overline{x_n}) \in \prod_{i=1}^{n} K_i$ a *multioptimum* for F if
$$F(\overline{x_1}, \ldots, \overline{x_n}) = \inf_{x_i \in K_i} \psi_i(x_i), i = 1, 2, \ldots, n,$$

The question that we asked was: Under what conditions is a multioptimum $(\overline{x_1}, \ldots, \overline{x_n})$ an optimum for F?

Remark 2.5. A game theoretic interpretation of the above Question 2 is as follows: Let the convex sets $K_i, i = 1, \ldots, n$ and the convex functional F be as before. Consider the co-operative game $(K, -F)$, where $K = \prod_{i=1}^{n} K_i$, denotes the set of strategy profiles corresponding to the strategy sets K_i for the players i. Given a strategy profile $\overline{x} = (\overline{x_1}, \ldots, \overline{x_n}) \in K$, let us denote as before
$$\psi_i(x_i) = F(x_i, \overline{x}_{-i}),\ i = 1, \ldots, n$$
the convex functionals defined on X_i. Here x_i denotes the strategy profile of player i and $\overline{x}_{-i}$ denotes the given strategy profile of all players except for player i. A strategy profile $\overline{x} \in K$ is called a *Nash equilibrium (NE)* for the game $(K, -F)$ provided
$$F(\overline{x}) \leq F(x_i, \overline{x}_{-i})\ \ \forall x_i \in K_i, i = 1, \ldots, n.$$

Clearly, Nash equilibrium for the payoff functions ψ_i's is precisely what has been called multioptimum before.

Theorem 2.6. *Let F be a proper convex function on $\prod_{i=1}^{n} X_i$, and that it is finite and continuous at $(\overline{x_1}, \ldots, \overline{x_n})$. Then in order that $(\overline{x_1}, \ldots, \overline{x_n})$ being a multioptimum for F imply that it is an optimum for F, it is sufficient that the following equality hold for the subdifferentials:*

$$\partial F(\overline{x_1}, \ldots, \overline{x_n}) = \prod_{i=1}^{n} \partial \psi_i(\overline{x_i}).$$

Remark 2.7.

- Let F satisfy the same hypothesis as in the last theorem. In general, it is easily seen that the following inclusion holds for the subdifferentials:

$$\partial F(\overline{x_1}, \ldots, \overline{x_n}) \subset \prod_{i=1}^{n} \partial \psi_i(\overline{x_i}).$$

- Under the same hypothesis as in the last theorem, F is Gâteaux differentiable at $(\overline{x_1}, \ldots, \overline{x_n})$ if and only if the functions ψ_i are Gâteaux differentiable at the points $\overline{x_i}, i = 1, \ldots, n$. In this case the equality for the subdifferentials as in the preceding theorem holds.
- Aside from the differentiable case of the preceding remark, another simple case, wherein this equality holds for the subdifferentials, is the following:

$$F(x_1, \ldots, x_n) = f_1(x_1) + \ldots, + f_n(x_n),$$

where $f_i \in conv(X_i), i = 1, \ldots, n$.

Theorem 2.8. *Let f be a proper convex function on X, and let it be finite and continuous at $\overline{u} - \overline{v}$. Then in order that $\overline{u} \in U, \overline{v} \in V$ being mutually f-nearest to each other from the respective set imply that they are f-prox, it is sufficient that f be Gâteaux-differentiable at $\overline{u} - \overline{v}$.*

Theorem 2.9. *Let f be a continuous guage function on X. Then in order that for arbitrarily given convex sets U, V in X and points $\overline{u} \in U, \overline{v} \in V$ such that $f(\overline{u} - \overline{v}) \neq 0, \overline{u}, \overline{v}$ being mutually f-nearest to each other from the respective set imply that they are f-prox, it is necessary and sufficient that f be Gâteaux-differentiable at each point $x \in X$, where $f(x) \neq 0$.*

2.3. Hyperspace topologies related to the prox map.

Definition 2.10. Let X be a metric space and $CL(X)$(resp. $CLB(X)$) denote the nonempty closed (resp. the nonempty closed and bounded) subsets of X.

- Given A, B in $CL(X)$ and $x \in X$, let $d(x, B)$ denote the distance between x and the set B, and let

$$D(A, B) := inf_{a \in A, b \in B} d(a, b) = inf_{a \in A} d(a, B)$$

denote the *gap* between the sets A, B.

- Any point $a \in A$ such that $d(a, B) = D(A, B)$ is called a *prox point* of B in A and a pair $(a, b) \in A \times B$ such that $d(a, b) = D(A, B)$ is called a *prox pair* of the pair (A, B) of sets.
- We denote by $Proximal(B, A)$ (resp.$Prox(B, A)$) the (possibly void) set of prox points of B in A (resp. prox pairs of the pair (B, A) of sets).
- Let us also recall that the *Hausdorff metric topology* τ_H on $CL(X)$ is the one induced by an infinite valued metric on $CL(X)$ (which when restricted to $CLB(X)$ is a finite valued metric), defined by

  $$H(A, B) = max\{sup_{a\in A} d(a, B), sup_{b\in B} d(b, A)\}, \quad A, B \in CL(X).$$

  It is well known (cf.,e.g. [15]), that

  $$H(A, B) = sup_{x\in X} |d(x, A) - d(x, B)|, \quad A, B \in CL(X).$$

A natural problem that arises in the context of continuity of the prox map is:

- To identify suitable families of sets $\mathcal{A}, \mathcal{B}$ in $CL(X)$ and appropriate topologies on them so as to ensure continuity of the gap functional

  $$(A, B) \to D(A, B) \text{ on } \mathcal{A} \times \mathcal{B}.$$

- A particular case was treated in [9] in the framework of a dual normed space.

2.4. **A unified approach to hyperspace topologies.** From the point of view of adressing the stability questions in optimization as one of the goals, a number of hyperspace topologies, the so-called *hypertopologies* were introduced by a number of authors, cf., e.g.,[12, 5, 11, 53, 13]. A simple unified approach to these hypertopologies has emerged in [13, 53], which is as follows: Given the families of sets $\mathcal{A}, \mathcal{B}$ in $CL(X)$, as before for each $B \in \mathcal{B}$, define

$$p_B(A, A') = |D(A, B) - D(A', B)|.$$

- Let us denote by $\tau(\mathcal{B})$ the topology on $\mathcal{A}$ corresponding to the uniformity generated by the family $\{p_B : B \in \mathcal{B}\}$ of pseudometrics on $\mathcal{A}$.
- Clearly, $\tau(\mathcal{B})$ is precisely the weak topology on $\mathcal{A}$ generated by the family of functionals $A \to D(B, A)$ obtained by varying B over $\mathcal{B}$.
- Clearly for the topology $\tau(\mathcal{B})$, a net $\{A_\lambda\}$ converges to A in $\mathcal{A}$, written $\tau(\mathcal{B}) - lim_\lambda A_\lambda = A$, if and only if

  $$lim_\lambda D(B, A_\lambda) = D(B, A), \forall B \in \mathcal{B}.$$

- This approach allows one to compare the various known hypertopologies on $\mathcal{A}$ in a natural manner.

2.5. **Hit-and-miss representations of hypertopologies.** In the sequel, the metric space X will sometimes be a normed linear space with X^* as its normed dual. In addition to $CL(X), CLB(X)$ mentioned in the introduction, we shall consider the following classes of sets:

- $S(X)$, the singletons of X;

- $K(X)$ will denote the nonempty compact subsets of X; In the linear case,
- $WCL(X)$ (resp.$W^*CL(X^*)$), will denote the nonempty weakly closed (resp. weak* closed) subsets of X (resp. X^*);
- $WK(X)$ (resp. $W^*K(X^*)$ will denote the nonempty weakly compact (resp. weak* compact) subsets of X (resp.X^*);
- $CLC(X)$ will denote the nonempty closed and convex subsets of X; and $CLBC(X)$ will denote the nonempty closed bounded and convex subsets of X.

For a subset E of X and $\mathcal{A} \subset CL(X)$, let us recall the following customary notation:

- $E^- := \{A \in \mathcal{A} : A \cap E \neq \emptyset\}$,
- $E^+ := \{A \in \mathcal{A} : A \subset E\}$,
- $E^{++} := \{A \in \mathcal{A} : D(A, E^c) > 0\}$.

Remark 2.11. In case A is in E^- (resp. E^+, resp. E^{++}), A is said to hit E (resp. miss E^c, resp. really miss E^c).

Remark 2.12.
- Following [13], one says that $\mathcal{B}$ is *stable under enlargements* if for each $B \in \mathcal{B}$ and $\epsilon > 0$, one has $V_\epsilon[B] \in \mathcal{B}$. Here $V_\epsilon[B] = \{x \in X : d(x, B) \leq \epsilon\}$ denotes the ϵ-*enlargement* of B.
- Clearly, the families $CL(X), CLB(X), CLC(X), W^*K(X^*)$ (resp. $WK(X)$ in case X is reflexive) are stable under enlargements.

The following theorem can be found in [13] and in a slightly different form in [53].

Theorem 2.13. *Suppose $\mathcal{B} \supset S(X)$ and that it is stable under enlargements. Then the topology $\tau(\mathcal{B})$ has a subbase consisting of all sets of the form V^- where V is open and all sets of the form $(B^c)^{++}$ where $B \in \mathcal{B}$.*

2.6. **Various hypertopologies.**

(i) Wijsman and proximal topologies
- *Wijsman topology*: Every family $\mathcal{B}$ such that $S(X) \subset \mathcal{B} \subset K(X)$ generates this topology. This topology τ_W has been widely studied on $\mathcal{A} = CL(X)$. See, e.g., [12, 13, 15].
- *Proximal topology*: Here $\mathcal{B} = CL(X)$. This topology τ_P on $\mathcal{A} = CL(X)$, which is known to be compatible with Fisher convergence [4] of sequences in $\mathcal{A}$, was introduced in [12]. For any $\mathcal{A} \subset CL(X)$, this topology is weaker than the Hausdorff metric topology τ_H and the two coincide precisely when X is totally bounded [12].

(ii) Bounded proximal and slice topologies
- *Bounded proximal topology*: Here $\mathcal{B} = CLB(X)$. This topology τ_{BP} on $\mathcal{A} = CL(X)$ is compatible with the bounded proximal convergence of sequences in $CL(X)$ introduced in [50, 49]. Clearly, this topology

is finer than τ_W and weaker than τ_P. This topology has been studied in detail in [13, 14].

- *Slice topology*: Here $\mathcal{B} = CLBC(X)$. This topology τ_{Sl} on $\mathcal{A} = CLC(X)$ was introduced in [11, 15, 52] as an alternative to the Mosco topology τ_M for nonreflexive spaces. When X is reflexive, $\tau_{Sl} = \tau_M$ on $CLC(X)$. Clearly τ_{Sl} is weaker than τ_M.

(iii) Mosco and dual Mosco topologies

- *Mosco topology (resp. the dual Mosco topology)*: Here $\mathcal{B} = WK(X)$ (resp.$\mathcal{B} = W^*K(X^*)$). This topology τ_M(resp.τ_{M^*}) is finer than τ_W, and in case X is reflexive (resp.the dual space X^*) and $\mathcal{A} \subset WCL(X)$(resp. $\mathcal{A} \subset W^*CL(X^*)$), this topology coincides with the restriction of the topology τ_M (resp.τ_{M^*}) defined in [5] (resp.[9])on $\mathcal{A}$. It is generated by all sets of the form V^-, where V is open and sets of the form $(B^c)^+$ where $B \in WK(X)$(resp. $B \in W^*K(X^*)$). For X reflexive and $\mathcal{A} = CLC(X)$, this topology is compatible with the classical Mosco convergence of sequences of sets (cf.[5]).

2.7. **Continuity of the gap functional.** As before, let X be a metric space and $\mathcal{A}, \mathcal{B}$ be given families in $CL(X)$.

Theorem 2.14. [39, 40] *Let X be a metric space and $\mathcal{A}, \mathcal{B}$ be families in $CL(X)$. Assume $\mathcal{B}$ contains singletons $S(X)$. Then the gap functional*

$$D : (\mathcal{B}, \tau_P) \times (\mathcal{A}, \tau(\mathcal{B})) \longrightarrow \mathbb{R}$$

is continuous. Moreover, if $\mathcal{A} = CL(X)$, then the proximal topology τ_P is the weakest topology τ on $\mathcal{B}$ such that

$$D : (\mathcal{B}, \tau) \times (\mathcal{A}, \tau(\mathcal{B})) \longrightarrow \mathbb{R}$$

is continuous.

Corollary 2.15. *In each of the following cases, the gap functional $D : (\mathcal{B}, \tau_P) \times (\mathcal{A}, \tau) \to \mathbb{R}$ is continuous.*

- *$\mathcal{B} = \mathcal{A} = CL(X)$ and $\tau = \tau_P$;*
- *$\mathcal{B} = CLB(X), \mathcal{A} = CL(X)$ and $\tau = \tau_{BP}$;*
- *$\mathcal{B} = CLBC(X), \mathcal{A} = CLC(X)$ and $\tau = \tau_{Sl}$;*
- *$\mathcal{B} = W^*K(X^*), \mathcal{A} = W^*CL(X^*)$ and $\tau = \tau_{M^*}$.*

2.8. **Upper semicontinuity of the proximal and prox maps.** As before, let X be a metric space and $\mathcal{A}, \mathcal{B}$ be families of sets in $CL(X)$. We explore topologies on them so as to yield upper semicontinuity of the proximal map over $\mathcal{B}$, and upper semicontinuity the prox map over $\mathcal{B} \times \mathcal{A}$.

Definition 2.16. [39] Let $\mathcal{B} \subset CL(X)$ and A be a nonempty subset of X. A is said to be *proximally compact* with respect to $\mathcal{B}$, if for each $B \in \mathcal{B}$, each net $< a_\lambda >$ in A satisfying $lim_\lambda d(a_\lambda, B) = D(A, B)$ has a convergent subnet to a point of A. In case X is a normed space, $\mathcal{B} \subset CL(X)$ the *w-proximal compactness* of A

with respect to $\mathcal{B}$ is defined similarly using w- convergence of the corresponding subnet.

The proximal map.

Theorem 2.17. *Let X be a metric space. Suppose $\mathcal{B} \subset CL(X)$ and $A \in CL(X)$ be such that each $B \in \mathcal{B}$ is proximinal with respect to A. If A is proximally compact with respect to $\mathcal{B}$, then for each $B \in \mathcal{B}$, $Proximal(B; A)$ is nonempty and the proximal map $B \to Proximal(B; A)$ is usco on $\mathcal{B}$ equipped with τ_P.*

In the following, we use the fact given below:

- *In case X is a normed space and $B \in CL(X)$ is convex or w-compact, then the function $x \to d(x, B)$ is w-l.s.c.*

Theorem 2.18. *Let X be a normed space. Suppose $\mathcal{B} \subset CL(X)$ and $A \in CL(X)$ are such that each $B \in \mathcal{B}$ is proximinal with respct to A, or that $\mathcal{B} \subset WK(X)$. If A is w-proximally compact with respect to $\mathcal{B}$, then for each $B \in \mathcal{B}$, $Proximal(B; A)$ is nonempty, and the proximal map $B \to Proximal(B; A)$ is w-usco on $\mathcal{B}$ equipped with τ_P.*

The prox map.

Theorem 2.19. [39] *Let X be a metric space and let $\mathcal{B} \subset CL(X)$ be such that each $B \in \mathcal{B}$ is boundedly compact. Let $\mathcal{A} \subset CL(X)$ be such that each $A \in \mathcal{A}$ is proximally compact with respect to $\mathcal{B}$. Then for each $A \in \mathcal{A}$ and $B \in \mathcal{B}, Prox(B, A) \neq \emptyset$ and the prox map: $(B, A) \to Prox(B, A)$ on $\langle \mathcal{B}, \tau_P \rangle \times \langle \mathcal{A}, \tau_P \rangle$ is usco.*

Theorem 2.20. [39] *Let X be a metric space. Let $\mathcal{B} \subset K(X)$ and $\mathcal{A}$ be a family of nonempty approximatively compact subsets of X. Then for each $A \in \mathcal{A}$ and $B \in \mathcal{B}, Prox(B, A) \neq \emptyset$ and the prox map:$(B, A) \to Prox(B, A)$ on $\langle \mathcal{B}, \tau_P \rangle \times \langle \mathcal{A}, \tau_{BP} \rangle$ is usco.*

The next theorem improves [10], Proposition 3.6(2).

Theorem 2.21. *Let X be a normed space. Let $\mathcal{B} \subset WK(X)$ and $\mathcal{A} \subset CL(X)$. If each $A \in \mathcal{A}$ is w-proximally compact with respect to $\mathcal{B}$, then for each $A \in \mathcal{A}$ and $B \in \mathcal{B}, Prox(B, A) \neq \emptyset$ and the prox map:$(B, A) \to Prox(B, A)$ is w-usco on $\langle \mathcal{B}, \tau_P \rangle \times \langle \mathcal{A}, \tau_{BP} \rangle$.*

Remark 2.22. (i) In the preceding theorem, if one replaces $\mathcal{B} \subset WK(X)$ by $\mathcal{B} \subset CL(X)$ such that each $B \in \mathcal{B}$ is boundedly compact, then the conclusion of this theorem holds with τ_{BP} replaced by τ_P on $\mathcal{A}$.

(ii) In case X is reflexive (resp. a dual normed space X^*), $\mathcal{B} = WK(X)$ $(resp. W^*K(X^*))$ and $\mathcal{A} = WCL(X)$(resp., $W^*CL(X^*)$) the preceding theorem holds with τ_{BP} replaced by the weaker topology $\tau_M (resp. \tau_{M^*})$ on $\mathcal{A}$. This improves [9], Theorem 3.3.

2.9. **Best approximation of convex-valued multifunctions.** Let us begin by recalling the following well known fixed point theorem for multifunctions.

Theorem 2.23. *(Himmelberg) Let C be a nonempty convex subset of a Hausdorff locally convex space X. Let $F : C \to C$ be an upper semicontinuous multifunction with closed convex values. If $F(C)$ is relatively compact, then F has a fixed point.*

Theorem 2.24. [39] *Let C be a nonempty convex subset of a normed space X. Let $F : C \to < CLC(X), \tau_P >$ be continuous, where C is equipped with the norm topology. Assume Fx is proximinal with respect to C for each $x \in C$. If $\mathcal{B} := \{Fx : x \in C\}$ is τ_P-relatively compact and C is proximally compact with respect to $\mathcal{B}$, then there exists $x \in \partial C$ (here ∂C denotes the boundary of C) such that*

$$d(x, Fx) = D(C, Fx).$$

Proof. For proving the above theorem, one considers the multifunction

$$G : C \to CLC(C)$$

defined by $G(x) = Proximal(Fx; C)$ and shows that G has a fixed point $x \in C$ applying Himmelberg's fixed point theorem. □

Theorem 2.25. [39] *Let C be a nonempty convex subset of a normed space X. Let*

$$F : C \to \langle CLC(X), \tau_P \rangle$$

be continuous, where C is equipped with the topology w. Assume Fx is proximinal with respect to C for each $x \in C$. If $\mathcal{B} := \{Fx : x \in C\}$ is τ_P-relatively compact and C is w-proximally compact with respect to $\mathcal{B}$, then there exists $x \in \partial C$ such that

$$d(x, Fx) = D(C, Fx).$$

2.10. **Generic uniqueness of prox maps. Baire category result**

Theorem 2.26. [9] *Let X be a separable reflexive space. Suppose $WKC(X)$ is equipped with the Hausdoff metric topology τ_H and $CLC(X)$ is equipped with τ_M. Then*

$$\Omega := \{(B, A) \in WKC(X) \times CLC(X) : D(B, A) > 0\}$$

as a subspace of $WKC(X) \times CLC(X)$ is open and completely metrizable.

Theorem 2.27. [9] *Let X be a separable reflexive space. Suppose $WKC(X)$ is equipped with τ_H and $CLC(X)$ is equipped with τ_M. Then there exists a dense and G_δ subset Ω_o of*

$$\Omega := \{(B, A) \in WKC(X) \times CLC(X) : D(B, A) > 0\}$$

such that for each $(B, A) \in \Omega_o, Prox(B, A)$ is a singleton.

3. On Stability of the Restricted Center Multifunction

3.1. **Restricted Chebyshev centers of sets and simultaneous best approximation.** Let X be a metric space, which will mostly be a normed linear space. We distinguish the following classes of normed spaces:

(Rf) := the reflexive Banach spaces,

(R) := the rotund (strictly convex) normed spaces,

(A) := the normed spaces for which the norm satisfies the Kadec property : w convergence of a sequence in $S(X)$ entails its norm convergence,

(UR) := the uniformly convex Banach spaces. We denote the class of spaces $(Rf) \cap (A)$ by (CD) and the class of spaces $(CD) \cap (R) = (Rf) \cap (R) \cap (A)$ by (D). Let the families $\mathcal{B} \subset CLB(X)$ and $\mathcal{A} \subset CL(X)$ be given. For $A \in \mathcal{A}, B \in \mathcal{B}$ and $x \in X$, let $r(B,x) := sup\{d(x,y) : y \in B\}$ denote the radius of the smallest closed ball centered at x covering B and let

$$rad(B;A) := inf\{r(B,x) : x \in A\},$$
$$Cent(B;A) := \{x \in A : r(B,x) = rad(B;A)\}.$$

The number $rad(B;A)$ is called the *restricted (Chebyshev) radius* of B in A. It is the *intrinsic error* in the problem of simultaneous approximation (global approximation) of the bounded *data set* B from the set A. Any element of the set $Cent(B;A)$ (possibly void) is called a *restricted (Chebyshev) center or best simultaneous approximant* of B in A. The problem of probing the continuity of the *restricted center multifunction* $\mathcal{B} \times \mathcal{A} : (B,A) \rightrightarrows Cent(B;A) \in CL(A)$ leads us to the problem of identifying suitable families $\mathcal{A}, \mathcal{B}$ of sets in question and appropriate topologies on them, so as to ensure continuity of the restricted radius functional $(B,A) \to rad(B;A)$ on $\mathcal{B} \times \mathcal{A}$.

3.2. **Bivariate continuity of the restricted radius functional.** As before, let X be a metric space, $\mathcal{B} \subset CLB(X), \mathcal{A} \subset CL(X)$. For $B, B' \in \mathcal{B}$ and $x \in X$, let $r_x(B,B') := |r(B,x) - r(B',x)|$. Likewise, for $A, A' \in \mathcal{A}$ and $B \in CLB(X)$, let $R_B(A,A') := |rad(B;A) - rad(B,A')|$. The topology corresponding to the uniformity on $\mathcal{B}$ (resp. $\mathcal{A}$) generated by the family of pseudometrics $\{r_x : x \in X\}$ (resp.$\{R_B : B \in \mathcal{B}\}$) will be denoted by τ_r (resp. $\tau_D(\mathcal{B})$). For X a normed space, $\mathcal{B} = CLB(X)$ and $\mathcal{A} = CLC(X)$, the topology $\tau_D(\mathcal{B})$ on $\mathcal{A}$ was called the *distal topology* in [8]. For arbitrary families $\mathcal{B} \subset CLB(X)$ and $\mathcal{A} \subset CL(X)$ such that $\mathcal{B}$ contains all finite subsets of X, the topology $\tau_D(\mathcal{B})$ is generated by all sets of the form V^-, where V is open and $(B^c)^+$, where B is an intersection of a finite family of closed balls with a common radius in each of the following cases;

(i) X is a metric space such that each closed ball in X that is a proper subset of X is compact;

(ii) X is a dual normed space.

Clearly, $\tau_D(\mathcal{B})$ is the weakest topology on $\mathcal{A}$ such that $A \to rad(B;A)$ is

continuous on $\mathcal{A}$ for each fixed $B \in \mathcal{B}$.

Lemma 3.1. *Let X be a metric space and $\mathcal{B} \subset CLB(X)$. Then the topology τ_r on $\mathcal{B}$ is weaker than the topology τ_P.*

Theorem 3.2. [39] *The restricted radius functional $rad : \langle \mathcal{B}, \tau_H \rangle \times \langle \mathcal{A}, \tau \rangle \to \mathbb{R}$ is continuous in each of the following cases:*

(i) X is a metric space, $\mathcal{B} = CLB(X), \mathcal{A} = CL(X)$ and $\tau = \tau_{BP}$;
(ii) X is a normed space, $\mathcal{B} = CLBC(X), \mathcal{A} = CLC(X)$ and $\tau = \tau_{Sl}$;
(iii) X^* is a dual normed space, $\mathcal{B} = W^*K(X^*), \mathcal{A} = W^*CL(X^*)$ and $\tau = \tau_{M^*}$.

Let $\mathcal{B} \subset CLB(X)$ and $\tilde{\tau}$ denote the supremum of the two topologies τ_r and τ_V^-, on $\mathcal{B}$.. By the previous lemma, we observe that $\tilde{\tau}$ is weaker than τ_P on $\mathcal{B}$.

Theorem 3.3. [39] *In each of the following cases, $\tilde{\tau}$ is the weakest topology τ_1 on $\mathcal{B}$ containing τ_V^- such that the restricted radius functional $rad : \langle \mathcal{B}, \tau_1 \rangle \times \langle \mathcal{A}, \tau \rangle \to \mathbb{R}$ is continuous.*

(i) X is a metric space, $\mathcal{B} = CLB(X), \mathcal{A} = BK(X)$ and $\tau = \tau_{BP}$;
(ii) X^* is a dual normed space, $\mathcal{B} = W^*K(X^*), \mathcal{A} = W^*CL(X^*)$ and $\tau = \tau_{M^*}$.

Remark 3.4. Clearly, $rad : \langle \mathcal{B}, \tau_P \rangle \times \langle \mathcal{A}, \tau \rangle \to \mathbb{R}$ is continuous in each of the two cases of the last theorem.

3.3. **Upper semicontinuity of the restricted center multifunction.**

Theorem 3.5. [39] *Let X be a metric space, $\mathcal{B} = CLB(X)$ and $\mathcal{A}$ be a family of nonempty boundedly compact subsets of X. Then for each $B \in \mathcal{B}$ and $A \in \mathcal{A}, Cent(B; A)$ is nonempty and the restricted center multifunction*

$$Cent : \langle \mathcal{B}, \tilde{\tau} \rangle \times \langle \mathcal{A}, \tau_{BP} \rangle \rightrightarrows K(X)$$

is usco.

Theorem 3.6. [39] *Let X be a normed linear space, $\mathcal{B} = CLB(X)$ and A be a family of nonempty boundedly w-compact subsets of X. Then for each $B \in \mathcal{B}$ and $A \in \mathcal{A}, Cent(B; A)$ is nonempty and the restricted center multifunction*

$$Cent : \langle \mathcal{B}, \tilde{\tau} \rangle \times \langle \mathcal{A}, \tau_{BP} \rangle \rightrightarrows WK(X)$$

is w-usco.

Let us note that in case $X \in (Rf)$ and $\mathcal{A} = WCL(X)$, the preceding theorem holds by replacing τ_{BP} by the weaker topology τ_M on $\mathcal{A}$.

Theorem 3.7. [39] *Let X^* be a dual normed space, $\mathcal{B} = CLB(X^*)$ and $\mathcal{A} = W^*CL(X^*)$. Then the restricted center multifunction*

$$Cent : \langle \mathcal{B}, \tilde{\tau} \rangle \times \langle \mathcal{A}, \tau_{M^*} \rangle \rightrightarrows W^*CL(X^*)$$

is w^*-usco.

3.4. **Simultaneous best approximation of convex-valued multifunctions.** We give below an extension of our earlier result ([33],Lemma 3.6) and also a fixed point result for multifunctions.

Theorem 3.8. [33, 39] *Let C be a nonempty boundedly compact (resp. boundedly w-compact) convex subset of a normed linear space X. Let $F : C \rightrightarrows \langle CLB(X), \tilde{\tau}\rangle$ be a continuous multifunction, where C is equipped with the norm topology (resp. the topology w). If $\mathcal{B} := \{Fx : x \in C\}$ is $\tilde{\tau}$-relatively compact, then there exists $x \in \partial C$ such that*

$$r(Fx, x) = rad(Fx; C).$$

Recall that for a subset C of X, the *inward cone* of C at x is this set

$$I_C(x) := \{z \in X : z = x + \alpha(u - x), \text{ for some } u \in C, \alpha \geq 0\}.$$

Theorem 3.9. [33, 39] *Let C and F be as in the last theorem. If for each $x \in \partial C$ such that $r(Fx, x) > 0$, we have*

$$rad(Fx; clI_C(x)) < r(Fx, x),$$

then F has a fixed point x, such that $Fx = \{x\}$.

4. Well-posedness of problems in approximation theory

4.1. **A review of some well-posedness notions for minimization problems.** Given a nonempty subset V of a metric space X and a function $I : E \to (-\infty, \infty]$ which is a proper extended real-valued function, let us review some well-posedness notions of the following abstract minimization problem:

$$\min I(v), v \in V,$$

which we denote by (V, I). Let $v_V(I) := \inf\{I(v) : v \in V\}$ denote the *optimal value function*. We assume I to be lower bounded on V, i.e., $v_V(I) > -\infty$, and let $\arg\min_V(I)$ denote the (possibly void) set $\{v \in V : I(v) = v_V(I)\}$ of optimal solutions of problem (V, I). For $\epsilon \geq 0$, let us also denote by ϵ- $\arg\min_V(I)$ the nonempty set $\{v \in V : I(v) \leq v_V(I) + \epsilon\}$ of ϵ-approximate minimizers of I. Recall (cf., e.g., [20], p.1) that problem (V, I) is said to be (i)*Tikhonov well-posed* if I has a unique global minimizer on V towards which every *minimizing sequence* (i.e., a sequence $\{v_n\} \subset V$, such that $I(v_n) \to v_V(I)$) converges. Put differently, there exists a point $v_0 \in V$ such that $\arg\min_V(I) = \{v_0\}$, and whenever a sequence $\{v_n\} \subset V$ is such that $I(v_n) \to I(v_0)$, one has $v_n \to v_0$.

It is said to be (ii) *generalized well-posed* (abbreviated g.w.p) if $\arg\min_V(I)$ is nonempty and every minimizing sequence for (V, I) has a subsequence convergent to an element of $\arg\min_V(I)$.

In case $V \in WCL(X)$, where X is a normed linear space, the problem (V, I) is said to be *w-T.w.p.* (resp. *w-g.w.p.*), if it is Tikhonov well-posed (resp. generalized well-posed) for w-convergence of sequences and simply T.w.p. (resp. g.w.p.) if

it is Tikhonov well-posed (resp. generalized well-posed) for strong convergence of sequences.

Proposition 4.1. [37] *Let $V \subset X$, a metric space (resp. $V \in WCL(X), X$ a normed space). Then problem (V, I) is T.w.p. (resp. w-T.w.p.) if and only if* $\arg\min_V(I)$ *is a singleton and (V, f) is g.w.p. (resp. w-g.w.p).*

4.2. **Well-posedness of best approximants and prox points.** Let X be a normed linear space over K (either $\mathbb{R}$ or $\mathbb{C}$), $V \in CL(X)$ and $x \in X$. The problem of finding a best approximant v_0 to x in $V : \|x-v_0\| = d(x,V) = \inf_{v\in V} \|x-v\|$, is the problem (V, I_x), where $I_x(v) = \|x - v\|$. Recall(cf., e.g., [28]) that the set V is called (i) *Chebyshev* if each $x \in X$ has a unique best approximant in V; It is called (ii) *almost Chebyshev* if each x in a dense and G_δ subset X_0 of X admits a unique best approximant in V; It is called (iii) *approximatively compact* (resp. *approximatively w-compact*) if each minimizing sequence has a subsequence convergent (resp. w-convergent) to an element of V. Here the multifunction $x \rightrightarrows P_V(x)$ of X to V, where $P_V(x) = \arg\min_V(I_x)$ is called the *metric projection* of X onto V.

Remark 4.2.
(i) the best approximation problems $(V, I_x), x \in X$ are all T.w.p. (resp. w-T.w.p) if and only if the set V is Chebyshev and approximatively compact (resp. approximatively w-compact).
(ii) A Banach space X is in the class $(D) = (Rf) \cap (R) \cap (A)$ if and only if each member of $CLC(X)$ is Chebyshev and approximatively compact. Hence, it follows from the first remark that:
(iii) A Banach space X is in the class (D) if and only if for each $V \in CLC(X)$, each problem $(V, I_x), x \in X$ is T.w.p.

Theorem 4.3. [46] *For a Banach space X the following statements are equivalent.*
(i) $X \in (CD) = (Rf) \cap (A)$.
(ii) For each $V \in CL(X)$, the family of problems $(V, I_x), x \in X\backslash V$ is generically g.w.p.

Let us now turn to well-posedness of the prox points.

It is easily seen that Prox $(B, A) \neq \emptyset$ whenever X is in (Rf) and (B, A) is in $WKC(X)\times CLC(X)$. In what follows, we consider the multifunction

$$\text{Prox} : WKC(X)\times CLC(X) \rightrightarrows X\times X.$$

As observed in [9], if $WKC(X)$ is equipped with the topology τ_H and $CLC(X)$ is equipped with τ_M, then the product space $WKC(X)\times CLC(X)$ is completely metrizable whenever X is reflexive and separable. The same thing can be said about its subspace $KC(X)\times CLC(X)$, since τ_H restricted to $KC(X)$ is complete. It is therefore meaningful to ask generic well-posedness questions about the multifunction Prox defined on $KC(X)\times CLC(X)$.

Let $B \in KC(X)$ and $A \in CLC(X)$. Let $I : B \times A \to \mathbb{R}$ be defined by: $I(b,a) = \|b-a\|$, $(b,a) \in B \times A$. We need the next lemma.

Lemma 4.4. *Let $X \in (Rf) \cap (A)$). If $(B,A) \in KC(X) \times CLC(X)$, then problem $(B \times A, I)$ is g.w.p.*

Theorem 4.5. [46] *Let $X \in (Rf) \cap (A)$ be separable. Suppose $KC(X)$ is equipped with τ_H and $CLC(X)$ is equipped with τ_M. Then there exists a dense G_δ subset Ω_0 of*

$$\Omega := \{(B,A) \in KC(X) \times CLC(X) : D(B,A) > 0\}$$

such that for each $(B,A) \in \Omega_0$, problem $(B \times A, I)$ is T.w.p.

4.3. **Well-posedness of restricted Chebyshev centers.** In this subsection, we adopt the same terminology and notation as in section 3.
Let us denote by $remote_V(X)$ the family of all sets in $CLB(X)$ which are 'remotal', w.r.t. V, i.e., possessing farthest points for points of V. For the next lemma and the following proposition, we refer the reader to [28](See Thorems 5 and 9 in Section 5.4, Chapter viii).

Lemma 4.6. *If $X \in (Rf) \cap (A)$ and $V \in CLC(X)$, then for each $F \in remote_V(X)$, problem (V, I_F) is g.w.p.*

Proposition 4.7. *(i) If $X \in (D), V \in CLC(X)$ and $F \in remote_V(X)$, then problem (V, I_F) is T.w.p.*
(ii) If $X \in (UR)$, $V \in CLC(X)$ and $F \in CLB(X)$, then problem (V, I_F) is T.w.p.

By ([5], Theorem 4.3), when X is reflexive and separable, $CLC(X)$ equipped with the Mosco topology τ_M is a Polish space (second countable and completely metrizable). Since $\langle CLC(X), \tau_M \rangle$ is a Baire space, it is of interest to consider the following generic theorem for Tikhonov well-posedness of restricted centers.

Theorem 4.8. [37, 46] *Let X in $(Rf) \cap (A))$ be separable. Let $K(X)$ be equipped with the topology τ_H, let $CLC(X)$ be equipped with the topology τ_M, and let the set*

$$\Omega = \{(F,V) \in K(X) \times CLC(X)) : Cent_X(F) \cap V = \emptyset\}$$

be equipped with the relative topology. Then there exists a dense G_δ subset Ω_0 of Ω such that for each (F,V) in Ω_0, the problem (V, I_F) is T.w.p.

References

[1] Attouch, H. and Beer, G., On the convergence of subdifferentials of convex functions, *Arch. Math.*, **60** (1993), 389–400.
[2] Attouch, H. and Beer, G., On some inverse stability problems for epigraphical sums, *Seminaire D'Analyse Convexe, Montpellier*, (1991), Exposé No.11.

[3] Attouch, H., Moudafi, A. and Riahi, H., Quantitative stability analysis for maximal monotone operators, *Seminaire D'Analyse Convexe, Montpellier*, (1991), Exposé No. 9.

[4] Baronti, M. and Papini, P. L., Convergence of sequence of sets, in *"Methods of Functional Analysis in Approximation Theory", ISNM* **76**, Birkhäuser-Verlag, Basel, 1986.

[5] Beer, G., On Mosco convergence of convex sets, *Bull. Austral. Math. Soc.*, **38** (1988), 239–253.

[6] Beer, G., Support and distance functionals for convex sets, *Numer. Funct. Anal. Optim.*, **10** (1989), 15–36.

[7] Beer, G. and Borwein, J. M., Mosco and slice convergence of level sets and graphs, *J. Math. Anal. Appl.*, **175** (1993), 53–67.

[8] Beer, G. and Pai, D., On convergence of convex sets and relative Chebyshev centers, *J. Approx. Theory*, **62** (1990), 147–179.

[9] Beer, G. and Pai, D., The prox map, *J. Math. Anal. Appl.*, **156** (1991), 428–443.

[10] Beer, G. and Pai, D., Proximal maps, prox maps and coincidence points, *Numer. Funct. Anal. Optim.*, **11** (1990), 429–448.

[11] Beer, G., The slice topology: a viable alternative to Mosco convergence in nonreflexive spaces, *Nonlinear Anal.*, **19** (1992), 271–290.

[12] Beer, G., Lechicki, S., Levi, S. and Naimpally, S., Distance functionals and suprema of hyperspace topologies, *Annali Mat. Pura. Appl.*, **162** (1992), 367–381.

[13] Beer, G. and Lucchetti, R., Weak topologies for the closed subsets of a metric space, *Trans. Amer. Math. Soc.*, **335** (1993), 805–822.

[14] Beer, G. and Lucchetti, R., Well-posed optimization problems and a new topology for the closed subsets of a metric space, *Rocky Mountain J. Math.*, **23** (1993), 1197–1220.

[15] Beer, G., *Topologies on Closed Convex Sets*, Kluwer Academic Publishers Proc., Netherlands, (1993).

[16] Cheney, W. and Goldstein, A. A., Proximity maps for convex sets, *Proc. Amer. Math. Soc.*,**10** (1959), 448–450.

[17] De Blasi, F. S., Myjak, J. and Papini, P. L., On mutually nearest and mutually farthest points of sets in Banach spaces, *J. Approx. Theory*, **70**(2) (1992), 142–155.

[18] De Blasi, F. S., Myjak, J. and Papini, P. L., Porous sets in best approximation theory, *J. London Math. Soc.*, **44**(2) (1991), 135–142.

[19] Deshpande, B. M., "Hypertopologies on uniform spaces and stability in optimization", *Ph.D. Dissertation, Indian Institute of Technology, Bombay*, 1997.

[20] Dontchev, A. L. and Zolezzi, T., "Well-posed Optimization Problems", *Lecture Notes in Mathematics, No.***1543**, Springer-Verlag, Berlin, 1993.

[21] Indira, K. and Pai, D. V., Hausdoff strong uniqueness in simultaneous approximation. Part I, in: *"Approximation Theory XI: Gatlinburg 2004"*, Chui, C. K., Neamtu, M. and Schumaker, L. L.(eds), 101–118, Nashboro Press, Brentwood, TN, USA, 2005.

[22] Indira, K., "Restricted center multifunction in approximation theory", *Ph.D. Dissertation, Indian Institute of Technology, Bombay*, 2000.

[23] Laurent, P. J. and Pai, D. V., On simultaneous approximation, *Numer. Funct. Anal. and Optimiz.*,**19** (1998), 1045–1064.

[24] Li, C. and Ni, R. X., On well-posed mutually nearest and mutually furthest point problems, *Act. Math. Sinica*, **20**(1) (2004), 147–156.

[25] Li, C. and Xu, H-K., Porosity of mutually nearest and mutually furthest poits, *J. Approx. Theory*, **125** (2003), 10–25.

[26] Li, C., Strong uniqueness of the restricted Chebyshev center with respect to an RS-set in a Banach space, *J. Approx. Theory*, **135** (2005), 35–53.

[27] Li, C. and Luo, X-F., Restricted p-centers for sets in real locally convex spaces, *Numer. Funct. Anal. Optim.*, **26**(3) (2005), 407–426.

[28] Mhaskar, H. N. and Pai, D. V., *Fundamentals of Approximation Theory*, Narosa Publishing House, New Delhi, India, and CRC Press, Boca Raton, Florida, USA, 2000.

[29] Pai, D. V., Proximal points of convex sets in normed linear space, *Yokohama Math. J.*, **22** (1974), 52–78.

[30] Pai, D. V., A characterization of smooth normed linear spaces, *J. Approx. Theory*, **17** (1976), 315–320.

[31] Pai, D. V., Multioptimum of a convex functional, *J. Approx. Theory*, **19** (1977), 83–99.

[32] Pai, D. V., Multi-optimum d'une fonctionnelle convexe sur des ensembles réguliers, *C. R. Acad. Sc., Paris*, **280** (1975), 1185–1188.

[33] Pai, D. V., Abstract minimization problems, restricted centers and fixed points, in *"Methods of Functional Analysis in Approximation Theory", ISNM* **76**, Birkhäuser-Verlag, Basel, 1986.

[34] Pai, D. V. and Veeramani, P., Applications of fixed point theorems to problems in optimization and best approximation, in: *"Nonlinear analysis and applications"* (St. Johns, Nfld., 1981), 393–400, *Lecture Notes in Pure and Appl. Math.,* **80**, Dekker, New York, 1982. MR84b: 47074.

[35] Pai, D. V., Strong uniqueness of best simultaneous approximation, *J. Indian Math. Soc.*, **67** (2000), 201–215.

[36] Pai, D. V., Strong unicity of order two in simultaneous approximation, *The Mathematics Student* **72** (2003), 113–121.

[37] Pai, D. V., On well-posedness of some problems in approximation, *J. Indian Math. Soc.* **70** (2003), 1–16.

[38] Pai, D. V., Strong unicity of restricted p-centers, *Numer. Funct. Anal. Optimiz.*, **29** (2008), 638–659.

[39] Pai, D. V. and Deshpande, B. M., Topologies related to the prox map and the restricted center map, *Numer. Funct. Anal. Opt.*, **16** (1995), 1211–1231.

[40] Pai, D. V. and Deshpande, B. M., On continuity of the prox map and the restricted center map, in: Chui, C. K., Schumaker L. L.(eds.) *Approximation Theory VIII, Vol.* **I***: Approximation and Interpolation*, World Scientific, New Jersey (1995), 451–458.

[41] Pai, D. V. and Deshpande, B. M., On hypertopologies and the Young-Fenchel transform, *Numer. Funct. Anal. Optim.*, **21** (2000), 885–900.

[42] Pai, D. V. and Nowroji, P. T., On restricted centers of sets, *J. Approx. Theory* **66** (1991), 170–189.

[43] Pai, D. V. and Indira, K., On well-posedness of some problems in approximation theory, in: *"Advances in Constructive Approximation"*, Neamtu, M. and Saff, E. B. (eds), 371–392, Nashboro Press, Brentwood, TN, 2004.

[44] Pai, D. V., Indira, K., Hausdorff strong uniqueness in simultaneous approximation. Part II, in: *"Frontiers in Interpolation and Approximation"*, Govil, N. K., Mhaskar, H. N., Mohapatra, R. N., Nashed, Z. and Szabados, J. (eds), 365–380, Chapman & Hall/CRC, Boca Raton, USA, 2007.

[45] Pai, D. V., On lower semicontinuity of the restricted center multifunction, *Analyse Numérique et Théorie de L'Approximation(ANTA)*, **38**(1) (2009), 83–99.

[46] Pai, D. V., Well-posedness, regularization, and viscosity solutions of minimization problems, in: *"Nonlinear Analysis: Approximation Theory, Optimization and Applications"*, Ansari, Q. H. (ed),"Trends in Mathematics" series, 135–164, Birkhäuser, Springer, New Delhi, 2014.

[47] Pai, D. V., Continuity of the restricted center multifunction, *J. Indian Math. Soc., Special Centenary Volume (1907-2007)*, 131–148.

[48] Rockafellar, R. T. and Wets, R. J-B., *Variational Analysis*, Springer, Berlin, 1998.

[49] Shunmugaraj, P., "On stability aspects in optimization", *Ph.D. Dissertation, Indian Inst. Tech., Bombay*, 1990.

[50] Shunmugaraj, P. and Pai, D. V., On stability of approximate solutions of minimization problems, *Numer. Funct. Anal. Optim.*, **12** (1991), 599–610.

[51] Shunmugaraj, P. and Pai, D. V., On approximate minima of convex functional and lower semicontinuity of metric projections, *J. Approx. Theory* **64** (1991), 25–37.

[52] Singh, S. P. and Singh, M. R., Best approximation in nonlinear functional analysis, in: *"Nonlinear Analysis: Approximation Theory, Optimization and Applications"*, Ansari, Q. H. (ed),"Trends in Mathematics" series, 165–198, Birkhäuser, Springer, New Delhi, 2014.

[53] Sonntag, Y. and Zalinescu, C., Set convergences. An attempt of classification., *Trans. Amer. Math. Soc.*, **340** (1993), 199–226.

[54] Sonntag, Y. and Zalinescu, C., Set convergences. An attempt of classification., in: *"Differential equations and Control Theory"*, (ed) Barbu, V., Pitman Research Notes in Math. Series No. **250** (1991), 312–323.

[55] Veeramani, P. and Rajesh, S., Best proximity points, in: *"Nonlinear Analysis: Approximation Theory, Optimization and Applications"*, Ansari, Q. H. (ed), "Trends in Mathematics" series, 1–32, Birkhäuser, Springer, New Delhi, 2014.

D. V. Pai
Visiting Professor, Mathematics
Indian Institute of Technology Gandhinagar
Palaj, Gandhinagar, Gujarat-382355, India
E-mail: *dvp@iitgn.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 31-37

ON POWER CONTRACTIONS AND DISCONTINUITY AT THE FIXED POINT

RAVINDRA K. BISHT

(Received : 22 - 08 - 2016, Revised : 04 - 03 - 2017)

Abstract. In this paper, we show that power contractions of generalized Meir-Keeler type conditions do not force the mapping to be continuous at the fixed point. Thus we provide one more answer to the open question posed in [18] "whether there exists a contractive definition which is strong enough to generate a fixed point but which does not force the map to be continuous at the fixed point".

1. Introduction

A fixed point theorem is one which ensures the existence of a fixed point of a mapping under suitable assumptions both on the space and the mapping. Apart from ensuring the existence of a fixed point, it often becomes necessary to prove the uniqueness of the fixed point. Besides, from a computational point of view, a constructive algorithm for finding the fixed point is desirable. Often such algorithms involve iterates of the given mapping.

The questions about the existence, uniqueness and approximation of a fixed point provide three significant distinct features of a general fixed point theorem. Most of the fixed point theorems for contractive mappings answer all the three questions of existence, uniqueness and constructive algorithm convincingly [20].

In all that follows T is a self-mapping on metric space (X, d). In [5] Jachymski listed some Meir-Keeler type conditions and established relations between them. Further he gave some new Meir-Keeler type conditions ensuring a convergence of the successive approximations. For $i = 1, 2, 3, 4, 5$;

$[A_i]$ for a given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that, for any $x, y \in X$,
$$\epsilon \leq m_i(x, y) < \epsilon + \delta \text{ implies } d(Tx, Ty) < \epsilon;$$

$[B_i]$ for a given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that, for any $x, y \in X$,
$$\epsilon < m_i(x, y) < \epsilon + \delta \text{ implies } d(Tx, Ty) \leq \epsilon;$$

$[C_i]$ for any $x, y \in X$ with $m_i(x, y) > 0$, $d(Tx, Ty) < m_i(x, y)$, where

2010 Mathematics Subject Classification: Primary: 47H09, 54E50; Secondary: 47H10, 54E40

Keywords and Phrases: Fixed point, $(\epsilon - \delta)$ contractions, power contraction.

$$m_1(x,y) = d(x,y), \qquad m_2(x,y) = max\{d(x,Tx), d(y,Ty)\},$$
$$m_3(x,y) = max\{d(x,y), d(x,Tx), d(y,Ty)\},$$
$$m_4(x,y) = max\{d(x,y), d(x,Tx), d(y,Ty), [d(x,Ty) + d(y,Tx)]/2\},$$
$$m_5(x,y) = max\{d(x,y), d(x,Tx), d(y,Ty), d(x,Ty), d(y,Tx)\}.$$

Condition A_1 is studied by Meir-Keeler [11] and B_1 has been considered by Matkowski [10]. By considering the common features of various contractive definitions several authors introduced some new contractive definitions which, by using standard types of assumptions and arguments, yielded new fixed point theorems (for various contractive definitions of Meir-Keeler type one may see [5, 9, 12, 14]).

It is well-known that $A_1 \Longrightarrow A_3 \Longrightarrow A_4 \Longrightarrow A_5$ and $A_i \Longrightarrow (B_i \wedge C_i)$, for $i = 1, 2, 3, 4, 5$ but not conversely [5].

In this paper, we prove fixed point theorems under a more general condition which subsumes several conditions studied by Jachymski [5]. Further, we do not assume any kind of continuity condition on the mapping. It may be observed that an $(\epsilon - \delta)$ contractive condition does not ensure the existence of a fixed point. The following example [15] illustrates this fact.

Example 1.1. Let $X = [0, 2]$ and d be the usual metric on X. Define $T : X \to X$ by $T(x) = (1 + x)/2$ if $x \in [0, 1]$, $T(x) = 0$ if $x \in (1, 2]$.

Then T satisfies condition (i) of Theorem 2.1 (below) with $\delta(\epsilon) = 1$ for $\epsilon \geq 1$ and $\delta(\epsilon) = 1 - \epsilon$ for $\epsilon < 1$ but T is a fixed point free mapping.

Therefore, to ensure the existence of fixed points under condition (i) of Theorem 2.1 (below), some additional condition is necessarily required either on δ or on the mapping. These additional conditions may assume various forms:

(A) δ is assumed lower semicontinuous [6];

(B) δ is assumed nondecreasing [16];

(C) Assuming relatively strong conditions on the continuity of mapping [19];

(D) Assuming corresponding certain ϕ-contractive condition but without additional hypothesis on ϕ and ϵ [13].

In 1988, Rhoades [17] compared 250 contractive definitions and showed that majority of the contractive definitions does not require the mapping to be continuous in the entire domain. However, in all the cases the mapping is continuous at the fixed point. He further demonstrated that the contractive definitions force the mapping to be continuous at the fixed point though continuity was neither assumed nor implied by the contractive definitions. The question whether there exists a contractive definition which is strong enough to generate a fixed point but which does not force the map to be continuous at the fixed point was reiterated by Rhoades in [18] as an existing open problem.

In 1999, Pant [13] proved the following fixed point theorem and obtained the first result that exhibits discontinuity at the fixed point:

Theorem 1.2. *Let T be a self-mapping of a complete metric space (X, d) such that for any $x, y \in X$;*

(i) $d(Tx, Ty) \leq \phi(m_2(x, y))$, where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\phi(t) < t$ for each $t > 0$;

(ii) for a given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that $\epsilon < m_2(x, y) < \epsilon + \delta$ implies $d(Tx, Ty) \leq \epsilon$.

Then T has a unique fixed point, say z. Moreover, T is continuous at z iff $lim_{x \to z} m_2(x, y) = 0$.

Recently, the author and R. P. Pant [1] proved the following theorem wherein they also gave a contractive definition which does not force the map to be continuous at the fixed point.

Theorem 1.3. *Let (X, d) be a complete metric space. Let T be a self-mapping on X such that T^2 is continuous and satisfy the folowing conditions.*

(i). $d(Tx, Ty) \leq \phi(m_4(x, y))$, where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\phi(t) < t$ for each $t > 0$;

(ii). for a given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that $\epsilon < m_4(x, y) < \epsilon + \delta$ implies $d(Tx, Ty) \leq \epsilon$.

Then T has a unique fixed point, say z, and $T^n x \to z$ for each $x \in X$. Moreover, T is discontinuous at z iff $lim_{x \to z} m_4(x, z) \neq 0$.

In this paper, we consider a class of contractive definitions ensuring a convergence of successive approximations but not forcing the mapping to be continuous at the fixed point. It is important to note that contractive definitions considered by us are independent of the contractive definitions employed in above Theorems 1.1 and 1.2. Thus we provide one more answer to the open question posed in[18].

2. Main results

In what follows we use the following notations.

$$M(x, y) = max\{d(x, y), [d(x, Tx) + d(y, Ty)]/2, [d(x, Ty) + d(y, Tx)]/2\};$$

$$N(x, y) = max\{d(x, y), a[d(x, Tx) + d(y, Ty)]/2, b[d(x, Ty) + d(y, Tx)]/2\},$$

where $0 \leq a, b < 1$.

Theorem 2.1. *Let (X, d) be a complete metric space. Let T be a self-mapping on X such that for any $x, y \in X$;*

(i) for a given $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) > 0$ such that $\epsilon < M(x, y) < \epsilon + \delta$ implies $d(Tx, Ty) \leq \epsilon$;

(ii) $d(Tx, Ty) < N(x, y)$, whenever $N(x, y) > 0$.

Then T has a unique fixed point, say z, and $T^n x \to z$ for each $x \in X$. Moreover, T is continuous at z iff $lim_{x\to z} N(x,z) = 0$.

Proof. Existence of a fixed point: Let $x_0 \in X$. If $Tx_0 = x_0$ then we are done, hence assume $Tx_0 \neq x_0$. Define a sequence $\{x_n\}$ in X given by $x_{n+1} = Tx_n$ and put $c_n = d(x_n, x_{n+1})$ for all $n \in \mathbb{N}\bigcup\{0\}$. Let us first prove that for all $n \in \mathbb{N}$ we have

$$c_n < c_{n-1}. \tag{2.1}$$

To see this observe that

$$\begin{aligned} c_n = d(x_n, x_{n+1}) &= d(Tx_{n-1}, Tx_n) \\ &< N(x_{n-1}, x_n) = max\{d(x_{n-1}, x_n), a[d(x_{n-1}, Tx_{n-1}) + d(x_n, Tx_n)]/2, \\ &\qquad b[d(x_{n-1}, Tx_n) + d(x_n, Tx_{n-1})]/2\} \\ &\leq max\{c_{n-1}, a[c_{n-1} + c_n]/2, b[(c_{n-1} + c_n) + 0]/2\} \end{aligned}$$

from which (2.1) easily follows. Being strictly decreasing sequence of positive numbers, obviously c_n tends to a limit $c \geq 0$. We prove $c = 0$. If possible, suppose $c > 0$. Then there exists a positive integer $k \in \mathbb{N}$ such that $n \geq k$ implies

$$c < c_n < c + \delta, \;\; \delta = \delta(c). \tag{2.2}$$

Also, it follows from (i) and $c_n < c_{n-1}$ that $c_n \leq c$ for all $n \geq k$, which contradicts above inequality. Thus we have $c = 0$.

We shall now show that $\{x_n\}$ is a Cauchy sequence. Fix an $\epsilon > 0$. Since $c_n \to 0$, there exists $k \in \mathbb{N}$ such that $c_n < \delta/2$, for $n \geq k$. Without loss of generality, we may assume that $\delta = \delta(\epsilon) < \epsilon$. Following Jachymski [5] we shall use induction to show that, for any $n \in \mathbb{N}$,

$$d(x_k, x_{k+n}) < \epsilon + \delta/2. \tag{2.3}$$

Obviously, (2.3) holds for $n = 1$. Assuming (2.3) is true for some n we shall prove it for $n + 1$. By the triangle inequlaity, we have

$$d(x_k, x_{k+n+1}) \leq d(x_k, x_{k+1}) + d(x_{k+1}, x_{k+n+1}). \tag{2.4}$$

Observe that it suffices to show that

$$d(x_{k+1}, x_{k+n+1}) \leq \epsilon, \tag{2.5}$$

or, in view of (i), to show that $M(x_k, x_{k+n}) \leq \epsilon + \delta$, where

$$\begin{aligned} M(x_k, x_{k+n}) = max\{&d(x_k, x_{k+n}), [d(x_k, Tx_k) + d(x_{k+n}, Tx_{k+n})]/2, \\ &[d(x_k, Tx_{k+n}) + d(x_{k+n}, Tx_k)]/2\}. \end{aligned} \tag{2.6}$$

By the induction hypothesis, we get

$$d(x_k, x_{k+n}) < \epsilon + \delta/2, \quad (1/2)[d(x_k, x_{k+1}) + d(x_{k+n}, x_{k+n+1})] < \delta/2 \tag{2.7}$$

and

$$\begin{aligned} (1/2)[d(x_k, x_{k+n+1}) + d(x_{k+1}, x_{k+n})] \leq (1/2)&[d(x_k, x_{k+n}) + d(x_{k+n}, x_{k+n+1}) \\ &+ d(x_k, x_{k+1}) + d(x_k, x_{k+n})] < \epsilon + \delta. \end{aligned}$$

Thus $M(x_k, x_{k+n}) < \epsilon + \delta$. This completes the induction, proves (2.3) and shows that $\{x_n\}$ is a Cauchy sequence. Since X is complete, there exists a point $z \in X$ such that $x_n \to z$ as $n \to \infty$. Also $Tx_n \to z$. We claim that $Tz = z$. In fact, in view of (ii) we get

$$d(Tz, Tx_n)$$
$$< max\{d(z, x_n), a[d(z, Tz) + d(x_n, Tx_n)]/2, b[d(z, Tx_n) + d(x_n, Tz)]/2\},$$

and letting $n \to \infty$ this yields, $d(Tz, z) \leq max\{ad(Tz, z), bd(Tz, z)\}$. Hence $Tz = z$, and thus z is a fixed point of T.

Uniqueness of the fixed point: Uniqueness of the fixed point follows easily.

Continuity criteria: Let T be continuous at the fixed point z and $x_n \to z$. Then $Tx_n \to Tz = z$. Hence

$$\lim_n N(x_n, z)$$
$$= \lim_n max\{d(x_n, z), a[d(x_n, Tx_n) + d(z, Tz)]/2, b[d(x_n, Tz) + d(z, Tx_n)]/2\} = 0.$$

On the other hand, if $\lim_{x_n \to z} N(x_n, z) = 0$, then $d(x_n, Tx_n) \to 0$ as $x_n \to z$. This implies that $Tx_n \to z = Tz$, i.e., T is continuous at z. This completes the proof of the theorem. □

Remark 2.2. The last part of Theorems 2.1 can alternatively be stated as: T is discontinuous at z iff $lim_{x \to z} N(x, z) \neq 0$.

The following example illustrates the above theorem:

Example 2.3. Let $X = [0, 2]$ and d be the usual metric on X. Define $T : X \to X$ by $T(x) = 1$ if $x \in [0, 1]$, $T(x) = 0$ if $x \in (1, 2]$. Then T satisfies the conditions of Theorem 2.1 and has a unique fixed point $x = 1$ at which T is discontinuous. The mapping T satisfies condition (i) with $\delta(\epsilon) = 1$ for $\epsilon \geq 1$ and $\delta(\epsilon) = 1 - \epsilon$ for $\epsilon < 1$. It can also be easily seen that $lim_{x \to 1} N(x, 1) \neq 0$ and T is discontinuous at the fixed point $x = 1$.

The following theorem shows that power contraction allows the possibility of discontinuity at the fixed point. We use the following notations.

$$M'(x, y) = max\{d(x, y), [d(x, T^m x) + d(y, T^m y)]/2, [d(x, T^m y) + d(y, T^m x)]/2\},$$
$$N'(x, y) = max\{d(x, y), a[d(x, T^m x) + d(y, T^m y)]/2, b[d(x, T^m y) + d(y, T^m x)]/2\},$$

where $0 \leq a, b < 1$; $m \in \mathbb{N}$.

Theorem 2.4. *Let (X, d) be a complete metric space. Let T be a self-mapping on X such that for any $x, y \in X$;*

(i) for a given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that $\epsilon < M'(x, y) < \epsilon + \delta$ implies $d(T^m x, T^m y) \leq \epsilon$;

(ii) $d(T^m x, T^m y) < N'(x, y)$, whenever $N'(x, y) > 0$.

Then T has a unique fixed point, say z, and $T^n x \to z$ for each $x \in X$.

Proof. By Theorem 2.1, T^m has a unique fixed point, say $z, \in X$; so that $T^m(z) = z$. Then $T(z) = T(T^m(z)) = T^m(T(z))$ and hence $T(z)$ is a fixed point of T^m. Since T^m has unique fixed point we have $Tz = z$.
If y is another fixed point of T then $Ty = y$ and hence $T^m(y) = y$. But then by the uniqueness of the fixed point of T^m, we have $z = y$. It follows that z is the fixed point of T. □

Remark 2.5. The above theorems unify and improve the results due to Jachymski [5], Kuczma et al. [8], Maiti and Pal [9], Matkowski [10] and Pant [13].

Acknowledgment. The author is thankful to the referee for his valuable suggestions to improve the presentation of the paper.

References

[1] Bisht, Ravindra K. and Pant, R. P., A remark on discontinuity at fixed point, *J. Math. Anal. Appl.*, **445** (2017), 1239–1241.
[2] Boyd, D. W. and Wong, J. S., On nonlinear contractions, *Proc. Amer. Math. Soc.*, **20** (1969), 458–464.
[3] Ćirić, Lj. B., On contraction type mappings, *Math. Balkanica*, **1** (1971), 52–57.
[4] Ćirić, Lj. B., A generalization of Banach's contraction principle, *Proc. Amer. Math. Soc.*, **45** (2) (1974), 267–273.
[5] Jachymski, J., Equivalent conditions and Meir-Keeler type theorems, *J. Math. Anal. Appl.*, **194** (1995), 293–303.
[6] Jungck, G., Moon, K. B., Park, S. and Rhoades, B. E., On generalizations of the Meir-Keeler type contraction maps : Corrections, *J. Math. Anal. Appl.*, **180** (1993), 221–222.
[7] Kannan, R., Some results on fixed points-II, *Amer. Math. Mon.*, **76** (1969), 405–408.
[8] Kuczma, M., Choczewski, B. and Ger, R., *Iterative Functional Equations*, Encyclopedia of Mathematics and its Applications, **32**, Cambridge Univ. Press, Cambridge, UK, 1990.
[9] Maiti, M. and Pal, T. K., Generalizations of two fixed point theorems, *Bull. Cal. Math. Soc.*, **70** (1978), 59–61.
[10] J. Matkowski, Integrable solutions of functional equations, *Diss. Math.*, **127** (1975), 1–68.
[11] Meir, A. and Keeler, E., A theorem on contraction mappings, *J. Math. Anal. Appl.*, **28** (1969), 326–329.
[12] Nabieia, Mona and Ezzati, Taha, A novel fixed point theorem for the k-Meir-Keeler function, *Quaestiones Mathematicae*, **39** (2) (2016), 245–250.
[13] Pant, R. P., Discontinuity and fixed points, *J. Math. Anal. Appl.*, **240** (1999), 284–289.
[14] Pant, R. P., A comparison of contractive definitions, *J. Indian Math. Soc.*, **72** (2005), 241–249
[15] Pant, R. P., Discontinuity at fixed points, *Ganita*, **51** (2000), 135–142.
[16] Pant, R. P., Common fixed points of two pairs of commuting mappings, *Indian J. Pure Appl. Math.*, **17** (1986) 187–192.
[17] Rhoades, B. E., A comparison of various definitions of contractive mappings, *Trans. Amer. Math. Soc.*, **226** (1977), 257–290.
[18] Rhoades, B. E., Contractive definitions and continuity, *Contemporary Mathematics*, **72** (1988), 233–245.
[19] Park, S. and Bae, J. S., Extension of a fixed point theorem of Meir-Keeler, *Ark. Math.*, **19** (1991), 223–228.

[20] Subrahmanyam, P. V., Some aspects of fixed point theory, *Resonance*, **5** (issue 5) (2000), 36–44.

Ravindra K. Bisht
Department of Mathematics
National Defence Academy, Pune, India
E-mail: *ravindra.bisht@yahoo.com*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 39-50

NORM OR EXCEPTION?

KANNAPPAN SAMPATH AND B. SURY

(Received : 14 - 09 - 2016, Revised : 04 - 01 - 2017)

Abstract. In the study of class groups of real quadratic fields, one encounters norm form equations of the type $x^2 - dy^2 = k$. Apart from the usual approach from algebraic number theory, we discuss also how one uses methods from continued fractions. We demonstrate the methods through a particular example. The continued fraction method does not seem to be well-known apart from the basic theory used for the equations $x^2 - dy^2 = \pm 1$. This article could be useful to graduate students or researchers in number theory.

Introduction

In a first course on algebraic number theory, a typical homework problem may ask the student to determine the class group of a quadratic field. One is expected to determine the Minkowski constant and analyse the behaviour of the small primes not exceeding it. For instance, for $\mathbb{Q}(\sqrt{223})$, the primes up to 13 need to be considered. In this particular example, it is quite easily seen that 3 splits into the two prime ideals $P := (3, 1 + \sqrt{223})$ and $P' := (3, 1 - \sqrt{223})$, and that the ideal classes of prime ideals lying above the other primes are either trivial or, are equivalent to one of the primes P, P' dividing 3. Further, it is easy to show that P^3 is principal. The complete determination of the class group then boils down to checking whether there are elements of norm ± 3 in the ring of integers. Typically, when a solution is not easily visible, some congruence conditions rule out the existence of solutions. In the above example too, it is easy to see that

$$a^2 - 223b^2 = 3$$

has no integral solutions by looking at the equation modulo 4. However, it does not seem equally easy to prove that

$$a^2 - 223b^2 = -3$$

has no integral solutions. In this note, we look at this example and discuss two proofs. Both proofs have the potential to be applied more generally.

We discuss the first proof just for this example but, while giving the second proof, we take the opportunity to analyze the power of continued fractions. The employment of the continued fraction expansion of $\sqrt{d}$ (d positive non-square) to

2010 Mathematics Subject Classification: 11 D 57, 11 R 11
Keywords and Phrases: Norm equations; Continued fraction; Real quadratic fields.

determine the solutions of $x^2 - dy^2 = \pm 1$ is well-known. We point out that this amounts to looking for the units (equivalently, elements of norm ± 1, the norm being taken in quadratic field $\mathbb{Q}(\sqrt{d})$) in the ring $\mathbb{Z}[\sqrt{d}]$. But more generally, if we let ξ be an irrational real number satisfying a quadratic equation with coefficients in $\mathbb{Q}$ so that $\mathbb{Q}(\xi)$ is a real quadratic field (the so-called *real quadratic irrationalities*), then, the continued fraction of ξ can often be used to study the existence (or the lack thereof) of elements of $\mathbb{Z}[\xi]$ of "small" norm (as an element of $\mathbb{Q}(\xi)$) - see Theorem 2.10. It appears to us that these results are due to Lagrange and have been laid out carefully in Serret's seminal work on "higher" algebra [7, Chapitre II, Section I, §35, p.80]. In fact, at the time of writing this article, the only other text where we could find a discussion of this nature is the book [1] by Chrystal; here, one finds a thorough discussion of the less general Diophantine equation $x^2 - dy^2 = m$ for $m \neq \pm 1$. Chrystal alludes to the general case in Exercises XXXII, (52.); however, the formulation as it stands is incomplete and seems a little misleading. The underlying principle is that the elements of "*small*" norm, if there are any, must come from convergents of the continued fraction of ξ. The key point that is only implicit even in the references mentioned above, is the estimation of the number of convergents that we must compute before we can refute the existence of an element of a given "small" norm. Our exposition aims to make this very transparent (v. Lemma 2.13) while remaining short and self-contained.

The very general phenomenon outlined above does not seem so well-known; at any rate, this has not been expounded in most standard texts on algebraic number theory. After this article was written, we looked through recent texts and discovered a new book by Trifković ([8]) on algebraic number theory which also coincidentally discusses the very example above and we recommend this text to the reader interested in a more detailed study of the subject.

1. Class group of $\mathbb{Q}(\sqrt{223})$

Let us start with a more detailed discussion of the computation of the class group of the real quadratic field $k := \mathbb{Q}(\sqrt{223})$.

As $223 \equiv 3 \pmod 4$, we have $O_k = \mathbb{Z}[\sqrt{223}]$ and its discriminant equals 4×223. The Minkowski constant of k is $\sqrt{223}$. One looks at the splitting of the primes $2, 3, 5, 7, 11, 13$ in O_k. The prime 2 (as well as the prime 223) ramifies as it divides the discriminant; in fact,

$$2\mathbb{Z}[\sqrt{223}] = (2, 1 + \sqrt{223})^2$$

since the minimal polynomial $f = X^2 - 223$ becomes $X^2 - 1 = (X+1)^2 \bmod 2$. Also, f remains irreducible (equivalently, has no root) modulo $5, 7$ or 13; so, these primes remain inert. Further, modulo 3, we have $X^2 - 223 = X^2 - 1 = (X+1)(X-1)$ which shows

$$3\mathbb{Z}[\sqrt{223}] = (3, 1 + \sqrt{223})(3, -1 + \sqrt{223}).$$

Modulo 11, $X^2 - 223 = X^2 - 3 = (X + 5)(X - 5)$ so that

$$11\mathbb{Z}[\sqrt{223}] = (11, 5 + \sqrt{223})(11, -5 + \sqrt{223}).$$

Now, if $(2, 1 + \sqrt{223})$ is principal, it would be generated by an element of norm ± 2 because its square is $2\mathbb{Z}[\sqrt{223}]$ which has norm 4. It is easy locate an element of norm 2; viz., $15 + \sqrt{223}$. It is then straightforward to check that

$$(2, 1 + \sqrt{223}) = (15 + \sqrt{223}).$$

Further, we can easily locate an element of norm $3 \times 11 = 33$, viz., $16 + \sqrt{223}$. It is once again a straightforward task to check that

$$(3, 1 + \sqrt{223})(11, 5 + \sqrt{223}) = (16 + \sqrt{223}).$$

Indeed, $16 + \sqrt{223} = (1 + \sqrt{223})(2(5 + \sqrt{223}) - 11) - (3)(11)(13)$.

Now, let us find the order of $P = (3, 1 + \sqrt{223})$.

As P has norm 3, we look for an element of norm ± 9 to ascertain whether P^2 is principal. Inspection of small values does not produce a solution. The next step is to look for an element of norm $\pm 3^3$ which would possibly generate P^3. Sure enough, the easily located element $14 + \sqrt{223}$ of norm -27 satisfies

$$P^3 = (14 + \sqrt{223}).$$

What is left is to ascertain whether P itself is principal; if it is not, the class group is the cyclic group of order 3 generated by the class of P. We shall prove that P is not principal. If P is principal, say $P = (a + b\sqrt{223})$, then

$$a^2 - 223b^2 = \pm 3.$$

Clearly, there is no solution with the positive sign on the right since the left hand side is $0, 1$ or 2 modulo 4. However, the proof of the fact that the equation has no solution with the negative sign on the right hand side, is not straightforward. We give two proofs. The first one is due to Peter Stevenhagen (personal correspondence); our main aim is to discuss the second proof at length. In both proofs, we do not need to separate the cases 3 and -3.

The fundamental unit of $\mathbb{Q}(\sqrt{223})$ is $\eta = 224 + 15\sqrt{223}$. This can be found simply by hand but while discussing the second proof, we give the details.

1.1. **First proof.** Let, if possible, $(3, 1 + \sqrt{223}) = (x)$ for some $x \in \mathbb{Z}[\sqrt{223}]$. As $P^3 = (x^3) = (14 - sqrt223)$, we have

$$14 - \sqrt{223} = ux^3$$

for some unit u. Now, the fundamental unit

$$\eta = 224 + 15\sqrt{223} \equiv -1 \ mod\ 5\ \mathbb{Z}[\sqrt{223}].$$

In particular, η becomes a cube in the finite field $F := \mathbb{Z}[\sqrt{223}]/5\ \mathbb{Z}[\sqrt{223}]$ which has 5^2 elements. In particular, every unit (being a power of η) is a cube in this field. Hence, the image of $14 - \sqrt{223}$ is a cube. An element in the cyclic group F^* of order $5^2 - 1 = 24$ is a cube if, and only if, its 8-th power is 1. Let us compute the image of $(14 - \sqrt{223})^8$ in the field F:

$$(14-\sqrt{223})^8 = (-1-\sqrt{223})^8 \qquad = (1+223+2\sqrt{223})^4 = (-1+2\sqrt{223})^4$$
$$= (1+892-4\sqrt{223})^2 = (-2-4\sqrt{223})^2 \qquad = 4(1+2\sqrt{223})^2$$
$$= 4(1+892+4\sqrt{223}) \equiv 4(-2-\sqrt{223}) \qquad = 2+\sqrt{223}.$$

But, $\sqrt{223}+2$ is not 1 in the cyclic group $(\mathbb{Z}[\sqrt{223}]/5\ \mathbb{Z}[\sqrt{223}])^*$. Otherwise, $223 = 1 \bmod 5\mathbb{Z}[\sqrt{223}]$ which is absurd as 222 is co-prime to 5. This completes the proof that P cannot be principal S

We deduce:

The ideal class group of $\mathbb{Q}(\sqrt{223})$ is cyclic, of order 3, generated by the class of $(3, 1+\sqrt{223})$. Further, the equation $a^2 - 223b^2 = -3$ has no integer solutions.

1.2. **A non-square.** Before embarking on the discussion on continued fractions required for the second proof, we make an interesting remark.

Rewriting the above equation as $a^2 + 3 = 223b^2$, one may argue within the field $\mathbb{Q}(\sqrt{-3})$ generated by the cube roots of unity. Its ring of integers is $\mathbb{Z}[\omega]$, a unique factorization domain where $\omega = \frac{-1+\sqrt{-3}}{2}$. If a is odd and b is even then $a^2 + 3 = 223b^2$ becomes equivalent to an equation

$$A^2 - A + 1 = 223B^2.$$

That is, $(A+\omega)(A+\omega^2) = 223B^2$.

Writing the element 223 is a product of two irreducible elements:

$$223 = (17+11\omega)(17+11\omega^2),$$

one has $A+\omega = (17+11\omega)(u+v\omega)$ or $A+\omega = (17+11\omega^2)(u+v\omega)$. Comparing the imaginary parts, one may deduce that there exists a number of the form $223s^2 + 79s + 7$ which is a perfect square. Therefore, we deduce:

Observation. *For an integer s, the number $223s^2 + 79s + 7$ cannot be a perfect square.*

In fact, write $223s^2 + 79s + 7 = t^2$. Then,

$$(446s+79)^2 + 3 = 223(2t)^2.$$

This contradicts the fact that $a^2 - 223b^2 = -3$ has no solution.

It will be interesting to give a direct proof of the above observation.

2. Continued fractions and Small norms

2.1. **Continued fractions.** We recall the basic terminology of simple continued fractions relevant to our application to real quadratic fields. For a more elaborate discussion, we recommend the classical works [7, 2, 1, 3] and the recent text [8].

A *simple continued fraction* (S.C.F.) is an expression of the form

$$\lim_{n\to\infty}\left(a_0 + \frac{1}{a_1+}\frac{1}{a_2+}\cdots\frac{1}{a_n}\right)$$

where $a_0 \in \mathbb{Z}$ and $\{a_n\}_{n>0}$ is a sequence of positive integers. In other words, an S.C.F. is the limit of the sequence whose nth term is

$$\ell_n := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n}}}}. \tag{1}$$

One also writes the limit above as $a_0 + \frac{1}{a_1+}\frac{1}{a_2+}\frac{1}{a_3+}\cdots$ or as $[a_0; a_1, a_2, \cdots]$ symbolically. Truncating this process at finite stages, the successive quotients

$$\frac{p_0}{q_0} := \frac{a_0}{1}, \frac{p_1}{q_1} := a_0 + \frac{1}{a_1} = \frac{a_0a_1+1}{a_1}, \cdots$$

are called the *convergents* to the continued fraction. It can be proven by a straightforward induction that

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}.$$

Immediately, a consideration of determinants gives:

$$p_nq_{n-1} - p_{n-1}q_n = (-1)^{n-1}$$

$$p_nq_{n-2} - p_{n-2}q_n = (-1)^n a_n$$

The most important fact about continued fractions that we need is the following observation due to Legendre [3]:

THEOREM 2.1. *If α is a real number which is irrational, and satisfies*

$$\left|\alpha - \frac{r}{s}\right| < \frac{1}{2s^2}$$

where $s > 0$, then r/s is a convergent to the continued fraction of α.

In algorithm 2.4 below, we study the algorithm for the S.C.F. for quadratic irrational to work out the small norms in the ring of integers of a real quadratic field. As a precursor to the general discussion, let us recall the classical facts about S.C.F. for $\sqrt{N}$ for positive square-free integers N.

2.2. **The S.C.F. of $\sqrt{N}$.** Let N be a square-free positive integer. The S.C.F. $\sqrt{N} = b_0 + \frac{1}{b_1+}\frac{1}{b_2+}\cdots$ is determined as

$$b_0 = a_1 = [\sqrt{N}], \qquad r_1 = N - a_1^2$$

$$b_1 = \left[\frac{\sqrt{N} + a_1}{r_1}\right], \qquad \text{etc.}$$

More generally, we have

$$b_n = \left[\frac{\sqrt{N} + a_n}{r_n}\right]$$

where $a_n = b_{n-1}r_{n-1} - a_{n-1}$ and $r_{n-1}r_n = N - a_n^2$. One shows easily that $a_{n+1}, r_{n+1} > 0$. Further, *if we know* that some r_k (say r_{n+1}) equals 1, then

$$(-1)^{n-1} = p_n^2 - Nq_n^2.$$

This is indeed the case (see [2]) and the key facts are summarized as:

LEMMA 2.2.

(i) *The b_n's recur.*
(ii) *The S.C.F. of $\sqrt{N}$ looks like $[b_0; \overline{b_1, b_2, \cdots, b_{n-1}, 2b_0}]$.*
(iii) *The penultimate convergent p_{n-1}/q_{n-1} before the recurring period gives a solution of $x^2 - Ny^2 = (-1)^n$.*

Hence, the penultimate convergent of the S.C.F. of $\sqrt{N}$ gives a solution of $x^2 - Ny^2 = \pm 1$ where the sign is positive or negative according as to whether the period is even or odd.

2.3. **The S.C.F. of real quadratic irrationalities.** Let us discuss how the above facts carry over from $\sqrt{N}$ to any element of a real quadratic field, which we christen a real quadratic irrationality.

DEFINITION 2.3. A number $\xi \in \mathbb{C} \setminus \mathbb{Q}$ is said to be a *real quadratic irrationality* if it satisfies an equation of the form $\xi^2 + p\xi + q = 0$ for uniquely determined rational numbers p and q satisfying $p^2 - 4q^2 > 0$.

Let ξ' denote the Galois conjugate of ξ (equal to $-p - \xi$ under the above notation). We have the following algorithm to produce the S.C.F. of a general real quadratic irrationality.

ALGORITHM 2.4. Let $\xi = \frac{P_0 + \sqrt{D}}{Q_0}$ be a real quadratic irrationality where D, P_0, Q_0 are positive integers. We assume, without loss of generality, that Q_0 divides $P_0^2 - D$ (otherwise, we may multiply P_0, Q_0 by Q_0 and D by Q_0^2).

Then, define the sequences $\{a_n\}_{n \geqslant 0}$, $\{\xi_n\}_{n \geqslant 0}$, $\{P_n\}_{n \geqslant 1}$ and $\{Q_n\}_{n \geqslant 1}$ of numbers by the following rule:

$$a_0 = [\xi_0] \quad \text{and} \quad \xi_1 = \frac{1}{\xi_0 - a_0} = \frac{P_1 + \sqrt{D}}{Q_1}$$

$$a_1 = [\xi_1] \quad \text{and} \quad \xi_2 = \frac{1}{\xi_1 - a_1} = \frac{P_2 + \sqrt{D}}{Q_2}$$

In general,

$$a_{m-1} = [\xi_{m-1}] \text{ and } \xi_m = \frac{1}{\xi_{m-1} - a_{m-1}} = \frac{P_m + \sqrt{D}}{Q_m}.$$

Then, $\xi = [a_0; a_1, a_2, \cdots]$ is the S.C.F. of ξ.

The following observations on this algorithm are the most useful ones:

LEMMA 2.5. *Let $\{p_n/q_n\}$ be the sequence of convergents of a quadratic irrational ξ. With notations as above, all P_i, Q_i are integers and, the following equations hold:*

$$\xi = (a_0; a_1, \ldots, a_{n-1}, \xi_n), \quad n \geqslant 1; \tag{2}$$

$$P_{n+1} = a_n Q_n - P_n, \quad n \geqslant 0 \tag{3}$$

$$P_{n+1}^2 + Q_n Q_{n+1} = D, \quad n \geqslant 0 \tag{4}$$

$$Q_{n+1} = Q_{n-1} + Q_n(P_n - P_{n+1}) \tag{5}$$

$$(-1)^n Q_n/Q_0 = (p_{n-1} - \xi q_{n-1})(p_{n-1} - \xi' q_{n-1}) \tag{6}$$

It is the last equation that is the protagonist of this story: it tells us that $p_{n-1} - q_{n-1}\xi$ solves the norm-form equation $N(\mathfrak{z}) = (-1)^n Q_n/Q_0$ where $N(\cdot)$ stands for the norm on the quadratic field $\mathbb{Q}(\xi)$. We shall soon discover that with appropriate bound on H, a primitive solution (if it exists at all) to the norm form equation $N(\mathfrak{z}) = H$ with $\mathfrak{z} \in \mathbb{Z}[\xi]$ must arise from convergents (see Theorem 2.10).

Proof. We prove the equalities asserted in the lemma, from which it follows inductively that the P_i's and the Q_i's are integers. The first equality holds by definition. The next two equalities are consequences of the identity:

$$\frac{P_{n+1} + \sqrt{D}}{Q_{n+1}} = \frac{1}{\frac{P_n + \sqrt{D}}{Q_n} - a_n}.$$

Indeed, multiply out and equate rational and irrational parts.

To prove the penultimate equality, note that

$$Q_n Q_{n+1} = D - P_{n+1}^2 = D - (a_n Q_n - P_n)^2 = P_n^2 + Q_{n-1}Q_n - (a_n Q_n - P_n)^2$$

which gives $Q_{n+1} = Q_{n-1} + a_n(P_n - P_{n+1})$.

Finally, we prove that the last equality follows from certain properties of convergents as follows. We know that the complete quotients ξ_n give us

$$\frac{P_0 + \sqrt{D}}{Q_0} = \frac{p_{n-1}\xi_n + p_{n-2}}{q_{n-1}\xi_n + q_{n-2}}.$$

Using the expression $\xi_n = \frac{P_n + \sqrt{D}}{Q_n}$, we have

$$\frac{P_0 + \sqrt{D}}{Q_0} = \frac{p_{n-1}P_n + p_{n-2}Q_n + p_{n-1}\sqrt{D}}{q_{n-1}P_n + q_{n-2}Q_n + q_{n-1}\sqrt{D}}.$$

A comparison of rational and irrational parts gives us:

$$q_{n-1}P_n + q_{n-2}Q_n = Q_0 p_{n-1} - P_0 q_{n-1};$$

$$p_{n-1}P_n + p_{n-2}Q_n = P_0 p_{n-1} + \left(\frac{D - P_0^2}{Q_0}\right) q_{n-1}.$$

Using $p_{n-1}q_{n-2} - p_{n-2}q_{n-1} = (-1)^n$, we obtain

$$(-1)^n P_n = P_0(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + \left(\frac{D - P_0^2}{Q_0}\right) q_{n-1}q_{n-2} - Q_0 p_{n-1}p_{n-2};$$

$$(-1)^n Q_n = -2p_{n-1}q_{n-1}P_0 + \left(\frac{P_0^2 - D}{Q_0}\right) q_{n-1}^2 + Q_0 p_{n-1}^2.$$

As $\xi + \xi' = \frac{2P_0}{Q_0}$, $\xi\xi' = \frac{P_0^2 - D}{Q_0^2}$, we obtain $N(p_{n-1} - \xi q_{n-1}) = (-1)^n Q_n/Q_0$. □

The key periodicity feature of this algorithm is given in the following theorem (see Chrystal, Chapter XXXIII, §4):

THEOREM 2.6 (Euler-Lagrange). *The sequence $(a_n)_{n\geqslant 0}$ in the continued fraction of the quadratic irrationality ξ is eventually periodic; that is, there are positive integers t and h such that $a_{n+h} = a_n$ for all $n \geqslant t$ and h is the least positive integer such that $a_{n+h} = a_n$ for all sufficiently large n and t is the least positive integer such that $a_{t+h} = a_t$.*

It is evident that the integers h and t with properties in Theorem 2.6 are uniquely determined. The word $a_0 \dots a_{t-1}$ is called the *preperiod* and, the number t is called the *length of the preperiod* of the sequence $(a_n)_n$. The number h is called the *length of the period* of $(a_n)_n$ and is denoted by $\ell(\xi)$. A convenient shorthand for this situation is the following notation:

$$(a_n)_n := (a_0, \dots, a_{t-1}, \overline{a_t, \dots, a_{t+h-1}}).$$

For our purposes, it is necessary to be able to compute the length of the preperiod of quadratic irrationals. The proof we present is the one of the key parts of the proof of the periodicity theorem above and consequently the theorem itself is known but is seldom formulated this way:

THEOREM 2.7. *The following are equal for a quadratic irrational ξ:*

(1) The length of the preperiod of ξ.
(2) The least index t such that $\xi_t > 1$ and $-1 < \xi_t' < 0$.
(3) The least index t such that $0 < P_t < \sqrt{D}$ and $0 < Q_t < P_t + \sqrt{D}$.

Proof. It is clear that the numbers defined in (2) and (3) are equal. Let k be the preperiod of ξ. Then, it follows from the uniqueness theorem that $\xi_k = (\overline{a_k, \dots, a_{k+h-1}})$. Therefore, we have

$$\xi_k = (a_k, \dots, a_{k+h-1}, \xi_k). \tag{7}$$

Notice that $\xi_k > a_k = a_{k+h} \geqslant 1$. Moreover, (7) gives us a quadratic equation satisfied by ξ_k (and hence ξ_k'):

$$F(\xi_k) = q_{k+h-1}\xi_k^2 + (q_{k+h-2} - p_{k+h-1})\xi_k - p_{k+h-2} = 0.$$

Note now that $F(0) = -p_{k+h-2} < 0$ and $F(-1) = q_{k+h-1} + p_{k+h-1} - q_{k+h-2} - p_{k+h-2} > 0$ since the numerator and denominator of a convergent form a (strictly) increasing sequence. Therefore F has a root in $(-1, 0)$ which proves that $-1 < -\xi_k' < 0$. Thus, we have that $k \geqslant t$.
If $k > t$, we shall conclude that $a_{k-1} = a_{k+h-1}$ which will contradict the definition of preperiod. We use the following lemma which easily follows by induction.

LEMMA 2.8. *If ξ_t satisfies the conditions $\xi_t > 1$ and $-1 < \xi_t' < 0$, then, so does ξ_n for all $n \geqslant t$.*

The theorem follows from above as $a_n = [-1/\xi'_{n+1}]$ for all $n \geqslant t$ and, in particular, we have that $a_{k-1} = a_{k+h-1}$ taking $n = k-1$. □

Now we have the following corollary.

COROLLARY 2.9. *Let $D > 0$ be a square-free integer. The length of the preperiod of $\sqrt{D}$ and that of $\frac{-1+\sqrt{D}}{2}$ are both 1.*

Proof. Let $a_0 = [\xi]$ where ξ is one of the quadratic irrationalities in the statement. Then

$$P_1 = \begin{cases} a_0, & \text{if } \xi = \sqrt{D} \\ 2a_0 + 1, & \text{if } \xi = \frac{-1+\sqrt{D}}{2} \end{cases} \quad \text{and} \quad Q_1 = \begin{cases} D - a_0^2, & \text{if } \xi = \sqrt{D} \\ \frac{D-(2a_0+1)^2}{2}, & \text{if } \xi = \frac{-1+\sqrt{D}}{2} \end{cases}$$

It is now easily verified that the least index for which inequalities in (3) of the above theorem hold in each case is $t = 1$. □

2.4. **Small norms.** The existence of elements of norm H (where H is an integer) in $\mathbb{Z}[\xi]$ is equivalent to the existence of an integral solution to the equation

$$(X + \xi Y)(X + \xi' Y) = H. \tag{8}$$

Note first that if $x, y \in \mathbf{Z}$ are integers satisfying $(x + \xi y)(x + \xi' y) = H$, we may replace H by $H/(x, y)^2$, and obtain a new solution $X = x/(x, y), Y = y/(x, y)$ to (8), which are relatively prime. An integral solution to (8) with $(x, y) = 1$ is said to be primitive.

The key result which is relevant to our original problem is the following observation that primitive solutions come from convergents of ξ when ξ generates the ring of integers in a real quadratic field K:

THEOREM 2.10. *Let ξ be a real, quadratic irrational written in the form*

$$\xi = \frac{P + \sqrt{D}}{Q}.$$

Suppose that $\xi > 0 > \xi'$ (equivalently $Q > 0$ and $-\sqrt{D} < P < \sqrt{D}$). If x, y are relatively prime integers such that $(x + \xi y)(x + \xi' y) = H$ with $|H| < \frac{\xi - \xi'}{2} = \frac{\sqrt{D}}{Q}$, then x/y is a convergent to $-\xi'$.

Moreover, we need to look at only the first lcm $(2, \ell(\xi)) + 1$ convergents.

Proof. Suppose first that $H > 0$. Thus $x + \xi' y > 0$ and so $x + \xi y > (\xi - \xi')y$. So, we have

$$0 < x + \xi' y < \frac{H}{(\xi - \xi')y} < \frac{1}{2y}$$

from which it follows that x/y is a convergent to $-\xi'$.

Now, if $H < 0$, we then have that $x + y\xi' < 0$ so that we have

$$0 < y + \frac{x}{\xi'} = \frac{H}{\xi'(x + y\xi)} = \frac{-H}{-\xi'(x + y\xi)} < \frac{\xi - \xi'}{2(-\xi')(x + y\xi)} < \frac{1}{2x}$$

where the last inequality amounts to checking that

$$x(\xi - \xi') < (-\xi')(x + y\xi)$$

which holds since $\xi > 0$ and $x + y\xi' < 0$. This shows that y/x is a convergent of $-\xi'^{-1}$. But we note that if x has the S.C.F. $[a_0; a_1, a_2, \dots]$, then, x^{-1} has the S.C.F. $[0; a_0, a_1, \dots]$ if $a_0 > 0$ and $[a_1; a_2 \dots]$ if $a_0 = 0$. Therefore, every non-zero convergent of x is also a convergent of x^{-1}. Thus, we have that x/y is a convergent of $-\xi'$. We defer the proof of the last assertion to Lemma 2.13. □

The above theorem immediately yields the following corollary.

COROLLARY 2.11. *Let D be a square-free positive integer, let K denote the quadratic field $\mathbb{Q}(\sqrt{D})$, let d_K be its discriminant:*

$$d_K = \begin{cases} 4D, & \text{if } D \equiv 2, 3 \bmod 4 \\ D, & \text{if } D \equiv 1 \bmod 4 \end{cases} \tag{9}$$

Let ω_D denote the quadratic irrationality

$$\omega_D = \begin{cases} \sqrt{D}, & D \equiv 2, 3 \bmod 4 \\ \frac{-1+\sqrt{D}}{2}, & D \equiv 1 \bmod 4 \end{cases} \tag{10}$$

so that $\mathcal{O}_K = \mathbb{Z}[\omega_D]$. Suppose that $|H| < \frac{\sqrt{d_K}}{2}$. The primitive elements of norm H in K come from convergents of ω_D.

Here, it is important that we view $\mathcal{O}_K$ as $\mathbb{Z}$-module with respect to the basis $\{1, -\omega'_D\}$ as is customary.

REMARK 2.12. Note that the bound on H is reminiscent of Gauss's bound; that is, in any ideal class in a quadratic field K, there is an integral ideal whose norm is atmost $\frac{\sqrt{d_K}}{2}$.

To make this principle practical, one needs a bound on the number of convergents one has to compute. This is given in the following lemma.

LEMMA 2.13. *The fundamental primitive solutions of* (8)*, if they exist, are to be found among the first $\ell' + 1$ convergents where $\ell' = lcm(2, \ell(\xi))$.*

Proof. The key ingredient in the proof is Theorem 2.15 which discusses the induced periodicity in the sequence $((-1)^n Q_n)_n$. Let us first reduce this question to the periodicity of (Q_n). This is a consequence of the following simple lemma:

LEMMA 2.14. *Let $(u_n)_{n\geqslant 0}$ be an eventually periodic sequence with preperiod of length t and period h; further suppose that $u_n \neq 0$ for all $n \geqslant t$. Then, the sequence $(v_n := (-1)^n u_n)_{n\geqslant 0}$ is eventually periodic with preperiod of length t and period of length h' where h' is a divisor of $lcm(2, h)$.*
Furthermore, if h is odd, then, $h' = 2h$.

Now, we may summarize the above discussion in the theorem.

THEOREM 2.15. *Let ξ be a quadratic irrational. Let $(a_n)_{n\geqslant 0}$ be its continued fraction expansion. Then*

(i) *The sequence (a_n) is eventually periodic.*
(ii) *The sequence (ξ_n) is eventually periodic.*

(iii) *The sequence (Q_n) is eventually periodic.*

(iv) *The preperiod and period of the above sequences are all equal.*

Proof. (i) is precisely Theorem 2.6. (ii) (and hence (iii) and (iv)) follows by noting that for any integer $n \geqslant 0$, we have $\xi_n = (a_k)_{k \geqslant n}$ by Lemma 2.5. □

These observations now complete the proof of our theorem. □

3. Examples

We illustrate the results of the last section by showing that $\mathbb{Z}[\sqrt{223}]$ has no elements of norm -3.

Example 3.1. It is straightforward to verify that $\sqrt{223} = [14; \overline{1, 13, 1, 28}]$. Here is the full computation: we begin by noting that $14 < \sqrt{223} < 15$ so

$$\begin{aligned}
\sqrt{223} = 14 + \sqrt{223} - 14 &= 14 + \cfrac{1}{\frac{\sqrt{223}+14}{27}} \\
&= 14 + \cfrac{1}{1+\cfrac{1}{\frac{\sqrt{223}+13}{2}}} \\
&= 14 + \cfrac{1}{1+\cfrac{1}{13+\cfrac{1}{\frac{\sqrt{223}+13}{27}}}} \\
&= 14 + \cfrac{1}{1+\cfrac{1}{13+\cfrac{1}{1+\cfrac{1}{\sqrt{223}+14}}}} \\
&= 14 + \cfrac{1}{1+\cfrac{1}{13+\cfrac{1}{1+\cfrac{1}{28+\cfrac{1}{\frac{\sqrt{223}+14}{27}}}}}} = [14; \overline{1, 13, 1, 28}].
\end{aligned}$$

The convergents are easily computed to be

$$\frac{p_0}{q_0} = \frac{14}{1} \qquad \frac{p_1}{q_1} = \frac{15}{1}$$
$$\frac{p_2}{q_2} = \frac{209}{14} \qquad \frac{p_3}{q_3} = \frac{224}{15}$$

and we have (cf. Lemma 2.5 (6))

$$14^2 - 223 = -27; \qquad 15^2 - 223 = 2;$$
$$209^2 - 223 \cdot 14^2 = -27; \qquad 224^2 - 223 \cdot 15^2 = 1.$$

This shows that there are no elements of norm -3 in $\mathbb{Z}[\sqrt{223}]$. More precisely, we see that the set of norms H in $\mathbb{Z}[\sqrt{223}]$ with $|H| \leqslant 14$ is $\{1, 2, 4, 8\}$.

To illustrate the differences that occur in the case $D \equiv 1 \bmod 4$, let us study the small norms in $\mathbb{Q}(\sqrt{229})$.

Example 3.2. Let $K = \mathbb{Q}(\sqrt{229})$. From Corollary 2.11 and Lemma 2.13, we must work out the first few convergents of S.C.F. of $\omega := \omega_{229} = \frac{-1+\sqrt{229}}{2}$ to find the list of all norms H with $|H| < 8$ in $\mathcal{O}_K$. Recall that $\{1, \xi\}$ where $\xi = -\omega'$ is a $\mathbb{Z}$-basis for $\mathcal{O}_K$.

We compute the S.C. F. of ω:

$$\omega = 7 + \frac{\sqrt{229} - 15}{2} = 7 + \cfrac{1}{\frac{\sqrt{229}+15}{2}} = 7 + \cfrac{1}{15 + \frac{\sqrt{229}-15}{2}} \text{ etc., } = 7 + \cfrac{1}{15+}\cfrac{1}{15+}\cfrac{1}{15+}\cdots$$

By Lemma 2.13, we must work out the first 3 convergents. These are easily computed to be

$$\frac{p_0}{q_0} = \frac{7}{1}, \frac{p_1}{q_1} = \frac{106}{15}, \frac{p_2}{q_2} = \frac{1597}{226}.$$

From here, we have the following (cf. Lemma 2.5 (6)):

$$N(p_0 + \xi q_0) = -1, \quad N(p_1 + \xi q_1) = 1, \quad N(p_2 + \xi q_2) = -1.$$

Thus, we see that the only norms H with $|H| < 8$ are $\{\pm 1, \pm 4\}$. In particular, there are no elements of norm $\pm 2, \pm 3, \pm 5, \pm 6, \pm 7$.

While ± 2 and ± 7 are non-squares mod 229, one checks that ± 3 and ± 5 are squares mod 229; in particular, there are no obvious local obstructions for the norm form to represent these primes.

Acknowledgment. We would like to thank the referee for carefully going through the article.

References

[1] Chrystal, G., *Algebra: An Elementary Text-Book*, Volume II, 2nd Edition, Adam and Charles Black, London, England, 1900.

[2] Hall, H. S. and Knight, S. R., *Higher Algebra*, 4th Edition, Macmillan and Co., London, England, 1891.

[3] Hardy, G. H. and Wright, E. M., *Theory of Numbers*, Oxford, Clarendon, 1938.

[4] Lagrange, J. L., Additions au mémoire sur la ré solution des équations numériques, *Mém. Acad. Royale Sc. et belles-lettres*, Berlin, **24**, 1770 (= Œuvres II, 581–652).

[5] Matthews, K., The Diophantine equation $ax^2 + bx + cy^2 = N$, $D = b^2 - 4ac > 0$, *J. Théor. Nombres Bordeaux*, **14**(1): 257–270, 2002.

[6] Pavone, M., A Remark on a Theorem of Serret, *J. Number Theory*, **23**, 268–278, 1986.

[7] Serret, J. A., *Cours D'algébre Supérieure*, Tome Premier, Gauthier-Villars, Paris, 1877.

[8] Trifković, M., *Algebraic theory of quadratic numbers*, Springer-Verlag, New York, 2013.

Kannappan Sampath
Department of Mathematics, University of Michigan
Ann Arbor, Michigan 48105, USA
knsam@umich.edu

B. Sury
Statistics and Mathematics Unit
Indian Statistical Institute
8th Mile Mysore Road, Bangalore-560059, India
sury@ms.isibang.ac.in

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 51-54

CANTOR-LIKE SETS CONSTRUCTED IN TREMAS*

DIYATH NELAKA PANNIPITIYA

(Received : 26 - 09 - 2016, Revised : 19 - 01 - 2017)

Abstract. A Cantor-like set can be obtained by processing a specific inductive construction on the interval $[0,1]$. In that process, say *'A Cantor-like process'*, we remove infinitely many number of open intervals from $[0,1]$ and these removed intervals are called *tremas*. We answer the question: what kind of sets will be obtained by carrying-out a Cantor-like process on the closures of each of these tremas again and again?.

1. Introduction

This article was motivated by the problem [2]: *"Find a set A such that $m^*(A) > 0$ and $0 < m^*(A \cap I) < m(I)$ for all non-empty intervals $I \subseteq [0,1]$"*. Here m^* is the Lebesgue outer measure.

As we know the construction of a Cantor-like set is inductive. Let us construct a Cantor-like set as follows. Let $\alpha \in (0, 1/3)$ and fix it. This fixed α will be used throughout this paper. From $[0,1]$, remove the middle open interval of length α. This leaves two residual intervals, say $I_{1,1}$ and $I_{1,2}$, each of length $(1-\alpha)/2$. Suppose the n^{th} step has been completed, leaving 2^n number of closed intervals, say $I_{n,1},\ I_{n,2},\ \ldots,\ I_{n,2^n}$, each of length $(1 - 3\alpha + 2^n\alpha^{n+1})/(2^n(1-2\alpha))$. We carry-out the $(n+1)^{st}$ step by removing the middle open interval of length α^{n+1} from $I_{n,k}$, for $k = 1, \ldots,\ 2^n$. Because $(1-3\alpha+2^n\alpha^{n+1})/(2^n(1-2\alpha)) > \alpha^{n+1}$, we can do such a construction. Let $K_n = \bigcup_{k=1}^{2^n} I_{n,k}$ and let $K = \bigcap_{n=1}^{\infty} K_n$. The set K is a *Cantor-like set* [2].
It is easy to compute that

$$m(K) = (1-3\alpha)/(1-2\alpha). \tag{1.1}$$

Let

$$K^0_{[a,b]} = \{a + (b-a)x : x \in K\}.$$

This is a Cantor-like set constructed in $[a,b]$. Notice that

$$m(K^0_{[a,b]}) = \ell(1-3\alpha)/(1-2\alpha), \tag{1.2}$$

* The word 'Tremas' in Greek means 'holes' in English.

2010 Mathematics Subject Classification: Primary 28A05, 28A12; Secondary 03E15.
Keywords and Phrases: Lebesgue outer measure, Cantor set.

where $\ell = b - a$. Recalling that two sets are said to be *almost everywhere disjoint* if the Lebesgue outer measure of the intersection of the two sets is zero, it may be observed that $[a,b] \setminus K^0_{[a,b]}$ consists of 2^k number of almost everywhere disjoint tremas each of length

$$\ell\ \alpha^{k+1},\ for\ k \in \mathbb{N} \cup \{0\}. \tag{1.3}$$

Therefore $[a,b] \setminus K^0_{[a,b]}$ is a countable union of tremas, say

$$[a,b] \setminus K^0_{[a,b]} = \bigcup_{j\in\mathbb{N}} I_{0,j}, \tag{1.4}$$

where $I_{0,j}$ *is a trema in* $[a,b] \setminus K^0_{[a,b]}$.

We now construct a sequence of sets $\{K^i_{[a,b]}\}_{i=0}$ recursively as follows.

Suppose $K^n_{[a,b]}$ has been constructed. Denote by T_n the collection of tremas in $[a,b]\setminus K^n_{[a,b]}$. Then T_0 is countable in view of (1.4). Let

$$K^{n+1}_{[a,b]} = \bigcup_{I\in T_n} K^0_{\overline{I}}. \tag{1.5}$$

Notice that for any $I \in T_n$, denoting its closure by $\overline{I}$, we have

$$[a,b]\setminus K^0_{\overline{I}} = [([a,b]\setminus\overline{I}) \cup \overline{I}]\setminus K^0_{\overline{I}} = (([a,b]\setminus\overline{I})\setminus K^0_{\overline{I}}) \cup (\overline{I}\setminus K^0_{\overline{I}}) = ([a,b]\setminus\overline{I}) \cup (\overline{I}\setminus K^0_{\overline{I}})$$

This implies that $[a,b]\setminus K^0_{\overline{I}}$ is a countable union of almost everywhere disjoint intervals as (1.3) and (1.4) implies $\overline{I}\setminus K^0_{\overline{I}}$ is a countable union of almost everywhere disjoint tremas in $\overline{I}\setminus K^0_{\overline{I}}$. Since T_0 is countable and

$$[a,b]\setminus K^{n+1}_{[a,b]} = [a,b]\setminus \bigcup_{I\in T_n} K^0_{\overline{I}} \subseteq \bigcup_{I\in T_n} [a,b]\setminus K^0_{\overline{I}},$$

we can inductively show that T_n is countable for $n \in \mathbb{N}$. So let

$$T_n = \{I_{n,j} :\ I_{n,j}\ is\ a\ trema\ in\ [a,b] \setminus K^n_{[a,b]}\ and\ j \in \mathbb{N}\},\ for\ n \in \mathbb{N} \cup \{0\}.$$

Then

$$K^{n+1}_{[a,b]} = \bigcup_{j\in\mathbb{N}} K^0_{\overline{I}_{n,j}} \tag{1.6}$$

Observe that (for an intuition refer Figure 1)

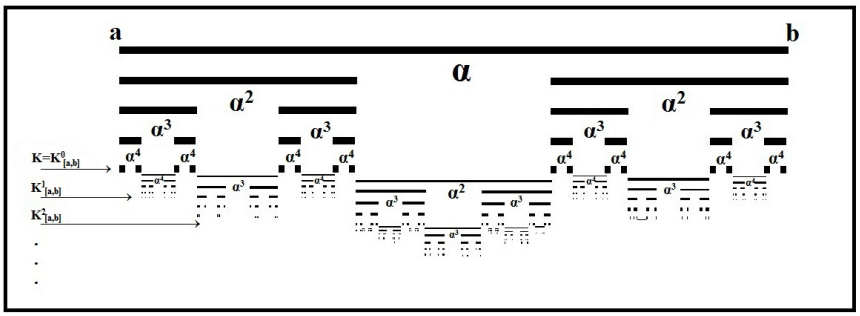

Figure 1

$$K^{n+1}_{[a,b]} = \bigcup_{j\in\mathbb{N}} K^0_{\bar{I}_{n,j}} = \bigcup_{j\in\mathbb{N}} K^1_{\bar{I}_{n-1,j}} = \cdots = \bigcup_{j\in\mathbb{N}} K^n_{\bar{I}_{0,j}} \tag{1.7}$$

In order to find $m(K^n_{[a,b]})$ we need following lemma which can be easily proved.

Lemma 1.1. Let $A = \{A_i\}_{i\in\mathbb{N}}$ be a collection of almost everywhere disjoint sets. Then

$$m(\textstyle\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i).$$

Observe that if we assume that

$$m(K^n_{[a,b]}) = \ell\alpha^n(1-3\alpha)/(1-2\alpha)^{n+1}, \tag{1.8}$$

then in view of (1.7), Lemma 1.1 and (1.3), we obtain

$$\begin{aligned} m(K^{n+1}_{[a,b]}) = m(\bigcup_{j\in\mathbb{N}} K^n_{\bar{I}_{0,j}}) = \sum_{j=0} m(K^n_{\bar{I}_{0,j}}) &= \sum_{j=0} 2^j \frac{(\ell\alpha^{j+1})\alpha^n(1-3\alpha)}{(1-2\alpha)^{n+1}} \\ &= \frac{(\ell\alpha^{n+1})(1-3\alpha)}{(1-2\alpha)^{n+1}} \sum_{j=0} (2\alpha)^j. \end{aligned}$$

Since $0 < \alpha < 1/3$, the series on the right side converges with sum $1/(1-2\alpha)$ and hence

$$m(K^{n+1}_{[a,b]}) = \ell\alpha^{n+1}(1-3\alpha)/(1-2\alpha)^{n+2}. \tag{1.9}$$

In view of (1.2), it therefore follows by the principle of induction that

$$m(K^n_{[a,b]}) = \ell\alpha^n(1-3\alpha)/(1-2\alpha)^{n+1}, \text{ for all } n = 0, 1, 2, \cdots. \tag{1.10}$$

Notice that for $n, m \in \mathbb{N}\cup\{0\}$, $n < m$,

$$K^n_{[a,b]} \cap K^m_{[a,b]} = \{\textit{end points of the tremas in } [a,b] \setminus K^n_{[a,b]}\} \tag{1.11}$$

Since the set on the right hand side of (1.11) is countable, it follows that $\bigcup_{n\in\mathbb{N}\cup\{0\}} K^n_{[a,b]}$ is a collection of almost everywhere disjoint sets.

Now let us take up the problem. For convenience, put $C_n = K^n_{[0,1]}$ for $n \in \mathbb{N}\cup\{0\}$ and $A = \bigcup_{n=0}^{\infty} C_{2n}$. Clearly A is non-empty and is a collection of almost everywhere disjoint sets such that

$$m(A) = m(\bigcup_{n=0}^{\infty} C_{2n}) = \sum_{n=0} m(C_{2n}) = \sum_{n=0} \frac{1.\alpha^{2n}(1-3\alpha)}{(1-2\alpha)^{2n+1}} = \frac{1-2\alpha}{1-\alpha} > 0$$

Let $I \subseteq [0,1]$ be a non-empty interval. Then observe that no-matter how small I is we still can find an $N \in \mathbb{N}$ such that there is a trema in $[0,1]\setminus C_{2N+1}$, say J, which is a proper sub-interval of I. Then, by Lemma 1.1 and (1,10), we have

$$\begin{aligned} m(A\cap J) = m(\bigcup_{n=0} C_{2n}\cap J) &= m(\emptyset \cup \bigcup_{n=N+1} C_{2n}\cap J) \\ &= m(\bigcup_{n=0} K^{2n}_{\bar{J}}) \\ &= \sum_{n=0} m(K^{2n}_{\bar{J}}) \\ &= \sum_{n=0} \frac{m(J)\alpha^{2n}(1-3\alpha)}{(1-2\alpha)^{2n+1}} \end{aligned}$$

$$= \frac{m(J)(1-3\alpha)}{1-2\alpha} \sum_{n=0} (\frac{\alpha}{1-2\alpha})^{2n}$$
$$= \frac{m(J)(1-2\alpha)}{1-\alpha}$$

Because $0 < \frac{m(J)(1-2\alpha)}{1-\alpha} < m(J)$, it is easy to see that neither $m(A \cap I) = 0$ nor $m(A \cap I) = m(I)$ can be happen.

Acknowledgements. The author is grateful to Dr. W. Ramasinghe of the university of Colombo for introducing the problem. The author is also thankful to the anonymous referee as well as to Dr. Dayal B. Dharmasena and Dr. Jayampathy Ratnayake of the university of Colombo for their valuable remarks and suggestions leading to the improvement of the article.

References

[1] Edgar, Gerald, *Measure, Topology and Fractal geometry*, Second Edition, Springer, (2008), 2, New York.

[2] De Barra, Gearoid, *An Introduction to measure theory*, Van Nostrand Reinhold, (1974).

Diyath Nelaka Pannipitiya
I T Unit-2, Faculty of Science
University of Colombo, Colombo-03, Sri Lanka
E-mail: *diyathnp@yahoo.com*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 55-62

FIBONACCI SEQUENCE WITH APPLICATIONS AND EXTENSIONS

SARANYA G. NAIR AND T. N. SHOREY

(Received : 10 - 11 - 2016, Revised : 10 - 01 - 2017)

Abstract. We give a survey of Fibonacci numbers and its relation and applications in several fields. An account of extensions of Fibonacci numbers together with their applications is also given.

1. Introduction

The original problem that Fibonacci investigated in the year 1202 is the following: By a pair of rabbits we always mean a male and a female rabbit. Suppose a newly born pair of rabbits are put in a field. Rabbits are able to mate at the age of one month so that at the end of the second month a female can produce another pair of rabbits. Assume that our rabbits never die and the female always produces new pair, every month from the second month onwards.

Fibonacci Puzzle: How many pairs of rabbits will there be in one year?
Solution:

- At the end of the first month, they mate but there is still only one pair.
- At the end of the second month, the female produces a new pair. So now there are two pairs of rabbits.
- At the end of the third month, the original female produces a second pair, making three in total.
- At the end of the fourth month, the original female has produced yet another pair, the female born after two months produces her first pair also, making five pairs in total.
- The number of pairs of rabbits at the end of a month is equal to the number of pairs of rabbits in the beginning of the next month.
- Thus the number of pairs of rabbits in the beginning of the first, second, third month and so on are given by $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \ldots$ respectively. This is the well-known *Fibonacci sequence* and there are 233 pairs of rabbits in one year. This answers Fibonacci Puzzle.

2010 Mathematics Subject Classification: 11B39
Keywords and Phrases: Fibonacci sequence, Pascal Triangle, Binary recursive sequences.

The above Rabbits problem is not a realistic one. There are such similar problems which are more realistic. For example, see E. Dudeney [5].

Fibonacci sequence is named after Fibonacci. He was also known as Leonardo da Pisa. Fibonacci was born in Pisa, Italy around 1175 AD. In his youth, he studied mathematics in North Africa when he was travelling with his father who was a diplomat. Once he returned to Italy, he wrote several works including his most famous work Liber Abaci (Book of Calculations) in 1202 and it contains his above explained Rabbits Puzzle with its solution. His work highlighted the benefits of Hindu-Arabic system over Roman numerals. Thus he was responsible for introducing Hindu-Arabic system to Europe. He died between 1240 and 1250. In the 1800's, the city of Pisa erected his statue in his memory.

In modern usage, we begin Fibonacci sequence with 0. Thus Fibonacci sequence denoted by F_n with $n \geq 0$, is as follows:

$$F_0 = 0, \quad F_1 = 1, \quad F_2 = 1, \quad F_3 = 2, \quad F_4 = 3, \quad F_5 = 5, \quad F_6 = 8,$$
$$F_7 = 13, \quad F_8 = 21, \quad F_9 = 34, \quad F_{10} = 55, \quad F_{11} = 89, \quad F_{12} = 144, \quad F_{13} = 233, \ldots$$

The first two members $F_0 = 0,\ F_1 = 1$ are called the initial terms of the Fibonacci sequence. Each subsequent member in the sequence is the sum of previous two. Thus Fibonacci sequence F_n with $n \geq 0$ is given by

$$F_0 = 0,\ F_1 = 1 \quad \text{and} \quad F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2.$$

Fibonacci sequence appears at several places in mathematics, so much so that there is an entire journal dedicated to the their study, *Fibonacci Quarterly*. For positive integers a and b, the greatest common divisor of a and b is the greatest positive integer dividing both a and b. The Fibonacci numbers are used in the computational run-time analysis of Euclid's algorithm to determine the greatest common divisor of two integers. For positive integers a and b with $a > b$, if the number of steps required in Euclid's algorithm is k, then we know that $a \geq F_{k+2}$ and this implies an explicit good upper bound for k in terms of a. Fibonacci numbers are used by Yuri Matiyasevich in 1970 in his original proof of Hilbert's Tenth problem. Hilbert's Tenth problem is on the famous list of Hilbert's problems of 1900. Its statement is as follows: For any polynomial equation with arbitrary number of variables and with integer coefficients, devise a process according to which it can be determined in a finite number of operations whether the equation has a solution in integers. The answer to Hilbert's Tenth problem is negative. In fact, it is an undecidable problem. Mathematical Probability theory has its origin in Pascal Triangle and it is related to Fibonacci numbers. More precisely, the sum of the terms of a diagonal of Pascal Triangle is a Fibonacci number. Now we introduce Pascal Triangle. Pascal Triangle is the infinite array of numbers where the n-th row with $n \geq 0$ has $n+1$ entries, namely, the coefficients in the binomial expansion of

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

where the binomial coefficient $\binom{n}{k}$ is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ for } 0 \leq k \leq n.$$

We observe that $\binom{n}{0} = \binom{n}{n} = 1$. Thus the entries in the fifth row of Pascal Triangle are

$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$

and in the 0-th row is 1. Each row begins with one and ends with one. Further, the entries in the $(n+1)$-th row are determined with the n-th row by the relation due to Pascal

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1} \tag{1}$$

In fact, Pascal was not the first to study this triangle. It dates backs to Khayyam Triangle from 11th century in Iran, Young Hai Triangle from 13th century in China and Tartaglia Triangle from 16th century in Italy. But Pascal made important contributions like relation (1) in the study of this triangle. Further, Pascal made an important breakthrough by using it to solve problems in Probability theory. In fact, he, together with Fermat and Christian Huygens, laid the mathematical foundation of Probability theory.

Besides mathematics, Fibonacci numbers find applications in several fields. Many plants show Fibonacci numbers, and sometimes consecutive Fibonacci numbers in the arrangement of leaves around the stem or branching in trees. Buttercups have $F_5 = 5$ petals and daisies can be found with $F_9 = 34, F_{10} = 55$ or even $F_{11} = 89$ petals. This is due to golden ratio

$$(1+\sqrt{5})/2 \approx 1.61803$$

which maximises the space for each leaf or average amount of light falling on each leaf. Golden ratio is used in art and architecture. Recall that in geometry, the Golden ratio is defined by sectioning a straight line segment in such a way that the ratio of the total length to the longer segment equals the ratio of the longer to the shorter segment. A Golden rectangle is a rectangle of sides whose quotient is equal to $\frac{F_{n+1}}{F_n}$ for some $n > 0$ and it is considered to be one of the most visually satisfying geometric forms. Leonardo da Vinci called it the "divine proportion" and it has featured in many of his paintings including Mona Lisa. Applications of Fibonacci numbers also include computer algorithms such as Fibonacci search techniques depending on Divide and Conquer algorithm breaking down a problem into two or more parts of the same type or related type and so on until these become simple enough to be solved directly. Further, Fibonacci numbers are used in running time analysis of Fibonacci heap data structure. Fibonacci Cube is an undirected graph

with a Fibonacci number of nodes and it has been used as a network topology for parallel computing. A theorem of Zeckendorf states that every positive integer can be written uniquely as the sum of one or more Fibonacci numbers such that no two in the sum are consecutive Fibonacci numbers. This theorem has been used in Fibonacci coding which encodes positive integers into binary code words such that each code ends with "11" and contains no other instance of "11" before the end.

In fact, Fibonacci only rediscovered Fibonacci sequence. It appears in Ancient Indian Mathematics in connection with Sanskrit prosody. This is a study central to the composition of Vedas. Development of Fibonacci sequence is attributed to Pingala (200 BC), Virhanka (700 AD), Gopala (1135 AD) and Hemachandra (1150 AD).

Now we give some properties of Fibonacci sequence. Each element of this sequence other than the first one is positive and the first one is zero. Further, every element is the sum of the previous two. Therefore, the elements keep on becoming larger. Another way of putting up this fact is that F_n tends to infinity as n tends to infinity and we write it as

$$F_n \to \infty \text{ as } n \to \infty, \quad \text{or,} \quad \lim_{n\to\infty} F_n = \infty.$$

The above statement continues to be valid for the greatest prime factor of elements of this sequence and the proof is non-trivial, not obvious like that of the previous statement. We have used above the notion of the greatest prime factors of an integer. Therefore it is relevant to recall the Fundamental theorem of Arithmetic. An integer greater than 1 is prime if it has no divisor other than one and itself. Thus 5 is prime but $6 = 2 \cdot 3$ is not prime. Further we observe that

$$10 = 2 \cdot 5, \quad 147 = 3 \cdot 7^2, \quad 240 = 2^4 \cdot 3 \cdot 5.$$

In fact, apart from sign, every integer other than 1 and -1 can be written as a product of prime powers. Further the factorization is unique apart from the order of factors. For an integer a with $|a| > 1$, the greatest prime factor of a is the largest prime occuring in the (unique) factorization of a. Finally we observe in this paragraph that Golden ratio is equal to $\lim_{n\to\infty}(F_{n+1}/F_n)$.

In the sequence F_n with $n \geq 0$, let us look at squares, cubes and higher powers. We observe that

$$F_0, \quad F_1, \quad F_2, \quad F_{12}$$

are squares and

$$F_0, \quad F_1, \quad F_2, \quad F_6$$

are cubes. How about finding a power in the Fibonacci sequence other than $F_0, F_1, F_2, F_6, F_{12}$. The answer to this question is that THERE IS NONE. This is a deep theorem of Bugeaud, Mignotte, Siksek [4] of 2006. Further, Florian Luca and

Shorey [8] proved that a product of two or more consecutive Fibonacci numbers F_n with $n > 0$ is never a power unless $F_1 \cdot F_2 = 1$. This is an analogue of an elegant and celebrated theorem of Erdös and Selfridge [6] that a product of two or more consecutive positive integers is not a power.

Let us consider more general sequence than Fibonacci sequence, namely, $u_0 = 0, u_1 = 1$ and $u_n = 2u_{n-1} - 3u_{n-2}$ for $n \geq 2$. This sequence is

u_0	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}	...
0	1	2	1	-4	-11	-10	13	56	73	-22	...

Here, unlike Fibonacci sequence, there are both positive and negative terms. Therefore, this sequence can not tend to infinity with n. Here we ask: Does

$$|u_n| \to \infty \text{ as } n \to \infty? \tag{2}$$

It is not obvious as in Fibonacci sequence since

$$|u_5| = 11, |u_6| = 10, |u_9| = 73, |u_{10}| = 22$$

and so on. For achieving (2), we need to be careful. We consider the sequence $u_0 = 0, u_1 = 1$ and $u_n = u_{n-1} - u_{n-2}$ for $n \geq 2$. Then the sequence is $0, 1, 1, 0, -1, -1, 0, 1, 1, \ldots$ and (2) is certainly not valid. Therefore, we need to exclude some possibilities. We have been considering sequences of the type $u_0 = 0, u_1 = 1$ and $u_n = ru_{n-1} + su_{n-2}$ for $n \geq 2$ where r and s are integers. Consider the polynomial

$$x^2 - rx - s$$

called the polynomial associated to the sequence. We assume that $s \neq 0$ and $r^2 + 4s \neq 0$ so that the roots α and β of the above polynomial are non-zero and distinct. Then

$$u_n = (\alpha^n - \beta^n)/(\alpha - \beta) \text{ for } n \geq 0. \tag{3}$$

The proof of (3) is by induction on n and using $\alpha + \beta = r, \alpha\beta = -s$. We check (3) for $n = 0$ and $n = 1$. Let $n \geq 2$ and we assume (3) for all $0 \leq m < n$. Then

$$\begin{aligned} u_n &= ru_{n-1} + su_{n-2} \\ &= (\alpha + \beta)\left(\frac{\alpha^{n-1} - \beta^{n-1}}{\alpha - \beta}\right) - \alpha\beta\left(\frac{\alpha^{n-2} - \beta^{n-2}}{\alpha - \beta}\right) \\ &= (\alpha^n - \beta^n)/(\alpha - \beta) \end{aligned}$$

Thus (3) is valid for n. Hence (3) holds for all $n \geq 0$ by induction on n. We must exclude $\alpha = -\beta$ since otherwise $u_n = 0$ for even n. Further, we observe that $(\frac{\alpha}{\beta})^1 = 1$ if $\alpha = \beta$ and $(\frac{\alpha}{\beta})^2 = 1$ if $\alpha = -\beta$. We assume that for every positive integer r

$$(\alpha/\beta)^r \neq 1$$

so that the above possibilities are excluded. If this happens, we call the sequence u_n with $n \geq 0$ non-degenerate which we assume from now onwards. If the sequence is non-degenerate, it is possible to prove (2). The proof is not immediate like the

Fibonacci sequence and further the greatest prime factor of u_n tends to infinity with n. The sequence that we have considered are called Lucas sequences and Fibonacci sequence is a particular example of a Lucas sequence. Lucas sequences are used in primality tests. They are also used in primality proofs. LUC is a public-key cryptosystem based on Lucas sequences. The encryption of the message in LUC is computed as term of certain Lucas sequence. In Lucas sequence, we take $u_0 = 0, u_1 = 1$. We can consider more general sequences by taking the initial terms arbitrary integers u_0, u_1 and each subsequent term is a linear combination of the previous two, as in Lucas sequence. We can define as above when these sequences are non-degenerate. These sequences are called binary recursive sequences (of order two). Petho [9] and Shorey and Stewart [10], independently, proved that there are only finitely many powers in a non-degenerate binary recursive sequence.

We have linear recursive sequences with constant coefficients of any order greater than or equal to two. Fibonacci numbers and more generally, recursive sequences can be extended to all integers. For example, the relation $F_n = F_{n+2} - F_{n+1}$ defines Fibonacci numbers for all integers n. Thus $F_{-1} = 1, F_{-2} = -1, F_{-3} = 2, F_{-4} = -3$ and so on. Further, we check by induction on n that $F_{-n} = (-1)^{n+1}F_n$ for all n. Recursive sequences are used in Population Dynamics in Biology, optimization problems in Economics, in algorithms in Computer Science, Cryptography and Digital signal processing.

Now, we show that the solutions of quadratic equations in integers are given by binary recursive sequences. We consider the equation

$$x^2 - 2y^2 = 1 \tag{4}$$

in integers $x > 0$ and $y > 0$. As was known to Ancient Indian mathematicians that the solutions of this equation are given by

$$x + \sqrt{2}y = (3 + 2\sqrt{2})^n,$$

where n is a positive integer. The above relation implies

$$x - \sqrt{2}y = (3 - 2\sqrt{2})^n.$$

Subtracting the second relation from the first one, we get

$$2\sqrt{2}y = (3 + 2\sqrt{2})^n - (3 - 2\sqrt{2})^n$$

That is,

$$y = \frac{4\sqrt{2}}{2\sqrt{2}} \left(\frac{(3 + 2\sqrt{2})^n - (3 - 2\sqrt{2})^n}{(3 + 2\sqrt{2}) - (3 - 2\sqrt{2})} \right).$$

Writing $\alpha = 3 + 2\sqrt{2}, \quad \beta = 3 - 2\sqrt{2}, \quad$ we have

$$y = y_n = 2(\alpha^n - \beta^n)/(\alpha - \beta).$$

As explained earlier for Lucas sequences, the sequence y_n with $n \geq 0$ satisfies $y_0 = 0, y_1 = 2$ and $y_n = 6y_{n-1} + y_{n-2}$ for $n \geq 2$ with associated polynomial

$$Y^2 - 6Y + 1.$$

Thus if (x, y) is an integer solution of (4), then $y = y_n$ is given by the above binary recursive sequence. Similarly, x is also given by a binary recursive sequence.

Now we apply the above observation to consider the following old and interesting problem solved by Baker [1] in 1968. We consider the integers

$$1 \quad 3 \quad 8 \quad 120$$

This quadruple satisfies the property that if 1 is added to the product of any two above distinct integers, then it is a square. For example

$$1 + 3 \cdot 8 = 5^2, \qquad 1 + 1 \cdot 120 = 11^2, \qquad 1 + 3 \cdot 120 = 19^2, \qquad 1 + 8 \cdot 120 = 31^2,$$
$$1 + 1 \cdot 3 = 2^2, \qquad 1 + 1 \cdot 8 = 3^2, \qquad 1 + 1 \cdot 120 = 11^2.$$

Now we ask the question whether we can find a positive integer x other than 120 such that $1, 3, 8, x$ satisfies the above property. Translating the problem into Diophantine equations, we have

$$3x^2 + 1 = y^2, \ 8x^2 + 1 = z^2.$$

As explained in the preceding paragraph, we see that x is given by two binary recursive sequences. Baker showed that the intersection of these sequences consists of precisely one element, namely 120. This answers the question in the negative.

The sum of the first $n > 1$ integers is equal to $\frac{n(n+1)}{2}$. By Størmer's method [12] on Pellian equations, we know that its greatest prime factor tends to infinity with n. Further, we show that it is not a cube or a higher power. Let

$$n(n+1)/2 = y^q, \ q > 2.$$

We assume that n is odd and the proof is similar when n is even. Since n and $(n+1)/2$ have no common factor, we get

$$n = y_1^q, \qquad (n+1)/2 = y_2^q$$

which implies that

$$2y_2^q - y_1^q = 1.$$

This is not possible by a theorem of Bennett [3]. Hence the sum of the first $n > 1$ positive integers is not a cube or higher power and further its greatest prime factor tends to infinity with n. Now we consider an analogue of the above result for Fibonacci numbers. We know that

$$F_1 + F_2 + \cdots + F_n = F_{n+2} - 1.$$

The greatest prime factor of the right hand side tends to infinity with n by a result of Baker [2] on linear forms in logarithms. Further it follows from Shorey and Stewart [11] (in fact the result of Kiss [7] suffices) that it is not a power whenever n is sufficiently large. Since

$$F_1 + F_3 + \cdots + F_{2n-1} = F_{2n}, \qquad F_2 + F_4 + \cdots + F_{2n} = F_{2n+1} - 1,$$
$$F_1 - F_2 + F_3 - F_4 + \cdots + (-1)^{n+1} F_n = (-1)^{n+1} F_{n-1} + 1,$$
$$F_1^2 + F_2^2 + \cdots + F_n^2 = F_n F_{n+1},$$

we conclude as above that each of the above sum is not a power and its greatest prime factor tends to infinity with n whenever n is sufficiently large. In fact, the

first and the last sum are never a power by the theorem of Bugeaud, Mignotte and Siksek [4] already stated since F_n and F_{n+1} have no factor in common.

Acknowledgements: We are thankful to Professor C. S. Aravinda for encouraging us to write this article. One of the authors (TNS) was getting INSA Senior Scientist award when this work was done.

References

[1] Baker, A., Linear forms in the logarithms of algebraic numbers, *Mathematika*, **15** (1968), 204 – 216.

[2] Baker, A., A sharpening of the bounds for linear forms in logarithms II, *Acta Arith.*, **24** (1973), 33 – 36.

[3] Bennett, M. A., Rational approximation to algebraic numbers of small height: The Diophantine equation $|ax^n - by^n| = 1$, *Jour. Reine Angew. Math.*, **535** (2001), 1 – 49.

[4] Bugeaud, Y., Mignotte, M. and Siksek, S., Classical and modular approaches to exponential diophantine equations I, Fibonacci and Lucas perfect powers, *Annals of Math*, *163* (2006), 969 – 1018.

[5] Dudeney, E., *Amusements in Mathematics*, Dover Press (1958).

[6] Erdös, P. and Selfridge, J. L., The product of consecutive positive integers is never a power *Illinois Jour. Math*, **19** (1975), 292 – 301.

[7] Kiss, P., Differences of the terms of linear recurrences, *Studi. Sci. Math. Hungarica*, **20** (1985), 285 – 293.

[8] Luca, Florian and Shorey, T. N., Diophantine equations with products of consecutive terms in Lucas sequences, *Jour. Number Theory*, **114** (2005), 541 – 560.

[9] Pethö, A., Perfect powers in second order linear recurrences, *Jour. Number Theory*, **15** (1982), 5 – 13.

[10] Shorey, T. N. and Stewart, C. L., On the diophantine equation $ax^{2t} + bx^t y + cy^2 = d$ and pure powers in recurrence sequences, -*Math. Scandinavica*, **52** (1983), 24 - 36.

[11] Shorey, T. N. and Stewart, C. L., Perfect powers in recurrence sequences and some related diophantine equations, *Jour. Number Theory*, **27** (1987), 324 – 352.

[12] Størmer, C., Quelques théormés sur l'equation de Pell $x^2 - Dy^2 = \pm 1$ et leurs applications, *Vid. Skr. I Math. Natur. Kl. (Christiana)* (1897) no. 2, 48 pp.

Saranya G. Nair
Indian Statistical Institute, 8th Mile, Mysore Road
RVCE Post, Bangalore 560059
E-mail: *sarug7@gmail.com*

T. N. Shorey
National Institute of Advanced Studies, IISc Campus, Bangalore 560012
E-mail: *shorey@math.iitb.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 63-76

GENERALIZED MITTAG-LEFFLER FUNCTION AND ITS PROPERTIES

B. V. NATHWANI AND B. I. DAVE

(Received : 28 - 11 - 2016 ; Revised : 15 - 03 - 2017)

Abstract. Motivated essentially by the success of the applications of the Mittag-Leffler functions in Science and Engineering, we propose here a unification of certain generalizations of Mittag-Leffler function including Saxena-Nishimoto's function, Bessel-Maitland function, Dotsenko function, Elliptic Function, etc. We obtain the order and type, asymptotic estimate, a differential equation, and Eigen function Property for the proposed unification. As a specialization, a generalized Konhauser polynomial is considered for which the series inequality relations and inverse series relations are obtained.

1. Introduction

In 1903, the Swedish mathematician Gosta Mittag-Leffler introduced the function

$$E_\alpha(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + 1)}, \tag{1}$$

where $z, \alpha \in \mathbb{C}$, $\Re(\alpha) > 0$, in connection with his method of summation of some divergent series ([12, 13]).

This function was generalized in the form

$$E_{\alpha,\,\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)} \tag{2}$$

by Wiman [24] in 1905 and studied by Humbert and Agarwal [7].

A further extension of this was introduced by Prabhakar [15] in the form :

$$E^{\gamma}_{\alpha,\,\beta}(z) = \sum_{n=0}^{\infty} \frac{(\gamma)_n}{\Gamma(\alpha n + \beta)} \frac{z^n}{n!}, \tag{3}$$

whereas Shukla and Prajapati [22] studied the generalization :

$$E^{\gamma\,q}_{\alpha,\,\beta}(z) = \sum_{n=0}^{\infty} \frac{(\gamma)_{qn}}{\Gamma(\alpha n + \beta)} \frac{z^n}{n!}. \tag{4}$$

These two versions are subject to the conditions that $\Re(\alpha, \beta, \gamma) > 0$, $(\gamma)_n$ is Pochhammer symbol with $(\gamma)_0 = 1$, $(\gamma)_n = (\gamma)(\gamma+1)(\gamma+2)...(\gamma+n-1) = \Gamma(\gamma+n)/\Gamma(\gamma)$ and, for (4), the generalized Pochhammer symbol $(\gamma)_{qn} = \Gamma(\gamma + qn)/\Gamma(\gamma)$, where $q \in (0,1) \cup \mathbb{N}$. Since the time of Wiman (1905), many researchers

2010 Mathematics Subject Classification : 33B15; 33E12; 33E99

Key words and phrases : Generalized Mittag-Leffler function, differential equation, eigen function, generalized Konhauser polynomial, series inequality relations.

have proposed and studied various generalizations of the Mittag-Leffler function [12] (see [4], [5], [8], [14], [15], [16], [17], [19], [20], [22], [24]).
We introduce here the function which is denoted and defined by

$$E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)=\sum_{n=0}^{\infty}\frac{[(\gamma)_{\delta n}]^s}{\Gamma(\alpha n+\beta)\ [(\lambda)_{\mu n}]^r}\frac{z^n}{n!}, \tag{5}$$

wherein the parameters $\alpha,\ \beta,\ \gamma,\ \lambda\ \in\mathbb{C}$ with $\Re(\alpha,\beta,\gamma,\lambda)>0,\ \delta,\ \mu>0,\ r\in\mathbb{N}\cup\{-1,0\}$ and $s\in\mathbb{N}\cup\{0\}$. We shall refer to this function as ***gml***.
The proposed *gml* can be viewed as a special case of the Wright function [23, Eq.(21), p.50]

$${}_p\Psi_q\left[\begin{array}{lll}(\alpha_1,A_1), & \ldots, & (\alpha_p,A_p);\\ & & \\ (\beta_1,B_1), & \ldots, & (\beta_q,B_q);\end{array}\ z\right]=\sum_{n=0}^{\infty}\frac{\prod_{i=1}^{p}\Gamma(\alpha_i+A_in)}{\prod_{j=1}^{q}\Gamma(\beta_j+B_jn)}\frac{z^n}{n!}, \tag{6}$$

when the parameters $\alpha_i=\gamma, A_i=\delta, \beta_1=\beta, B_1=\alpha, \beta_j=\lambda, B_j=\mu$ for all $i=1,2,\ldots,s$ and $j=2,3,\ldots,r$.
The function in (5), besides containing the above cited generalizations, also includes the following functions.
(i) Bessel-Maitland function [6, Eq.(1.7.8), p.19]:

$$J^{\mu}_{\nu}(z)=\sum_{n=0}^{\infty}\frac{(-1)^n}{\Gamma(\nu+n\mu+1)}\frac{z^n}{n!},$$

(ii) Dotsenko function [6, Eq.(1.8.9), p.24]:

$${}_2R_1(a,b;c,\omega;\nu;z)=\frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)\Gamma(b+n\frac{\omega}{\nu})}{\Gamma(c+n\frac{\omega}{\nu})}\frac{z^n}{n!},$$

(iii) a particular form ($m=2$) of extension of Mittag-Leffler function due to Saxena and Nishimoto [21] given by

$$E_{\gamma,K}[(\alpha_j,\beta_j)_{1,2};z]=\sum_{n=0}^{\infty}\frac{(\gamma)_{Kn}}{\Gamma(\alpha_1n+\beta_1)\Gamma(\alpha_2n+\beta_2)}\frac{z^n}{n!},$$

where $z,\gamma,\alpha_j,\beta_j\in\mathbb{C},\Re(\alpha_1+\alpha_2)>\Re(K)-1,\Re(K)>0$, and
(iv) the Elliptic function [11, Eq.(1), p.211] :

$$K(k)=\frac{\pi}{2}\ {}_2F_1\left(\begin{array}{ll}\frac{1}{2}, & \frac{1}{2};\\ 1; & \end{array}\ k^2\right).$$

The reducibility of the *gml* to the above mentioned functions is tabulated below.

Table - 1

Function	**r**	**s**	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	$\boldsymbol{\delta}$	$\boldsymbol{\lambda}$	$\boldsymbol{\mu}$
Mittag-Leffler	0	1	α	1	1	1	-	-
Wiman	0	1	α	β	1	1	-	-

Function	**r**	**s**	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	$\boldsymbol{\delta}$	$\boldsymbol{\lambda}$	$\boldsymbol{\mu}$
Prabhakar	0	1	α	β	γ	1	-	-
Shukla and Prajapati	0	1	α	β	γ	q	-	-
Bessel-Maitland	0	0	μ	$\nu+1$	-	-	-	-
Dotsenko	-1	1	ω/ν	c	a	1	b	ω/ν
Saxena-Nishimoto	1	1	α_1	β_1	γ	K	β_2	α_2
Elliptic	-1	1	1	1	$\frac{1}{2}$	1	$\frac{1}{2}$	1

Table - 1 (complete)

It is noteworthy that the role of parameter $'s'$ is special; as is seen in the last section.

2. Main Results

In this section, we prove the following results.

2.1. **Convergence.**

Theorem 2.1. *Let* $\Re(\alpha,\beta,\gamma,\lambda)>0, \Re(\alpha+r\mu-s\delta+1)>0, \delta,\mu>0, r\in\mathbb{N}\cup\{-1,0\}$ *and* $s\in\mathbb{N}\cup\{0\}$. *Then* $E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)$ *is an entire function of order* $\varrho=1/(\Re(\alpha+r\mu-s\delta+1))$ *and type* $\sigma=(1/\varrho)(\delta^{s\delta}/\{\Re(\alpha)\}^{\Re(\alpha)}\,\mu^{r\mu})^{\varrho}$.

Proof. Let us take

$$u_n=\frac{[(\gamma)_{\delta n}]^s}{\Gamma(\alpha n+\beta)\;[(\lambda)_{\mu n}]^r\;n!}.=\frac{[\Gamma(\gamma+\delta n)]^s\;[\Gamma(\lambda)]^r}{[\Gamma(\gamma)]^s\;\Gamma(\alpha n+\beta)\;[\Gamma(\lambda+\mu n)]^r\;\Gamma(n+1)}$$

in (5) so that

$$E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)=\sum_{n=0}^{\infty}u_n\;z^n.$$

Then in view of the Stirling's asymptotic formula of the Γ-function [3] given by

$$\Gamma(z)\sim\sqrt{2\pi}\;e^{-z}\;z^{z-\frac{1}{2}}, \tag{7}$$

for large $|z|$, we get

$$\begin{aligned}u_n\quad\sim\quad&\frac{\left(\sqrt{2\pi}e^{-(\gamma+\delta n)}(\gamma+\delta n)^{\gamma+\delta n-1/2}\right)^s}{\left(\sqrt{2\pi}e^{-\gamma}\gamma^{\gamma-1/2}\right)^s\;\left(\sqrt{2\pi}e^{-(\alpha n+\beta)}(\alpha n+\beta)^{\alpha n+\beta-1/2}\right)}\\&\times\frac{\left(\sqrt{2\pi}e^{-\lambda}\lambda^{\lambda-1/2}\right)^r}{\left(\sqrt{2\pi}e^{-(\lambda+\mu n)}(\lambda+\mu n)^{\lambda+\mu n-1/2}\right)^r\;\left(\sqrt{2\pi}e^{-(n+1)}(n+1)^{n+1/2}\right)}\end{aligned}$$

$$= \frac{e^{-\delta ns}\ (\delta n)^{s(\gamma+\delta n-1/2)}\left(1+\frac{\gamma}{\delta n}\right)^{s(\gamma+\delta n-1/2)}}{\gamma^{s(\gamma-1/2)}\sqrt{2\pi}e^{-(\alpha n+\beta)}(\alpha n)^{\alpha n+\beta-1/2}\left(1+\frac{\beta}{\alpha n}\right)^{\alpha n+\beta-1/2}} \times \frac{\lambda^{r(\lambda-1/2)}\left(1+\frac{\lambda}{\mu n}\right)^{-r(\lambda+\mu n-1/2)}}{e^{-\mu nr}(\mu n)^{r(\lambda+\mu n-1/2)}\sqrt{2\pi}e^{-n}n^{n+1/2}\left(1+\frac{1}{n}\right)^{n+1/2}}.$$

Hence if R is radius of convergence of the series of $E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)$, then with the use of Cauchy-Hadamard formula:

$$\frac{1}{R} = \lim_{n\to\infty}\sup\sqrt[n]{|u_n|} = \lim_{n\to\infty}\sup\left|\frac{e^{\alpha+r\mu-s\delta+1}\delta^{s\delta}}{\alpha^{\alpha}\mu^{r\mu}}\right|\left|n^{s\delta-\alpha-r\mu-1}\right| = \lim_{n\to\infty}\sup\frac{e^{\Re(\alpha+r\mu-s\delta+1)}\delta^{s\delta}}{\{\Re(\alpha)\}^{\Re(\alpha)}\,\mu^{r\mu}}\ n^{\Re(s\delta-\alpha-r\mu-1)} = 0,$$

when $\Re(\alpha+r\mu-s\delta+1)>0$. Therefore, the function (5) turns out to be an *entire* function. In order to determine its order, we use the result [1, Eq.(1.1)] which states that if $f(z)=\sum_{n=0}^{\infty}a_nz^n$ is an entire function then the order $\varrho(f)$ of f is given by [1, Eq.(1.2)]

$$\varrho(f) = \lim_{r\to\infty}\sup\frac{logM(r;f)}{logr} = \lim_{n\to\infty}\sup\frac{n\ \log n}{\log(1/|u_n|)}.$$

By choosing $f(z)=E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)$, this particularizes to

$$\varrho = \varrho(E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)) = \lim_{n\to\infty}\sup\frac{n\ \log n}{\log(1/|u_n|)}.$$

Here,

$$\begin{aligned}
\log\left(\frac{1}{|u_n|}\right) &= \log\left|\frac{[\Gamma(\gamma)]^s\ \Gamma(\alpha n+\beta)\ [\Gamma(\lambda+\mu n)]^r\ \Gamma(n+1)}{[\Gamma(\gamma+\delta n)]^s\ [\Gamma(\lambda)]^r}\right| \\
&\sim \log\left|\frac{\left(\sqrt{2\pi}e^{-(\alpha n+\beta)}(\alpha n+\beta)^{\alpha n+\beta-1/2}\right)\ \left(\sqrt{2\pi}e^{-\gamma}\gamma^{\gamma-1/2}\right)^s}{\left(\sqrt{2\pi}e^{-(\gamma+\delta n)}(\gamma+\delta n)^{\gamma+\delta n-1/2}\right)^s}\right. \\
&\quad \left.\times\frac{\left(\sqrt{2\pi}e^{-(\lambda+\mu n)}(\lambda+\mu n)^{\lambda+\mu n-1/2}\right)^r\ \left(\sqrt{2\pi}e^{-(n+1)}(n+1)^{n+1/2}\right)}{\left(\sqrt{2\pi}e^{-\lambda}\lambda^{\lambda-1/2}\right)^r}\right| \\
&= \log\left|2\pi\ \gamma^{s(\gamma-1/2)}\ e^{s\delta n-\alpha n-\beta-r\mu n-n}\ (\alpha n+\beta)^{(\alpha n+\beta-1/2)}\right. \\
&\quad \left.\times(\lambda+\mu n)^{r(\lambda+\mu n-1/2)}\ (n+1)^{n+1/2}\right| \\
&\quad -\log\left|(\gamma+\delta n)^{s(\gamma+\delta n-1/2)}\lambda^{r(\lambda-1/2)}\right| \\
&= \log\Big(2\pi\ |\gamma|^{s\Re(\gamma-1/2)}\ e^{\Re(s\delta n-\alpha n-\beta-r\mu n-n)}\ |\alpha n+\beta|^{\Re(\alpha n+\beta-1/2)} \\
&\quad \times|\lambda+\mu n|^{r\Re(\lambda+\mu n-1/2)}\ (n+1)^{n+1/2}\Big) \\
&\quad -\log\Big(|\gamma+\delta n|^{s\Re(\gamma+\delta n-1/2)}|\lambda|^{r\Re(\lambda-1/2)}\Big)
\end{aligned}$$

$$
\begin{aligned}
= \quad & \log(2\pi) + s\Re(\gamma - 1/2)\log|\gamma| + \Re(s\delta n - \alpha n - \beta - r\mu n - n) \\
& + \Re(\alpha n + \beta - 1/2)\log|\alpha n + \beta| + r\Re(\lambda + \mu n - 1/2)\log|\lambda + \mu n| \\
& + (n + 1/2)\log(n+1) - s\Re(\gamma + \delta n - 1/2)\log|\gamma + \delta| \\
& - r\Re(\lambda - 1/2)\log|\lambda|. \qquad (8)
\end{aligned}
$$

Hence, $1/\varrho = \lim_{n\to\infty} \sup\{\log(1/|u_n|)/(n\log n)\} = \Re(\alpha + r\mu - s\delta + 1)$.
Thus the order of *gml* is

$$\varrho = 1/\Re(\alpha + r\mu - s\delta + 1). \qquad (9)$$

The type σ of the *gml* [10] is given by

$$\sigma(E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)) = (1/e\ \varrho)\ \lim_{n\to\infty} \sup\left\{n\,|u_n|^{\varrho/n}\right\}, \qquad (10)$$

where

$$
\begin{aligned}
|u_n| \quad = \quad & \left| \frac{[\Gamma(\gamma+\delta n)]^s\ [\Gamma(\lambda)]^r}{[\Gamma(\gamma)]^s\ \Gamma(\alpha n+\beta)\ [\Gamma(\lambda+\mu n)]^r\ \Gamma(n+1)} \right| \\
\sim \quad & \left| \frac{\left(\sqrt{2\pi}e^{-(\gamma+\delta n)}(\gamma+\delta n)^{\gamma+\delta n-1/2}\right)^s \left(\sqrt{2\pi}e^{-\gamma}\gamma^{\gamma-1/2}\right)^{-s}}{\left(\sqrt{2\pi}e^{-(\alpha n+\beta)}(\alpha n+\beta)^{\alpha n+\beta-1/2}\right)} \right. \\
& \left. \times \frac{\left(\sqrt{2\pi}e^{-\lambda}\lambda^{\lambda-1/2}\right)^r}{\left(\sqrt{2\pi}e^{-(\lambda+\mu n)}(\lambda+\mu n)^{\lambda+\mu n-1/2}\right)^r \left(\sqrt{2\pi}e^{-(n+1)}(n+1)^{n+1-1/2}\right)} \right| \\
= \quad & \left| \frac{1}{2\pi}\ \frac{e^{\alpha n+\beta+r\mu n-s\delta n+n-1}\ (\delta n)^{s(\gamma+\delta n-1/2)}}{(\alpha n)^{\alpha n+\beta-1/2}\left(1+\frac{\beta}{\alpha n}\right)^{\alpha n+\beta-1/2}\gamma^{s(\gamma-1/2)}} \right. \\
& \left. \times \frac{\lambda^{r(\lambda-1/2)}\ \left(1+\frac{\gamma}{\delta n}\right)^{s(\gamma+\delta n-1/2)}}{(\mu n)^{r(\lambda+\mu n-1/2)}\left(1+\frac{\lambda}{\mu n}\right)^{r(\lambda+\mu n-1/2)}\ (n+1)^{n+1/2}} \right|.
\end{aligned}
$$

On substituting this on the right hand side of (10) and then using (9), we get

$$
\begin{aligned}
\lim_{n\to\infty} \sup\left\{n\,|u_n|^{\varrho/n}\right\} \quad = \quad & \left(\delta^{s\delta}/\left\{\Re(\alpha)\right\}^{\Re(\alpha)}\mu^{r\mu}\right)^{\varrho} e^{\Re(\alpha+r\mu-s\delta+1)\varrho} \\
& \times \lim_{n\to\infty} n^{\Re(s\delta-\alpha-r\mu-1)\varrho+1}.
\end{aligned}
$$

This gives

$$\sigma(E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)) = (1/\varrho)\left((\delta^{s\delta}/(\{\Re(\alpha)\}^{\Re(\alpha)}\mu^{r\mu}\right)^{\varrho}. \qquad (11)$$

For every positive ϵ, the asymptotic estimate [10]

$$\left|E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)\right| < exp\left((\sigma+\epsilon)\ |z|^{\varrho}\right),\ |z| \geq r_0 > 0 \qquad (12)$$

holds with ϱ, σ as in (9), (11) for $|z| \geq r_0(\epsilon)$, and for sufficiently large $r_0(\epsilon)$. □

2.2. **Differential Equation.** Let us take

$$\frac{\delta^{s\delta}}{\alpha^{\alpha}\ \mu^{r\mu}} = p,\ \frac{d}{dz} = D,\ zD = \theta,\ \prod_{j=0}^{a-1}\left[\left(\theta + \frac{b+j}{a}\right)\right]^m = \Delta_j^{(a,b;m)},$$

$$\prod_{j=0}^{a-1}\left[\left(\theta+\frac{b+j}{a}-1\right)\right]^m=\Upsilon_j^{(a,b;m)},\ \prod_{j=0}^{a-1}\left[\left(-\theta+\frac{b+j}{a}-1\right)\right]^m=\Theta_j^{(a,b;m)},\quad(13)$$

and

$$p^{-1}\ D\ \Theta_m^{(\delta,\gamma;-s)}\Upsilon_k^{(\mu,\lambda;r)}\Upsilon_j^{(\alpha,\beta;1)}=\Omega_{\Theta;\Upsilon}.\quad(14)$$

Here the operators $\Theta_m^{(\delta,\gamma;-s)},\ \Upsilon_k^{(\mu,\lambda;r)},\ \Upsilon_j^{(\alpha,\beta;1)}$ in (14) are not commutative with the operator D.

With these notations, we now derive the differential equation satisfied by (5).

Theorem 2.2. *Let $\alpha,\mu,\delta\in\mathbb{N}$ then $y=E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)$ satisfies the equation*

$$\left[\Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ \theta-z\frac{\delta^{s\delta}}{\alpha^\alpha\ \mu^{r\mu}}\ \Delta_m^{(\delta,\gamma;s)}\right]y=0.\quad(15)$$

Proof. We have

$$\begin{aligned}
y &= \sum_{n=0}^{\infty}\frac{[(\gamma)_{\delta n}]^s\ z^n}{\Gamma(\alpha n+\beta)\ [(\lambda)_{\mu n}]^r\ n!}=\frac{1}{\Gamma(\beta)}\sum_{n=0}^{\infty}\frac{[(\gamma)_{\delta n}]^s\ z^n}{(\beta)_{\alpha n}\ [(\lambda)_{\mu n}]^r\ n!}\\
&= \frac{1}{\Gamma(\beta)}\sum_{n=0}^{\infty}\frac{\delta^{s\delta n}[(\frac{\gamma}{\delta})_n]^s\ [(\frac{\gamma+1}{\delta})_n]^s...[(\frac{\gamma+\delta-1}{\delta})_n]^s z^n}{\alpha^{\alpha n}(\frac{\beta}{\alpha})_n\ (\frac{\beta+1}{\alpha})_n...(\frac{\beta+\alpha-1}{\alpha})_n}\\
&\qquad\times\frac{1}{\mu^{r\mu n}[(\frac{\lambda}{\mu})_n]^r\ [(\frac{\lambda+1}{\mu})_n]^r...[(\frac{\lambda+\mu-1}{\mu})_n]^r\ n!}\\
&= \frac{1}{\Gamma(\beta)}\sum_{n=0}^{\infty}\frac{\delta^{s\delta n}}{\alpha^{\alpha n}\ \mu^{r\mu n}}\frac{\left\{\prod_{m=0}^{\delta-1}[(\frac{\gamma+m}{\delta})_n]^s\right\}}{\left\{\prod_{j=0}^{\alpha-1}(\frac{\beta+j}{\alpha})_n\right\}\left\{\prod_{k=0}^{\mu-1}[(\frac{\lambda+k}{\mu})_n]^r\right\}n!}\ z^n.
\end{aligned}\quad(16)$$

Now take

$$\frac{1}{\Gamma(\beta)}\prod_{m=0}^{\delta-1}\left[\left(\frac{\gamma+m}{\delta}\right)_n\right]^s=A_n,\ \prod_{j=0}^{\alpha-1}\left(\frac{\beta+j}{\alpha}\right)_n=B_n,\ \prod_{k=0}^{\mu-1}\left[\left(\frac{\lambda+k}{\mu}\right)_n\right]^r=C_n,$$

then the *gml* (5) takes the form $y=\sum_{n=0}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ n!}z^n$. Now,

$$\theta\ y=\sum_{n=0}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ n!}\ \theta\ z^n=\sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ (n-1)!}\ z^n.$$

Further,

$$\begin{aligned}
\Upsilon_j^{(\alpha,\beta;1)}\ \theta\ y &= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ (n-1)!}\prod_{j=0}^{\alpha-1}\left(\theta+\frac{\beta+j}{\alpha}-1\right)z^n\\
&= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ (n-1)!}\prod_{j=0}^{\alpha-1}\left(n+\frac{\beta+j}{\alpha}-1\right)z^n\\
&= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_{n-1}\ C_n\ (n-1)!}\ z^n.
\end{aligned}$$

Finally,

$$\begin{aligned}\Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ \theta\ y &= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_{n-1}\ C_n\ (n-1)!}\ \prod_{k=0}^{\mu-1}\left[\left(\theta+\frac{\lambda+k}{\mu}-1\right)\right]^r\ z^n\\ &= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_{n-1}\ C_n\ (n-1)!}\ \prod_{k=0}^{\mu-1}\left[\left(n+\frac{\lambda+k}{\mu}-1\right)\right]^r\ z^n\\ &= \sum_{n=1}^{\infty}\frac{A_n\ p^n}{B_{n-1}\ C_{n-1}\ (n-1)!}\ z^n.\end{aligned}$$

Thus,

$$\Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ \theta\ y=\sum_{n=0}^{\infty}\frac{A_{n+1}\ p^{n+1}}{B_n\ C_n\ n!}\ z^{n+1}. \tag{17}$$

On the other hand,

$$\begin{aligned}\Delta_m^{(\delta,\gamma;s)}\ y &= \sum_{n=0}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ n!}\ \prod_{m=0}^{\delta-1}\left[\left(\theta+\frac{\gamma+m}{\delta}\right)\right]^s\ z^n\\ &= \sum_{n=0}^{\infty}\frac{A_n\ p^n}{B_n\ C_n\ n!}\ \prod_{m=0}^{\delta-1}\left[\left(n+\frac{\gamma+m}{\delta}\right)\right]^s z^n=\sum_{n=0}^{\infty}\frac{A_{n+1}\ p^n}{B_n\ C_n\ n!}\ z^n,\end{aligned}$$

that is,

$$p\ z\ \Delta_m^{(\delta,\gamma;s)}\ y=\sum_{n=0}^{\infty}\frac{A_{n+1}\ p^{n+1}}{B_n\ C_n\ n!}\ z^{n+1}. \tag{18}$$

On comparing (17) and (18), we get (15). □

2.3. **Eigen function property.**

Theorem 2.3. *Let $\alpha,\mu,\delta\in\mathbb{N}$ then $E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(z;s,r)$ is an eigen function with respect to the operator $\Omega_{\Theta;\Upsilon}$. That is,*

$$\Omega_{\Theta;\Upsilon}\left(E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(\zeta z;s,r)\right)=\ \zeta\ E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(\zeta z;s,r). \tag{19}$$

Proof. We first note that

$$w=E^{\gamma,\delta}_{\alpha,\beta,\lambda,\mu}(\zeta z;s,r)=\sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_n\ C_n\ n!}z^n.$$

Now in view of (13),

$$\begin{aligned}\Upsilon_j^{(\alpha,\beta;1)}\ w &= \sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_n\ C_n\ n!}\ \prod_{j=0}^{\alpha-1}\left(\theta+\frac{\beta+j}{\alpha}-1\right)\ z^n\\ &= \sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_n\ C_n\ n!}\ \prod_{j=0}^{\alpha-1}\left(n+\frac{\beta+j}{\alpha}-1\right)\ z^n=\sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_n\ n!}\ z^n.\end{aligned}$$

Next

$$\begin{aligned}\Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ w &= \sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_n\ n!}\ \prod_{k=0}^{\mu-1}\left[\left(\theta+\frac{\lambda+k}{\mu}-1\right)\right]^r\ z^n\\ &= \sum_{n=0}^{\infty}\frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_n\ n!}\ \prod_{k=0}^{\mu-1}\left[\left(n+\frac{\lambda+k}{\mu}-1\right)\right]^r\ z^n\end{aligned}$$

$$= \sum_{n=0}^{\infty} \frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_{n-1}\ n!}\ z^n.$$

Further, from (13),

$$\begin{aligned}
\Theta_m^{(\delta,\gamma;-s)}\ \Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ w &= \sum_{n=0}^{\infty} \frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_{n-1}\ n!}\ \Theta_m^{(\delta,\gamma;-s)}\ z^n \\
&= \sum_{n=0}^{\infty} \frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_{n-1}\ n!} \prod_{j=0}^{\delta-1} \left[\left(-\theta + \frac{\gamma+j}{\delta} - 1\right)\right]^{-s}\ z^n \\
&= \sum_{n=0}^{\infty} \frac{A_n\ (\zeta p)^n}{B_{n-1}\ C_{n-1}\ n!} \prod_{j=0}^{\delta-1} \left[\left(n + \frac{\gamma+j}{\delta} - 1\right)\right]^{-s}\ z^n \\
&= \sum_{n=0}^{\infty} \frac{A_{n-1}\ (\zeta p)^n}{B_{n-1}\ C_{n-1}\ n!}\ z^n.
\end{aligned}$$

Finally, using (14) we get

$$\begin{aligned}
\Omega_{\Theta;\Upsilon}^{(\delta,\gamma,s;\alpha,\beta,\mu,\lambda,r)}\ w &= p^{-1}\ D\ \Theta_m^{(\delta,\gamma;-s)}\ \Upsilon_k^{(\mu,\lambda;r)}\ \Upsilon_j^{(\alpha,\beta;1)}\ w \\
&= \sum_{n=0}^{\infty} \frac{A_{n-1}\ \zeta^n\ p^{n-1}}{B_{n-1}\ C_{n-1}\ n!}\ D\ z^n \\
&= \sum_{n=1}^{\infty} \frac{A_{n-1}\ \zeta^n\ p^{n-1}}{B_{n-1}\ C_{n-1}\ (n-1)!}\ z^{n-1} \\
&= \sum_{n=0}^{\infty} \frac{A_n\ \zeta^{n+1}\ p^n}{B_n\ C_n\ n!}\ z^n \\
&= \zeta \sum_{n=0}^{\infty} \frac{A_n\ \zeta^n\ p^n}{B_n\ C_n\ n!}\ z^n = \zeta\ E_{\alpha,\beta,\lambda,\mu}^{\gamma,\delta}(\zeta z; s, r).
\end{aligned}$$

□

The properties corresponding to the special cases listed in Table-1 may be deduced by suitably specializing the parameters involved in the above derived properties of *gml*.

2.4. **Generalized Konhauser polynomial.** The well known Konhauser polynomial [9]

$$Z_n^{\mu}(x;k) = \frac{\Gamma(kn+\mu+1)}{\Gamma(n+1)} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \frac{x^{kj}}{\Gamma(kj+\mu+1)}, \tag{20}$$

with $\Re(\mu) > -1$, admits a generalization by means of the *gml* as follows.

Taking $\alpha, \beta, \lambda > 0,\ \delta(=m), \mu, r, s \in \mathbb{N}, \gamma =$ a negative integer: $-n$, $n^* = [n/m]$ the greatest integer part, replacing β by $\beta+1$, and z by a real variable x^k and denoting the polynomial thus obtained by $B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; s, r)$, we get

$$\begin{aligned}
E_{\alpha,\beta+1,\lambda,\mu}^{-n,m}(x^k; s, r) &= \sum_{j=0}^{n^*} \frac{[(-n)_{mj}]^s\ x^{kn}}{\Gamma(\alpha j+\beta+1)\,[(\lambda)_{\mu n}]^r\ j!} \\
&= \frac{(n!)^s}{\Gamma(\alpha n+\beta+1)}\ B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; s, r), \qquad (21)
\end{aligned}$$

where

$$B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;s,r)=\frac{\Gamma(\alpha n+\beta+1)}{(n!)^s}\sum_{j=0}^{[n/m]}\frac{[(-n)_{mj}]^s\,x^{kj}}{\Gamma(\alpha j+\beta+1)\,[(\lambda)_{\mu j}]^r\,j!}. \tag{22}$$

The presence of parameter $'s'$ yields the *unusual* inverse series relations involving the inequalities. In fact, for $s=1$ the usual inverse series relations occur whereas for other values of s the inverse inequality relations occur. This is shown in the following theorems.

Theorem 2.4. *Let $f(x,n;s)$ and $g(x,n;s)$ be real valued functions, $\alpha,\beta,\lambda>0$, and $\mu,k\in\mathbb{N},\ r\in\mathbb{N}\cup\{0\}$, then*

$$f(x,n;s)<B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;s,r) \tag{23}$$

implies

$$x^{kn}>\frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s}\sum_{j=0}^{mn}\frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\ f(x,j;s); \tag{24}$$

and

$$x^{kn}<\frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s}\sum_{j=0}^{mn}\frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\ g(x,j;s), \tag{25}$$

implies

$$g(x,n;s)>B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;s,r). \tag{26}$$

Proof. In order to prove (23) implies (24), assume that the inequality (23) holds. Denote the right hand side of (24) by ϕ_n then

$$\phi_n\ =\ \frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s}\sum_{j=0}^{mn}\frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\ f(x,j;s).$$

Now substituting for $f(x,j;s)$ from (23), we get

$$\begin{aligned}\phi_n\ &<\ \frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s\ (n!)^s}\sum_{j=0}^{mn}\frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\\ &\quad\times\sum_{i=0}^{[j/m]}\frac{[(-j)_{mi}]^s\,x^{ki}}{\Gamma(\alpha i+\beta+1)\,[(\lambda)_{\mu i}]^r\,i!}\\ &=\frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s\ (n!)^s}\sum_{j=0}^{mn}\frac{(-1)^{sj}(mn!)^s}{[(mn-j)!]^s\,\Gamma(\alpha j+\beta+1)}\\ &\quad\times\sum_{i=0}^{[j/m]}\frac{(-1)^{smi}(j!)^s x^{ki}}{[(j-mi)!]^s\,\Gamma(\alpha i+\beta+1)\,[(\lambda)_{\mu i}]^r\,i!}\\ &=\ \sum_{j=0}^{mn}\sum_{i=0}^{[j/m]}\frac{(-1)^{smi+sj}\ (j!)^s\ \Gamma(\alpha n+\beta+1)\ [(\lambda)_{\mu n}]^r\ n!\,x^{ki}}{[(j-mi)!]^s\ [(mn-j)!]^s\ \Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ i!}.\end{aligned}$$

In view of the double series relation

$$\sum_{j=0}^{mn}\sum_{i=0}^{[j/m]} A(j,i) = \sum_{i=0}^{n}\sum_{j=0}^{mn-mi} A(j+mi,i), \tag{27}$$

we further have

$$\begin{aligned}
\phi_n &< \sum_{i=0}^{n}\sum_{j=0}^{mn-mi} \frac{(-1)^{sj}\ (j!)^s\ \Gamma(\alpha n+\beta+1)\ [(\lambda)_{\mu n}]^r\ n!\ x^{ki}}{[(j)!]^s\ [(mn-mi-j)!]^s\ \Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ i!} \\
&= x^{kn} + \sum_{i=0}^{n-1} \frac{\Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ n!\ x^{ki}}{[(mn-mi)!]^s\ \Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ i!} \\
&\quad \times \sum_{j=0}^{mn-mi} (-1)^{sj} \binom{mn-mi}{j}^s \\
&\leq x^{kn} + \sum_{i=0}^{n-1} \frac{\Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ n!\ x^{ki}}{[(mn-mi)!]^s\ \Gamma(\alpha i+\beta+1)\ [(\lambda)_{\mu i}]^r\ i!} \\
&\quad \times \left(\sum_{j=0}^{mn-mi} (-1)^{j} \binom{mn-mi}{j}\right)^s.
\end{aligned}$$

Since the inner series on the right hand side vanishes, it follows that $\phi_n < x^{kn}$, furnishing the inequality (24). □

The proof of (25) implies (26) is similar and therefore omitted here for the sake of brevity.

Towards the converse of these inequality relations, we have the following theorem.

Theorem 2.5. *Let $f(x,n;s)$ and $g(x,n;s)$ be real valued functions, $\alpha,\beta,\lambda>0$, and $\mu,k\in\mathbb{N},\ r\in\mathbb{N}\cup\{0\}$, then*

$$x^{kn} > \frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s} \sum_{j=0}^{mn} \frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\ f(x,j;s); \tag{28}$$

implies

$$f(x,n;s) < B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;s,r) \tag{29}$$

and

$$g(x,n;s) > B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;s,r), \tag{30}$$

implies

$$x^{kn} < \frac{\Gamma(\alpha n+\beta+1)\,[(\lambda)_{\mu n}]^r\,n!}{(mn!)^s} \sum_{j=0}^{mn} \frac{[(-mn)_j]^s}{\Gamma(\alpha j+\beta+1)\ j!}\ g(x,j;s). \tag{31}$$

The proof runs parallel to that of Theorem 2.4, hence is omitted.

Now, for $s=1$, we obtain the inverse series relations for the polynomial (22) which is stated as

Theorem 2.6. *For $\alpha, \beta, \lambda > 0, m, \mu, k \in \mathbb{N}, r \in \mathbb{N} \cup \{0\}$,*

$$B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r) = \frac{\Gamma(\alpha j + \beta + 1)}{j!} \sum_{i=0}^{[j/m]} \frac{(-j)_{mi}\ x^{ki}}{\Gamma(\alpha i + \beta + 1)\,[(\lambda)_{\mu i}]^r\ i!} \tag{32}$$

if and only if

$$\frac{x^{kn}}{n!} = \frac{\Gamma(\alpha n + \beta + 1)\,[(\lambda)_{\mu n}]^r}{(mn)!} \sum_{j=0}^{mn} \frac{(-mn)_j}{\Gamma(\alpha j + \beta + 1)} B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r), \tag{33}$$

and for $n \neq ml,\ l \in \mathbb{N}$,

$$\sum_{j=0}^{n} \frac{(-n)_j}{\Gamma(\alpha j + \beta + 1)} B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r) = 0. \tag{34}$$

Proof. (32) implies (33): Let us denote the right hand side of (33) by Ω_n. Then, substitution for $B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r)$ in view of (32) gives

$$\Omega_n = \frac{\Gamma(\alpha n + \beta + 1)\,[(\lambda)_{\mu n}]^r}{(mn)!} \sum_{j=0}^{mn} \frac{(-mn)_j}{j!} \sum_{i=0}^{[j/m]} \frac{(-j)_{mi}\ x^{ki}}{\Gamma(\alpha i + \beta + 1)\,[(\lambda)_{\mu i}]^r\ i!}.$$

In view of the double series relation (27), this further takes the form

$$\begin{aligned}
\Omega_n &= \sum_{j=0}^{mn} \sum_{i=0}^{[j/m]} \frac{(-1)^{j+mi}\ \Gamma(\alpha n + \beta + 1)\ [(\lambda)_{\mu n}]^r\ x^{ki}}{(mn - j)!\ (j - mi)!\ \Gamma(\alpha i + \beta + 1)\ [(\lambda)_{\mu i}]^r\ i!} \\
&= \sum_{i=0}^{n} \sum_{j=0}^{mn-mi} \frac{(-1)^j\ \Gamma(\alpha n + \beta + 1)\ [(\lambda)_{\mu n}]^r}{(mn - mi - j)!\ j!\ \Gamma(\alpha i + \beta + 1)\ [(\lambda)_{\mu i}]^r\ i!}\ x^{ki} \\
&= \frac{x^{kn}}{n!} + \sum_{i=0}^{n-1} \frac{\Gamma(\alpha n + \beta + 1)\ [(\lambda)_{\mu n}]^r\ x^{ki}}{\Gamma(\alpha i + \beta + 1)\ [(\lambda)_{\mu i}]^r\ (mn - mi)!\ i!} \\
&\quad \times \sum_{j=0}^{mn-mi} (-1)^j \binom{mn - mi}{j}.
\end{aligned}$$

Here the inner sum in the second term on the right hand side vanishes and consequently we arrive at $\Omega_n = \frac{x^{kn}}{n!}$.

To show further that (32) also implies (34), observe that substituting right side expression of (32) for $B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r)$ in the left hand member of (34) and making use of the formula

$$\sum_{j=0}^{n} \sum_{i=0}^{[j/m]} A(i, j) = \sum_{i=0}^{[n/m]} \sum_{j=0}^{n-mi} A(i, j + mi), \qquad (n \neq ml)$$

we find that

$$\begin{aligned}
&\sum_{j=0}^{n} \frac{(-n)_j}{\Gamma(\alpha j + \beta + 1)}\ B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k; 1, r) \\
&= \sum_{j=0}^{n} \frac{(-1)^j\ n!}{(n - j)!} \sum_{i=0}^{[j/m]} \frac{(-1)^{mi}\ x^{ki}}{\Gamma(\alpha i + \beta + 1)\ ((\lambda)_{\mu i})^r\ (j - mi)!\ i!}
\end{aligned}$$

$$= \sum_{i=0}^{[n/m]} \frac{n!\ x^{ki}}{\Gamma(\alpha i+\beta+1)\ ((\lambda)_{\mu i})^r\ (n-mi)!\ i!} \sum_{j=0}^{n-mi} (-1)^j \binom{n-mi}{j} = 0$$

if $n \neq ml,\ l \in \mathbb{N}$, thus completing the first part. The proof of converse part runs as follows [2]. In order to show that the series (33) and the condition (34) together imply the series (32), we first note the simplest inverse series relations [18, Eq.(1), p.43]:

$$\omega_n = \sum_{j=0}^{n} \frac{(-n)_j}{j!}\ \rho_j \ \Leftrightarrow\ \rho_n = \sum_{j=0}^{n} \frac{(-n)_j}{j!}\ \omega_j.$$

Here putting

$$\rho_j = \frac{j!}{\Gamma(\alpha j+\beta+1)}\ B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r),$$

and considering one sided relation that is, the series on the left hand side implies the series on the right side, we get

$$\omega_n = \sum_{j=0}^{n} \frac{(-n)_j}{\Gamma(\alpha j+\beta+1)}\ B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r) \tag{35}$$

implies

$$B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r) = \frac{\Gamma(\alpha n+\beta+1)}{n!} \sum_{j=0}^{n} \frac{(-n)_j}{j!}\ \omega_j. \tag{36}$$

Since the condition (34) holds, $\omega_n = 0$ for $n \neq ml,\ l \in \mathbb{N}$, whereas

$$\omega_{mn} = \sum_{j=0}^{mn} \frac{(-mn)_j}{\Gamma(\alpha j+\beta+1)}\ B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r).$$

But since the series (33) also holds true,

$$\omega_{mn} = \frac{(mn)!\ x^{kn}}{n!\ \Gamma(\alpha n+\beta+1)\ ((\lambda)_{\mu n})^r}.$$

Consequently, the inverse pair (35) and (36) assume the form:

$$\frac{x^{kn}}{n!} = \frac{\Gamma(\alpha n+\beta+1)((\lambda)_{\mu n})^r}{(mn)!} \sum_{j=0}^{mn} \frac{(-mn)_j}{\Gamma(\alpha j+\beta+1)}\ B_{j^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r)$$

implies

$$\begin{aligned} B_{n^*}^{(\alpha,\beta,\lambda,\mu)}(x^k;1,r) &= \frac{\Gamma(\alpha n+\beta+1)}{n!} \sum_{j=0}^{[n/m]} \frac{(-n)_{mj}}{(mj)!}\ \omega_{mj} \\ &= \frac{\Gamma(\alpha n+\beta+1)}{n!} \sum_{j=0}^{[n/m]} \frac{(-n)_{mj}\ x^{kj}}{\Gamma(\alpha j+\beta+1)\ ((\lambda)_{\mu j})^r\ j!}, \end{aligned}$$

subject to the condition (34). □

References

[1] Annaby, M. H. and Mansour, Z. S., *q−Fractional Calculus and Equations*, Springer-Verlag, Berlin-Heidelberg, 2012.

[2] Dave, B. I. and Dalbhide, M., Gessel-Stanton's inverse series and a system of q-polynomials, *Bull. Sci. math.*, **138**, (2014), 323–334.

[3] Erdélyi, A., et al. (Eds.), *Higher Transcendental Functions* McGraw-Hill-New York, 1953.

[4] Gorenflo, R., Kilbas, A. A. and Rogosin, S. V., On the Generalized Mittag - Leffler type function, *Integral Transforms Spec. Funct.*, **7**(1998), 215–224.

[5] Gupta, I. S. and Debnath, L., Some properties of the Mittag-Leffler functions, *Integral Trans. Spec. Funct.*, **18**(2007), 329–336.

[6] Haubold, H. J., Mathai, A. M. and Saxena, R. K., *The H-Function: Theory and Applications*, Centre for Mathematical Sciences- Pala Campus, Kerala, India, 2008.

[7] Humbert, P. and Agarwal, R. P., Sur la fonction de Mittag-Leffler et quelques unes de ses generalizations, *Bulletin of Science and Mathematics Series II*, **77**(1953), 180–185.

[8] Kilbas, A. A., Saigo, M. and Saxena, R. K., Generalized Mittag-Leffler function and generalized fractional calculus operators, *Integral Transforms Spec. Funct.*, **15**(2004), 31–49.

[9] Konhauser, J. D. E., Biorthogonal polynomials suggested by the Laguerre polynomial, *Pacific J. Math.*, **21** (1967), No. 9, 303–314.

[10] Kiryakova, V. S., Multiple (multiindex) Mittag-Leffler functions and relations to generalized fractional calculus, *Journal of Computational and Applied Mathematics*, **118**(2000), 241–259.

[11] Luke, Y. L., *The Special Functions and their approximations*, Academic Press-New York, London, 1969.

[12] Mittag-Leffler, G. M., Sur la nouvelle fonction $E_\alpha(x)$, *C. R. Acad. Sci. Paris*, **137**(1903), 554–558.

[13] Mittag-Leffler G. M., Une generalisation de l'integrale de Laplace-Abel, *Comptes Rendus de l'Académie des Sciences Série II*, **137** (1903), 537–539.

[14] Nagai, A., *Discerete Mittag-Leffler function and its applications*, Deparment of Mathematical Sciences, Graduate School of Engineering Sciences, Osaka University **1302**(2003), 1–20.

[15] Prabhakar, T. R., A singular equation with a generalized Mittag-Leffler function in the kernel, *Yokohama Math. J.*, **19**(1971), 7–15.

[16] Prajapti, J. C., Dave, B. I. and Nathwani, B. V., On a unification of generalized Mittag-Leffler Function and family of Bessel Functions, *Advances in Pure Mathematics*, **3**(2013), 127–137.

[17] Prajapati, J. C. and Nathwani, B. V., Fractional calculus of a unified Mittag-Leffler function, *Ukrainian Mathematical Journal*, **66**(2015), 1267–1280.

[18] Riordan, J., *Combinatorial Identities*, Robert E. Krieger Publishing Co. Inc., New York, 1979.

[19] Saigo, M. and Kilbas, A. A., On Mittag leffler type function and applications, *Integral Transforms Spec. Funct.*, **7**(1998), 97–112.

[20] Saxena, R. K. and Kalla, S. L., Multivariate analogue of generatized Mittag-Leffer function, *Integral Transform. Spec. Funct.*, **22**(2011), 533–548.

[21] Saxena, R. K. and Nishimoto, K. N., fractional calculus of generalized Mittag-Leffler functions, *J. Fract Calc.*, **37**(2010), 43–52.

[22] Shukla, A. K. and Prajapati, J. C., On a generlization of Mittag-Leffler functions and its properties, *J. Math. Anal. Appl.*, **337**(2007), 797–811.

[23] Srivastava, H. M. and Manocha, H. L., *A Treatise on Generating Functions*, Ellis Horwood Ltd., Chichester, England, 1984.

[24] Wiman, A., *Über* die nullstellen der funktionen $E_\alpha(x)$, *Acta Math.*, **29**(1905), 217–234.

B. V. Nathwani
Department of Mathematics, Faculty of Science
The Maharaja Sayajirao University of Baroda
Vadodara-390 002, Gujarat, India
E-mail: *bharti.nathwani@yahoo.com*

B. I. Dave
Department of Mathematics, Faculty of Science
The Maharaja Sayajirao University of Baroda
Vadodara-390 002, Gujarat, India
E-mail: *bidavemsu@yahoo.co.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 77-86

REVISITING EISENTEIN-TYPE CRITERION OVER INTEGERS

AKASH JENA* AND BINOD KUMAR SAHOO

(Received : 14 - 12 - 2016, Revised : 13 - 03 - 2017)

Abstract. The following result, a consequence of Dumas criterion for irreducibility of polynomials over integers, is generally proved using the notion of Newton diagram:

> "Let $f(x)$ be a polynomial with integer coefficients and k be a positive integer relatively prime to the degree of $f(x)$. Suppose that there exists a prime number p such that the leading coefficient of $f(x)$ is not divisible by p, all the remaining coefficients are divisible by p^k, and the constant term of $f(x)$ is not divisible by p^{k+1}. Then $f(x)$ is irreducible over $\mathbb{Z}$".

For $k = 1$, this is precisely the well-known Eisenstein criterion. The aim of this article is to give an alternate proof, accessible to the undergraduate students, of this result for $k \in \{2, 3, 4\}$ using basic divisibility properties of integers.

1. Introduction

Let $\mathbb{Z}[x]$ be the ring of polynomials with coefficients from the ring $\mathbb{Z}$ of integers. A nonconstant polynomial $f(x) \in \mathbb{Z}[x]$ is said to be *reducible* over $\mathbb{Z}$ if it can be written as a product of two nonconstant polynomials in $\mathbb{Z}[x]$, otherwise, $f(x)$ is called *irreducible* over $\mathbb{Z}$. There is no universal criterion which can be applied to determine the reducibility/irreducibility of all the polynomials in $\mathbb{Z}[x]$. However, many criteria exist in the literature each of which give this information for some particular class of polynomials. One such criterion, the so called "Eisenstein criterion", is due to Gotthold Eisenstein (1823–1852), a German mathematician. This is perhaps the most well-known criterion which gives a sufficient condition for a polynomial in $\mathbb{Z}[x]$ to be irreducible.

> *Eisenstein criterion.* Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$ of positive degree. Suppose that there exists a prime number p such that the

* The first author is a final year student of the Integrated M.Sc. program in Mathematics. He is a recipient of INSPIRE student fellowship awarded by the Department of Science and Technology, Government of India.

2010 Mathematics Subject Classification: Primary: 11C08, 13M10

Keywords and Phrases: Irreducible polynomial, Eisenstein criterion, Dumas criterion, Newton diagram.

leading coefficient of $f(x)$ is not divisible by p, all the remaining coefficients are divisible by p, and the constant term is not divisible by p^2. Then $f(x)$ is irreducible over $\mathbb{Z}$.

[As mentioned in [4, p.49], one can reverse the roles of the constant term and the leading coefficient of $f(x)$ to get another version of the Eisenstein criterion. More precisely, *if the constant term of $f(x)$ is not divisible by p, all the remaining coefficients are divisible by p, and the leading coefficient of $f(x)$ is not divisible by p^2, then $f(x)$ is irreducible over $\mathbb{Z}$.*]

A polynomial satisfying the conditions of Eisenstein criterion for some prime is called an *Eisenstein polynomial*. In practice, it may happen that the original polynomial $f(x)$ is not Eisenstein for any prime, but the criterion is applicable (with respect to some prime) to the polynomial obtained after transforming $f(x)$ by some substitution for x. The fact that the polynomial after substitution is irreducible then allows to conclude that the original polynomial itself is irreducible.

To test for the irreducibility of a polynomial, Eisenstein criterion is a special case of the general technique of "reducing the coefficients modulo a prime". To illustrate this technique, let us consider the polynomial

$$f(x) = x^{p-1} + x^{p-2} + \cdots + x + 1 \in \mathbb{Z}[x],$$

where p is a given prime number. Recall that the map from $\mathbb{Z}[x]$ to $\mathbb{Z}_p[x]$ defined by $g(x) \mapsto \overline{g(x)}$ is a surjective ring homomorphism, where $\overline{g(x)}$ is the polynomial in $\mathbb{Z}_p[x]$ obtained from $g(x)$ by reducing each of the coefficients of $g(x)$ modulo p. If $h(x) = x^p - 1 \in \mathbb{Z}[x]$, then $h(x) = (x-1)f(x)$. Since $\overline{h(x)} = (x-1)^p$ in $\mathbb{Z}_p[x]$, we get

$$\overline{f(x)} = \frac{\overline{h(x)}}{x-1} = (x-1)^{p-1}.$$

Suppose that $f(x) = a(x)b(x)$, where $a(x), b(x)$ are polynomials in $\mathbb{Z}[x]$ of positive degree. Then $\overline{a(x)} = (x-1)^r$ and $\overline{b(x)} = (x-1)^s$ for some integers r, s, where $1 \leq r, s < p-1$ and $r + s = p - 1$. Putting $x = 1$, we see that

$$p = f(1) = a(1)b(1). \tag{1}$$

We have $a(1) \equiv \overline{a(1)} \mod p$, and $b(1) \equiv \overline{b(1)} \mod p$. Since $\overline{a(1)} = (1-1)^r = 0$ and $\overline{b(1)} = (1-1)^s = 0$, it follows that p divides both $a(1)$ and $b(1)$. Thus p^2 divides the right hand side of (1). As a consequence, p^2 divides p, leading to a contradiction. Hence $f(x)$ is irreducible. In the usual proof of the irreducibility of $f(x)$, we substitute x by $x+1$ and obtain that the polynomial $f(x+1)$ has constant term p and $f(x+1) \equiv x^{p-1} \mod p$, so that Eisenstein criterion can be applied to $f(x+1)$ with respect to the prime p.

We learn Eisenstein criterion generally at the undergraduate level as a part of our mathematics training. At that time, realizing its power and simplicity, students try to generalize the statement of the criterion and ask the following natural question:

> Suppose that there exists a prime number p and an integer $k \geq 2$ such that the leading coefficient of $f(x)$ is not divisible by p, all the remaining coefficients are divisible by p^k, and the constant term is not divisible by p^{k+1}. Is $f(x)$ necessarily irreducible over $\mathbb{Z}$?

The answer is certainly No!. For example, one can have the following factorizations:

$$x^2 - p^2 = (x-p)(x+p),\ \ x^3 - p^3 = (x-p)(x^2+px+p^2), \text{ etc.}$$

However, the answer could be affirmative if one adds an extra condition connecting k and the degree of $f(x)$, see Theorem 1.2 below.

1.1. **Dumas Criterion.** The second best known irreducibility criterion based on divisibility of the coefficients by a prime is probably the so called "Dumas Criterion", due to Gustave Dumas (1872–1955), a Swiss mathematician. To state this criterion, it is necessary to recall the notion of 'Newton diagram' of a polynomial over integers with respect to a given prime number.

Let p be a fixed prime number and let $f(x) \in \mathbb{Z}[x]$ be a polynomial of degree $n \geq 1$. We refer to [3, Section 2.2.1] or [2, Page 96] for the construction of the Newton diagram of $f(x)$ with respect to p. Let

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where the constant term a_0 is nonzero (otherwise, $f(x)$ would be reducible for $n \geq 2$). Every nonzero coefficient a_i of $f(x)$ can be written in the form

$$a_i = \bar{a}_i p^{\alpha_i},$$

where $\bar{a}_i$ is an integer not divisible by p, that is, α_i is the maximum power of p such that $p^{\alpha_i} \mid a_i$. Set

$$X = \{(i, \alpha_i) : a_i \neq 0\}.$$

Call the elements of X as vertices and plot them in the plane. Since $f(x)$ is of positive degree, there are at least two vertices: the *initial* vertex $(0, \alpha_0)$ and the *terminal* vertex (n, α_n). Note that there is no vertex corresponding to a zero coefficient of $f(x)$. The construction of the *Newton diagram of* $f(x)$ *with respect to* p is as follows.

Start with the initial vertex $v_0 = (0, \alpha_0)$. Then find the vertex $v_1 = (i_1, \alpha_{i_1})$, where $i_1 \neq 0$ is the largest integer for which there is no vertex of X below the line through v_0 and v_1. It may happen that the line segment v_0v_1 joining v_0 and v_1 contain vertices from X which are different from v_0 and v_1. Then find the vertex $v_2 = (i_2, \alpha_{i_2})$, where i_2 $(\neq i_1)$ is the largest integer for which there is no vertex below the line through v_1 and v_2. Proceed in this way to draw the line segments v_0v_1, v_1v_2 etc. one by one. The very last line segment is of the form $v_{k-1}v_k$, where $v_k = (n, \alpha_n)$ is the terminal vertex. Then the Newton diagram of $f(x)$ with respect to p consists of the line segments $v_{j-1}v_j$, $1 \leq j \leq k$. It has at least one line segment. We say that a line segment $v_{i-1}v_i$ is *simple* if v_{i-1} and v_i are the only points on it with integer coordinates.

We now state the irreducibility criterion by Dumas, a proof of which can be found in [3, Section 2.2]. The original proof by Dumas appeared in 1906 in the paper [1].

> *Dumas criterion.* Let $f(x) \in \mathbb{Z}[x]$ be a polynomial of positive degree. Suppose that there exists a prime p for which the Newton diagram of $f(x)$ consists of exactly one simple line segment. Then $f(x)$ is irreducible over $\mathbb{Z}$.

Observe that if p satisfies the three conditions of Eisenstein criterion, then the Newton diagram of $f(x)$ with respect to p consists of one simple line segment with end vertices $(0,1)$ and $(n,0)$ and so $f(x)$ is irreducible. Thus Dumas criterion can be considered as a generalization of Eisenstein criterion.

Example 1.1. *The Newton diagram of $f(x) = x^4 + 12$ with respect to $p = 2$ consists of one line segment through the initial vertex $(0,2)$ and the terminal vertex $(4,0)$. It contains the point $(2,1)$ with integer coordinates and so Dumas criterion can not be applied with respect to 2. However, $f(x)$ is Eisenstein for $p = 3$ and hence irreducible over $\mathbb{Z}$.*

Now let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$. Suppose that there exists a positive integer k and a prime number p such that

$$p \nmid a_n,\ p^k \mid a_j\ (0 \le j \le n-1) \text{ and } p^{k+1} \nmid a_0.$$

Then the Newton diagram of $f(x)$ with respect to p consists of exactly one line segment uv, where $u = (0,k)$ and $v = (n,0)$. The equation of the line through u and v is

$$kX + nY = nk.$$

If k and n are relatively prime, then it can be seen that there is no integer coordinate points on the line segment uv different from u and v. So uv is a simple line segment and hence $f(x)$ is irreducible by Dumas criterion. Thus, we have the following result which is related to the question mentioned before.

Theorem 1.2. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$ be of degree n, and k be a positive integer relatively prime to n. Suppose that there exists a prime p such that $p \nmid a_n$, $p^k \mid a_j$ for $0 \le j \le n-1$ and $p^{k+1} \nmid a_0$. Then $f(x)$ is irreducible over $\mathbb{Z}$.*

For $k = 1$, Theorem 1.2 is simply the Eisenstein criterion. The aim of this article is to give an elementary proof, which is accessible to the undergraduate students, of Theorem 1.2 for $k \in \{2,3,4\}$ using basic divisibility properties of integers. One can use similar argument for other small values of k, but more steps will be involved. For $k \ge 2$, it can be observed from the Newton diagram of $f(x)$ with respect to p that the condition $p^k \mid a_j$ for $0 \le j \le n-1$ is much stronger. It can further be relaxed for higher value of j. For example, for $k = 2$, this condition can be

replaced by that $p \mid a_j$ for $j \leq n-1$ and $p^2 \mid a_i$ for $0 \leq i \leq \lfloor n/2 \rfloor$ (see Theorem 2.2).

2. For $k = 2$

We start with the following lemma which essentially proves the Eisenstein criterion, but stated in a different way as per our requirement. This result is useful in all the cases of $k \in \{2, 3, 4\}$.

Lemma 2.1. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$. Suppose that there exists a prime p such that $p \nmid a_n$ and $p \mid a_i$ for $0 \leq i \leq n-1$. If $f(x) = g(x)h(x)$ for two nonconstant polynomials $g(x)$, $h(x)$ in $\mathbb{Z}[x]$, then p divides all the coefficients, except the leading ones, of $g(x)$ and $h(x)$. In particular, if $p^2 \nmid a_0$, then $f(x)$ is irreducible over $\mathbb{Z}$.*

Proof. Let $g(x) = b_k x^k + \cdots + b_1 x + b_0$ and $h(x) = c_l x^l + \cdots + c_1 x + c_0$, where $k, l \geq 1$. We first show that b_0 and c_0 are divisible by p. Since $a_n = b_k c_l$ and $p \nmid a_n$, we have $p \nmid b_k$ and $p \nmid c_l$. Since $a_0 = b_0 c_0$ and $p \mid a_0$, we have $p \mid b_0$ or $p \mid c_0$. We may assume that $p \mid b_0$. Let r, $1 \leq r \leq k$, be the smallest integer such that $p \nmid b_r$. Considering the coefficient a_r in $f(x)$, we have

$$a_r = b_r c_0 + b_{r-1}c_1 + \cdots + b_0 c_r.$$

Since $p \mid a_r$ and $p \mid b_i$ for $0 \leq i \leq r-1$, it follows that $p \mid b_r c_0$. Then $p \mid c_0$ as $p \nmid b_r$.

Now consider r as above and let s, $1 \leq s \leq l$, be the smallest integer such that $p \nmid c_s$. Considering the coefficient a_{r+s} in $f(x)$, we have

$$a_{r+s} = b_{r+s}c_0 + \cdots + b_{r+1}c_{s-1} + b_r c_s + b_{r-1}c_{s+1} + \cdots + b_0 c_{r+s}.$$

Note that $p \nmid b_r c_s$ and all the remaining terms in the above expression of a_{r+s} are divisible by p. So $p \nmid a_{r+s}$. Since $p \mid a_i$ for $0 \leq i \leq n-1$, we get $a_{r+s} = a_n = a_{k+l}$. So $r = k$ and $s = l$. □

We now prove the following result which is an improved version of Theorem 1.2 for $k = 2$.

Theorem 2.2. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$. Suppose that there exists a prime p such that $p \nmid a_n$, $p \mid a_i$ for $i \leq n-1$, $p^2 \mid a_j$ for $j \leq \lfloor n/2 \rfloor$ and $p^3 \nmid a_0$. Then the following hold:*

(1) If n is odd, then $f(x)$ is irreducible over $\mathbb{Z}$.

(2) If n is even, then either $f(x)$ is irreducible over $\mathbb{Z}$, or $f(x)$ is a product of exactly two irreducible polynomials in $\mathbb{Z}[x]$ of equal degree which are Eisenstein with respect to p.

Proof. Suppose that $f(x) = g(x)h(x)$ for some nonconstant polynomials $g(x), h(x)$ in $\mathbb{Z}[x]$, where

$$g(x) = b_r x^r + b_{r-1}x^{r-1} + \cdots + b_1 x + b_0,$$
$$h(x) = c_s x^s + c_{s-1}x^{s-1} + \cdots + c_1 x + c_0.$$

Since $p \nmid a_n$, we have $p \nmid b_r$ and $p \nmid c_s$. By Lemma 2.1, b_i and c_j are divisible by p for $0 \leq i \leq r-1$ and $0 \leq j \leq s-1$. Since $p^3 \nmid a_0$, we have $p^2 \nmid b_0$ and $p^2 \nmid c_0$. Thus $g(x)$ and $h(x)$ both are Eisenstein with respect to p and hence irreducible over $\mathbb{Z}$. In order to complete the proof, it is enough to show that $r = s$.

First suppose that $r > s$. We shall get a contradiction by showing that $p \mid c_s$. Considering the coefficient a_s in $f(x)$, we have

$$a_s = b_s c_0 + b_{s-1} c_1 + \cdots + b_1 c_{s-1} + b_0 c_s.$$

Since $r > s$, each term in the above expression of a_s, different from $b_0 c_s$, is divisible by p^2. Also, $\lfloor n/2 \rfloor = \lfloor (r+s)/2 \rfloor \geq s$ implies that $p^2 \mid a_s$. Then it follows that $p^2 \mid b_0 c_s$. Since $p^2 \nmid b_0$, we get $p \mid c_s$, a contradiction. If $s > r$, then similar argument holds to get a contradiction that $p \mid b_r$. □

We give examples below to show that both the possibilities in Theorem 2.2 may occur for even degree polynomials in $\mathbb{Z}[x]$.

Example 2.3. *(1) For any prime p, the polynomial $f(x) = x^2 + p^2 \in \mathbb{Z}[x]$ satisfies the conditions of Theorem 2.2 with respect to p. So it is irreducible over $\mathbb{Z}$.*

(2) The polynomial $f(x) = x^4 + 5x^3 + 25x^2 + 50x + 150$ satisfies the conditions of Theorem 2.2 with $p = 5$. But it is reducible over $\mathbb{Z}$, as we have the factorization: $f(x) = (x^2 + 10)(x^2 + 5x + 15)$.

3. For $k = 3$

The following elementary result is useful for us. We include a proof of it for the sake of completeness.

Lemma 3.1. *Let p be a prime and u, v be integers which are not divisible by p. If $p \mid xy$ and $p \mid (ux + vy)$ for some integers x and y, then $p \mid x$ and $p \mid y$.*

Proof. Since p is a prime and $p \mid xy$, we have $p \mid x$ or $p \mid y$. Assume that $p \mid x$. Then $p \mid (ux + vy)$ implies that $p \mid vy$. Then $p \mid y$ as $p \nmid v$. □

Theorem 3.2. *Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$, where n is not divisible by 3. Suppose that there exists a prime p such that the leading coefficient a_n is not divisible by p, the remaining coefficients are divisible by p^3 and the constant term a_0 is not divisible by p^4. Then $f(x)$ is irreducible over $\mathbb{Z}$.*

Proof. Suppose that $f(x)$ is reducible over $\mathbb{Z}$. Let $f(x) = g(x)h(x)$ for some nonconstant polynomials $g(x), h(x)$ in $\mathbb{Z}[x]$, where

$$g(x) = b_r x^r + b_{r-1} x^{r-1} + \cdots + b_1 x + b_0,$$
$$h(x) = c_s x^s + c_{s-1} x^{s-1} + \cdots + c_1 x + c_0.$$

Since $p \nmid a_n$, $p \nmid b_r$ and $p \nmid c_s$. By Lemma 2.1, $p \mid b_i$ for $0 \leq i \leq r-1$ and $p \mid c_j$ for $0 \leq j \leq s-1$. Since $p^3 \mid a_0$ and $p^4 \nmid a_0$, either $b_0 = up^2$ and $c_0 = vp$, or $b_0 = up$ and $c_0 = vp^2$ for some integers u, v which are not divisible by p. Without loss, we may assume that $b_0 = up^2$ and $c_0 = vp$.

Claim 3.2.1. *$r > s$.*

On the contrary, suppose that $s \geq r$. Considering the coefficient a_r in $f(x)$, we have

$$b_r vp + b_{r-1}c_1 + \cdots + b_1 c_{r-1} + c_r up^2 = a_r \equiv 0 \mod p^2.$$

Since b_i and c_i are divisible by p for $1 \leq i \leq r-1$, it follows that $p^2 \mid b_r vp$. Then $p \nmid v$ implies that $p \mid b_r$, a contradiction.

Claim 3.2.2. *$p^2 \mid b_l$ for $0 \leq l \leq s-1$.*

We shall prove by induction on l. This is clear for $l = 0$, since $b_0 = up^2$. So assume that $1 \leq l \leq s-1$ and that $p^2 \mid b_i$ for $0 \leq i \leq l-1$. The coefficient a_l in $f(x)$ is divisible by p^3 and so

$$b_l vp + b_{l-1}c_1 + \cdots + b_1 c_{l-1} + c_l up^2 \equiv 0 \mod p^3.$$

Using the induction hypothesis and the fact that $p \mid c_i$ for $1 \leq i \leq l$, it follows that $p^3 \mid b_l vp$. Then $p \nmid v$ implies that $p^2 \mid b_l$.

Claim 3.2.3. *$r \geq 2s$.*

Suppose that $r \leq 2s-1$. Since the coefficient a_r in $f(x)$ is divisible by p^2, we have

$$b_r vp + b_{r-1}c_1 + \cdots + b_{r-s+1}c_{s-1} + b_{r-s}c_s \equiv 0 \mod p^2.$$

Note that $r - s \leq s-1$ as $r \leq 2s-1$ by our assumption, and so $p^2 \mid b_{r-s}$ by Claim 3.2.2. Since $p \mid c_i$ for $1 \leq i \leq s-1$ and $p \mid b_j$ for $j \leq r-1$, it follows that $p^2 \mid b_r vp$. Then $p \nmid v$ implies that $p \mid b_r$, a contradiction.

Claim 3.2.4. *$r \geq 2s+1$.*

By Claim 3.2.3, we have $r \geq 2s$. Since $n = r + s$, the hypothesis that $3 \nmid n$ implies $r \neq 2s$. So $r \geq 2s+1$.

Now the coefficient $a_s = b_s vp + b_{s-1}c_1 + \cdots + b_1 c_{s-1} + c_s up^2$ of x^s in $f(x)$ is divisible by p^3. Using Claim 3.2.2, it follows that

$$\bar{b}_s v + c_s u \equiv 0 \mod p, \tag{2}$$

where $b_s = \bar{b}_s p$. Considering the coefficient a_{2s} of x^{2s} in $f(x)$ which is divisible by p^2, we have

$$b_{2s} vp + b_{2s-1}c_1 + \cdots + b_{s+1}c_{s-1} + b_s c_s \equiv 0 \mod p^2.$$

Note that $p \mid b_j$ for $j \leq 2s$ as $r \geq 2s+1$. It follows that $b_s c_s$ is divisible by p^2 and this gives

$$\bar{b}_s c_s \equiv 0 \mod p. \tag{3}$$

Then the congruence relations (2), (3) and Lemma 3.1 together imply that $p \mid c_s$, a final contradiction to our assumption that $f(x)$ is reducible. This completes the proof. □

4. For $k = 4$

Theorem 4.1. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$, where n and 4 are relatively prime. Suppose that there exists a prime p such that the leading*

coefficient a_n is not divisible by p, the remaining coefficients are divisible by p^4 and the constant term a_0 is not divisible by p^5. Then $f(x)$ is irreducible over $\mathbb{Z}$.

Proof. Suppose that $f(x)$ is reducible over $\mathbb{Z}$. Let $f(x) = g(x)h(x)$ for some nonconstant polynomials $g(x), h(x)$ in $\mathbb{Z}[x]$, where

$$g(x) = b_r x^r + b_{r-1}x^{r-1} + \cdots + b_1 x + b_0,$$
$$h(x) = c_s x^s + c_{s-1}x^{s-1} + \cdots + c_1 x + c_0.$$

Then $p \nmid b_r$ and $p \nmid c_s$, since $p \nmid a_n$. By Lemma 2.1, $p \mid b_i$ for $0 \leq i \leq r-1$ and $p \mid c_j$ for $0 \leq j \leq s-1$. Since $p^4 \mid a_0$ and $p^5 \nmid a_0$, we have the following cases for some integers u, v which are not divisible by p:

(1) either $b_0 = up^3$ and $c_0 = vp$, or $b_0 = up$ and $c_0 = vp^3$.
(2) $b_0 = up^2$ and $c_0 = vp^2$.

Case-(1). Without loss, we may assume that $b_0 = up^3$ and $c_0 = vp$. Applying the argument as in the proof of Claim 3.2.1, we get $r > s$. Then applying similar arguments as in the proof of Claims 3.2.2 and 3.2.3, we have the following facts:

$$p^3 \mid b_l \text{ for } 0 \leq l \leq s-1 \text{ and } r \geq 2s.$$

The coefficient $a_s = b_s vp + b_{s-1}c_1 + \cdots + b_1 c_{s-1} + c_s up^3$ in $f(x)$ is divisible by p^4. It follows that $p^2 \mid b_s$ and that

$$\bar{b}_s v + c_s u \equiv 0 \mod p, \tag{4}$$

where $b_s = \bar{b}_s p^2$. Since $p \nmid u$ and $p \nmid c_s$, (4) implies that $p \nmid \bar{b}_s$.

Claim 4.1.1. *$p^2 \mid b_{s+t}$ for $0 \leq t \leq s-1$.*

We prove this by induction on t. For $t = 0$, we have obtained above that $p^2 \mid b_s$. So assume that $1 \leq t \leq s-1$ and that $p^2 \mid b_{s+i}$ for $0 \leq i \leq t-1$. We have

$$a_{s+t} = b_{s+t}vp + b_{s+t-1}c_1 + \cdots + b_{t+1}c_{s-1} + b_t c_s.$$

Note that b_t is divisible by p^3 as $t \leq s-1$. Using the induction hypotheses, it follows that all the terms, different from the first one, in the above expression of a_{s+t} are divisible by p^3. Since $p^3 \mid a_{s+t}$, we get $p^3 \mid b_{s+t}vp$ and so $p^2 \mid b_{s+t}$ as $p \nmid v$.

Claim 4.1.2. *$r \geq 3s+1$.*

First suppose that $r \leq 3s-1$. Considering the coefficient a_r of x^r in $f(x)$ which is divisible by p^2, we have

$$b_r vp + b_{r-1}c_1 + \cdots + b_{r-s+1}c_{s-1} + b_{r-s}c_s \equiv 0 \mod p^2.$$

Note that $r - s \leq 2s-1$ as $r \leq 3s-1$ by our assumption, and so $p^2 \mid b_{r-s}$ by Claim 4.1.1. Since $p \mid c_i$ for $1 \leq i \leq s-1$ and $p \mid b_j$ for $j \leq r-1$, it follows that $p^2 \mid b_r vp$. Then $p \nmid v$ implies $p \mid b_r$, a contradiction. Thus $r \geq 3s$. Since $n = r+s$, and 4 and n are relatively prime, we get $r \geq 3s+1$.

The coefficient $a_{2s} = b_{2s}vp + b_{2s-1}c_1 + \cdots + b_{s+1}c_{s-1} + b_s c_s$ of x^{2s} in $f(x)$ is divisible by p^3. Using Claim 4.1.1, it follows that

$$\bar{b}_{2s} v + \bar{b}_s c_s \equiv 0 \mod p, \tag{5}$$

where $b_{2s} = \bar{b}_{2s}p$ and $\bar{b}_s$ is as before. Considering the coefficient a_{3s} in $f(x)$ which is divisible by p^2, we have

$$b_{3s}vp + b_{3s-1}c_1 + \cdots + b_{2s+1}c_{s-1} + b_{2s}c_s \equiv 0 \mod p^2.$$

Note that $p \mid b_j$ for $j \leq 3s$ as $r \geq 3s+1$. It follows that $b_{2s}c_s$ is divisible by p^2 and this gives

$$\bar{b}_{2s}c_s \equiv 0 \mod p. \tag{6}$$

Since $p \nmid v$ and $p \nmid \bar{b}_s$, Lemma 3.1 applying to the congruence relations (5) and (6) gives $p \mid c_s$, a contradiction. This completes the proof of Case-(1).

Case-(2). Here $b_0 = up^2$ and $c_0 = vp^2$. Without loss, we may assume that $r \geq s$. Since 4 and n are relatively prime, we must have $r > s$.

Claim 4.1.3. *b_l and c_l are divisible by p^2 for $0 \leq l \leq \lfloor (s-1)/2 \rfloor$.*

We shall prove by induction on l. This is clear for $l = 0$, since $b_0 = up^2$ and $c_0 = vp^2$. So assume that $1 \leq l \leq \lfloor (s-1)/2 \rfloor$ and that b_i, c_i are divisible by p^2 for $0 \leq i \leq l-1$. The coefficient a_l in $f(x)$ is divisible by p^4 and so

$$b_lvp^2 + b_{l-1}c_1 + \cdots + b_1c_{l-1} + c_lup^2 \equiv 0 \mod p^4.$$

Using the induction hypothesis, we get

$$\bar{b}_lv + \bar{c}_lu \equiv 0 \mod p, \tag{7}$$

where $b_l = \bar{b}_lp$ and $c_l = \bar{c}_lp$. For the coefficient a_{2l} in $f(x)$, we have

$$a_{2l} = b_{2l}vp^2 + b_{2l-1}c_1 + \cdots + b_{l+1}c_{l-1} + b_lc_l + b_{l-1}c_{l+1} + \cdots + b_1c_{2l-1} + c_{2l}up^2.$$

Since $a_{2l} \equiv 0 \mod p^3$, again using the induction hypothesis, it follows that

$$\bar{b}_l\bar{c}_l \equiv 0 \mod p. \tag{8}$$

Then (7), (8) and Lemma 3.1 together imply that $p \mid \bar{b}_l$ and $p \mid \bar{c}_l$ and so the claim follows.

If s is odd, say $s = 2k+1$ for some k, then $\lfloor (s-1)/2 \rfloor = k$. Considering the coefficient $a_s = a_{2k+1}$ in $f(x)$, we have

$$b_{2k+1}vp^2 + b_{2k}c_1 + \cdots + b_{k+1}c_k + b_kc_{k+1} + \cdots + b_1c_{2k} + c_{2k+1}up^2 \equiv 0 \mod p^3.$$

This gives $p \mid c_{2k+1}$, that is, $p \mid c_s$, a contradiction.

If s is even, say $s = 2k$ for some k, then $\lfloor (s-1)/2 \rfloor = k-1$. For the coefficient $a_s = a_{2k}$ in $f(x)$, we have

$$a_{2k} = b_{2k}vp^2 + b_{2k-1}c_1 + \cdots + b_{k+1}c_{k-1} + b_kc_k + b_{k-1}c_{k+1} + \cdots + b_1c_{2k-1} + c_{2k}up^2.$$

Since $r \geq s+1 = 2k+1$ and $a_s = a_{2k} \equiv 0 \mod p^3$, it follows that

$$\bar{b}_k\bar{c}_k + c_{2k}u \equiv 0 \mod p, \tag{9}$$

where $b_k = \bar{b}_kp$ and $c_k = \bar{c}_kp$. Now, for the coefficient a_{3k} in $f(x)$, we have

$$a_{3k} = b_{3k}vp^2 + b_{3k-1}c_1 + \cdots + b_{2k+1}c_{k-1} + b_{2k}c_k + \cdots + b_{k+1}c_{2k-1} + b_kc_{2k}.$$

Each term, different from the last one, in the above expression is divisible by p^2. Since $a_{3k} \equiv 0 \mod p^2$, we get $p^2 \mid b_kc_{2k}$ and so

$$\bar{b}_kc_{2k} \equiv 0 \mod p. \tag{10}$$

Then, as $p \nmid u$, the congruence relations (9) and (10) give $p \mid c_{2k}$, that is, $p \mid c_s$, a contradiction. This completes the proof. □

References

[1] Dumas, G., Sur quelques cas d'irréductibilité des polynomes à coefficients rationnels, *J. Math. Pure Appl.* **2** (1906), 191–258.

[2] Oleinikov, V. A., *Irreducibility and irrationality*, [Kvant 1986, no. 11, 12–16]. Kvant selecta: algebra and analysis, II, 95–103, Math. World, **15**, Amer. Math. Soc., Providence, RI, 1999.

[3] Prasolov, V. V., *Polynomials*, Translated from the 2001 Russian second edition by Dimitry Leites, Algorithms and Computation in Mathematics, 11 Springer-Verlag, Berlin, 2010.

[4] Sury, B., Polynomials with integer values, *Resonance*, **6** (9) (2001), 46–60.

Akash Jena
School of Mathematical Sciences
National Institute of Science Education and Research, Bhubaneswar (HBNI)
P.O.- Jatni, Dist- Khurda, Odisha - 752050, India
E-mail: *akash.jena@niser.ac.in*

Binod Kumar Sahoo
School of Mathematical Sciences
National Institute of Science Education and Research, Bhubaneswar (HBNI)
P.O.- Jatni, Dist- Khurda, Odisha - 752050, India
E-mail: *bksahoo@niser.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 87-101

GENERALIZED LAGURRE POLYNOMIALS WITH APPLICATIONS

SARANYA G. NAIR AND T. N. SHOREY

(Received : 30 - 01 - 2017, Revised : 10 - 05 - 2017)

Abstract. We give an account of important features of Generalised Laguerre Polynomials and their applications in several directions. Further we give a survey of algebraic properties including irreducibility of these polynomials.

1. Introduction

For a complex number z and integer $\nu > 0$, let

$$\binom{z}{\nu} = \frac{z(z-1)\cdots(z-\nu+1)}{\nu!}$$

and we put $\binom{z}{0} = 1$. If $z > 0$ is an integer, we observe that $\binom{z}{\nu} = 0$ whenever $\nu > z$. If $f(x)$ is a polynomial of degree m, we denote by $C(f(x), r)$ with $0 \le r \le m$ the coefficient of x^r in $f(x)$. Thus $C(f(x), m)$ is the leading coefficient of $f(x)$ and $C(f(x), 0)$ is the constant term of $f(x)$. For real number α and integer $n \ge 1$, the Generalised Laguerre Polynomial GLP is defined by

$$L_n^{(\alpha)}(x) = \sum_{j=0}^{n} \binom{n+\alpha}{n-j} \frac{(-x)^j}{j!}. \tag{1.1}$$

It is a polynomial with real coefficients of degree n such that

$$C(L_n^{(\alpha)}(x), n) = \frac{(-1)^n}{n!},\ C(L_n^{(\alpha)}(x), 0) = \binom{n+\alpha}{\alpha}.$$

It is also called associated Laguerre Polynomial or Sonine Polynomial after the name of its discoverer Nikolay Yakovlevich Sonin. The GLP with $\alpha = 0$ is called Laguerre Polynomial after its inventor Edmond Laguerre (1834-1886) and we denote $L_n^{(0)}(x)$ by $L_n(x)$. These polynomials were discovered around 1880. They satisfy second order linear differential equation

$$xy'' + (\alpha + 1 - x)y' + ny = 0 \tag{1.2}$$

with $y = L_n^{(\alpha)}(x)$. The left hand side of (1.2) is a polynomial of degree n and we denote it by $g(x)$. Then

2010 Mathematics Subject Classification: 11C08 (11A41 11B25 11N05).
Keywords and Phrases: Lagurre polynomials, Orthogonal polynomials, Irreducibility, Galois Group, Arithmetic Progression, Primes.

$$C(g(x),n) = -C(y^{'},n-1) + nC(y,n) = -n\frac{(-1)^n}{n!} + n\frac{(-1)^n}{n!} = 0.$$

In fact

$$C(g(x),r) = 0 \text{ for } 0 \leq r \leq n$$

and hence (1.2) follows. Further they satisfy the difference equation

$$L_n^{(\alpha)}(x) - L_n^{(\alpha-1)}(x) = L_{n-1}^{(\alpha)}(x)$$

and recurrence relation

$$L_{n+1}^{(\alpha)}(x) = \frac{(2n+1+\alpha-x)L_n^{(\alpha)}(x) - (n+\alpha)L_{n-1}^{(\alpha)}(x)}{n+1}$$

for $n \geq 1$ where

$$L_0^{(\alpha)}(x) = 1, L_1^{(\alpha)}(x) = -x+1+\alpha.$$

Further $\frac{1}{(1-t)^{\alpha+1}}e^{-\frac{tx}{1-t}}$ is a generating function for $L_n^{(\alpha)}(x)$ with $n \geq 1$. This means

$$\sum_{n=0}^{\infty} L_n^{(\alpha)}(x)t^n = \frac{1}{(1-t)^{\alpha+1}}e^{-\frac{tx}{1-t}}. \tag{1.3}$$

The right hand side of (1.3) is equal to

$$\frac{1}{(1-t)^{\alpha+1}}\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\frac{(tx)^j}{(1-t)^j} = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\frac{(tx)^j}{(1-t)^{j+\alpha+1}}$$

$$= \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}(tx)^j\sum_{m=0}^{\infty}\frac{(j+\alpha+1)\cdots(j+\alpha+m)}{m!}t^m$$

by $|t| < 1$ which we may suppose. Thus the right hand side of (1.3) is equal to

$$\sum_{j=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-x)^j}{j!}\frac{(j+\alpha+1)\cdots(j+\alpha+m)}{m!}t^{j+m}$$

$$= \sum_{n=0}^{\infty}t^n\sum_{j=0}^{n}\frac{(n+\alpha)\cdots(j+1+\alpha)}{(n-j)!}\frac{(-x)^j}{j!} = \sum_{n=0}^{\infty}L_n^{(\alpha)}(x)t^n.$$

2. Rodrigues formula

We show that GLP satisfies

$$L_n^{(\alpha)}(x) = \frac{x^{-\alpha}e^x}{n!}\frac{d^n}{dx^n}x^{n+\alpha}e^{-x}. \tag{2.1}$$

By Leibniz theorem, the right hand side of (2.1) is equal to

$$\frac{x^{-\alpha}e^x}{n!}\sum_{j=0}^{n}\binom{n}{j}(n+\alpha)(n+\alpha-1)\cdots(n+\alpha-j+1)x^{n+\alpha-j}(-1)^{n-j}e^{-x} \tag{2.2}$$

$$= \frac{1}{n!}\sum_{j=0}^{n}\binom{n}{j}(n+\alpha)(n+\alpha-1)\cdots(j+1+\alpha)(-x)^j \tag{2.3}$$

by writing j for $n-j$ in (2.2). Now we conclude (2.1) since each term in the preceeding sum (2.3) satisfies

$$\frac{1}{n!}\binom{n}{j}(n+\alpha)(n+\alpha-1)\cdots(j+1+\alpha)$$
$$=\frac{(n+\alpha)(n+\alpha-1)\cdots(j+1+\alpha)}{j!(n-j)!}=\frac{1}{j!}\binom{n+\alpha}{n-j}.$$

Since $e^x\frac{d^n}{dx^n}x^{n+\alpha}e^{-x}=\left(\frac{d}{dx}-1\right)^n x^{n+\alpha}$, we derive from (2.1) that

$$L_n^{(\alpha)}(x)=\frac{x^{-\alpha}}{n!}\left(\frac{d}{dx}-1\right)^n x^{n+\alpha}. \tag{2.4}$$

By combining (2.1) and (2.4), we have

$$L_n^{(\alpha)}(x)=\frac{x^{-\alpha}e^x}{n!}\frac{d^n}{dx^n}x^{n+\alpha}e^{-x}=\frac{x^{-\alpha}}{n!}\left(\frac{d}{dx}-1\right)^n x^{n+\alpha} \tag{2.5}$$

known as Rodrigues formula for $L_n^{(\alpha)}(x)$.

By putting $\alpha=0$ in (2.5), we write Rodrigues formula for Laguerre Polynomial

$$L_n(x)=\frac{e^x}{n!}\frac{d^n}{dx^n}x^ne^{-x}=\frac{1}{n!}\left(\frac{d}{dx}-1\right)^n x^n.$$

Thus

$$\frac{d}{dx}L_n(x)=\frac{1}{n!}\frac{d}{dx}\left(\frac{d}{dx}-1\right)^n x^n$$
$$=\frac{1}{n!}\left(\frac{d}{dx}-1\right)^n\frac{d}{dx}x^n$$
$$=\frac{1}{(n-1)!}\left(\frac{d}{dx}-1\right)\left(\frac{d}{dx}-1\right)^{n-1}x^{n-1}=\left(\frac{d}{dx}-1\right)L_{n-1}(x).$$

By repeated application of the above formula, we get

$$L_{n-1}(x)=-\sum_{i\geq 1}\left(\frac{d}{dx}\right)^i L_n(x).$$

Therefore $\mathcal{L}_n(x)=n!L_n(x)$ with $n\geq 1$ is a sequence of Scheffer Polynomials. This is also true similarly for $\mathcal{L}_n^{(\alpha)}(x)=n!L_n^{(\alpha)}(x)$. Scheffer Polynomials serve as a mathematical tool to unify and structure certain kinds of problems like recursions and expansions in Statistical Sciences.

3. Orthogonality

Let $w(x)$ be a non-negative integrable function on an interval $[a,b]$, which may be infinte, such that

$$\int_a^b w(x)dx>0.$$

We call $w(x)$ a weight function. Let $p_n(x)$ with $n\geq 1$ be a sequence of polynomials with real coefficients such that the degree of p_n is equal to n. Then they are called orthogonal polynomials on $[a,b]$ with respect to weight function $w(x)$ if

$$\int_a^b w(x)p_m(x)p_n(x)dx=\begin{cases}>0 & \text{if } m=n\\ 0 & \text{otherwise.}\end{cases}$$

The Generalised Laguerre Polynomials $L_n^{(\alpha)}(x)$ with real $\alpha \geq 0$ are orthogonal on $[0, \infty)$ with respect to weight function $x^\alpha e^{-x}$. More precisely

$$\int_0^\infty x^\alpha e^{-x} L_m^{(\alpha)}(x) L_n^{(\alpha)}(x) dx = \frac{\Gamma(n+\alpha+1)}{n!} \delta_{m,n} \tag{3.1}$$

where $\Gamma(u) = \int_0^\infty e^{-t} t^{u-1} dt > 0$ for $u > 0$ and $\delta_{m,n} = 1$ if $m = n$ and 0 otherwise. If $\alpha = 0$, we re-write (3.1) as

$$\int_0^\infty e^{-x} L_m(x) L_n(x) dx = \delta_{m,n} \tag{3.2}$$

since $\Gamma(n+1) = n!$.

Now we give a proof of (3.2). We may assume $m \leq n$. For non- negative $m \leq n$, we consider the integral $\int_0^\infty e^{-x} x^m L_n(x) dx$. Integrating by parts, we derive from (2.1) with $\alpha = 0$ that

$$\int_0^\infty e^{-x} x^m L_n(x) dx = \frac{(-1)^m m!}{n!} \int_0^\infty \frac{d^{n-m}}{dx^{n-m}} (x^n e^{-x}) dx.$$

But the right hand side is equal to 0 whenever $m < n$. Since the degree of $L_m(x)$ is m, we have

$$\int_0^\infty e^{-x} L_m(x) L_n(x) dx = 0 \text{ if } m < n.$$

Then

$$\int_0^\infty e^{-x} L_m(x) L_n(x) dx = \frac{(-1)^n}{n!} \int_0^\infty e^{-x} x^n L_n(x) dx$$

since $C(L_n(x), n) = \frac{(-1)^n}{n!}$. Now we derive from (2.1) as above that the integral on the right hand side is equal to

$$\frac{(-1)^n n!}{n!} \int_0^\infty x^n e^{-x} dx = (-1)^n n!.$$

Hence

$$\int_0^\infty e^{-x} L_m(x) L_n(x) dx = \frac{(-1)^n}{n!} (-1)^n n! = (-1)^{2n} = 1.$$

This completes the proof of (3.2).

4. Applications

Hermite Polynomials are given by

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = (2x - \frac{d}{dx})^n .1.$$

These polynomials are defined by Laplace in 1810. They were studied by Chebyshev in 1859 and Hermite in 1864 but they are named after Hermite. They arise in Probability(Edgeworth series), Combinatorics (Appell sequence obeying umbral calculus), Numerical analysis (Gaussian quadrature), Physics (Quantum harmonic

oscillator) and Systems theory (Gaussian noise). In fact Hermite polynomials, apart from a constant factor, are particular cases of GLP as follows:

$$H_{2n}(x) = (-1)^n 2^{2n} n! L_n^{(-\frac{1}{2})}(x^2) \text{ and } H_{2n+1}(x) = (-1)^n 2^{2n+1} n! x L_n^{(\frac{1}{2})}(x^2).$$

GLP are used in the radial part of the solution of Schrodinger equation for a one-electron atom in the static Wigner functions of oscillator systems in quantum mechanics in phase space and in the quantum mechanics of the Morse potential and 3D isotropic harmonic oscillator. GLP are used for approximating the values of the integrals of the form $\int_0^\infty e^{-x} f(x)dx$ where $f(x)$ is continuous or more generally smooth function. More precisely, it is shown by Gauss-Laguerre quadrature method that

$$\int_0^\infty e^{-x} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

where x_i is the ith root of Laguerre Polynomial $L_n(x)$ and w_i is given by

$$w_i = \frac{x_i}{(n+1)^2 (L_{n+1}(x_i))^2} \text{ for } 1 \leq i \leq n.$$

Gauss-Laguerre quadrature method is an extension of Gauss quadrature method.

Let $\mathcal{B}$ be a $m \times n$ rectangular board consisting of m rows and n columns and it looks like a chess board. If $m = n = 8$, then $\mathcal{B}$ is the ordinary chess board. Let B be a subset of $\mathcal{B}$. Assume that we can place a rook in each square of B. Two rooks are called non-attacking if they are neither in the same row nor in the same column. We assume that every pair of rooks is non- attacking. If a rook is placed at the intersection of mth row and nth column, we say that rook is at (m, n). The rook polynomial is defined as

$$R_B(x) = \sum_{k=0}^{\infty} r_k(B) x^k,$$

where $r_k(B)$ is the number of ways k rooks can be arranged in the squares of B. We observe that $r_k(B) = 0$ for $k > \min(m, n)$ since if a rook is placed at (m_1, n_1), then there is no other rook in the m_1th row as well as in n_1th row. Thus

$$R_B(x) = \sum_{k=0}^{\min(m,n)} r_k(B) x^k.$$

Rook polynomials were introduced by Kaplansky and Riordan and developed by Riordan [24]. Rook polynomials are used in pure and applied combinatorics, group theory, number theory and statistical physics. If $\mathcal{B} = B$, we write $R_{m,n}(x)$ for $R_B(x)$ and $R_n(x)$ for $R_{m,n}(x)$ if $m = n$. Thus

$$R_n(x) = \sum_{k=0}^{n} r_k(n) x^k.$$

Let $n = 3, k = 2$ and we calculate $r_2(3)$. If a rook is at $(1, 1)$, then there is no other rook in the first row and also in the first column. Thus there are four

possibilities for the second rook, namely $(2,2),(2,3),(3,2),(3,3)$. Similarly there are 4 possibilities when the rook is at any other place. Therefore there are $9\times4=36$ possibilities. We observe that possibility $(2,2)$ is also counted when the rook is at $(1,3)$. In fact every possibility is taken twice in the above counting. Hence $r_2(3)=\frac{36}{2}=18$. We calculate

$R_1(x)=x+1,\quad R_2(x)=2x^2+4x+1,\quad R_3(x)=6x^3+18x^2+9x+1$ and $R_4(x)=24x^4+96x^3+72x^2+16x+1$.

By (1.1), we calculate $L_4(x)=(1/24)\left(x^4-16x^3+72x^2-96x+24\right)$. Then we observe that $4!x^4L_4(-x^{-1})=R_4(x)$. More generally, we have

$$R_n(x)=n!x^nL_n(-x^{-1}) \tag{4.1}$$

and

$$R_{m,n}(x)=n!x^nL^{(m-n)}(-x^{-1}) \text{ for } m\geq n. \tag{4.2}$$

Thus, Rook polynomials upto a constant factor, can be obtained from GLP by elementary changes of variables. The zeros of GLP are positive and simple. Therefore we see from (4.2) that the zeros of $R_{m,n}(x)$ with $m\geq n$ are negative and simple. For an account of GLP, we refer to [29], [31], [32] and [33]. The Tricomi-Carlitz polynomials are given by

$$l_n(x)=(-1)^nL_n(x-n)$$

and they are related to random walks on positive integers.

5. Algebraic properties of GLP

Schur [26],[27] was the first to study algebraic properties of these polynomials. He gave a formula for the discrimnant of these polynomials. He studied whether these polynomials are irreducible. Further he determined their Galois group when they are irreducible. Let $f(x)$ be a polynomial with rational coefficient and deg $f=n$. By irreducibility of a polynomial, we shall always mean its irreducibility over rationals. We observe that if f has a factor of degree $k<n$, then it has a factor of degree $n-k$. Therefore given a polynomial of degree n, we always consider factors of degree k where $1\leq k\leq\frac{n}{2}$. For having a familiarity with GLP, we consider some particular cases of $L_n^{(\alpha)}(x)$ by restricting α. We shall always restrict α to rational numbers in this section. Every $\alpha\in\mathbb{Q}$ with denominator $d\geq1$, written in the reduced form, can be uniquely written as

$$\alpha=\alpha(u)=u+(a/d) \tag{5.1}$$

where $u,a\in\mathbb{Z}$ with $a=0$ if $d=1$ and $1\leq a<d$, $\gcd(a,d)=1$ if $d>1$. Thus $\alpha=u$ if $d=1$.

5.1. Well-known examples of $L_n^{(\alpha)}(x)$.

(i) Let $\alpha=-n-1$. Then

$$L_n^{(\alpha)}(x) = \sum_{j=0}^{n} \frac{(-1)(-2)\cdots(-(n-j))}{(n-j)!j!}(-x)^j = (-1)^n \sum_{j=0}^{n} \frac{x^j}{j!}$$

which is, upto sign, truncated exponential polynomial.

(ii) Let $\alpha = a$ where $a \geq 0$ is an integer. Then

$$L_n^{(a)}(x) = \sum_{j=0}^{n} \frac{(n+a)(n-1+a)\cdots(j+1+a)}{(n-j)!j!}(-x)^j$$

and

$$n!L_n^{(a)}(x) = \sum_{j=0}^{n}(-1)^j \binom{n}{j} \frac{(n+a)!}{(j+a)!}x^j = (n+a)!\sum_{j=0}^{n}(-1)^j \binom{n}{j} \frac{x^j}{(j+a)!}.$$

We put $a_j = (-1)^j \binom{n}{j}$. We observe that $a_0, a_1, \ldots, a_n$ are integers such that $|a_0| = |a_n|$=1. Thus

$$L_n^{(a)}(x) = \frac{(n+a)!}{n!} \sum_{j=0}^{n} a_j \frac{x^j}{(j+a)!}.$$

The irreducibility of $L_n^{(a)}(x)$ is equivalent to the irreducibility of $\sum_{j=0}^{n} a_j \frac{x^j}{(j+a)!}$ with $a_j = (-1)^j \binom{n}{j}$. In fact we shall consider irreducibility of more general polynomials $\sum_{j=0}^{n} a_j \frac{x^j}{(j+a)!}$ where $a_0, a_1, \ldots, a_n \in \mathbb{Z}$ with $|a_0| = |a_n| = 1$. These are called Generalised Schur Polynomials.

(iii) Let $\alpha = -2n - 1$. Then

$$\begin{aligned} L_n^{(-2n-1)}(x) &= \sum_{j=0}^{n} \frac{(n-2n-1)\cdots(j+1-2n-1)}{(n-j)!j!}(-x)^j \\ &= (-1)^n \sum_{j=0}^{n} \frac{(n+1)\cdots(2n-j)}{(n-j)!j!}x^j \\ &= (-1)^n \sum_{j=0}^{n} \frac{(n+1)\cdots(n+j)}{(n-j)!j!}x^{n-j} = \frac{(-1)^n}{n!} \sum_{j=0}^{n} \frac{(n+j)!}{(n-j)!j!}x^{n-j}. \end{aligned}$$

Bessel polynomials of degree n are given by

$$y_n(x) = \sum_{j=0}^{n} \frac{(n+j)!}{2^j(n-j)!j!}x^j.$$

We have

$$z_n(x) := x^n y_n\left(\frac{2}{x}\right) = \sum_{j=0}^{n} \frac{(n+j)!}{(n-j)!j!}x^{n-j} = (-1)^n n! L_n^{(-2n-1)}(x).$$

Thus Bessel polyomials $y_n(x)$ are irreducible if and only if $L_n^{(-2n-1)}(x)$ are irreducible. This connection between Bessel polynomials and GLP is due to Hajir [14].

(iv) Let $\alpha = -a$ be a negative integer. Then $a \geq 1$. Further the constant term of $L_n^{(\alpha)}(x)$ is equal to

$$\frac{(n+\alpha)\cdots(1+\alpha)}{n!} = \frac{(n-a)\cdots(1-a)}{n!}$$

We observe that $n-a \geq 0$ if $n \geq a = |\alpha|$ and $1-a \leq 0$. Therefore $L_n^{(\alpha)}(x)$ with $n \geq |\alpha|$ is reducible.

5.2. **Three remarkable results on GLP.**

(i) Combining the earlier work of Schur and Gow[11], Filaseta and Trifonov [7] proved that $L_n^{(-2n-1)}(x)$ is irreducible. Therefore, by example (iii), Bessel polynomials are irreducible. Then their Galois group is S_n by a result of Grosswald [12],[13].

(ii) Filaseta, Kidd and Trifonov [6] proved that for every $n > 2$, there exists α such that $L_n^{(\alpha)}(x)$ is irreducible and the Galois group of $L_n^{(\alpha)}(x)$ is the alternating group A_n. In fact, we can take

$$\alpha = \begin{cases} 1 & \text{if } n \equiv 1 \pmod 2 \\ -n-1 & \text{if } n \equiv 0 \pmod 4 \\ n & \text{if } n \equiv 2 \pmod 4 \end{cases}$$

For $n = 2$, we can also take $\alpha = n$. In this case $L_2^{(2)}(x) = \frac{1}{2}(x-6)(x-2)$ and its Galois group is $A_2 = \{e\}$. The cases $n \equiv 0 \pmod 4$ and $n \equiv 1 \pmod 4$ were already settled by Schur [26], [27]. These results settled the inverse Galois problem for A_n explicitly that for every positive integer $n > 1$, there exists an explicit Laguerre polynomial of degree n whose Galois group is the alternating group A_n. This remains open for an arbitrary group.

(iii) Filaseta and Lam [5] proved that for a fixed rational number α which is not a negative integer, $L_n^{(\alpha)}(x)$ is irreducible for all but finitely many n. As already mentioned in example (iv), the assumption that α is not a negative integer is necessary.

5.3. **Irreducibility of GLP and its extensions.** Let $d = 1$ i.e $\alpha \in \mathbb{Z}$. Then Laishram and Shorey [16] proved that for integers α with $0 \leq \alpha \leq 50$, $L_n^{(\alpha)}(x)$ is irreducible for all n except for $n = 2, \alpha \in \{2, 7, 14, 23, 34, 47\}$ and $n = 4, \alpha \in \{5, 23\}$ where it has a linear factor. In fact

$$L_2^{(2)}(x) = (1\ 2)(x-2)(x-6), \qquad L_2^{(7)}(x) = (1\ 2)(x-6)(x-12),$$
$$L_2^{(14)}(x) = (1\ 2)(x-12)(x-20), \qquad L_2^{(23)}(x) = (1\ 2)(x-20)(x-30),$$
$$L_2^{(34)}(x) = (1\ 2)(x-30)(x-42), \qquad L_2^{(47)}(x) = (1\ 2)(x-42)(x-56),$$
$$L_4^{(5)}(x) = (1\ 24)(x-6)(x^3 - 30x^2 + 252x - 504),$$
$$L_4^{(23)}(x) = (1\ 24)(x-30)(x^3 - 78x^2 + 1872x - 14040).$$

The cases $0 \leq \alpha \leq 1$ and $2 \leq \alpha \leq 10$ of the above result on GLP were already settled by Schur [26] and Filaseta, Finch and Leidy [8], respectively. We observe from (4.2) that Rook polynomial $R_{m,n}(x)$ with $m \geq n$ is irreducible if and only if $L_n^{(m-n)}(x)$ is irreducible. Therefore, we derive that Rook polynomials $R_{m,n}(x)$ with $0 \leq m-n \leq 50$ are irreducible except when $(m,n) \in \{(4,2),(9,2),(16,2),(25,2),(36,2),(49,2),(9,4),(27,4)\}$.

Next we consider GLP with α negative. As already stated in Example (iv) that $L_n^{(\alpha)}(x)$ is reducible with $n \geq -\alpha$. Therefore we restrict $-\alpha > n$ i.e $\alpha < -n$. We write $\alpha = -1-n-r$ where $r \geq 0$ is an integer. Nair and Shorey [23] proved that $L_n^{(-1-n-r)}(x)$ is irreducible for $0 \leq r \leq 22$ and for $23 \leq r \leq 60$ by Jindal, Laishram and Sarma in [15]. The above result with $r = 0, r = 2$ and $r = 1, 3 \leq r \leq 8$ were already proved by Schur [26], Sell [28] and Hajir [14], respectively. Hajir [14] conjectured that $L_n^{(-1-n-r)}(x)$ is irreducible for $r \geq 0$. Hajir [14] confirmed the conjecture when $n > e^{r!+\frac{1}{2}}(1-\frac{1}{r!})^{-r!}$ and Nair and Shorey [23] with $n > \frac{r}{1.63}e^{r(1+\frac{r}{\log r})}$ and Jindal, Laishram and Sarma in [15] when $n > re^{r(1+\frac{r}{\log r})}$. We do not know whether the right hand side of the above inequality can be replaced by an absolute constant. As already stated, the case $r = n$ was settled by Filaseta and Trifonov [7]. If we want to prove only that $L_n^{(\alpha)}(x)$ does not have a factor of large degree, it is possible to consider more general values of α. Let s and t be fixed integers given by either

$$\alpha = -tn - s - 1 \text{ with } t \geq 2 \qquad \text{or} \qquad \alpha = tn + s \text{ with } t \geq 1.$$

If $L_n^{(\alpha)}(x)$ has a factor of degree k, then Fuchs and Shorey [10] proved that k is bounded by an effectively computable number depending only on s and t. Further we refer to [10] for a qualitative version of the above result.

Now we turn to the case $d = 2$. i.e $\alpha = u + \frac{1}{2}$ where u is an integer. We have

$$H_{2n}(x) = (-1)^n 2^{2n} n! L_n^{(-\frac{1}{2})}(x^2) \text{ and } H_{2n+1}(x) = (-1)^n 2^{2n+1} n! x L_n^{(\frac{1}{2})}(x^2)$$

where H_{2n} and H_{2n+1} are Hermite polynomials. Schur [26], [27] proved that $L_n^{(-\frac{1}{2})}(x^2)$ and $L_n^{(\frac{1}{2})}(x^2)$ are irreducible implying the irreducibility of $H_{2n}(x)$ and $H_{2n+1}(x)/x$. Further Laishram, Nair and Shorey [19] proved that $L_n^{(\alpha)}(x^2)$ with $1 \leq u \leq 45$ are irreducible except when $(u,n) = (10,3)$ in which case $L_3^{(21/2)}(x^2) = -\frac{1}{48}(2x^2-15)(4x^4-132x^2+1035)$.

Let $d \in \{3,4\}$. Then Laishram and Shorey [18] proved that $L_n^{(\alpha)}(x)$ is irreducible whenever $\alpha \in \{\pm\frac{1}{3}, \pm\frac{2}{3}, \pm\frac{1}{4}, \pm\frac{3}{4}\}$ except for $(n,\alpha) = (4, \frac{21}{4})$. There is no result for $d > 4$ is available in the literature.

In Example (ii), we introduced Generalised Schur polynomials and we observe that they are extensions of GLP with integer $\alpha \geq 0$ upto a constant factor. Now, for every $\alpha \in \mathbb{Q}$, we consider a polynomial whose irreducibility implies the irreducibility of $L_n^{(\alpha)}(x)$. Let $\alpha \in \mathbb{Q}$ be given by (5.1) and $a_0, a_1, \cdots, a_n \in \mathbb{Z}$ with $|a_0| = |a_n| = 1$. We consider

$$\begin{aligned} G_n^{(\alpha)}(x) =& G_n^{(\alpha)}(x; a_0, a_1, \ldots, a_n) \\ =& \sum_{j=0}^{n} a_j (n+\alpha)(n-1+\alpha)\cdots(j+1+\alpha) d^{n-j} x^j \qquad (5.2) \\ =& \sum_{j=0}^{n} a_j x^j \Big(\prod_{i=j+1}^{n} (a+(u+i)d) \Big) \end{aligned}$$

since $(i+\alpha)d = a + (i+u)d$. We observe that

$$G_n^{(\alpha)}(x) = d^n n! L_n^{(\alpha)}(\frac{x}{d}) \text{ if } a_j = (-1)^j \binom{n}{j} \qquad (5.3)$$

and therefore the irreducibility of $G_n^{(\alpha)}(x)$ with $a_j = (-1)^j \binom{n}{j}$ implies the irreducibility of $L_n^{(\alpha)}(x)$. Further for an integer $\alpha \geq 0$, we have

$$G_n^{(\alpha)}(x) = (n+\alpha)! \sum_{j=0}^{n} a_j \frac{x^j}{(j+\alpha)!}.$$

Now we consider the irreducibility of $G_n^{(\alpha)}(x)$. Let $d = 1$. The first result on $G_n^{(\alpha)}(x)$ is due to Schur [26] who proved that $G_n^{(\alpha)}(x)$ with $\alpha \in \{-1, 0\}$ is irreducible unless $\alpha = 0$ and either $n+1$ is a power of 2 where it may have a linear factor or $n = 8$ where it may have a quadratic factor.

The assumption $|a_n| = 1$ has been relaxed in the above result of Schur. Let $\alpha = 0$. Filaseta [4] proved that $G_n^{(\alpha)}(x)$ with $|a_0| = 1$ and $0 < |a_n| < n$ is irreducible unless $(a_n, n) \in \{(\pm 5, 6), (\pm 7, 10)\}$ in which case $G_n^{(\alpha)}(x)$ is either irreducible or is a product of two irreducible polynomials of same degree. The assumption $0 < |a_n| < n$ is necessary. For this, we take $a_n = n, a_{n-1} = -1, a_{n-2} = \cdots = a_2 = 0, a_1 = -1, a_0 = 1$, then $x - 1$ is a factor of $G_n^{(\alpha)}(x)$. This example is given in [4]. Let $\alpha = 1$. Allen and Filaseta [1] extended the above result as follows. Let $n+1 = k' 2^u$ with $2 \nmid k', (n+1)n = k" 2^v 3^w$ with $2 \nmid k", 3 \nmid k"$ and $M = \min(k', k")$. Then $G_n^{(\alpha)}(x)$ with $\alpha = 1, |a_0| = 1$ and $0 < |a_n| < M$ is irreducible. Here the assumption $0 < |a_n| < M$ is best possible in the sense that for $|a_0| = 1$ and $|a_n| = M$, there exist integers $a_1, a_2, \ldots, a_{n-1}$ such that $G_n^{(\alpha)}(x)$ is irreducible. For example as in [1] when $n = 5$, we have $M = 3$ and $G_n^{(\alpha)}(x)$ has a factor $x - 2$ if $a_5 = 3, a_4 = -1, a_3 = a_2 = 0, a_1 = 1, a_0 = -1$.

The result of Filaseta and Lam [5] already stated for $L_n^{(\alpha)}(x)$ is also valid for $G_n^{(\alpha)}(x)$ when α is not a negative integer. Further Laishram and Shorey [16] showed that for $k \geq 2$, $G_n^{(\alpha)}(x)$ with $|a_0 a_n| = 1$ has no factor of degree k when α is an integer satisfying $0 \leq \alpha \leq 40$ if $k = 2$ and $0 \leq \alpha \leq 50$ if $k \geq 3$ except for an explicitly given finite set of triples (n, k, α) and we refer to [16] for a complete list of exceptions. In fact it has no factor of degree ≥ 5 unless $(n, k, \alpha) \in \{(17, 5, 11), (19, 5, 9), (40, 5, 12)\}$. The cases $0 \leq \alpha \leq 10$ if $k \in \{3, 4\}$ and $0 \leq \alpha \leq 30$ if $k \geq 5$ were already covered by Shorey and Tijdeman [30]. There

are some reasons to be unhappy with the above result. They give no information when $G_n^{(\alpha)}(x)$ has a linear factor. Further the set of exceptions is large in [17]. Therefore we consider a polynomial $\psi_n^{(\alpha)}(x)$ which is a particular case of $G_n^{(\alpha)}(x)$ by restricting a_j to $a_j\binom{n}{j}$ for $0 \le j \le n$ but extends $L_n^{(\alpha)}(x)$. Let

$$\psi_n^{(\alpha)}(x) = \sum_{j=0}^{n} a_j \binom{n}{j} (n+\alpha)\cdots(j+1+\alpha)x^j.$$

It is clear that the irreducibility of $G_n^{(\alpha)}(x)$ implies the irreducibility of $\psi_n^{(\alpha)}(x)$. Further, since

$$\psi_n^{(\alpha)}(x) = n!L_n^{(\alpha)}(x) \text{ if } a_j = (-1)^j,$$

we observe that $L_n^{(\alpha)}(x)$ is irreducible whenever $\psi_n^{(\alpha)}(x)$ is irreducible. The following result of Laishram, Nair and Shorey [19] takes care of the flaws mentioned above of the results of [17] and [30] stated above. Let

$$\begin{aligned}\Omega =&\{(2,2),(2,7),(2,14),(2,23),(2,34),(2,47),(3,24),(4,4),(4,5),(4,14),(4,20),\\ &(4,23),(6,44),(8,8),(8,41),(12,24),(16,16),(16,20),(16,24),(16,29),(24,8),\\ &(24,24),(30,24),(32,32),(32,48),(40,24),(48,24),(112,48),(120,24)\}.\end{aligned}$$

Let $0 \le \alpha \le 50$ be an integer and $|a_0a_n| = 1$. Then $\psi_n^{(\alpha)}(x)$ is irreducible except when $(n,\alpha) \in \Omega$ where it may have a linear factor. Further for every $(n,\alpha) \in \Omega - \{(3,24)\}$, there exist integers $a_0, a_1, \ldots, a_n$ with $|a_0| = |a_n| = 1$ such that $\psi_n^{(\alpha)}(x)$ has a linear factor. For $0 \le \alpha \le 10$, the above result was already proved by Filaseta, Finch and Leidy [8].

Let $d = 2$. i.e $\alpha = u + \frac{1}{2}$. Schur [26] proved that $G_n^{(\alpha)}(x^2)$ is irreducible when $u = -1$ and $u = 0$ unless $2n+1$ is a power of 3 where it may have a linear or quadratic factor. Allen and Filaseta [2] extended this result for $u = -1$ and $|a_0| = 1, 0 < |a_n| < 2n-1$. Further the assumption $0 < |a_n| < 2n-1$ is best possible. For example, if $a_n = \pm(2n-1), a_{n-1} = -(1\cdot 3\cdots(2n-3)) \mp 1, a_{n-2} = \cdots a_1 = 0, a_0 = 1$, then $G_n^{(\alpha)}(x^2)$ with $u = -1$ has a factor $x^2 - 1$. Further he proved that if $|a_n| = 2n-1$ and $G_n^{(\alpha)}(x^2)$ with $u = -1$ is reducible, then $G_n^{(\alpha)}(x^2)$ must have a factor of degree ≤ 4. Finch and Saradha [9] showed that for $1 \le u \le 13$, the polynomials $G_n^{(\alpha)}(x)$ with $a_0, a_n \in A$ have no factor of degree ≥ 2 except for $(u,n) \in \{(1,121),(8,59),(8,114),(9,4),(9,113),(9,163),(9,554)\}$ where it may have either a linear factor or a quadratic factor. Let $S = \{(1,121),(8,59),(8,114),(9,4),(9,113),(9,163),(9,554),(15,23),(15,107),(16,106),(20,102),(21,101),(26,155),(26,287),(30,92),(36,86),(43,1158),(44,716)\}$. Laishram, Nair and Shorey [19] proved that for $1 \le u \le 45$ and $P(a_0a_n) \le 2$, $G_n^{(\alpha)}(x^2)$ has no factor of degree ≥ 3 except when $(u,n) \in \{(1,12),(6,7),(9,113),(10,3),(21,101)\}$ or $(u,n) \in S$ or $(u,n) = (44,79)$ where it may have a factor of degree 3 or 4 or 6, respectively.

Since irreducibility of $G_n^{(\alpha)}(x^2)$ implies the irreducibility of $G_n^{(\alpha)}(x)$, the above result implies that for $1 \leq u \leq 45$ and $P(a_0a_n) \leq 2$, $G_n^{(\alpha)}(x)$ has no factor of degree ≥ 2 except when $(u,n) \in S$ or $(u,n) = (44,79)$ where it may have a factor of degree 2 or 3, respectively. Further the assumptions $(u,n) \in S$ is necessary. They also gave a bound for the degree l of any factor of $G_n^{(\alpha)}(x^2)$ in terms of u given by $l < 1.49u + 1.8$ except for $(u,n) \in \{(1,12),(1,121)\}$. Next we give an analogue of above result on $G_n^{(\alpha)}(x^2)$ for $\psi_n^{(\alpha)}(x^2)$. Let

$$\Omega_1 = \{(2,2),(2,8),(2,2^9),(6,2^4),(9,4),(9,2^6),(10,3),(10,12),(10,24),(10,192),$$
$$(21,2^4),(24,2^4),(30,2^6),(35,2^5),(35,2^9),(37,12),(37,36),(37,144),(44,2^{12})\}.$$

Let $\alpha = u + \frac{1}{2}$, where $0 \leq u \leq 45$ is an integer. Then Laishram, Nair and Shorey [20] proved that $\psi_n^{(\alpha)}(x^2)$ with $|a_0a_n| = 1$ is irreducible except when $(u,n) \in \Omega_1$ where it may have a quadratic factor. Further for every $(u,n) \in \Omega_1$ except for $(u,n) = (44,2^{12})$, there exist integers $a_0, a_1, \ldots, a_n$ with $|a_0| = |a_n| = 1$ such that $\psi_n^{(\alpha)}(x^2)$ has a quadratic factor. Recently Nair and Shorey proved results for $L_n^{(\alpha)}(x^2)$ and $G_n^{(\alpha)}(x^2)$ when $u \leq -2$. They proved that $L_n^{(\alpha)}(x^2)$ with $-18 \leq u \leq -2$ are irreducible and $G_n^{(\alpha)}(x^2)$ with $u = -2$ and $n > 2$ is irreducible. The assumption $n > 2$ in the above result is necessary since

$$G_2^{(-3/2)}(x^2) = (x^2-1)^2 \text{ when } a_0 = 1, a_1 = -2, a_2 = -1.$$

Further they extended the results of Schur for $u = -1$ and $u = 0$ stated in the beginning of this paragraph for $G_n^{(\alpha)}(x^2)$ to $-6 \leq u \leq -3$.

Let $d \in \{3,4\}$ and $u \in \{-1,0\}$. Then Laishram and Shorey [18] proved irreducibility results on $G_n^{(\alpha)}(x)$ and $G_n^{(\alpha)}(x^d)$ which we state now. Let $u = -1$. Then $\alpha \in \{\frac{-1}{3}, \frac{-2}{3}\}$ if $d = 3$ and $\alpha \in \{\frac{-1}{4}, \frac{-3}{4}\}$ if $d = 4$. Assume that $\alpha \neq \frac{-1}{4}$ if $d = 4$. Then $G_n^{(\alpha)}(x^d)$ is irreducible except when $d = 3, \alpha = \frac{-2}{3}, n = 2$ or $d = 3, \alpha = \frac{-1}{3}, n = 43$ where it may have a factor of degree 3 or 5, respectively. Consequently, $G_n^{(\alpha)}(x^d)$ is irreducible unless $d = 3, \alpha = \frac{-2}{3}, n = 2$ where $G_n^{(\alpha)}(x)$ may have a linear factor and $G_n^{(\alpha)}(x^3)$ may have a cubic factor. Let $u = 0$ and $d = 3$. Then $\alpha \in \{\frac{1}{3}, \frac{2}{3}\}$. Now $G_n^{(\alpha)}(x)$ and $G_n^{(\alpha)}(x^d)$ are irreducible except possibly when

(*i*) $1 + 3n = 2^b$ where $G_{\frac{1}{3}}(x)$ may have a linear factor and $G_{\frac{1}{3}}(x^3)$ may have a quadratic or cubic factor.

(*ii*) $2 + 3n = 2^b$ and $n \neq 42$ where $G_{\frac{2}{3}}(x^3)$ may have a quadratic factor.

(*iii*) $2 + 3n = 2^b \cdot 5^c$ with $c > 0$ where $G_{\frac{2}{3}}(x)$ may have a linear factor or $G_{\frac{2}{3}}(x^3)$ may have a cubic factor.

(*iv*) $n = 42$ where $G_{\frac{2}{3}}(x)$ may have a quadratic factor and $G_{\frac{2}{3}}(x^3)$ may have a factor of degree in $\{2,4,5,6\}$.

Let $u = 0, d = 4$. Then $\alpha \in \{\frac{1}{4}, \frac{3}{4}\}$ and $G_n^{(\alpha)}(x^d)$ is irreducible except when $1 + 4n = 3^b \cdot 5^c, \alpha = \frac{1}{4}$ or $3 + 4n = 7^y, \alpha = \frac{3}{4}$ where $G_n^{(\alpha)}(x)$ may have a linear

factor and $G_n^{(\alpha)}(x^4)$ may have a factor of degree 4. This is also the case if $3+4(n-1)$ is a power of 3 and $\alpha = -\frac{1}{4}$. Further $G_n^{(\alpha)}(x^4)$ and hence $G_n^{(\alpha)}(x)$ is irreducible if $\alpha = -\frac{1}{4}$, $3+4(n-1)$ is not a power of 3 or $\alpha = -\frac{3}{4}$. Let $u = -1, d = 4$ and $\alpha = \frac{-1}{4}$. Assume that $3+4(n-1)$ is not a power of 3. Then $G_n^{(\alpha)}(x^d)$ and hence $G_n^{(\alpha)}(x)$ are irreducible. The assumption $|a_0 a_n| = 1$ has been relaxed in the above results. For example, we may assume that $P(a_0 a_n) \leq 3$ when $a + 3(n+u)$ is not a power of 2 in the case $d = 3$ and $u \in \{-1, 0\}$. It follows from the above results that $G_n^{(\alpha)}(x)$ with $\alpha \in \{\pm\frac{1}{3}, \pm\frac{2}{3}, \pm\frac{1}{4}, \pm\frac{3}{4}\}$ is either irreducible or linear polynomial times an irreducible polynomial of degree $n-1$. If $d > 4$, there is no analogous result available in the literature.

5.4. **Galois groups of GLP.** We denote by $G_n(\alpha)$ the Galois group of $L_n^{(\alpha)}(x)$. Schur [26], [27] proved that $G_n(0) = S_n, G_n(1) = S_n$ whenever $n \equiv 0 \pmod 2$ such that $n+1$ is not a square and A_n otherwise, $G_n(-n-1) = S_n$ when $n \equiv 0 \pmod 4$ and A_n otherwise. The Galois groups of $L_n^{(\alpha)}(x)$ with $2 \leq \alpha \leq 10$ have been determined by Banerjee, Filaseta, Finch and Leidy [3] . From the results of Schur stated above, we see that every positive integer n with n not congruent to $2 \pmod 4$, there exists $\alpha \in \{1, -n-1\}$ such that $G_n^{(\alpha)} = A_n$. As stated in Section 5.2, we have $G_n(n) = A_n$ for $n \equiv 2 \pmod 4$. Further it is proved in Banerjee, Filaseta, Finch and Leidy [3] that there are only finitely many n satisfying $G_n(\alpha) = A_n$ with integer $\alpha \neq n$ and it has been conjectured there that $G_n(\alpha) = A_n$ with α integer implies $\alpha = n$.

Next we consider $G_n(-n-1-r)$ where r is a positive integer. Hajir [14] conjectured that $G_n(-n-1-r)$ contains A_n for all n and r. Hajir [14] and Sell[28] proved that $G_n(-n-1-r)$ contains A_n whenever $1 \leq r \leq 8, r \neq 2$ and $r = 2$, respectively, and $G_n(-n-1-r)$ with $9 \leq r \leq 22$ has been determined in Nair and Shorey [23]. For $r \geq 9$, Hajir [14] proved that $G_n(-n-1-r)$ contains A_n when $n \geq B(r)$ where

$$B(r) = e^{r!+\frac{1}{2}} \left(1 - \frac{1}{r!}\right)^{-r!}.$$

The value of $B(r)$ was reduced to $\frac{r}{1.63} e^{r\left(1+\frac{1.2762}{\log r}\right)}$ and $e^{r\left(1+\frac{1.2762}{\log r}\right)}$ in Nair and Shorey [23] and Jindal, Laishram and Sarma [15] where the Galois groups of $L_n^{(-n-1-r)}$ with $9 \leq r \leq 22$ and $23 \leq r \leq 60$ have been determined, respectively, confirming the conjecture of Hajir stated above.

It is proved in Laishram, Nair and Shorey [19] that for $\alpha = u + \frac{1}{2}$ with $-1 \leq u \leq 45$, the Galois group of $L_n^{(\alpha)}(x)$ is S_n except when $(u, n) = (10, 3)$ in which case the Galois group is $\mathbb{Z}_2$. The above result with $u \in \{-1, 0\}$ was already proved by Saradha and Shorey [25] when $n \geq 182$ and for all n by Laishram [21]. Further Saradha and Shorey [25] proved that $G_n(\alpha) = S_n$ for $n \geq 876$ if $\alpha \in \{\pm\frac{1}{3}, \pm\frac{2}{3}\}$ and $n \geq 1325$ if $\alpha \in \{\pm\frac{1}{4}, \pm\frac{3}{4}\}$. Finally Laishram [21] proved that

$G_n(\alpha) = S_n$ if $\alpha \in \{\pm\frac{1}{3}, \pm\frac{2}{3}, \pm\frac{1}{4}, \pm\frac{3}{4}\}$ unless $(\alpha, n) \in \{(-\frac{2}{3}, 11), (\frac{2}{3}, 7)\}$ where $G_n(\alpha) = A_n$ and $(\alpha, n) = (\frac{1}{4}, 2)$ where $G_n(\alpha) = S_1$.

The main ingredients in the proofs are Newton polygons and lower bounds for the greatest prime factor of positive consecutive terms in arithmetic progression. We refer to [17], [18], [19], [22], [30] for the latter and to [29] for showing that these two ingredients lead to a proof of irreducibility of GLP.

One of the authors (TNS) was getting INSA Senior Scientist award when this work was done.

References

[1] Allen, M. and Filaseta, M., A generalization of a second irreducibility theorem of I. Schur, *Acta Arith.*, **109** (2003), 65–79.

[2] Allen, M. and Filaseta, M., A generalization of a third irreducibility theorem of I. Schur, *Acta Arith.*, **114** (2004), 183–197.

[3] Banerjee, P., Filaseta, M., Finch, E. and Leidy, J., On classifying Laguerre polynomials which have Galois group the alternating group, *J. Thor. Nombres Bordeaux* **25** (2013), 1–30.

[4] Filaseta, M., A generalization of an irreducibility theorem of I. Schur, Analytic Number Theory: *Proceedings of a Conference in Honor of Heini Halberstam*, Vol. **1** (1996), 371–396.

[5] Filaseta, M. and Lam, T. Y., On the irreducibility of the generalized Laguerre polynomials, *Acta Arith.*, **105** (2002), 177–182.

[6] Filaseta, M., Kidd, T. and Trifonov, O., Laguerre polynomials with Galois group A_m for each m, *J. Number Theory* **132** (2012), 776 – 805.

[7] Filaseta, M. and Trifonov, O., The irreducibility of the Bessel polynomials, *J. Reine Angew. Math.*, **550** (2002), 125–140.

[8] Filaseta, M., Finch, C. and Leidy, J. R., T. N. Shorey's influence in the theory of irreducible polynomials, *Diophantine Equations*, Narosa Publ., New Delhi (2008), 77–102.

[9] Finch, C. and Saradha, N., On the irreducibility of a certain polynomials with coefficients that are products of terms in an arithmetic progression, *Acta Arith.*, **143** (2010), 211–226.

[10] Fuch, C. and Shorey, T. N., Divisibility properties of generalized Laguerre polynomials, *Indag. Math.*, **20** (2009), 217–231.

[11] Gow, R., Some generalized Laguerre polynomials whose Galois groups are the Alternating groups, *J. Number Theory*, **31** (1989), 201–207.

[12] Grosswald, E., *Bessel Polynomials*, Lecture Notes in Mathematics **698**, Springer Verlag, Berlin (1978).

[13] Grosswald, E., Math.Review 89a: 11105, Math Reviews, Issue 89a (January 1989), 84.

[14] Hajir, F., Algebraic properties of a family of generalised Laguerre polynomials, *Canad. J. Math.*, **61** (2009), 583–603.

[15] Jindal, A., Laishram, S. and Sarma, R., Irreducibility and Galois groups of Generalised Laguerre polynomials $L_n^{-1-n-r}(x)$, to appear.

[16] Laishram, S. and Shorey, T. N., Extensions of Schur's irreducibility results, *Indag. Math.*, **21** (2011), 87–105.

[17] Laishram, S. and Shorey, T. N., Irreducibility of generalized Hermite-Laguerre Polynomials, *Functiones et Approximatio*, **47** (2012), 51–64.

[18] Laishram, S. and Shorey, T. N., Irreducibility of generalized Hermite-Laguerre Polynomials III, *J. Number Theory*, **164** (2016), 303–322.

[19] Laishram, S., Nair, S. G. and Shorey, T. N., Irreducibility of Generalized Laguerre Polynomials $L_n^{(\frac{1}{2}+u)}(x)$ with integer u, *J. Number Theory*, **160** (2016), 76–107.

[20] Laishram, S., Nair, S. G. and Shorey, T. N., Irreducibility of extensions of Laguerre Polynomials, to appear.

[21] Laishram, S., On the Galois groups of generalized Laguerre Polynomials, *Hardy- Ramanujan Journal*, **37** (2015), 8–12.

[22] Nair, S. G. and Shorey, T. N., Lower bounds for the greatest prime factor of product of consecutive positive integers, *J. Number Theory*, **159** (2016), 307-328.

[23] Nair, S. G. and Shorey, T. N., Irreducibility of Laguerre Polynomial $L_n^{(-1-n-r)}(x)$, *Indag. Math.*, **26** (2015), 615–625.

[24] Riordan, John, *An Introduction to Combinatorial Analysis*, Princeton University Press (1980).

[25] Saradha, N. and Shorey, T. N., Squares in blocks from an arithmetic progression and Galois group of Laguerre polynomials, *Int. J. Number Theory*, **11** (2015), 233–250.

[26] Schur, I., *Gleichungen ohne Affekt.* in: Gesammelte Abhandlungen, Band III, Springer-Verlag, Berlin, 1973, 191–197.

[27] Schur, I., *Affektlose Gleichungen in der Theorie der Laguerreschen und Hermiteschen Polynome* in: Gesammelte Abhandlungen, Band III, Springer-Verlag, Berlin-New York, 1973, 227–233.

[28] Sell, E. A., On a certain family of generalized Laguerre polynomials, *J. Number Theory* **107** (2004), 266–281.

[29] Shorey,T. N., Theorems of Sylvester and Schur, *Math. Student, Special Centenary Volume*, (2008), 135–145.

[30] Shorey, T. N. and Tijdeman, R., Generalisations of some irreducibility results by Schur, *Acta Arith.*, **145** (2010), 341–371.

[31] Szego, G., *Orthogonal polynomials*, Amer.Math. Soc. Colloq. **23**(1975).

[32] Weisstein, Eric. W., *Laguerre polynomials*, Wolfrom MathWorld (2003)

[33] *Laguerre polynomials*, https// en.wikipedia. org

Saranya G. Nair
Stat-Math Unit, Indian Statistical Institute
8th Mile Mysore Road, Bangalore-560059, Karnataka, India
E-mail: *saranya_vs@isibang.ac.in*

T. N. Shorey
National Institute of Advanced Studies
IISc Campus, Bangalore - 560012, Karnataka, India
E-mail: *shorey@math.iitb.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 103-113

REPRESENTATION OF NUMBERS BY QUATERNARY QUADRATIC FORMS

ARPITA KAR

(Received : 16 - 03 - 2017 ; Revised : 27 - 04 - 2017)

ABSTRACT. In this note, using the theory of modular forms, we will sketch a general method to find explicit formulas for the number of representations of a positive integer as $ax^2 + by^2 + cz^2 + dw^2$. This method had been used earlier to obtain results, but to the best of our knowledge, there is no exposition explaining the approach in general. That is the main goal of this article.

1. INTRODUCTION

The interest in representing numbers as a sum of squares of non-negative integers is very old and has led to several celebrated results. A classical conjecture of Fermat from 1640 asserts that any prime $p \equiv 1 (\text{mod } 4)$ is a sum of two squares of integers. Fermat claimed that he had a proof but the first formal proof was given by Euler. Fermat also conjectured that for each $n \in \mathbb{N}$, $8n + 3$ is a sum of three squares (of odd integers). In 1796, Gauss proved that every positive integer n is the sum of 3 triangular numbers; and this statement is essentially equivalent to Fermat's conjecture. A year later, Legendre proved that any positive integer n can be written as a sum of three squares of integers iff $n \neq 4^i(8k + 7)$ for any non-negative integers i and k (see [9]). Based on some work of Euler, in 1722, Lagrange showed that every natural number is a sum of four squares of integers. In connection with Lagrange's theorem, Ramanujan raised the problem of determining all the positive integers a, b, c, d such that every natural number n is representable in the form $ax^2 + by^2 + cz^2 + du^2$. Such forms are called universal diagonal quaternary quadratic forms in the literature. He proved that there exist only 54 such quadruples (a, b, c, d) with $1 \leq a \leq b \leq c \leq d$ (see [14]) (in fact, Ramanujan actually listed 55 such forms which was later corrected to 54 since one of the forms failed to represent 15, for more details, see [4] and p.341 of [6]). While we can ask about which integers can be represented by a given quadratic form, it is also interesting to ask in how many ways a certain integer m can be represented by that quadratic form. This will be the theme in this article.

2010 Mathematics Subject Classification : 11M06, 20C15
Key words and phrases : Quaternary quadratic forms, Representations, Theta functions, Eta quotients, Modular forms.

For a positive integer k, let $r_k(n)$ denote the number of representations of the non-negative integer n as a sum of k squares of integers, that is, $r_k(n)$ is the number of solutions of the Diopantine equation

$$x_1^2 + \cdots x_k^2 = n \ (x_i \in \mathbb{Z}, 1 \leq i \leq k). \tag{1.1}$$

We observe that $r_2(1) = 4$. For $k = 2$, Euler proved that (1.1) is solvable iff each prime divisor p of n, for which $p \equiv 3 (\text{mod } 4)$ occurs in n to an even power. Later the formula

$$r_2(n) = 4\{ \sum_{\substack{d|n \\ d\equiv 1(4)}} 1 - \sum_{\substack{d|n \\ d\equiv 3(4)}} 1\}$$

was established independently by Gauss using the arithmetic of $\mathbb{Z}[i]$ and by Jacobi using elliptic functions. By similar methods, using theta functions, Jacobi found formulas for $r_4(n)$, $r_6(n)$, $r_8(n)$ (see for eg. p.244 of [15]). Liouville found formulas for the number of ways of representing an integer as $x^2 + y^2 + 2z^2 + 2u^2$, $x^2 + y^2 + z^2 + 2u^2$, $x^2 + 2y^2 + 2z^2 + 2u^2$ (see [10] and [11]) by a different method which is elementary and enigmatic.

Another approach to study the number of representations of a number using a certain quadratic form is using the theory of modular forms. It has been used extensively (see [1], [2], [3]), but in theory, to the best of our knowledge, there is no article which gives a sketch of the method in general. In this note, the general method to approach this problem using modular forms is being sketched and an example to illustrate the method is given. We believe Theorem 4.4 and Corollary 4.5 giving an explicit formula for the number of solutions of $n = x^2 + y^2 + z^2 + 3w^2$ are new and have not been stated explicitly in literature before.

2. Preliminaries and definitions

Let $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{C}$ denote the sets of positive integers, non-negative integers, integers, rational numbers and complex numbers, respectively. The sum of divisors function $\sigma(n)$ for $n \in \mathbb{N}$ is given by

$$\sigma(n) = \sum_{m|n} m.$$

This function will appear often in our formulas.

For $a, b, c, d \in \mathbb{N}$, $n \in \mathbb{N}_0$, we define

$$N(a, b, c, d; n) = |\{(x, y, z, w) \in \mathbb{Z}^4 : n = ax^2 + by^2 + cz^2 + dw^2\}|.$$

We here determine explicit formulas for $N(a, b, c, d; n)$ and then apply the formula to the case where $a = 1$, $b = 1$, $c = 1$, $d = 3$. That is, we give an explicit formula for $N(1, 1, 1, 3; n)$.

Let $\theta(z)$ denote Ramanujan's theta function defined by

$$\theta(z) = \sum_{n=-\infty}^{\infty} e^{2\pi i n^2 z}$$

for $z \in \mathfrak{H}$ where $\mathfrak{H} = \{z \in \mathbb{C} | \text{Im}(z) > 0\}$.

The Dedekind eta function $\eta(z)$ is the holomorphic function defined on the upper half plane $\mathfrak{H}$ by

$$\eta(z) = e^{\pi iz/12} \prod_{n=1}^{\infty} (1 - e^{2\pi inz}).$$

If we take $q = q(z) = e^{2\pi iz}$ with $z \in \mathfrak{H}$ and so $|q| < 1$, we get

$$\eta(z) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n).$$

We will see later that $\eta(z)$ and $\theta(z)$ are related.

Now it is easy to see that for $q \in \mathbb{C}$, writing $q = e^{2\pi iz}$, we have

$$\sum_{n=0}^{\infty} N(a, b, c, d; n) q^n = \theta(az)\theta(bz)\theta(cz)\theta(dz),$$

where we define $N(a, b, c, d; 0) = 1$.

A Dirichlet character of modulus N is a homomorphism

$$\chi : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*.$$

This implies that $\chi(1) = 1$.

Let χ and ψ be Dirichlet characters. For $n \in \mathbb{N}$, we define $\sigma_{\chi,\psi}(n)$ by

$$\sigma_{\chi,\psi}(n) = \sum_{1 \le m, m|n} \psi(m)\chi(n/m)m.$$

For $N \in \mathbb{N}$ and a Dirichlet character χ of modulus N, the modular subgroup $\Gamma_0(N)$ is defined by

$$\Gamma_0(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) : c \equiv 0 (\bmod N) \right\},$$

where $SL_2(\mathbb{Z})$ is the set of all 2×2 matrices with integer entries which have determinant 1.

Let $k \in \mathbb{Z}$. $M_k(\Gamma_0(N), \chi)$ denotes the vector space of modular forms of weight k with character χ for $\Gamma_0(N)$, $E_k(\Gamma_0(N), \chi)$ and $S_k(\Gamma_0(N), \chi)$ denotes the Eisenstein subspace and the subspace of cusp forms respectively. It is known that

$$M_k(\Gamma_0(N), \chi) = S_k(\Gamma_0(N), \chi) \oplus E_k(\Gamma_0(N), \chi).$$

See [18]. One can review the basic theory of modular forms from [13]. For further reading, see [5], [7], [8], [12], [15], [16] and [17].

3. Sketch of the General Method

For $q \in \mathbb{C}$ and $z \in \mathfrak{H}$, writing $q = e^{2\pi iz}$, we have

$$\sum_{n=0}^{\infty} N(a, b, c, d; n) q^n = \theta(az)\theta(bz)\theta(cz)\theta(dz), \tag{3.1}$$

where $N(a, b, c, d; 0) = 1$. $\theta(z)$ has the following infinite product expansion

$$\theta(z) = \frac{\eta^5(2z)}{\eta^2(z)\eta^2(4z)},$$

(see p.81 of [13]). So, we see that $\sum_{n=0}^{\infty} N(a, b, c, d; n) q^n$ is given by a certain eta quotient.

We will now use the following theorem to determine if certain eta quotients are modular forms. For a proof, see p.99 of [7]

Theorem 3.1. *(Ligozat's Criterion) Let $f(z)$ be the eta quotient given by*

$$f(z) = \prod_{\delta} \eta^{r_\delta}(\delta z)$$

where δ runs through a finite set of positive integers and r_δ are non-zero integers and there exists a positive integer N which satisfy the following conditions :

(L1) $\sum_{\delta|N} \delta \cdot r_\delta \equiv 0 \ (mod\ 24)$.

(L2) $\sum_{\delta|N} \frac{N}{\delta} \cdot r_\delta \equiv 0 \ (mod\ 24)$.

(L3) For each $d|N$, $\sum_{\delta|N} \frac{gcd(d,\delta)^2 \cdot r_\delta}{\delta} \geq 0$.

Then $f(z) \in M_k(\Gamma_0(N), \chi)$, *where the character* χ *is given by* $\chi(m) = (\frac{(-1)^k s}{m})$ *with weight* $k = \frac{1}{2}\sum_{\delta|N} r_\delta$ *and* $s = \prod_{\delta|N} \delta^{r_\delta}$.

If (L3) is replaced by

(L4) For each $d|N$, $\sum_{\delta|N} \frac{gcd(d,\delta)^2 \cdot r_\delta}{\delta} > 0$,

then $f(z) \in S_k(\Gamma_0(N), \chi)$, *where the character* χ *is given by* $\chi(m) = (\frac{(-1)^k s}{m})$ *with weight* $k = \frac{1}{2}\sum_{\delta|N} r_\delta$ *and* $s = \prod_{\delta|N} \delta^{r_\delta}$.

Thus Ligozat's criterion explicitly determines the values of k, N and χ such that $\theta(az)\theta(bz)\theta(cz)\theta(dz) \in M_k(\Gamma_0(N), \chi)$ where $\theta(az)\theta(bz)\theta(cz)\theta(dz)$ is as in (3.1). Now the next step is to calculate the dimension of $M_k(\Gamma_0(N), \chi)$.

Fix a positive integer N. Let ϵ be a Dirichlet character modulo N. To find a set of canonical generators for the group $(\mathbb{Z}/N\mathbb{Z})^*$, write $N = \prod_{i=0}^{n} p_i^{e_i}$ where $p_0 < p_1 < \cdots < p_n$ are the prime divisors of N. Each factor $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ is a cyclic group $C_i = \langle g_i \rangle$, except if $p_0 = 2$ and $e_0 \geq 3$, in which case $(\mathbb{Z}/p_0^{e_0}\mathbb{Z})^*$ is a product of the cyclic group $C_0 = \langle -1 \rangle$ of order 2 with the cyclic subgroup $C_1 = \langle 5 \rangle$. In all cases we have

$$(\mathbb{Z}/N\mathbb{Z})^* \cong \prod_{0 \leq i \leq n} C_i.$$

For i such that $p_i > 2$, choose the generator g_i of C_i to be the element of $\{2, 3, \cdots, p_i^{e_i} - 1\}$ that is smallest and generates C_i. Finally use the Chinese Remainder Theorem to lift each g_i to an element in $(\mathbb{Z}/N\mathbb{Z})^*$, also denoted g_i, that is 1 modulo each $p_j^{e_j}$ for $j \neq i$. Now we will describe how one can compute the conductor of a character. As a reference for these facts, see page 70 of [18].

The following is the algorithm for computing the order of a Dirichlet character.

(Order of Character). This algorithm computes the order of a Dirichlet character ϵ modulo N.

- Compute the order r_i of each $\epsilon(g_i)$, for each minimal generator g_i of $(\mathbb{Z}/N\mathbb{Z})^*$. The order of $\epsilon(g_i)$ is a divisor of $n = |(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*|$ so we can compute its order by considering the divisors of n.
- Compute and output the least common multiple of the integers r_i.

The next algorithm factors a character ϵ as a product of "local" characters.

(Factorisation of Character). Given a Dirichlet character ϵ modulo N, with $N = \prod_{i=0}^{n} p_i^{e_i}$, this algorithm finds Dirichlet characters ϵ_i modulo $p_i^{e_i}$, such that for all $a \in (\mathbb{Z}/N\mathbb{Z})^*$,we have $\epsilon(a) = \prod \epsilon_i(a(\text{mod } p_i^{e_i}))$. If $2|N$, the steps are as follows:

- Let g_i be the minimal generators of $(\mathbb{Z}/N\mathbb{Z})^*$, so ϵ is given by a list $[\epsilon(g_0), \cdots, \epsilon(g_n)]$
- For $i = 2, \cdots, n$, let ϵ_i be the character modulo $p_i^{e_i}$ defined by the singleton list $[\epsilon(g_i)]$.
- Let ϵ_1 be the character modulo 2^{e_1} defined by the list $[\epsilon(g_0), \epsilon(g_1)]$ of length 2. Output the ϵ_i and terminate.

If $2 \not| N$, then omit the third step and include all i in the second step.

To proceed further, we need to recall the definition of conductor of a Dirichlet character.

Definition 1 (Conductor)**.** *The conductor of a Dirichlet character ϵ modulo N is the smallest positive divisor $c|N$ such that there is a character ϵ' modulo c for which $\epsilon(a) = \epsilon'(a)$ for all $a \in \mathbb{Z}$ with $(a, N) = 1$. A Dirichlet character is primitive if its modulus equals its conductor. The character ϵ' associated to ϵ with modulus equal to the conductor of ϵ is called the primitive character associated to ϵ.*

(Conductor). The following algorithm computes the conductor of a Dirichlet character modulo N.

1. [Factor Conductor] Find characters χ_i whose product is χ.
2. [Computing order] Compute order r_i for each χ_i.
3. [Conductor of factors] For each i, either set c_i to be 1 if χ_i is the trivial character or set $c_i = p_i^{ord_{p_i}(r_i)+1}$, where $ord_p(n)$ denotes the largest power of p that divides n.
4. [Finished] compute product of the c_i.

Once we have computed the conductor of the character, we can use it to compute the dimension of the space $M_k(\Gamma_0(N), \chi)$. The following theorem gives the formulae to compute the dimensions of $E_k(\Gamma_0(N), \chi)$ and $S_k(\Gamma_0(N), \chi)$. For a reference, see pg. 98 of [18].

Theorem 3.2. *We have*

$$\dim S_k(\Gamma_0(N), \chi) - \dim M_{2-k}(\Gamma_0(N), \chi) = \frac{k-1}{12}\mu_0(N) - 1/2 \prod_{p|N} \lambda(p, N, \nu_p(c)) + \gamma_4(k) \sum_{x \in A_4(N)} \chi(x) + \gamma_3(k) \sum_{x \in A_3(N)} \chi(x), \quad (3.2)$$

where

$$\mu_0(N) = \prod_{p|N} (p^{\nu_p(N)} + p^{\nu_p(N)-1}),$$

$$A_4(N) = \{x \in \mathbb{Z}/N\mathbb{Z} : x^2 + 1 = 0\},$$
$$A_3(N) = \{x \in \mathbb{Z}/N\mathbb{Z} : x^2 + x + 1 = 0\},$$
$$\gamma_4(k) = \begin{cases} -1/4 & \text{if } k \equiv 2 \ (mod\ 4) \\ 1/4 & \text{if } k \equiv 0 \ (mod\ 4) \\ 0 & \text{if } k \text{ odd}, \end{cases}$$
$$\gamma_3(k) = \begin{cases} -1/3 & \text{if } k \equiv 2 \ (mod\ 3) \\ 1/3 & \text{if } k \equiv 0 \ (mod\ 3) \\ 0 & \text{if } k \equiv 1 \ (mod\ 3), \end{cases}$$

and, for $p|N$, *if we put* $r = \nu_p(N)$ *then*

$$\lambda(p, N, \nu_p(c)) = \begin{cases} p^{r/2} + p^{r/2-1} & \text{if } 2 \cdot \nu_p(c) \leq r, 2|r \\ 2 \cdot p^{(r-1)/2} & \text{if } 2 \cdot \nu_p(c) \leq r, 2 \nmid r \\ 2 \cdot p^{r-\nu_p(c)} & \text{if } 2 \cdot \nu_p(c) > r. \end{cases}$$

Also,

$$dim\ E_k(\Gamma_0(N), \chi) = dim\ M_k(\Gamma_0(N), \chi) - dim\ S_k(\Gamma_0(N), \chi),$$

where

$$\dim M_k(\Gamma_0(N), \chi) = -\left(\frac{1-k}{12}\mu_0(N) - 1/2 \prod_{p|N} \lambda(p, N, \nu_p(c)\right) + \gamma_4(2-k) \sum_{x \in A_4(N)} \chi(x) + \gamma_3(2-k) \sum_{x \in A_3(N)} \chi(x)). \quad (3.3)$$

Note: Here c denotes the conductor of χ.

Once the dimension of $M_k(\Gamma_0(N), \chi)$ is computed, our next goal is to compute a basis for this vector space.

Let χ and ψ be primitive Dirichlet characters with conductors L and R respectively. Let

$$E_{k,\chi,\psi}(z) = c_0 + \sum_{m \geq 1} (\sum_{n|m} \psi(n)\chi(m/n)n^{k-1}) e^{2\pi i m z}, \quad (3.4)$$

where

$$c_0 = \begin{cases} 0 & \text{if } L > 1 \\ -\frac{B_{k,\psi}}{2k} & \text{if } L = 1. \end{cases}$$

When $\chi = \psi = 1, k \geq 4$, then $E_{k,\chi,\psi} = E_k$.

Theorem 3.3. *Let* $t > 0$ *be an integer and* χ, ψ *be as above, and let* k *be a positive integer such that* $\chi(-1)\psi(-1) = (-1)^k$. *Except when* $k = 2$ *and* $\chi = \psi = 1$, *the power series* $E_{k,\chi,\psi}(tz)$ *defines an element of* $M_k(RLt, \chi/\psi)$. *If* $\chi = \psi = 1, k = 2, t > 1$, *and* $E_2(z) = E_{k,\chi,\psi}(z)$, *then* $E_2(z) - tE_2(tz)$ *is a modular form in* $M_2(\Gamma_0(t))$.

Theorem 3.4. *The Eisenstein series in* $M_k(\Gamma_0(N), \epsilon)$ *coming from the previous theorem with* $RLt|N$ *and* $\chi/\psi = \epsilon$ *form a basis for* $E_k(\Gamma_0(N), \epsilon)$.

For a reference to Theorem 3.3 and Theorem 3.4, see chapter 7 of [12].
Once we have a basis for $E_k(\Gamma_0(N), \chi)$, one can compute a basis of $S_k(\Gamma_0(N), \chi)$ using Ligozat's criterion and combining the two, get a basis for $M_k(\Gamma_0(N), \chi)$. Thus we can write $\sum_{n=0}^{\infty} N(a,b,c,d;n)q^n$ as a linear combination of the basis elements. Finally comparing coefficients of q^n on both sides, we can derive an explicit formula for $N(a,b,c,d;n)$.

It is worth mentioning here that for any positive integer n, $N(a,b,c,d;n) \neq 0$ for (a,b,c,d) being one of the 54 quadruples that Ramanujan listed in [14]. The example in the next section will illustrate this fact.

4. Computing the number of representations of an integer as $x^2 + y^2 + z^2 + 3u^2$

In this section, we will determine a formula for $N(1,1,1,3;n)$ using the theory outlined in the previous section. Since we are interested in $N(1,1,1,3;n)$, we need to consider $f(z) = \theta^3(z)\theta(3z)$.

Theorem 4.1. *$f(z) = \theta^3(z)\theta(3z) \in M_2(\Gamma_0(12), \chi)$ for $\chi(d) = (\frac{2^4.3}{d})$.*

Proof. Using the infinite product representation for $\theta(z)$, we get that

$$\theta^3(z)\theta(3z) = \frac{\eta^{15}(2z)\eta^5(6z)}{\eta^6(z)\eta^6(4z)\eta^2(3z)\eta^2(12z)}.$$

Now, using Ligozat's Criterion for $N = 12$ and $f(z) = \theta^3(z)\theta(3z)$, we get that

$$f(z) \in M_2(\Gamma_0(12), \chi)$$

for $\chi(d) = (\frac{2^4.3}{d})$. □

Now our goal is to find a basis for $M_2(\Gamma_0(12), \chi)$ so that we can write $f(z) = \theta^3(z)\theta(3z)$ in terms of the basis elements.

To apply the formula mentioned in the previous section, we first need to compute the conductor of χ. Firstly, we see that $\chi(d) = (\frac{2^4.3}{d})$ is a character modulo 12.

d	0	1	2	3	4	5	6	7	8	9	10	11
$\chi(d)$	0	1	0	0	0	-1	0	-1	0	0	0	1

Table 1: Table for values of χ.

To factorise χ, we do the following: first we note that

$$(\mathbb{Z}/12\mathbb{Z})^* \cong (\mathbb{Z}/2^2\mathbb{Z})^* \times (\mathbb{Z}/3\mathbb{Z})^*.$$

Since $(\mathbb{Z}/4\mathbb{Z})^*$ is generated by $\{1,3\}$ and $(\mathbb{Z}/3\mathbb{Z})^*$ is generated by $\{1,2\}$, the minimal generators for $(\mathbb{Z}/12\mathbb{Z})^*$ are x_1 and x_2 such that x_1 is the lift of $(1,2) \in (\mathbb{Z}/2^2\mathbb{Z})^* \times (\mathbb{Z}/3\mathbb{Z})^*$ and x_2 is the lift of $(3,1) \in (\mathbb{Z}/2^2\mathbb{Z})^* \times (\mathbb{Z}/3\mathbb{Z})^*$ respectively to $(\mathbb{Z}/12\mathbb{Z})^*$. Using the Chinese remainder Theorem, we get that $x_1 = 5$ and $x_2 = 7$.
Now we use the algorithm from the previous section and note that $\chi(5) = -1$

has order 2 in $(\mathbb{Z}/3\mathbb{Z})^*$ and $\chi(7) = -1$ has order 2 in $(\mathbb{Z}/4\mathbb{Z})^*$. Thus, $c_1 = 2^{ord_2(2)+1} = 4$, $c_2 = 3^{ord_3(2)+1} = 3$.

Hence, we get that the conductor of χ is 12. Also,

$$\mu_0(12) = 12, \qquad \lambda(2, 12, \nu_2(12)) = 2,$$

$$\lambda(3, 12, \nu_3(12)) = 2, \qquad A_4(12) = \varnothing, \qquad A_3(12) = \varnothing.$$

Since $\dim M_0(\Gamma_0(12), \chi) = 0$, applying the dimension formulas, we get that

$$\dim S_2(\Gamma_0(12), \chi) = 0 \qquad \text{and} \qquad \dim E_2(\Gamma_0(12), \chi) = 4.$$

Thus, $\dim M_2(\Gamma_0(12), \chi) = 4$.

We will construct a basis for $E_2(\Gamma_0(12), \chi)$.

Since $|(\mathbb{Z}/12\mathbb{Z})^*| = 4$, there are 4 Dirichlet characters of modulus 12 over $\mathbb{R}$. Also since we know that $\{5, 7\}$is the set of minimal generators for $(\mathbb{Z}/12\mathbb{Z})^*$, the Dirichlet characters modulo 12 are given by $\epsilon_1 = \chi$, ϵ_2, ϵ_3 and ϵ_4 which are defined as follows :

- $\epsilon_1(5) = -1, \quad \epsilon_1(7) = -1, \quad \epsilon_2(5) = 1, \quad \epsilon_2(7) = 1,$
- $\epsilon_3(5) = -1, \quad \epsilon_3(7) = 1, \quad \epsilon_4(5) = 1, \quad \epsilon_4(7) = -1.$

Evaluating the conductors of these characters as before, we get that ϵ_2, ϵ_3, ϵ_4 have conductors $1, 3, 4$ respectively. Thus, ϵ_2, ϵ_3, ϵ_4 are primitive Dirichlet characters modulo $1, 3, 4$ respectively.

Theorem 4.2. *For $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ as defined earlier, we define the following power series :*

$$E_{\epsilon_1,\epsilon_2}(z) = \sum_{n=1}^{\infty} \sigma_{\epsilon_1,\epsilon_2}(n) e^{2\pi i n z}, \qquad E_{\epsilon_2,\epsilon_1}(z) = \sum_{n=1}^{\infty} \sigma_{\epsilon_2,\epsilon_1}(n) e^{2\pi i n z},$$

$$E_{\epsilon_4,\epsilon_3}(z) = \sum_{n=1}^{\infty} \sigma_{\epsilon_4,\epsilon_3}(n) e^{2\pi i n z}, \qquad E_{\epsilon_3,\epsilon_4}(z) = \sum_{n=1}^{\infty} \sigma_{\epsilon_3,\epsilon_4}(n) e^{2\pi i n z}.$$

Then these forms $E_{\epsilon_1,\epsilon_2}(z)$, $E_{\epsilon_2,\epsilon_1}(z)$, $E_{\epsilon_3,\epsilon_4}(z)$ and $E_{\epsilon_4,\epsilon_3}(z)$ form a basis for $E_2(\Gamma_0(12), \chi)$ for $\chi(d) = (\frac{2^4.3}{d})$.

Proof. First, write $q = e^{2\pi i z}$. Then, we consider the following 4 cases :

Case 1: For $\chi = \frac{\epsilon_1}{\epsilon_2}$. $R = 1, L = 12, t = 1, k = 2$,

$$E_{2,\epsilon_1,\epsilon_2}(z) = c_0 + \sum_{m \geq 1} \Big(\sum_{n|m} \epsilon_2(n)\epsilon_1(m/n)n \Big) q^m = q + 2q^2 + 3q^3 + 4q^4 + \cdots .$$

Case 2: For $\chi = \frac{\epsilon_2}{\epsilon_1}$. $R = 12, L = 1, t = 1, k = 2$,

$$E_{2,\epsilon_2,\epsilon_1}(z) = c_0 + \sum_{m \geq 1} \Big(\sum_{n|m} \epsilon_1(n)\epsilon_2(m/n)n \Big) q^m = -1 + q + q^2 + q^3 + q^4 + \cdots .$$

Case 3: For $\chi = \frac{\epsilon_4}{\epsilon_3}$. $R = 3, L = 4, t = 1, k = 2$,

$$E_{2,\epsilon_4,\epsilon_3}(z) = c_0 + \sum_{m \geq 1} \Big(\sum_{n|m} \epsilon_3(n)\epsilon_4(m/n)n \Big) q^m = q - 2q^2 - q^3 + 4q^4 + \cdots .$$

Case 4: For $\chi = \frac{\epsilon_3}{\epsilon_4}$. $R = 4, L = 3, t = 1, k = 2$,

$$E_{2,\epsilon_3,\epsilon_4}(z) = c_0 + \sum_{m\geq 1}(\sum_{n|m}\epsilon_4(n)\epsilon_3(m/n)n)q^m = q - q^2 - 3q^3 + q^4 + \cdots .$$

Then using Theorem 3.4, these four forms form a basis for $E_2(\Gamma_0(12), \chi)$ for $\chi(d) = (\frac{2^4.3}{d})$. □

Corollary 4.3. *For $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ as defined earlier and the power series $E_{\epsilon_1,\epsilon_2}(z)$, $E_{\epsilon_2,\epsilon_1}(z)$, $E_{\epsilon_4,\epsilon_3}(z)$, $E_{\epsilon_3,\epsilon_4}(z)$ as defined in Theorem 4.2, $\{E_{\epsilon_1,\epsilon_2}(z), E_{\epsilon_2,\epsilon_1}(z), E_{\epsilon_4,\epsilon_3}(z), E_{\epsilon_3,\epsilon_4}(z)\}$ form a basis for $M_2(\Gamma_0(12), \chi)$ for $\chi(d) = (\frac{2^4.3}{d})$.*

Proof. Since $\dim S_2(\Gamma_0(12), \chi) = 0$ and $M_k(\Gamma_0(N), \chi) = E_k(\Gamma_0(N), \chi) \oplus S_k(\Gamma_0(N), \chi)$, we have

$$\dim M_2(\Gamma_0(12), \chi) = \dim E_2(\Gamma_0(12), \chi) = 4,$$

from which the result follows in view of previous theorem. □

Theorem 4.4. $f(z) = \theta^3(z)\theta(3z) = 6E_{\epsilon_1,\epsilon_2}(z) - E_{\epsilon_2,\epsilon_1}(z) - 2E_{\epsilon_4,\epsilon_3}(z) + 3E_{\epsilon_3,\epsilon_4}(z)$.

Proof. We have

$$f(z) = \theta^3(z)\theta(3z) = (\sum_{n=-\infty}^{\infty} e^{2\pi i n^2 z})^3(\sum_{n=-\infty}^{\infty} e^{2\pi i 3n^2 z})$$
$$= (1 + 2\sum_{n\geq 1} e^{2\pi i n^2 z})^3(1 + 2\sum_{n\geq 1} e^{2\pi i 3n^2 z}).$$

Using our basis, this is equal to

$$aE_{\epsilon_1,\epsilon_2}(z) + bE_{\epsilon_2,\epsilon_1}(z) + cE_{\epsilon_4,\epsilon_3}(z) + dE_{\epsilon_3,\epsilon_4}(z),$$

for certain a, b, c and d. Then writing $q = e^{2\pi i z}$ and from the proof of Theorem 4.2, comparing coefficients of q^0, q^1, q^2 and q^3 on both sides of the equality, we get $-b = 1$ *or* $b = -1$, $a + b + c + d = 6$, $2a + b - 2c - d = 12$, $3a + b - c - 3d = 10$. Solving these equations for a, b, c, d we obtain $a = 6$, $b = -1$, $c = -2$ and $d = 3$ which proves the theorem. □

Corollary 4.5. $N(1, 1, 1, 3; n) = 6\sigma_{\epsilon_1,\epsilon_2}(n) - \sigma_{\epsilon_2,\epsilon_1}(n) - 2\sigma_{\epsilon_4,\epsilon_3}(n) + 3\sigma_{\epsilon_3,\epsilon_4}(n)$.

Proof. Since for $q = e^{2\pi i z}$, we have $\sum_{n=0}^{\infty} N(1, 1, 1, 3; n)q^n = \theta(z)^3\theta(3z)$ and using Theorem 4.4, we have

$$\theta^3(z)\theta(3z) = 6E_{\epsilon_1,\epsilon_2}(z) - E_{\epsilon_2,\epsilon_1}(z) - 2E_{\epsilon_4,\epsilon_3}(z) + 3E_{\epsilon_3,\epsilon_4}(z),$$

the result follows. □

Now, let us illustrate in an example that the formula indeed works. Consider the case when $n = 10$. Let us try to compute $N(1, 1, 1, 3; 10)$.

$\sigma_{\epsilon_1,\epsilon_2}(10) = \epsilon_2(1)\epsilon_1(10) \cdot 1 + \epsilon_2(2)\epsilon_1(5) \cdot 2 + \epsilon_2(5)\epsilon_1(2) \cdot 5 + \epsilon_2(10)\epsilon_1(1) \cdot 10 = 8.$

$\sigma_{\epsilon_2,\epsilon_1}(10) = \epsilon_1(1)\epsilon_2(10) \cdot 1 + \epsilon_1(2)\epsilon_2(5) \cdot 2 + \epsilon_1(5)\epsilon_2(2) \cdot 5 + \epsilon_1(10)\epsilon_2(1) \cdot 10 = -4.$

$$\sigma_{\epsilon_4,\epsilon_3}(10) = \epsilon_3(1)\epsilon_4(10)\cdot 1 + \epsilon_3(2)\epsilon_4(5)\cdot 2 + \epsilon_3(5)\epsilon_4(2)\cdot 5 + \epsilon_3(10)\epsilon_4(1)\cdot 10 = 8.$$
$$\sigma_{\epsilon_3,\epsilon_4}(10) = \epsilon_4(1)\epsilon_3(10)\cdot 1 + \epsilon_4(2)\epsilon_3(5)\cdot 2 + \epsilon_4(5)\epsilon_3(2)\cdot 5 + \epsilon_4(10)\epsilon_3(1)\cdot 10 = -4.$$

Then, by Corollary 4.5,

$$N(1,1,1,3;10) = (6\cdot 8) - (-4) - (2\cdot 8) + (3\cdot(-4)) = 24.$$

Let us explicitly write down the representations of 10 as $x^2+y^2+z^2+3u^2$. Firstly, note that $u = 0$. This is because u cannot be greater than equal to 2. If $u = 1$, then 7 must be written as a sum of three squares which is not possible. Thus, the possibilities are $(x,y,z,u) = (0,1,3,0)$, $(x,y,z,u) = (0,-1,3,0)$, $(x,y,z,u) = (0,1,-3,0)$ and $(x,y,z,u) = (0,-1,-3,0)$. But also, since u remains fixed, the values of x,y,z can be permuted in 3! ways. Thus, total number of representations of 10 as $x^2+y^2+z^2+3u^2$ is $4\cdot(3!)$ which equals 24 which is what we got using Corollary 4.5.

Concluding remarks

The example illustrates that in theory, it is possible to derive explicit formulas for the number of representations of n as $ax^2+by^2+cz^2+du^2$ for each (a,b,c,d) in Ramanujan's list ([14]). One can find numerous papers in the literature that address sporadic cases of this strategy (see for eg. [1], [2], [3]). The purpose of this paper was to acquaint the student with the theoretical framework through which all of these papers can be understood.

Acknowledgment. I would like to thank Professor Ram Murty for introducing me to this problem and guiding me in my Master's study during which I worked on this problem and obtained the results of section 4. I also thank the referee for detailed comments on the original version of this article.

References

[1] Alaca, A., Alaca, S., Lemire, M. F. and Williams, K. S., Nineteen quaternary quadratic forms, *Acta Arith.*, **130** (2007), no. 3, 277–310.

[2] Alaca, A., Alaca, S., Lemire, M. F. and Williams, K. S., Theta function identities and representations by certain quaternary quadratic forms II, *Int. Math. Forum* **3** (2008), no. 9-12, 539–579.

[3] Alaca, A. and Alanazi, J., Representations by quaternary quadratic forms with coefficients 1, 2, 7 or 14, *Integers* **16** (2016), Paper No. A55, 16 pp.

[4] Bhargava, M., On the Conway-Schneeberger fifteen theorem, *Quadratic forms and their applications* (Dublin, 1999), 27–37; Contemp. Math., **272**, Amer. Math. Soc., Providence, RI, 2000.

[5] Diamond, F. and Shurman, J., *A First Course in Modular Forms*, Graduate Text in Mathematics **228**, Springer-Verlag, 2004.

[6] Hardy, G. H., Seshu Aiyar, P. V. and Wilson, B. M., *Collected papers of Srinivasa Ramanujan*, 310–321, AMS Chelsea Publ., Providence, RI, 2000.

[7] Kilford, L. J. P., *Modular Forms, A Classical and Computational Introduction*, Imperial College Press, London, 2008.

[8] Koblitz, N., *Introduction to Elliptic Curves and Modular Forms*, Springer-Verlag, New York, 1984.

[9] Legendre, A. M., *Essai sur la théorie des nombres*, Duprat, Paris, 1798.

[10] Liouville, J., Sur les deux formes $x^2+y^2+2(z^2+t^2)$, *J. Pures Appl. Math.*, **5** (1860), 269–272.

[11] Liouville, J., Sur les deux formes $x^2 + y^2 + z^2 + 2t^2$, $x^2 + 2(y^2 + z^2 + t^2)$, *J. Pures Appl. Math.*, **6** (1861), 225–230.
[12] Miyake, T., *Modular Forms, Springer Monographs in Mathematics*, Springer-Verlag, Berlin, English edition, 2006. Translated from the 1976 Japanese original by Yoshitaka Maeda.
[13] Ram Murty, M., Dewar, M. and Graves, H., *Problems in the Theory of Modular Forms*, IMSc Lecture Notes No.**1**, Hindustan Book Agency, 2015.
[14] Ramanujan, S., On the expression of a number in the form $ax^2 + by^2 + cz^2 + du^2$, *Proceedings of the Cambridge Philosophical Society, XIX*, (1917), 11–21.
[15] Rankin, R. A. *Modular Forms and Functions*, Cambridge University Press, Cambridge, 1977.
[16] Lang, S., *Introduction to Modular Forms*, Springer-Verlag, Berlin, Heidelberg, 1976.
[17] Serre, J.-P., *A Course in Arithmetic*, Springer-Verlag, New York, 1973.
[18] Stein, W., *Modular Forms, a Computational Approach*, Graduate Studies in Mathematics, American Math Society, Volume **79**, Rhode Island, 2007.

Arpita Kar
Department of Mathematics, Queen's University
Kingston, Ontario K7L 3N6, Canada.
E-mail: *arpita@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 115-119

EULER'S PARTITION IDENTITY AND TWO PROBLEMS OF GEORGE BECK

GEORGE E. ANDREWS

(Received : 23 - 04 - 2017 ; Revised : 26 - 04 - 2017)

Abstract. Euler's famous partition identity asserts that the number of partitions of an integer n into odd parts equals the number of partitions of n into distinct parts. This paper examines what happens if one even part might be allowed among the odd parts or one part might be repeated thrice among distinct parts. This study yields proofs of two conjectures of George Beck.

1. Introduction

Our starting point is sequence A090867 in the On-Line Encyclopedia of Integer Sequences [3]. The sequence in question, $a(n)$, counts the number of partitions of n such that the set of even parts has only one element. Thus $a(5) = 4$ where the relvant partitions are $4+1$, $3+2$, $2+2+1$ and $2+1+1+1$.

The sequence $a(n)$ is a natural one to investigate in the light of Euler's theorem [1, p. 5, Cor. 1.2]:

The number of partitions of n into odd parts equals the number of partitions of n into distinct parts.

Thus, the partitions counted by $a(n)$ are much like Euler's partitions with odd parts except now a single even number occurs as a part (possibly repeated).

Also on the page for A090867, we find the following conjecture by George Beck:

Conjecture. $a(n)$ is also the difference between the number of parts in the odd partitions of n and the number of parts in the distinct partitions of n (offset 0). For example, if $n = 5$, there are 9 parts in the odd partitions of 5 (5, 311, 11111) and 5 parts in the distinct partitions of 5 (5, 41, 32), with difference 4.

– George Beck, Apr 22, 2017

Let us define $b(n)$ to be the difference between the number of parts in the odd partitions of n and the number of parts in the distinct partitions of n.

While we are at it, let us define $c(n)$ to be the number of partitions of n in which exactly one part is repeated. Thus $c(5) = 4$ with the relevant partitions being $3+1+1$, $2+2+1$, $2+1+1+1$, and $1+1+1+1+1$.

Theorem 1. *For all $n \geq 1$, $a(n) = b(n) = c(n)$.*

2010 Mathematics Subject Classification : 11P83
Key words and phrases : Euler's partition identity, partitions.

So far we have seen that the theorem is true for $n = 5$. Our proof relies on generating functions and differentiation. There are many combinatorial proofs of Euler's Theorem (cf. [2]). Can one prove this new theorem combinatorially?

In section 3, we treat a further conjecture of George Beck also related to Euler's theorem.

2. Proof of the Theorem 1

Proof. As mentioned in the introduction, we require the generating functions for our sequences:

$$A(q) = \sum_{n\geq 0} a(n)q^n, \tag{2.1}$$

$$B(q) = \sum_{n\geq 0} b(n)q^n, \tag{2.2}$$

and

$$C(q) = \sum_{n\geq 0} c(n)q^n. \tag{2.3}$$

To prove our theorem, we shall show that each of $A(q)$, $B(q)$ and $C(q)$ is equal to

$$\prod_{n=1}^{\infty} \frac{1}{1-q^{2n-1}} \sum_{m=1}^{\infty} \frac{q^{2m}}{1-q^{2m}}. \tag{2.4}$$

The equality of the generating functions then proves the theorem.

Throughout our proof we shall require the following elegant, elementary identity of Euler used by him to prove his theorem [1, p.5, eq. (1.2.5)] for $|q| < 1$,

$$\prod_{n=1}^{\infty} \frac{1}{1-q^{2n-1}} = \prod_{n=1}^{\infty} \frac{(1-q^{2n})}{(1-q^n)} = \prod_{n=1}^{\infty} (1+q^n) \tag{2.5}$$

Let us do the easiest part first.

$$\begin{aligned} C(q) &= \sum_{n=1}^{\infty} (q^{2n} + q^{2n+2n} + q^{2n+2n+2n} + \cdots) \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \\ &= \left(\sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}} \right) \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}}, \end{aligned}$$

and we have establish that $C(q)$ is the expression in (2.4).

The next easiest part is $A(q)$. Clearly $A(q)$ is the coefficient of z in

$$\prod_{n=1}^{\infty} (1 + q^n + zq^{n+n} + zq^{n+n+n} + \cdots) = \prod_{n=1}^{\infty} \left(1 + q^n + \frac{zq^{2n}}{1-q^n} \right),$$

and the coefficient of z is:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^n} \prod_{\substack{m=1 \\ m\neq n}}^{\infty} (1+q^m) &= \prod_{m=1}^{\infty} (1+q^m) \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^n} \cdot \frac{1}{1+q^n} \\ &= \prod_{m=1}^{\infty} (1+q^m) \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}} \end{aligned}$$

$$= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}}$$

by (2.5). Hence $A(q)$ is also equal to the expression in (2.4).

Finally, we have the trickier problem of $B(q)$. In the infinite product

$$\prod_{n=1}^{\infty} \frac{1}{1-zq^{2n-1}},$$

the coefficient of $z^M q^N$ is the number of partitions of N into M odd parts, and in the infinite product

$$\prod_{n=1}^{\infty} (1+zq^n),$$

the coefficient of $z^M q^N$ is the number of partitions of N into M distinct parts [1, Ch. 2, p. 16].

So if we differentiate each of these functions with respect to z we will then be counting each partition with M parts with weight M.

Consequently

$$\begin{aligned}
B(q) &= \frac{\partial}{\partial z}\Big|_{z=1} \left(\prod_{m=1}^{\infty} \frac{1}{(1-zq^{2m-1})} - \prod_{m=1}^{\infty} (1+zq^m) \right) \\
&= \sum_{n=1}^{\infty} \frac{q^{2n-1}}{(1-q^{2n-1})^2} \prod_{\substack{m=1 \\ m\neq n}}^{\infty} \frac{1}{1-q^{2m-1}} - \sum_{n=1}^{\infty} q^n \prod_{\substack{m=1 \\ m\neq n}}^{\infty} (1+q^m) \\
&= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} - \prod_{m=1}^{\infty} (1+q^m) \sum_{n=1}^{\infty} \frac{q^n}{1+q^n} \\
&= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \left(\sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} - \sum_{n=1}^{\infty} \frac{q^n}{1+q^n} \right) \quad \text{(by (2.5))} \\
&= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \left(\sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} - \sum_{n=1}^{\infty} \frac{q^n(1-q^n)}{1-q^{2n}} \right) \\
&= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}} \\
&\quad + \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \left(\sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} - \sum_{n=1}^{\infty} \frac{q^n}{1-q^{2n}} \right), \\
&= \prod_{m=1}^{\infty} \frac{1}{1-q^{2m-1}} \sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}}, \qquad (2.6)
\end{aligned}$$

because

$$\sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} q^{m(2n-1)} = \sum_{m=1}^{\infty} \frac{q^m}{1-q^{2m}},$$

and (2.6) establishes that $B(q)$ is also equal to the expression in (2.4). □

3. George Beck's Second Problem

In the On-Line Encyclopedia of Integer Sequences, we also find sequence A265251. The sequence in question, $a_1(n)$, is the number of partitions of n such that there is exactly one part ocurring three times while all other parts occur only once. George Beck made the following:

Conjecture. $a_1(n)$ is also the difference between the number of parts in the distinct partitions of n and the number of distinct parts in the odd partitions of n (offset 0). For example, if $n = 5$, there are 5 parts in the distinct partitions of 5 (5, 41, 32) and 4 distinct parts in the odd partitions of 5 (namely, 5, (3, 1), 1 in 5, 311, 11111) with difference 1.

–George Beck, Apr 22, 2017

Here we define $b_1(n)$ to be the difference between the total number of parts in the partitions of n into distinct parts and the total number of different parts in the partitions of n into odd parts.

Theorem 2. $a_1(n) = b_1(n)$.

Proof. We let

$$A_1(q) = \sum_{n\geq 0} a_1(n)q^n$$

and

$$B_1(q) = \sum_{n\geq 0} b_1(n)q^n$$

. As in the proof of Theorem 1, we see that $A_1(q)$ is the coefficient of z in

$$\prod_{n=1}^{\infty}(1+q^n+zq^{3n}).$$

Hence

$$A_1(q) = \prod_{n=1}^{\infty}(1+q^n)\sum_{n=1}^{\infty}\frac{q^{3n}}{1+q^n}. \tag{3.1}$$

Next, as in our treatment of $B(q)$, we see that $B_1(q)$ must be

$$\begin{aligned} B_1(q) &= \frac{\partial}{\partial z}\Big|_{z=1}\left(\prod_{n=1}^{\infty}(1+zq^n) - \prod_{n=1}^{\infty}\left(1+\frac{zq^{2n-1}}{1-q^{2n-1}}\right)\right) \\ &= \prod_{n=1}^{\infty}(1+q^n)\sum_{m=1}^{\infty}\frac{q^m}{1+q^m} - \prod_{n=1}^{\infty}\frac{1}{1-q^{2n-1}}\sum_{m=1}^{\infty}q^{2m-1} \end{aligned}$$

Hence by (2.5)

$$\begin{aligned} &A_1(q) - B_1(q) \\ &= \prod_{n=1}^{\infty}(1+q^n)\left(\sum_{n=1}^{\infty}\frac{q^{3n}}{1+q^n} - \sum_{n=1}^{\infty}\frac{q^n}{1+q^n} + \frac{q}{1-q^2}\right) \end{aligned}$$

$$
\begin{aligned}
&= \prod_{n=1}^{\infty}(1+q^n)\left(-\sum_{n=1}^{\infty}\frac{q^n(1-q^{2n})}{1+q^n}+\frac{q}{1-q^2}\right)\\
&= \prod_{n=1}^{\infty}(1+q^n)\left(-\sum_{n=1}^{\infty}q^n(1-q^n)+\frac{q}{1-q^2}\right)\\
&= \prod_{n=1}^{\infty}(1+q^n)\left(-\frac{q}{1-q}+\frac{q^2}{1-q^2}+\frac{q}{1-q^2}\right)\\
&= 0,
\end{aligned}
$$

and Theorem 2 is proved. □

4. Conclusion

It would be very interesting to provide bijective proofs of any of the assertions in our theorems. As we noted, there are many bijective proofs of Euler's theorem.

It might also be interesting to examine what would happen if we were to allow repetitions of two different parts or appearances of two even parts, but the differentiation technique suggests that the resulting theorems would be messy and somewhat unattractive.

References

[1] Andrews, G. E., *The Theory of Partitions*, Addison-Wesley, Reading, 1976 (Reprinted: Cambridge University Press, Cambridge, 1985).

[2] Andrews, G. E., Euler's partition identity-finite version, *The Math. Student*, **85**, Nos. 3-4, (2016), 99–102.

[3] *The On-Line Encyclopedia of Integer Sequences*, Sequence A090867, https://oeis.org.

[4] *The On-Line Encyclopedia of Integer Sequences*, Sequence A265251, https://oeis.org.

George E. Andrews
Department of Mathematics
The Pennsylvania State University
University Park, PA 16802, USA
E-mail *gea1@psu.edu*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 121-131

THE ART OF RESEARCH

M. RAM MURTY

As we all know, there is no simple algorithm for research. There is no recipe for making new discoveries. It is a mysterious and inscrutable process. However, we know that this process has some guiding principles. It is the purpose of this article[1] to discuss these principles in a general way, illustrating them with examples from science and mathematics. Naturally, these examples will have some personal bias.

So let us begin. What exactly is research? In one sentence, it can be said to be the art of asking good questions. In our search for understanding, the SOCRATIC method of questioning is the way.

Let us observe that the word 'Question' has as root word 'Quest'. In our quest for understanding, the method of questioning seems to be the only way. Socrates taught Plato that all ideas must be examined critically and fundamental questions must be asked and pursued in order to gain proper understanding. Buddha instructed his disciple Ananda to question, to reflect deeply. As most of you know, Buddha advocated clear thinking. Socrates was adamant about definitions and in mathematics too, definitions are very important.

This method is not infallible. But it is the only way we have available. Some basic questions seem to defy simple answers. But this doesn't stop us from asking them. Often, the inquiry is a good exercise for the mind, and maybe it is the exercise that is the most important thing rather than complete understanding. Nevertheless, one can enquire into the nature of understanding itself. But then, this would take us into the realm of philosophy. This is not our goal here.

Our goal here is to explore how asking proper questions leads to some form of knowledge and understanding. We must keep in mind that each person brings their own past knowledge and experience to deal with the question. Each one brings their own methodology. Let us take an example

In a room, there are five people: an engineer, a physicist, a mathematician, a philosopher and an accountant. They are asked the simple question: what is 2+2?

* Research partially supported by an NSERC Discovery grant.

[1] This article is based on a public lecture given at the Tata Institute of Fundamental Research in Mumbai, India, some years ago. It is presented here in the hope that students may profit from it so that the ideas can gain a wider circulation

The engineer takes out his calculator and says the answer is 3.99. The physicist runs an experiment and finds the answer is between 3.8 and 4.2. The mathematician says he doesn't know the answer but can show that it exists. The philosopher asks for the meaning of the question. The accountant closes all doors and asks, 'What would you like the answer to be?'

Let us begin with some famous questions. What is life? This is the most basic question. It is related to 'What is consciousness?' and we still don't have a satisfactory answer. What is time? This is one of the most difficult questions to deal with and invariably takes us simultaneously into physics and philosophy. What is space? What is light? These questions have baffled the physicists for centuries and much of modern physics is the outcome of this inquiry. What is a number? This question has led to the development of mathematics. There are other questions that seem unrelated to any of these and are seemingly simple, like 'What is a knot?' Yet, on inquiry, we find it is related to the notion of number and the notion of light, as I will briefly indicate at the end of this article.

As I said, existential questions take us into the realm of philosophy. But there are other perhaps simpler questions one can ask and the inquiry into them quickly leads us to some understanding. So how to ask 'good' questions? What is a 'good' question? It is one that leads us to new discoveries. Below, we will present eight methods of generating good questions and we call this the 'Eightfold way', to borrow a phrase from Buddhist philosophy.

The simplest method of generating good questions is the survey method. This method consists of two steps. After selecting the topic to survey, we gather all the facts about the topic and then organize them. Arrangement of ideas leads to understanding. The amazing thing is that in this process, what is missing is also revealed. The method quickly leads to fundamental questions.

A good example is given by the discovery of the periodic table. Dimitri Mendeleev organized the existing knowledge of the elements and was surprised to find a periodicity in the properties of the elements. Mendeleev was born in 1834 in Siberia and was the last of 17 children. Those who think that writing graduate level textbooks is not research perhaps should think again! In the process of writing a student text in chemistry, Mendeleev decided to gather all the facts then known about the elements and organize them according to atomic weight. In this way, he was able to predict the existence of new elements. In 1875, six years after Mendeleev published his periodic table, the first of his predicted elements was discovered. This was gallium, which is an essential component of the electronics industry. For example, the liquid crystal displays in digital watches and calculators are based on gallium technology. Shortly after the discovery of gallium,

scandium was found. Then, germanium was discovered. These names also suggest the nationalities of the discoverers. The race for finding the missing elements soon became a form of national pride! And the race was on! Every time someone discovered a missing element, they got the Nobel Prize in chemistry! Today more than a century after Mendeelev suggested the periodic table, it was finally complete. It now sits as the presiding deity in all chemistry laboratories.

A similar survey method can be found in the writings of the mathematician David Hilbert. Born in 1862, Hilbert studied under Lindemann (who first proved that π is transcendental) and obtained his doctorate in 1885 from the University of Göttingen. Hilbert's list of 69 doctoral students is quite illustrious and includes Courant, Hecke, Takagi, Weyl and Zermelo.

Hilbert's approach to mathematics has always been that of an organizer of knowledge. He would set out to write a definitive textbook on a specific area of mathematics and invariably would find new and fundamental questions the answers to which led to rudimentary discoveries in mathematics. In 1900, at the International Congress of Mathematicians in Paris, Hilbert organized 23 problems which he considered important in mathematics. Six of these problems deal with the notion of number and have acted as a catalyst in the development of number theory.

- The 7th problem led to the development of transcendental number theory.
- The 8th problem is the Riemann hypothesis that plays a major role in analytic number theory.
- The 9th problem led to the development of reciprocity laws in algebraic number theory.
- The 10th problem led to the development of logic and diophantine set theory.
- The 11th problem led to the theory of quadratic forms and the 12th to class field theory, which began as a part of algebraic number theory and now is expanding into the realm of representation theory.

At the dawn of the 21st century, a similar program was launched. In the year 2000, the Clay Mathematical Institute designated 7 problems of mathematics as millenium problems and is offering a prize of one million dollars U.S. for the solution of any of the following problems:

- 1. $P = NP$;
- 2. The Riemann hypothesis;
- 3. The Birch and Swinnerton-Dyer conjecture;
- 4. The Poincaré conjecture;
- 5. The Hodge conjecture;
- 6. The Navier-Stokes equations;

- 7. The Yang-Mills theory.

As far as I know, only one of these problems has been solved and that is the Poincaré conjecture. In 2003, Grigori Perelman in a series of papers posted on the arxiv [4], [5], [6] settled the Poincaré conjecture but turned down the million dollar prize! In 2006, he was awarded the Fields Medal but again turned it down saying that "I am not interested in money or fame. I don't want to be on display like an animal in a zoo!" [1]

Further details about the Clay problems can be found at www.claymath.org. There the reader will find survey lectures in video format. You will undoubtedly find many more questions that need to be answered from these surveys. Thus, we see that the survey method is a powerful way to generate fundamental research questions.

The next method is the method of observations. Careful observations lead to patterns and patterns lead to the question why? In physics, the famous 1887 experiment of Michelson and Morley to determine the speed of light, first with reference to a stationary frame of reference and next with reference to a moving frame of reference was based on careful observations. They found that the velocity of light is constant and no evidence for the postulated ether. This was revolutionary and led to the special theory of relativity by Albert Einstein.

Another example of the power of observation leading to discovery is the apocryphal story of Archimedes. King Hiro commissioned a goldsmith to make a crown and was wondering if the goldsmith had stolen some gold. So he asked Archimedes to find out without destroying the crown. Archimedes started to ponder this and went to take a bath and noticed that the volume of water displaced was proportional to his weight. Immediately, his mind made the connection. The amount of gold given by the king should displace the same amount of water as the king's crown. If not, the goldsmith had taken some of the gold and replaced it with a baser metal. He was so happy with this discovery that he went running through the streets of Syracuse shouting "Eureka" (I have found it!) and forgot that he was taking a bath! Incidentally, the story is that the goldsmith had indeed cheated the king of some gold!

Often, we are unable to determine what impact our discovery will have. The role of the scientist is simply to investigate and report.

Careful observations lead to the discovery of patterns and consequently to conjectures. Certain conjectures gain prominence and act as powerful inducements for the development of a subject. Fermat's last theorem is a good example.

In 1637, Pierre de Fermat, who was a lawyer by profession, was reading Bachet's translation of the work of Diophantus when he came across the discussion

[1] BBC News, March 24, 2010.

of Pythagorean triples. This led him to wonder if the same works for higher powers and he was led to conjecture that for any $n > 2$, one cannot find three positive integers a, b, c such that

$$a^n + b^n = c^n.$$

Then, he wrote in the margin of the book that he had a wonderful proof of this fact, but the margin was too narrow to contain it!

Let us look more closely at Fermat's marginal note. First, the tome that Fermat was reading had very wide margins and so if Fermat had a proof, it must have been very long! One of my students told me that Fermat must have had a proof since he was a lawyer by profession and lawyers always tell the truth!

As most of you know, Fermat's last theorem was finally solved by a galaxy of mathematicians, culminating in the work of Andrew Wiles[11] in 1996. This proof certainly cannot be the one Fermat may have had in mind since it uses many ideas with which Fermat was unfamiliar with and hadn't been discovered yet. To trace the development of these new ideas, we look at another great mathematician who had the uncanny ability of making powerful and incisive conjectures.

This is Srinivasa Ramanujan, who discovered the importance of the τ-function and isolated it for further study. Ramanujan was born in the city of Erode in present-day Tamil Nadu. He is certainly one of the wonders to come out of the dust of India since he more or less educated himself by reading books and doing problems.

Ramanujan was never averse to making extensive calculations on his slate, since he didn't have much paper. Most of his findings, he would store in his brain. He studied the τ-function defined as follows. $\tau(n)$ is the coefficient of q^n in the infinite product expansion

$$q \prod_{r=1}^{\infty} (1 - q^r)^{24}.$$

He computed (see [7] and the table on the next page) by hand the first 30 values of the τ-function. What he observes is that τ is multiplicative and he makes his three famous conjectures concerning its behaviour.

(1) $\tau(mn) = \tau(m)\tau(n)$ for $(m, n) = 1$.
(2) for p prime, and $a \geqslant 1$, we have $\tau(p^{a+1}) = \tau(p)\tau(p^a) - p^{11}\tau(p^{a-1})$.
(3) $|\tau(p)| \leqslant 2p^{11/2}$, p prime.

The first two of his conjectures were proved by Mordell [2] the year after Ramanujan made the conjecture and third one defied attempts by many celebrated mathematicians, until 1974, when Deligne[1] solved it as a consequence of the Weil conjectures. For this work, Deligne was awarded the Fields medal. As we shall see, these conjectures play a vital role in the final solution of Fermat's last theorem. For more details, the reader is referred to the forthcoming monograph [3].

n	$\tau(n)$	n	$\tau(n)$
1	1	16	987136
2	-24	17	-6905934
3	252	18	2727432
4	-1472	19	10661420
5	4830	20	-7109760
6	-6048	21	-4219488
7	-16744	22	-12830688
8	84480	23	18643272
9	-113643	24	21288960
10	-115920	25	-25499225
11	534612	26	13865712
12	-370944	27	-73279080
13	-577738	28	24647168
14	401856	29	128406630
15	1217160	30	-29211840

The fourth method of generating questions (and perhaps the most difficult one) is by the method of re-interpretation. Why I have listed it here will become apparent by the end of the article. This method tries to examine what is known from a new vantage point. An excellent example is given by gravitation.

As most of you know, Isaac Newton first formulated the mathematical theory of universal gravitation. However, much of his theory relied on careful observations that Tycho Brahe and Johannes Kepler made concerning planetary orbits.

For Isaac Newton, gravity is a force and he was able to formulate the inverse square law $F = Gm_1m_2/r^2$ familiar to all of us from high school. On the other hand, for Albert Einstein, gravity is curvature of space.

So let us see how Einstein re-interpreted Newton's theory of gravitation. The surface of the universe is a 3-dimensional manifold. A sun or a planet kind of sits on this surface and consequently distorts the space around it depending on how massive it it. This view has serious implications to the behaviour of light. Thus, one of the consequences is the study of light in such gravitational fields. Light, as it travels on this surface must be therefore influenced by the distortions of space caused by massive gravitational fields and so it was predicted that such a phenomenon can probably be observed during a total eclipse of the sun. So in 1919, scientists were able to verify this phenomenon.

This was one spectacular victory for the theory of relativity. Its mathematical predictions were verified by numerous experiments. Perhaps the most spectacular illustration of this bending of light was the discovery in 1979 of a twin quasar.

It was long predicted that if a massive galaxy was in the line of sight between us and a quasar (quasi-stellar object of the size of our solar system) we would see a "double" and sure enough, this was verified in 1979 and this bending of light phenomenon is now a powerful tool in astrophysics.

Another anomaly resolved by relativity that Newton's theory could not explain is the orbit of Mercury, which was noticed to be not a perfect ellipse. It doesn't quite close upon itself and is called the precession of the perihelion of Mercury. Einstein's theory of relativity could explain this coming from the gravitational field curvature, since Mercury was closest to the sun, and so would feel this effect more than the other planets. This too has now been verified.

Perhaps the most powerful prediction of relativity theory is the existence of black holes, and it was in 1928 that Subramanyan Chandrasekhar worked out this consequence as a graduate student. As we all know, we see objects because light is reflected off of them. When a star dies, it can do one of three things: it can become very cold and become what is called a white dwarf; or it could explode and be a nova, or it can collapse into itself and become a black hole. All of these discoveries were possible only by re-interpreting gravity as curvature.

Let us look at an example of the method of re-interpretation in mathematics. Everyone is familiar with the unique factorization theorem. This says that every natural number can be written as a product of prime numbers uniquely. Euler reformulated this fact in an analytic fashion by introducing the zeta function. He did this by considering the Dirichlet series:

$$\sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Since every natural number can be written as a product of prime numbers uniquely, this series can be written as an infinite product over prime numbers:

$$\prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Euler gave an analytic proof of the infinitude of primes by noting that both sides converge absolutely for $\Re(s) > 1$ and when we take the limit as $s \to 1^+$, the series diverges and so the product must also diverge, showing the infinitude of primes.

However, it was Riemann who stressed that the zeta function must be studied as a function of a complex variable so that we can gain a better understanding of the distribution of prime numbers. As we shall see, this reformulation of the unique factorization theory re-emerges as a theory of Euler products in the famous Langlands program.

Yet another example of re-interpretation occurs in the work of Dedekind in algebraic number theory. Following some early work of Kummer, it was clear that the unique factorization theorem did not generally extend to the rings of

integers of algebraic number fields. Dedekind realized that one needed to replace the notion of a number by the notion of an ideal. He was led to this idea by re-interpreting divisibility. A natural number d divides n if and only if

$$d\mathbb{Z} \supseteq n\mathbb{Z}.$$

"To contain is to divide" became the aphorism for Dedekind's development of algebraic number theory. This re-interpretation transformed number theory and propelled it to major advances in the 19th and 20th centuries.

Another dynamic method for research is the method of analogy. When two theories are analogous, or exhibit some similarities, we try to see if ideas in one theory have analogous counterparts. For instance, the zeta function, and the Ramanujan's zeta function exhibit similarities in that they both have Euler products and functional equations. This analogy was first pushed by Erich Hecke in his study of the theory of modular forms. It also signalled the beginning of a general theory of L-functions and connected representation theory with number theory in a fundamental way. Building on the work of Harish-Chandra, Langlands showed how one can attach L-functions to representations of adele groups. This is the foundation for the Langlands program.

Another profound example of the method of analogy from physics is the Doppler effect. When a train approaches you, the pitch of sound is high and as it moves away, the pitch gets lower. This behaviour with sound waves was extended to light waves by Doppler and used to explain the red shift of stars. When the stars are approaching us, there is a red shift in their spectra and when they are moving away from us, there is a blue shift. This discovery was fundamental in explaining the expansion of the universe.

In a much more down-to-earth application of the Doppler effect, we see that police radar really makes use of the Doppler effect to record the speed of cars.

The strength of analogy is best illustrated in mathematics by the discovery of arithmetic of function fields over finite fields. Hilbert and others already noticed there was an analogy between complex function theory and algebraic number theory. But at the dawn of the 20th century, beginning with the doctoral work of Emil Artin, a new kind of zeta function was discovered which showed structural similarity to the Riemann zeta function but was much simpler to study. Artin conjectured the analog of the Riemann hypothesis for his zeta function and this was proved later by Hasse. But these reflections led Weil to study the zeta functions attached to curves and show the Riemann hypothesis held for these functions as well. Finally, in his epochal paper of 1949 [8], he formulated what became known as the Weil conjectures and these were settled by Deligne [1] in 1974, a part of which led to the proof of the Ramanujan conjecture. In his reflective essay [9], Weil records how he was led to his conjectures. "In 1947, in Chicago, I felt bored

and depressed, and, not knowing what to do, I started reading Gauss's two memoirs on biquadratic residues, which I have never read before. The first one deals with the number of solutions of $ax^4 - by^4 = 1$ over finite fields and the second one with $ax^3 - by^3 = 1$. Then I noticed similar principles can be applied to all equations of the form $ax^m + by^n + cz^r + \cdots = 0$ and this implies the truth of the so-called Riemann hypothesis for diagonal equations."

The Rosetta stone was discovered in 1799 and inscribed in the stone were three scripts: hieroglyphics, demotic and ancient Greek. Since scholars knew ancient Greek, they could decipher the other two scripts. It was in this way, the Egyptian hieroglyphs were decoded. Weil makes the analogy to the Rosetta stone when he compares the analogy between the number field, the function field over the finite field case and the complex function theoretical frame with its rich legacy of algebraic topology. The fascinating account is recorded in [10].

The method of transfer is to transfer an idea from one area of study to another. Again, a good example is again of the Doppler effect used in weather prediction. Microwaves are bounced off clouds to see if there are particles there that will cause precipitation and if so, how fast these clouds will be approaching us. As we all know, this is not a fool-proof method but it is approximately true and is a good illustration of the principle of transfer.

A seventh method is induction. This is essentially the method of generalization. Here is a simple example of how one uses the method of induction.

$$1^3 + 2^3 + \cdots n^3 = (1 + 2 + \cdots n)^2 .$$

A more sophisticated example is from the theory of L-functions alluded to earlier. $GL(1)$ and $GL(2)$ are two layers of a larger hierarchy. The Langlands program was largely suggested by induction.

The converse method of generating questions is simple enough. Whenever A implies B, we can ask if B implies A. This is called the converse question. A good example occurs in physics in the discovery of electromagnetism.

Around 1820, Oersted performed a historic experiment to show that an electric current creates a magnetic field. It was only a question of time before some one asked if the converse is true? That someone was Michael Faraday. Shortly after, he showed by experiment that the converse was true. A magnetic field creates an electric current.

The story is that when Faraday gave a public lecture demonstrating electromagnetic induction, the prime minister asked him of what use is it. Faraday responded by saying, "I don't know, but I am sure that someday you will figure out a way to tax it!" And he was right!

The converse method was also fundamental in the resolution of Fermat's last theorem. We have seen that the Riemann zeta function and the Ramanujan zeta functions have similar properties. We also learned that Langlands constructs

many more zeta functions from automorphic representations. The question of whether all such objects arise from automorphic representations is called converse theory in the Langlands program. Langlands proved a 2-dimensional special case of a prediction of this theory.

This was the starting point for the proof by Wiles of Fermat's last theorem. To compress three centuries of history is difficult. However many mathematicians played a vital role in the genesis of the solution: Fermat, Euler, Kummer, Riemann, Ramanujan, Hecke, Rankin, Selberg, Taniyama, Shimura, Weil, Iwasawa, Frey, Serre, Mazur, Ribet, Langlands, Taylor and Wiles. Inspired by some earlier work of Hasse, Taniyama predicted that L-functions attached to elliptic curves come from automorphic representations. This was made a bit more precise by Shimura and Weil. Then in 1985, Frey (and independently Hellegouarch), noticed that such a conjecture may imply Fermat's last theorem. This connection was then made more precise in some fundamental conjectures of Serre and Mazur, and then Ribet proved a special case of these conjectures. Ribet then showed that Taniyama's conjecture implies Fermat's last theorem. It was at this point that Wiles was inspired to prove the Taniyama conjecture. Beginning with the fundamental work of Langlands, he showed how one can construct a modular form whose L-series is the same as the L-series of a given elliptic curve over the rationals. At first, his announced proof of 1995 had a gap in it which was subsequently corrected in a joint paper of his with Taylor. With this, the proof of FLT was complete.

What are the future directions of research? In the last two decades, some new connections have been discovered linking Feynman diagrams, knot theory, zeta functions, and more generally, multiple zeta functions. This is a novel theme linking number theory and physics and will undoubtedly inspire many more discoveries. This is the way science progresses: through small steps by innumerable researchers. This gives us hope. We can all join in the adventure of expanding human knowledge.

To summarise, we have seen there are eight methods of generating good questions in the art of research. They are

- Survey,
- Observations,
- Conjectures,
- Re-interpretation,
- Analogy,
- Transfer,
- Induction and
- Converse,

giving us the acronym of SOCRATIC.

Science has come very far in expanding our vision. The Hubble telescope has been able to look very deep into outer space, as far as the coma cluster of galaxies whose movement subtantiates the claim for dark matter. If we are able to see this far, it is not that we have stood on the shoulders of giants, but rather because it is the power of the human mind to question, to inquire, that we have exercised.

REFERENCES

[1] Deligne, P., La conjecture de Weil. I, *Publications Mathématiques de l'IHÉS,* **43** (1974), 273–307.

[2] Mordell, L. J., On Mr. Ramanujan's empirical expansions of modular functions, *Proceedings of the Cambridge Philosophical Society,* **19** (1917), 117–124.

[3] Ram Murty, M., Panorama Lectures: The Ramanujan conjectures and zeta functions, *Ramanujan Mathematical Society*, (2017), to appear.

[4] Perelman, G., The entropy formula for the Ricci flow and its geometric applications, n arXiv.math.DG/0211159, November 11, 2002.

[5] Perelman, G., Ricci flow with surgery on three-manifolds, n arXiv.math.DG/0303109, March 10, 2003.

[6] Perelman, G., Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, n arXiv.math.DG/0307245, July 17, 2003

[7] Ramanujan, S., On certain arithmetical functions, *Transactions of the Cambridge Philosophical Society,* **22** (1916), 159–184.

[8] Weil, A., Number of solutions of equations in finite fields, *Bulletin of the American Math. Society,* **55** (1949), 497–508.

[9] Weil, A., Two lectures on number theory, past and present, *L'Enseignement Mathématique,* **20** (1974), 87–110.

[10] Weil, A., A 1940 Letter of André Weil on Analogy in Mathematics, Translated by Martin H. Krieger, *Notices of the American Math. Society,* **52** (2005), 334–341.

[11] Weil, A., Modular elliptic curves and Fermat's last theorem, *Annals of Mathematics,* **141** (1995), no. 3, 443–551.

M. Ram Murty
Department of Mathematics, Queen's University
Kingston, Ontario, K7L 3N6, Canada
E-mail: *murty@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 133-138

WEIERSTRASS INTERPOLATION OF HECKE EISENSTEIN SERIES

TIM HUBER AND MATTHEW LEVINE[1]

(Received : 23 - 04 - 2017, Revised : 08 - 05 - 2017)

Abstract. The Eisenstein series for the full modular group satisfy a well known recursion. We present a recursive formulation for Hecke Eisenstein series of odd level in terms of the Weierstrass zeta-function using only elementary facts about Gauss sums and L-functions. The Eisenstein series associated with primitive Dirichlet characters χ are expressed as linear combinations of Weierstrass zeta-values at division points.

1. Introduction

The Weierstrass $\wp$ function

$$\wp := \wp(z \mid \tau) = \frac{1}{z^2} + \sum_{n=1}^{\infty} (2n+1) G_{2n}(\tau) z^{2n}. \tag{1.1}$$

generates the holomorphic Eisenstein series on $SL(2,\mathbb{Z})$ with Fourier expansion

$$G_{2k}(\tau) = 2\zeta(2k) + \frac{4\zeta(2k)}{\zeta(1-2k)} \sum_{n=1}^{\infty} \frac{n^{k-1}q^n}{1-q^n}, \quad q = e^{2\pi i \tau},\ Im\ \tau > 0, \tag{1.2}$$

where ζ denotes the analytic continuation of the Riemann zeta function. Analytic properties of $\wp$ translate to properties for Eisenstein series. For example, the equation

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3, \quad g_2 = 60G_4(\tau), \quad g_3 = 140G_6(\tau), \tag{1.3}$$

demonstrates the recursion for Eisenstein series, with $d_k = k! \cdot (2k+3)G_{2k+4}$,

$$\frac{2n+9}{3n+6} d_{n+2} = \sum_{k=0}^{n} \binom{n}{k} d_k d_{n-k}, \qquad d_0 = 3G_4(\tau), \quad d_1 = 5G_6(\tau). \tag{1.4}$$

In this note, we provide an elementary formulation of a twisted analogue of (1.4) for normalized Eisenstein series of weight k on $\Gamma_0(N)$ twisted by the Dirichlet character χ modulo N [2, p. 17]. The Fourier expansion for these Eisenstein series is given by

2010 Mathematics Subject Classification: Primary 30B10; Secondary 11M36
Keywords and Phrases: Eisenstein series, L-functions, Gauss sums, Dirichlet characters, Weierstrass elliptic functions.
[1] The second author was supported by the UTPA Undergraduate Research Initiative.

$$E_{k,\chi}(q) = 1 + \frac{2}{L(1-k,\chi)} \sum_{n=1}^{\infty} \chi(n) \frac{n^{k-1} q^n}{1-q^n}, \tag{1.5}$$

where $L(1-k,\chi)$ is the analytic continuation of the associated L-series and $\chi(-1) = (-1)^k$. Although twisted Eisenstein series are of fundamental importance in number theory, explicit exposition of recursion formulas induced by their relation to elliptic functions does not appear in the extensive literature on elliptic modular functions.

2. Preliminaries

Define the Weierstrass zeta function with quasi-periods $\omega_1 = 2\pi, \omega_2 = 2\pi\tau$ by

$$\zeta(\theta) = \frac{1}{2} \cot \frac{\theta}{2} + \frac{\theta}{12} - 2\theta \sum_{n=1}^{\infty} \frac{nq^n}{1-q^n} + 2 \sum_{n=1}^{\infty} \frac{q^n \sin n\theta}{1-q^n}. \tag{2.1}$$

For $\{w_n\}_{n=0}^{k-1} \subset \mathbb{C}$, there exist k uniquely determined finite Fourier coefficients $\{a_n : n = 0, 1, ..., k-1\}$ such that

$$w_m = \sum_{n=0}^{k-1} a_n e^{2\pi i m n/k}, \qquad a_m = \frac{1}{k} \sum_{n=0}^{k-1} w_n e^{-2\pi i m n/k}, \qquad 0 \le m \le k-1. \tag{2.2}$$

Let χ be a Dirichlet character modulo k. The sum [1, p. 165]

$$G(n,\chi) = \sum_{m=1}^{k} \chi(m) e^{2\pi i m n/k} \tag{2.3}$$

is called the Gauss sum associated with χ. The Gauss sum is said to be separable if

$$G(n,\chi) = \overline{\chi}(n) G(1,\chi). \tag{2.4}$$

3. Main Results

Theorem 3.1. *Let $\zeta(\theta)$ denote the Weirstrass ζ function. For each odd $p \in \mathbb{N}$ and nonprincipal primitive Dirichlet character χ modulo p with $\chi(-1) = (-1)^k$, we have*

$$E_{k,\chi}(q) = \frac{-2}{(-i)^k p \cdot L(1-k,\chi)} \sum_{m=1}^{\frac{p-1}{2}} G(m,\chi) \zeta^{(k-1)} \left(\frac{2m\pi}{p} \right), \qquad k > 1. \tag{3.1}$$

If $k = 1$, the corresponding Eisenstein series satisfies

$$E_{1,\chi}(q) = \frac{2}{ipL(0,\chi)} \left(-\frac{\pi E_2(q) G(1,\chi)}{6p} \sum_{m=1}^{\frac{p-1}{2}} m\overline{\chi}(m) + \sum_{m=1}^{\frac{p-1}{2}} G(m,\chi) \zeta \left(\frac{2m\pi}{p} \right) \right), \tag{3.2}$$

where E_2 denotes the Eisenstein series for $SL(2,\mathbb{Z})$ with Fourier expansion

$$E_2(q) = 1 - 24 \sum_{n=1}^{\infty} \frac{nq^n}{1-q^n}.$$

Proof. From the partial fraction expansion for the cotangent,

$$\cot\theta = \sum_{n=1}^{\infty}\left(\frac{1}{\theta - n\pi} + \frac{1}{\theta + (n-1)\pi}\right), \tag{3.3}$$

we derive

$$\cot^{(\ell)}\theta = (-1)^{\ell}\ell!\sum_{n=1}^{\infty}\left(\frac{1}{(\theta - n\pi)^{\ell+1}} + \frac{1}{(\theta + (n-1)\pi)^{\ell+1}}\right). \tag{3.4}$$

If χ is a nonprincipal primitive Dirichlet character modulo p and $\chi(-1) = (-1)^{\ell+1}$, then (3.4) implies

$$\begin{aligned}
&\sum_{m=1}^{\frac{p-1}{2}}\chi(m)\cot^{(\ell)}\left(\frac{m\pi}{p}\right) \qquad (3.5)\\
&\quad = (-1)^{\ell}\ell!\frac{p^{\ell+1}}{\pi^{\ell+1}}\sum_{m=1}^{\frac{p-1}{2}}\sum_{n=1}^{\infty}\left(\frac{\chi(-m)}{(pn-m)^{\ell+1}} + \frac{\chi(m)}{(m+p(n-1))^{\ell+1}}\right)\\
&\quad = (-1)^{\ell}\ell!\frac{p^{\ell+1}}{\pi^{\ell+1}}\sum_{n=1}^{\infty}\frac{\chi(n)}{n^{\ell+1}} = (-1)^{\ell}\ell!\frac{p^{\ell+1}}{\pi^{\ell+1}}L(\ell+1,\chi).
\end{aligned}$$

The L-functions, for primitive Dirichlet characters modulo p, satisfy the functional equation

$$L(1-s,\chi) = \frac{p^{s-1}\Gamma(s)}{(2\pi)^s}\{e^{-\pi is/2} + \chi(-1)e^{\pi is/2}\}G(1,\chi)L(s,\overline{\chi}). \tag{3.6}$$

For any primitive Dirichlet character modulo p, we have

$$\sum_{h=1}^{p}\overline{\chi}(h)e^{2\pi inh/p} = \chi(n)G(1,\overline{\chi}). \tag{3.7}$$

Here $\overline{\chi}$ denotes the complex conjugate of the character $\overline{\chi}$. This formula implies that for any primitive character, the corresponding Gauss sum is separable [1, p. 165], so

$$G(m,\chi) = \overline{\chi}(m)G(1,\chi). \tag{3.8}$$

Therefore, with the assumption that χ and ℓ have opposite parity, we may use (3.5) and (3.6) to derive

$$\begin{aligned}
\sum_{m=1}^{\frac{p-1}{2}}\chi(m)cot^{(\ell)}\left(\frac{m\pi}{p}\right) &= (-1)^{\ell}\frac{p^{\ell+1}}{\pi^{\ell+1}}L(\ell+1,\chi)\ell!\\
&= (-1)^{\ell}\frac{p2^{\ell+1}L(-\ell,\overline{\chi})\ell!}{\Gamma(\ell+1)\{e^{-\pi i(\ell+1)/2} + \overline{\chi}(-1)e^{\pi i(\ell+1)/2}\}G(1,\overline{\chi})}\\
&= (-1)^{\ell}\frac{p2^{\ell+1}}{-2\cdot(-i)^{\ell}iG(1,\overline{\chi})}L(-\ell,\overline{\chi}) = \frac{-(-i)^{\ell+1}p2^{\ell}}{G(1,\overline{\chi})}L(-\ell,\overline{\chi}). \qquad (3.9)
\end{aligned}$$

For each integer k with $0 \leq k \leq (p-1)/2$, we have

$$\cos\left(\frac{2\pi k}{p}\right) = \cos\left(\frac{2\pi(p-k)}{p}\right), \tag{3.10}$$

$$\sin\left(\frac{2\pi k}{p}\right) = -\sin\left(\frac{2\pi(p-k)}{p}\right). \tag{3.11}$$

Therefore, from the above identities and (2.2)

$$\chi(m) = \sum_{n=0}^{p-1} a_n e^{2\pi i m n/p} = \begin{cases} 2\sum_{n=1}^{\frac{p-1}{2}} a_n \cos\left(\frac{2mn\pi}{p}\right), & \chi \text{ even}, \\ 2i\sum_{n=1}^{\frac{p-1}{2}} a_n \sin\left(\frac{2mn\pi}{p}\right), & \chi \text{ odd}, \end{cases} \tag{3.12}$$

where

$$a_n = \frac{1}{p}\sum_{n=0}^{p-1} \chi(n) e^{-2\pi i m n/p} = \begin{cases} \frac{2}{p}\sum_{n=1}^{\frac{p-1}{2}} \chi(n)\cos\left(\frac{2mn\pi}{p}\right), & \chi \text{ even}, \\ -\frac{2i}{p}\sum_{n=1}^{\frac{p-1}{2}} \chi(n)\sin\left(\frac{2mn\pi}{p}\right), & \chi \text{ odd}. \end{cases} \tag{3.13}$$

It readily follows that $G(m,\chi) = \chi(-1)pa_m$. Note the trivial equality

$$\frac{d^{k-1}}{d\theta^{k-1}}\sin(n\theta) = \begin{cases} -i^k n^{k-1}\cos(n\theta), & k \textit{ even}, \\ i^{k-1}n^{k-1}\sin(n\theta), & k \textit{ odd}. \end{cases} \tag{3.14}$$

Thus, if $\chi(-1) = (-1)^k$, and k is odd, we derive

$$2\sum_{n=1}^{\infty}\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\frac{n^{k-1}q^n \sin^{(k-1)}\left(\frac{2mn\pi}{p}\right)}{1-q^n} \tag{3.15}$$

$$= 2\sum_{n=1}^{\infty}\sum_{m=1}^{\frac{p-1}{2}} (-pa_m)\frac{n^{k-1}q^n \sin^{(k-1)}\left(\frac{2mn\pi}{p}\right)}{1-q^n}$$

$$= i^k p\sum_{n=1}^{\infty}\sum_{m=1}^{\frac{p-1}{2}} 2ia_m \sin\left(\frac{2mn\pi}{p}\right)\frac{n^{k-1}q^n}{1-q^n}$$

$$= -(-i)^k p\sum_{n=1}^{\infty}\frac{\chi(n)n^{k-1}q^n}{1-q^n}. \tag{3.16}$$

Similarly when $\chi(-1) = (-1)^k = 1$, the extreme sides of (3.15) and (3.16) are equal. Therefore, from the definition (2.1) of $\zeta(\theta)$ and by (3.9), for $k > 1$

$$\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\zeta^{(k-1)}\left(\frac{2m\pi}{p}\right)$$
$$= \frac{G(1,\chi)}{2\cdot 2^{k-1}} \sum_{m=1}^{\frac{p-1}{2}} \overline{\chi}(m)\cot^{(k-1)}\left(\frac{m\pi}{p}\right) + 2\sum_{n=1}^{\infty}\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\frac{n^{k-1}q^n \sin^{(k-1)}\left(\frac{2mn\pi}{p}\right)}{1-q^n}$$
$$= \frac{-(-i)^k pL(1-k,\chi)}{2}\left(1+\frac{2}{L(1-k,\chi)}\sum_{n=1}^{\infty}\frac{n^{k-1}\chi(n)q^n}{1-q^n}\right).$$

For $k=1$, we have

$$\begin{aligned}\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\zeta\left(\frac{2m\pi}{p}\right) &= \frac{1}{2}\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\cot\left(\frac{m\pi}{p}\right) + \frac{\pi}{6p}E_2(q)\sum_{m=1}^{\frac{p-1}{2}} mG(m,\chi)\\ &\quad + 2\sum_{m=1}^{\frac{p-1}{2}} G(m,\chi)\sum_{n=1}^{\infty}\frac{q^n\sin\left(\frac{2mn\pi}{p}\right)}{1-q^n}\\ &= \frac{1}{2}G(1,\chi)\sum_{m=1}^{\frac{p-1}{2}}\overline{\chi}(m)\cot\left(\frac{m\pi}{p}\right)\\ &\quad + \frac{\pi}{6p}E_2(q)G(1,\chi)\sum_{m=1}^{\frac{p-1}{2}} m\overline{\chi}(m) + ip\sum_{n=1}^{\infty}\frac{\chi(n)q^n}{1-q^n}\\ &= \frac{\pi}{6p}E_2(q)G(1,\chi)\sum_{m=1}^{\frac{p-1}{2}} m\overline{\chi}(m) + \frac{1}{2}ipL(0,\chi)E_{1,\chi}(q).\end{aligned}$$

This is equivalent to the claimed identity (3.2). □

Theorem 3.1 demonstrates that Hecke Eisenstein series of weight k on $\Gamma_0(p)$ may be expressed as a linear combination of $(k-1)$th order derivatives of the Weierstrass zeta-function at division points. The relation between the Weierstrass functions $\zeta'(\theta) = -\wp(\theta)$ and Equation (1.3) leads to

$$\zeta^{(n+3)}(\theta) = -6\sum_{k=1}^{n}\binom{n}{k}\zeta^{(k)}(\theta)\zeta^{(n-k+2)}(\theta), \qquad n\geq 1. \tag{3.17}$$

In particular, for p an odd prime, a recursive relation between the twisted Eisenstein series for $\Gamma_0(p)$ may be determined by inverting the relation from Theorem 3.1 between the $(p-1)/2$ Weierstrass zeta-values at division points and the $(p-1)/2$ linearly independent Eisenstein series. Although straightforward, the recursion is not as elegant as (1.4) for the Eisenstein series on $SL(2,\mathbb{Z})$. We conclude with an example from [3, 4] for $p=5$.

Theorem 3.2. *Let $\chi_{2,5}$ and $\chi_{4,5}$ be the odd Dirichlet characters modulo 5. For $k\geq 1$,*

$$E_{2k+1,\chi_{2,5}}(q) = 2(-1)^k G(1,\chi_{2,5}) \frac{\zeta^{(2k)}(4\pi/5) + i\zeta^{(2k)}(2\pi/5)}{5L(-2k,\chi_{2,5})}, \tag{3.18}$$

$$E_{2k+1,\chi_{4,5}}(q) = 2(-1)^{k-1} G(1,\chi_{4,5}) \frac{\zeta^{(2k)}(4\pi/5) - i\zeta^{(2k)}(2\pi/5)}{5L(-2k,\chi_{4,5})}, \tag{3.19}$$

where (3.17) *may be used to reduce the order of derivatives in* (3.18), (3.19), *and*

$$\zeta^{(2k)}(2\pi/5) = \frac{5(-1)^k}{4i}\left(\frac{L(-2k,\chi_{2,5})}{G(1,\chi_{2,5})} E_{2k+1,\chi_{2,5}}(q) + \frac{L(-2k,\chi_{4,5})}{G(1,\chi_{4,5})} E_{2k+1,\chi_{4,5}}(q)\right),$$

$$\zeta^{(2k)}(4\pi/5) = \frac{5(-1)^k}{4}\left(\frac{L(-2k,\chi_{2,5})}{G(1,\chi_{2,5})} E_{2k+1,\chi_{2,5}}(q) - \frac{L(-2k,\chi_{4,5})}{G(1,\chi_{4,5})} E_{2k+1,\chi_{4,5}}(q)\right).$$

References

[1] Apostol, T. M., *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976.

[2] Bruinier, J. H., Geer, G. van der, Harder, G. and Zagier, D., *The 1-2-3 of modular forms*, Universitext. Springer-Verlag, Berlin, 2008. Lectures from the Summer School on Modular Forms and their Applications held in Nordfjordeid, June 2004, Edited by Kristian Ranestad.

[3] Charles, R., Huber, T. and Mendoza, A., Parameterizations for quintic Eisenstein series. *Journal of Number Theory*, **133** (1) (2013), 195–214.

[4] Huber, T., On quintic Eisenstein series and points of order five of the Weierstrass elliptic functions, *The Ramanujan Journal*, **28** (2) (2012), 273–308.

Tim Huber
School of Mathematical and Statistical Sciences
University of Texas Rio Grande Valley
1201 West University Avenue, Edinburg, Texas 78539, USA
E-mail: *timothy.huber@utrgv.edu*

Matthew Levine
School of Mathematical and Statistical Sciences
University of Texas Rio Grande Valley
1201 West University Avenue, Edinburg, Texas 78539, USA
E-mail: *malevine@broncs.utpa.edu*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), 139-150

EVALUATION OF THE QUADRATIC GAUSS SUM

M. RAM MURTY[1] **AND SIDDHI PATHAK**

(Received : 06 - 05 - 2017 ; Revised : 17 - 05 - 2017)

ABSTRACT. For any natural number n and $(m,n)=1$, we analyse the eigenvalues and their multiplicities of the matrix $\mathcal{A}(n,m) := (\zeta_n^{mrs})$ for $0 \leq r,s \leq n-1$. As a consequence, we evaluate the quadratic Gauss sum and derive the law of quadratic reciprocity using only elementary methods.

1. INTRODUCTION

For natural numbers n and k, a general Gauss sum is defined as

$$\mathcal{G}(k) := \sum_{j=0}^{n-1} e^{2\pi i j^k/n}. \tag{1.1}$$

When $k=1$, (1.1) reduces to the sum of all n-th roots of unity, which is a geometric sum and can be easily evaluated to be zero. The case $k=2$ turns out to be more difficult, and it took Gauss several years to determine (1.1) when n is an odd prime in order to prove the law of quadratic reciprocity. For further reading on Gauss sums, we refer the reader to [2] and [3].

In this article, we focus on the quadratic Gauss sum, namely,

$$\mathcal{G}(2) = \sum_{j=0}^{n-1} e^{2\pi i j^2/n}. \tag{1.2}$$

It can be shown that

Theorem 1.1. *For a natural number n,*

$$\mathcal{G}(2) = \begin{cases} \sqrt{n} & \text{if } n \equiv 1 \bmod 4, \\ 0 & \text{if } n \equiv 2 \bmod 4, \\ i\sqrt{n} & \text{if } n \equiv 3 \bmod 4, \\ (1+i)\sqrt{n} & \text{if } n \equiv 0 \bmod 4. \end{cases}$$

There are many proofs of Theorem 1.1 in the literature. But most of the proofs use advanced tools. For example, [4] uses the theory of Fourier series, while [7] proves it using the truncated Poisson summation formula. The novelty of this article is that it utilizes only elementary methods, thus making the proof of Theorem 1.1 accessible to high school students. This linear algebra approach to

2010 Mathematics Subject Classification: 11L05, 11A15.

Key words and phrases : Gauss sum, quadratic reciprocity law.

[1] Research of the first author was partially supported by an NSERC Discovery grant.

evaluating (1.2) was initiated by Schur [9] in 1921 and simplified by Waterhouse [10] in 1970, to prove Theorem 1.1 when n is an odd prime. It was later expanded upon by the first author [8] to prove Theorem 1.1 for all n odd. The case n even was left open. In this note, we use a slight generalization of the method in these earlier works to prove Theorem 1.1 for even natural numbers n, hence determining (1.2) for all natural numbers n using only linear algebra and elementary number theory.

For clarity and continuity of exposition, we include the proof of Theorem 1.1 for n odd and the law of quadratic reciprocity in the earlier sections. We then use these results to prove Theorem 1.1 for n even.

2. Preliminary Results

Let n be a natural number and $\zeta_n := e^{2\pi i/n}$. For $(m,n) = 1$, we define the $n \times n$ matrix, $\mathcal{A}(n,m) = (\zeta_n^{mrs})$ for $0 \leq r, s \leq n-1$.

The motivation behind defining this matrix is the observation that

$$\operatorname{Tr} \mathcal{A}(n,1) = \sum_{j=0}^{n-1} \zeta_n^{j^2} = \mathcal{G}(2).$$

Let $\mathcal{A}(n,m)_{r,s}$ denote the (r,s)-th entry of $\mathcal{A}(n,m)$. Since the trace of a matrix is the sum of its eigenvalues counted with multiplicities, it suffices to find the eigenvalues of $\mathcal{A}(n,m)$ and their multiplicities. In order to compute the eigenvalues, observe that for $0 \leq k, l \leq n-1$,

$$\left(\mathcal{A}(n,m)\right)^2_{k,l} = \sum_{j=0}^{n-1} \zeta_n^{mkj} \zeta_n^{mjl} = \sum_{j=0}^{n-1} \zeta_n^{mj(k+l)}, \tag{2.1}$$

which is zero unless $m(k+l) \equiv 0 \pmod{n}$. Since $(m,n) = 1$, this is equivalent to the condition that $(k+l) \equiv 0 \pmod{n}$, in which case the sum is n because it is a geometric sum. In other words,

$$\mathcal{A}(n,m)^2 = \begin{bmatrix} n & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & n \\ 0 & 0 & 0 & \dots & n & 0 \\ \vdots & \vdots & \vdots & \iddots & \vdots & \\ 0 & 0 & n & \dots & 0 & 0 \\ 0 & n & 0 & \dots & 0 & 0 \end{bmatrix}. \tag{2.2}$$

Therefore,

$$\left(\mathcal{A}(n,m)\right)^4_{r,s} = \sum_{k=0}^{n-1} \left(\mathcal{A}(n,m)\right)^2_{r,k} \left(\mathcal{A}(n,m)\right)^2_{k,s},$$

and the summand is n^2 if $r + k \equiv 0 \pmod{n}$ and $s + k \equiv 0 \pmod{n}$ and zero otherwise. Thus, the summand is non-zero only when $r = s$ in which case it is n^2. This shows that

$$\left(\mathcal{A}(n,m)\right)^4 = n^2 I.$$

Hence, the eigenvalue of $\big(\mathcal{A}(n,m)\big)^4$ is n^2. By elementary linear algebra, we get that the eigenvalues of $\big(\mathcal{A}(n,m)\big)^2$ are n and $-n$. Consequently, the eigenvalues of $\mathcal{A}(n,m)$ are $\pm\sqrt{n}$ and $\pm i\sqrt{n}$. Let a,b,c,d be the multiplicities of $\sqrt{n}, -\sqrt{n}, i\sqrt{n}$ and $-i\sqrt{n}$ respectively. Thus,

$$\operatorname{Tr}\mathcal{A}(n,m) = \sqrt{n}\big((a-b)+i(c-d)\big), \tag{2.3}$$

for some natural numbers a,b,c and d.

Now, if $[x_0, x_1, \cdots, x_{n-1}]$ is an eigenvector of $\big(\mathcal{A}(n,m)\big)^2$ with eigenvalue n, then due to (2.1), it satisfies $x_i = x_{n-i}$ for $1 \le i \le n-1$. Hence, the dimension of the eigenspace corresponding to the eigenvalue n of $\big(\mathcal{A}(n,m)\big)^2$ is $(n+1)/2$ if n is odd and $n/2+1$ if n is even. Since the n-eigenspace of $\big(\mathcal{A}(n,m)\big)^2$ comprises of the $\pm\sqrt{n}$-eigenspace of $\mathcal{A}(n,m)$, we get the relations

$$a+b = (n+1)/2 \qquad \text{and} \qquad c+d = (n-1)/2, \tag{2.4}$$

when n is odd and

$$a+b = (n/2)+1 \quad \text{and} \quad c+d = (n/2)-1, \tag{2.5}$$

when n is even. Before proceeding, we prove the following lemma:

Lemma 2.1. *For any natural number n and $(m,n)=1$, let $\mathcal{A}(n,m) = (\zeta_n^{mrs})$ with $0 \le r,s \le n-1$. Then we have*

$$\big|\operatorname{Tr}\mathcal{A}(n,m)\big| = \begin{cases} \sqrt{n} & \text{if } n \text{ is odd,} \\ \sqrt{2n} & \text{if } n \equiv 0 \bmod 4, \\ 0 & \text{if } n \equiv 2 \bmod 4. \end{cases}$$

Proof. Observe that

$$\begin{aligned} \big|\operatorname{Tr}\mathcal{A}(n,m)\big|^2 &= \big(\operatorname{Tr}\mathcal{A}(n,m)\big)\big(\overline{\operatorname{Tr}\mathcal{A}(n,m)}\big) \\ &= \Big(\sum_{k=0}^{n-1}\zeta_n^{m\,k^2}\Big)\Big(\sum_{l=0}^{n-1}\zeta_n^{-m\,l^2}\Big) \\ &= \sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\zeta_n^{m(k^2-l^2)} \quad = \quad \sum_{l=0}^{n-1}\sum_{k=0}^{n-1}\zeta_n^{m(k-l)(k+l)}. \end{aligned}$$

The above sums depend only on the residue class of k and l modulo n and run over all residue classes mod n. Thus, for each fixed l mod n, we can make the linear change of variable $j = k-l$, which again runs over all residue classes mod n. Therefore, we have

$$\big|\operatorname{Tr}\mathcal{A}(n,m)\big|^2 = \sum_{l=0}^{n-1}\sum_{j=0}^{n-1}\zeta_n^{mj(j+2l)} = \sum_{j=0}^{n-1}\zeta_n^{mj^2}\sum_{l=0}^{n-1}\zeta_n^{2mjl}.$$

Since $(m,n)=1$, the inner sum is non-zero only if $2j \equiv 0 \pmod n$. If n is odd, then the only value of j which satisfies this congruence is $j=0$. Thus, $|\operatorname{Tr}\mathcal{A}(n,m)|$ evaluates to n when n is odd. If n is even, there are two values of j that satisfy the congruence, namely, $j=0$ and $j=n/2$. Hence, the sum becomes

$$\big|\operatorname{Tr}\mathcal{A}(n,m)\big|^2 = \big(1+\zeta_n^{mn^2/4}\big)n = (1+i^{mn})n,$$

which is zero if $n \equiv 2 \pmod 4$ and $2n$ if $n \equiv 0 \pmod 4$. This proves the lemma. □

Note that Lemma 2.1 proves Theorem 1.1 when $n \equiv 2 \pmod 4$. As a consequence of the above lemma, we have

Corollary 1. *For an odd natural number n,*

$$\operatorname{Tr} \mathcal{A}(n,m) = \begin{cases} \pm\sqrt{n} & \text{if } n \equiv 1 \bmod 4, \\ \pm i\sqrt{n} & \text{if } n \equiv 3 \bmod 4. \end{cases}$$

Proof. From (2.3),

$$\left|\operatorname{Tr} \mathcal{A}(n,m)\right| = \sqrt{n}\left((a-b)^2 + (c-d)^2\right)^{1/2}.$$

When n is odd, Lemma 2.1 leads us to deduce that either

$$(1)\ a-b = \pm 1 \text{ and } c = d, \qquad \text{or} \qquad (2)\ a = b \text{ and } c-d = \pm 1.$$

In Case (1), equation (2.4) implies that $c+d = 2d = (n-1)/2$, i.e, $d = (n-1)/4 \in \mathbb{N}$ and hence $n \equiv 1 \pmod 4$. In Case (2), equation (2.4) implies that $a+b = 2b = (n+1)/2$, i.e, $b = (n+1)/4 \in \mathbb{N}$ so that $n \equiv 3 \pmod 4$. □

We observe that the quadratic Gauss sums have the following multiplicative property.

Lemma 2.2. *For a natural number $n = n_1 n_2$ with $(n_1, n_2) = 1$ and $(m,n) = 1$, define $\mathcal{A}(n,m) = (\zeta_n^{mrs})$ for $0 \le r, s \le n-1$. Then we have,*

$$\operatorname{Tr} \mathcal{A}(n,m) = \operatorname{Tr} \mathcal{A}(n_1, mn_2) \operatorname{Tr} \mathcal{A}(n_2, mn_1).$$

Proof. The right hand side can be simplified as follows

$$\begin{aligned} \operatorname{Tr} \mathcal{A}(n_1, mn_2) \operatorname{Tr} \mathcal{A}(n_2, mn_1) &= \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e^{2\pi i m n_2 j^2/n_1}\, e^{2\pi i m n_1 k^2/n_2} \\ &= \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e^{2\pi i m ({n_2}^2 j^2 + {n_1}^2 k^2)/n_1 n_2} \\ &= \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e^{2\pi i m (n_2 j + n_1 k)^2/n}, \end{aligned}$$

as $e^{2\pi i m (2jk n_1 n_2)/n} = 1$. Now, since $(n_1, n_2) = 1$, the Chinese remainder theorem gives that as j and k range from 0 to $n_1 - 1$ and 0 to $n_2 - 1$ respectively, $n_2 j + n_1 k$ ranges over all residue classes modulo n. Hence, we have $\operatorname{Tr} \mathcal{A}(n_1, mn_2) \operatorname{Tr} \mathcal{A}(n_2, mn_1) = \sum_{r=0}^{n-1} e^{2\pi i m r^2/n} = \operatorname{Tr} \mathcal{A}(n,m)$. □

3. Proof of Theorem 1.1 for n odd

As seen earlier, $\mathcal{G}(2) = \operatorname{Tr} \mathcal{A}(n,1)$. Thus, we consider the case $m = 1$ in this section. Corollary 1 gives the value of the desired sum up to sign. To determine the sign in each case, we consider the determinant of the matrix $\mathcal{A}(n,1)$, which is the product of its eigenvalues counted with multiplicities.

Lemma 3.1. *Let $\mathcal{A} := \mathcal{A}(n,1) = (\zeta_n^{rs})$ for $0 \le r, s \le n-1$. Then*

$$\det \mathcal{A} = \begin{cases} i^{\binom{n}{2}}\, n^{n/2} & \text{if } n \text{ is odd}, \\ i^{\binom{n}{2}+1}\, n^{n/2} & \text{if } n \text{ is even}. \end{cases} \tag{3.1}$$

Proof. Observe that $\mathcal{A}$ is a Vandermonde matrix, that is, $\mathcal{A}$ is of the form

$$\begin{bmatrix} 1 & x_1 & {x_1}^2 & {x_1}^3 & \dots & {x_1}^{n-1} \\ 1 & x_2 & {x_2}^2 & {x_2}^3 & \dots & {x_1}^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & {x_n}^2 & {x_n}^3 & \dots & {x_n}^{n-1} \end{bmatrix}.$$

The determinant of an $n \times n$ Vandermonde matrix is well-known to be

$$\prod_{1 \le i < j \le n} (x_j - x_i).$$

Hence, we have that

$$\det \mathcal{A} = \prod_{0 \le r < s \le n-1} (\zeta_n^s - \zeta_n^r). \tag{3.2}$$

From the explicit computation of $\mathcal{A}(n,1)^2$ in (2.2), we see that this matrix is nI up to interchanging of rows. Moreover, interchanging 2 rows of a matrix only changes the sign of the determinant. Hence, $\det \mathcal{A}^2 = \pm n^n$. In particular, the number of row interchanges to transform $\mathcal{A}^2$ to nI is $(n-1)/2$ when n is odd and $(n-2)/2$ when n is even. This is because we need to interchange rows corresponding to r and $n-r$ for $1 \le r \le n-1$ to get nI. This is precisely $(n-1)/2$ number of distinct changes for odd n. When n is even, the row corresponding to $r = n/2$ has its diagonal entry as n, which need not be changed. For reasons evident from the calculations below, we write this as

$$\det \mathcal{A} = \begin{cases} \pm i^{\binom{n}{2}}\, n^n & \text{if } n \text{ is odd,} \\ \pm i^{\binom{n}{2}+1}\, n^n & \text{if } n \text{ is even.} \end{cases} \tag{3.3}$$

To determine the sign in the above computation, we calculate product in equation (3.2) in another way. For notational convenience, we will write $r < s$ for $0 \le r < s \le n-1$ and simplify (3.2) as follows -

$$\begin{aligned} \det \mathcal{A} = \prod_{r<s} \left(\zeta_n^s - \zeta_n^r\right) &= \prod_{r<s} \left(e^{2\pi i s/n} - e^{2\pi i r/n}\right) \\ &= \prod_{r<s} e^{i\pi s/n} e^{i\pi r/n} \left(e^{(i\pi s - i\pi r)/n} - e^{-(i\pi s - i\pi r)/n}\right) \\ &= i^{\binom{n}{2}} \prod_{r<s} \left[e^{i\pi(r+s)/n}\right] \prod_{r<s} \left[2\sin\left(\frac{(s-r)\pi}{n}\right)\right], \end{aligned} \tag{3.4}$$

as $\sin\theta = (e^{i\theta} - e^{-i\theta})/2i$. Now, note that

$$\begin{aligned} \sum_{\substack{r,s=0, \\ r \ne s}}^{n-1} (r+s) = \sum_{r=0}^{n-2} \sum_{s=r+1}^{n-1} (r+s) &= \sum_{s=1}^{n-1} \sum_{r=0}^{s-1} (r+s) \\ &= \sum_{s=1}^{n-1} \left(\frac{s(s-1)}{2} + s^2\right) \end{aligned}$$

$$= \sum_{s=1}^{n-1} \frac{3s^2 - s}{2}$$
$$= \frac{3}{2}\,\frac{(n-1)n(2n-1)}{6} - \frac{1}{2}\,\frac{n(n-1)}{2} \quad = \quad 2n\left(\frac{n-1}{2}\right)^2.$$

Therefore, the first product in (3.4) becomes

$$e^{i\pi(\sum_{r<s}(r+s))/n} = e^{i\pi(2n(n-1)^2/4n)} = i^{(n-1)^2},$$

which is 1 when n is odd and i when n is even. Since $0 < (s-r)\pi/n < \pi$, the second product in (3.4) is a positive quantity. Thus, the determinant becomes $i^{\binom{n}{2}}n^n$ when n is odd and $i^{\binom{n}{2}+1}n^n$ when n is even. □

Since the determinant of a matrix is the product of its eigenvalues, we have

$$\det \mathcal{A} = (\sqrt{n})^a(-\sqrt{n})^b(i\sqrt{n})^c(-i\sqrt{n})^d = i^{2b+c+3d}\, n^{n/2}.$$

Comparing this with Lemma 3.1 and noting that $3 \equiv -1 \bmod 4$, we get the conditions that

$$2b + c - d \equiv {}^nC_2 \bmod 4, \tag{3.5}$$

when n is odd. We will use this congruence to determine a, b, c, d as follows.

Suppose n is odd and $n \equiv 1 \pmod 4$. By Corollary 1, we know that $a-b = \pm 1$ and $c - d = 0$. Thus, (2.4) and (3.5) lead to

$$a - b = a + b - 2b \equiv \frac{n+1}{2} - \frac{n(n-1)}{2} \bmod 4 \equiv \frac{n+1-n+1}{2} \bmod 4 \equiv 1 \bmod 4.$$

Therefore, $a - b = 1$, which proves that $\mathcal{G}(2) = \operatorname{Tr}\mathcal{A}(n,1) = \sqrt{n}$ when $n \equiv 1 \pmod 4$. Now, suppose $n \equiv 3 \pmod 4$. Corollary 1 tells us that $a = b$ and $c - d = \pm 1$. Thus, (2.4) and (3.5) give

$$c - d \equiv \frac{n(n-1)}{2} - 2b \bmod 4 \equiv \frac{3(n-1)}{2} - \frac{n+1}{2} \bmod 4$$
$$\equiv \frac{3n-3-n-1}{2} \bmod 4 \equiv \frac{2n-4}{2} \bmod 4 \equiv 1 \bmod 4.$$

Therefore, when $n \equiv 3 \bmod 4$, we deduce that $\mathcal{G}(2) = \operatorname{Tr}\mathcal{A}(n,1) = i\sqrt{n}$ as in Theorem 1.1.

4. The law of quadratic reciprocity

Let a be a natural number and p be an odd prime. The Legendre symbol is defined as

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p \mid a, \\ 1 & \text{if } x^2 \equiv a \bmod p \text{ has a solution}, \\ -1 & \text{if } x^2 \equiv a \bmod p \text{ has no solution}. \end{cases}$$

We connect the quadratic Gauss sum $\mathcal{G}(2)$ with the Legendre symbol in the following lemma.

Lemma 4.1. *Let p be an odd prime and $(m,p)=1$. Define $\mathcal{A}(p,m)=(\zeta_p^{mrs})$ for $0\le r,s\le p-1$. Then*

$$\operatorname{Tr}\mathcal{A}(p,m)=\left(\frac{m}{p}\right)\operatorname{Tr}\mathcal{A}(p,1),$$

where $(\frac{m}{p})$ is the Legendre symbol.

Proof. We note that

$$\left(\frac{k}{p}\right)+1=\begin{cases}1 & \text{if } p\mid k,\\ 2 & \text{if } p\nmid k,\ k \text{ is a quadratic residue } \bmod p,\\ 0 & \text{otherwise.}\end{cases}$$

For any $0\le k\le p-1$, the polynomial x^2-k has at most two roots in $\mathbb{F}_p$, the finite field with p elements. Also, if j is a root of this polynomial, then so is $p-j$ (which is distinct from j as p is odd). Hence, for each $k\in\mathbb{F}_p$ and $k\neq 0$, there are either 2 values of j satisfying $j^2\equiv k \pmod p$ or none. Thus, the quadratic Gauss sum can be rewritten as

$$\begin{aligned}\operatorname{Tr}\mathcal{A}(p,m)=\sum_{j=0}^{p-1}e^{2\pi imj^2/p}&=\sum_{k=0}^{p-1}\left[\left(\frac{k}{p}\right)+1\right]e^{2\pi imk/p}\\&=\sum_{k=0}^{p-1}\left[e^{2\pi imk/p}\right]+\sum_{k=0}^{p-1}\left(\frac{k}{p}\right)e^{2\pi imk/p}\\&=\sum_{k=0}^{p-1}\left(\frac{k}{p}\right)e^{2\pi imk/p},\end{aligned}\tag{4.1}$$

as the first sum is the sum of all p-th roots of unity and vanishes. Since the Legendre symbol is multiplicative, we multiply the second sum by $1=(\frac{m}{p})^2$ and have

$$\begin{aligned}\operatorname{Tr}\mathcal{A}(p,m)=\left(\frac{m}{p}\right)^2\sum_{k=0}^{p-1}\left(\frac{k}{p}\right)e^{2\pi imk/p}&=\left(\frac{m}{p}\right)\sum_{k=0}^{p-1}\left(\frac{km}{p}\right)e^{2\pi ikm/p}\\&=\left(\frac{m}{p}\right)\sum_{j=0}^{p-1}\left(\frac{j}{p}\right)e^{2\pi ij/p}=\left(\frac{m}{p}\right)\operatorname{Tr}\mathcal{A}(p,1),\end{aligned}$$

by taking $m=1$ in (4.1). □

The law of quadratic reciprocity can be stated as follows.

Theorem 4.2. *Let p and q be distinct odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{(p-1)(q-1)/4}.$$

Proof. For convenience of notation, we define

$$e(n)=\begin{cases}1 & \text{if } n\equiv 1 \bmod 4,\\ i & \text{if } n\equiv 3 \bmod 4.\end{cases}$$

Thus, Theorem 1.1 states that for odd n, $\operatorname{Tr}\mathcal{A}(n,1) = e(n)\sqrt{n}$. Therefore, taking $n = pq$, we have

$$e(pq)\sqrt{pq} = \operatorname{Tr}\mathcal{A}(pq,1) = \Big(\operatorname{Tr}\mathcal{A}(p,q)\Big)\Big(\operatorname{Tr}\mathcal{A}(q,p)\Big),$$

by Lemma 2.2. Using Lemma 4.1, we get

$$e(pq)\sqrt{pq} = \left(\frac{p}{q}\right)\left(\frac{q}{p}\right)\Big(\operatorname{Tr}\mathcal{A}(p,1)\Big)\Big(\operatorname{Tr}\mathcal{A}(q,1)\Big) = \left(\frac{p}{q}\right)\left(\frac{q}{p}\right)e(p)e(q)\sqrt{pq},$$

which implies that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \frac{e(pq)}{e(p)e(q)}.$$

We observe that the right hand side is 1 if at least one of p or q is congruent to 1 (mod 4) and -1 otherwise. This is precisely as stated in the law of quadratic reciprocity. □

5. Evaluation of $\operatorname{Tr}\mathcal{A}(n,m)$ for odd n

In Section 3, we evaluated $\operatorname{Tr}\mathcal{A}(n,1)$ for odd natural numbers n. We use this computation to determine $\operatorname{Tr}\mathcal{A}(n,m)$ for n odd and $(m,n) = 1$ in general. Before proceeding, we recall the Jacobi symbol, which is a generalization of the Legendre symbol. For any positive integer a and an odd natural number $n = p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k}$, where p_j are distinct odd primes, the Jacobi symbol $\left(\frac{a}{n}\right)$ is defined as a product of the Legendre symbols,

$$\left(\frac{a}{n}\right) = \prod_{j=1}^{k}\left(\frac{a}{p_j}\right)^{\alpha_j}.$$

Recall that the law of quadratic reciprocity extends to the Jacobi symbol by elementary number theory considerations.

Lemma 5.1. *For an odd natural number n and $(m,n) = 1$, we have*

$$\operatorname{Tr}\mathcal{A}(n,m) = \left(\frac{m}{n}\right)\operatorname{Tr}\mathcal{A}(n,1),$$

where $\left(\frac{m}{n}\right)$ is the Jacobi symbol.

Proof. Let $\omega_3(n)$ be the number of prime divisors p of n with $p \equiv 3 \pmod 4$ counted with multiplicity. We claim that for any odd n,

$$\operatorname{Tr}\mathcal{A}(n,1) = \delta(n)\prod_{p|n}\operatorname{Tr}\mathcal{A}(p,1), \tag{5.1}$$

where $\delta(n) = \pm 1$ and the product is over primes dividing n repeated with multiplicity. Indeed, we know that the product on the right hand side of (5.1) can be evaluated by the already proven cases of Theorem 1.1 to be

$$\prod_{p|n}\operatorname{Tr}\mathcal{A}(p,1) = i^{\omega_3(n)}\sqrt{n}.$$

Also, since

$$n \equiv \begin{cases} 1 \bmod 4 & \text{if } \omega_3(n) \text{ is even,} \\ 3 \bmod 4 & \text{if } \omega_3(n) \text{ is odd,} \end{cases}$$

and the results from Section 3, the left hand side of (5.1) is $\sqrt{n}$ if $\omega_3(n)$ is even and $i\sqrt{n}$ if $\omega_3(n)$ is odd. Thus, $\operatorname{Tr}\mathcal{A}(n,1)$ and the product agree up to sign (which of course depends on n) so that(5.1) is immediate.

Writing (5.1) explicitly,

$$\sum_{j=0}^{n-1} e^{2\pi i j^2/n} = \delta(n) \prod_{p|n} \left[\sum_{k=0}^{p-1} e^{2\pi i k^2/p} \right] = \delta(n) \prod_{p|n} \left[\sum_{k=0}^{p-1} \left(e^{2\pi i k^2/n} \right)^{n/p} \right],$$

we observe that all terms in (5.1) lie in the n-th cyclotomic field, $\mathbb{Q}(\zeta_n)$. Thus, by applying the Galois automorphism that sends ζ_n to ζ_n^m, and noting that this automorphism fixes the rationals (and hence $\delta(n)$), (5.1) becomes

$$\sum_{j=0}^{n-1} e^{2\pi i m j^2/n} = \delta(n) \prod_{p|n} \left[\sum_{k=0}^{p-1} \left(e^{2\pi i m k^2/n} \right)^{n/p} \right] = \delta(n) \prod_{p|n} \left[\sum_{k=0}^{p-1} e^{2\pi i m k^2/p} \right].$$

Each term in the above product is $\operatorname{Tr}\mathcal{A}(p,m)$ for an odd prime p and $(m,p)=1$. Hence, by Lemma 4.1, we get

$$\operatorname{Tr}\mathcal{A}(n,m) = \delta(n) \left[\prod_{p|n} \left(\frac{m}{p} \right) \right] \left[\prod_{p|n} \operatorname{Tr}\mathcal{A}(p,1) \right].$$

Thus, by (5.1), $\operatorname{Tr}\mathcal{A}(n,m) = \left(\frac{m}{n}\right) \operatorname{Tr}\mathcal{A}(n,1)$. □

6. Proof of Theorem 1.1 for n even

The case $n \equiv 2 \pmod 4$ was settled in Lemma 2.1. Thus, we assume that $4|n$. We begin with the following elementary result which will be proved by induction.

Lemma 6.1. *Let r and s be natural numbers with $r \geq 2$ and s odd. Then*

$$\operatorname{Tr}\mathcal{A}(2^r, s) = \left(\frac{2^r}{s} \right) (1 + i^s) \sqrt{2^r}.$$

Proof. We proceed by induction on r. The base cases are $r = 2$ and $r = 3$. For $r = 2$,

$$\operatorname{Tr}\mathcal{A}(4,s) = 1 + e^{2\pi i s/4} + e^{2\pi i s} + e^{2\pi i s 9/4} = 2(1 + i^s).$$

For $r = 3$,

$$\operatorname{Tr}\mathcal{A}(8,s) = 2(1 + (-1)^s + 2e^{i\pi s/4}),$$

by considering squares modulo 8. Thus, $\operatorname{Tr}\mathcal{A}(8,s) = 4e^{i\pi s/4}$. Using $e^{i\theta} = \cos\theta + i\sin\theta$, we get that $\operatorname{Tr}\mathcal{A}(8,s) = 4(\cos(s\pi/4) + i\sin(s\pi/4))$, which is $2\sqrt{2}(1+i)$ if $s \equiv 1 \pmod 4$ and $-2\sqrt{2}(1-i)$ if $s \equiv 3 \pmod 4$. Hence, we see that Lemma 6.1 is true when $r = 2, 3$.

Suppose that $r \geq 4$ and Lemma 6.1 holds for all $2 \leq \alpha \leq r-1$. To prove it for r, we note that

$$\begin{aligned}\operatorname{Tr}\mathcal{A}(2^r, s) &= \sum_{j=1}^{2^r} e^{2\pi i s j^2/2^r} \\ &= \sum_{\substack{j=1,\\ j-\text{odd}}}^{2^r} e^{2\pi i s j^2/2^r} + \sum_{\substack{j=1,\\ j-\text{even}}}^{2^r} e^{2\pi i s j^2/2^r} \\ &= \frac{1}{2}\Bigg(\sum_{\substack{j=1,\\ j-\text{odd}}}^{2^r} e^{2\pi i s j^2/2^r} + e^{2\pi i s (j+2^{r-2})^2/2^r} \Bigg) + \sum_{k=1}^{2^{r-1}} e^{2\pi i s k^2/2^{r-2}},\end{aligned}$$

where in the first sum, we pair the terms corresponding to j and $j+2^{r-2}$ (which are distinct as j is odd) and in the second sum, we change the index of summation by setting $j = 2k$. Now, each summand of the first term is

$$\begin{aligned}e^{2\pi i s j^2/2^r} + e^{2\pi i s (j+2^{r-2})^2/2^r} &= e^{2\pi i s j^2/2^r} + e^{2\pi i s j^2/2^r}\, e^{2\pi i s (2^{r-1} j)/2^r} \\ &= e^{2\pi i s j^2/2^r} - e^{2\pi i s j^2/2^r} \quad = \quad 0,\end{aligned}$$

as j is odd. We recognize the second term as $2\operatorname{Tr}\mathcal{A}(2^{r-2})$, which is equal to

$$(2^{r-2}/s)\ 2\ (1+i^s)\sqrt{2^{r-2}}$$

by the induction hypothesis. Thus, the principle of mathematical induction implies that

$$\operatorname{Tr}\mathcal{A}(2^r, s) = (2^r/s)(1+i^s)\sqrt{2^r}$$

for all $r \geq 2$. □

We now derive Theorem 1.1 in the case $4|n$ as a consequence of the proposition below.

Proposition 6.2. *Let n be natural number with $4|n$ and $(m,n)=1$. Then*

$$\operatorname{Tr}\mathcal{A}(n, m) = (n/m)\,(1+i^m)\,\sqrt{n}.$$

Proof. Write $n = 2^u v$, with $u \geq 2$ and v odd. We would like to evaluate $\operatorname{Tr}\mathcal{A}(2^u v, m)$ for $(m,n)=1$. By Lemma 2.2, we get

$$\operatorname{Tr}\mathcal{A}(2^u v, m) = \Big(\operatorname{Tr}\mathcal{A}(2^u, vm) \Big) \Big(\operatorname{Tr}\mathcal{A}(v, 2^u m) \Big). \tag{6.1}$$

We note that v is odd and $(v, 2^u m) = 1$. Hence, by Lemma 5.1,

$$\operatorname{Tr}\mathcal{A}(v, 2^u m) = \left(\frac{2^u m}{v}\right) \operatorname{Tr}\mathcal{A}(v, 1), \tag{6.2}$$

which is known by the results in Section 3. Since $4|n$ and $(m,n)=1$, m is odd. Therefore, vm is odd and $(2^u, vm) = 1$. To determine $\operatorname{Tr}\mathcal{A}(2^u, vm)$, we use Lemma 6.1.

Thus, by (6.1), (6.2) and Lemma 6.1, we have

$$\operatorname{Tr}\mathcal{A}(2^u v, m) = \sqrt{2^u} \left(\frac{2^u m}{v}\right) \left(\frac{2^u}{vm}\right) (1+i^{vm}) \operatorname{Tr}\mathcal{A}(v, 1),$$

which can be simplified using the multiplicativity of the Jacobi symbol to

$$\operatorname{Tr}\mathcal{A}(2^u v, m) = \sqrt{2^u}\ (2^u/m)\ \epsilon_{v,m},$$

where

$$\epsilon_{v,m} = (m/v)\ (1 + i^{vm})\operatorname{Tr}\mathcal{A}(v, 1).$$

Hence we see that the value of trace depends on whether v and m are congruent to 1 or 3 modulo 4. We remark that the case of odd n from Theorem 1.1 can be re-written as

$$\operatorname{Tr}\mathcal{A}(n, 1) = i^{(n-1)^2/4}\sqrt{n}.$$

Since both v and m are odd, we can use the law of quadratic reciprocity to deduce

$$\left(\frac{m}{v}\right)\left(\frac{v}{m}\right) = \begin{cases} -1 & \text{if both } m, v \equiv 3 \bmod 4, \\ 1 & \text{otherwise.} \end{cases}$$

Therefore, we have the following table of values of $\epsilon_{v,m}$:

v \ m	1 mod 4	3 mod 4
1 mod 4	$\left(\frac{v}{m}\right)(1+i)\sqrt{v}$	$\left(\frac{v}{m}\right)(1-i)\sqrt{v}$
3 mod 4	$\left(\frac{v}{m}\right)(1-i)i\sqrt{v}$	$-\left(\frac{v}{m}\right)(1+i)i\sqrt{v}$

Observe that $i(1-i) = (1+i)$ and $i(1+i) = -(1-i)$. Thus, whenever $4|n$, we have $\operatorname{Tr}\mathcal{A}(n, m) = (n/m)(1 + i^m)\sqrt{n}$. □

In particular, for $m = 1$, Proposition 6.2 implies Theorem 1.1 for $n \equiv 0 \bmod 4$.

7. Concluding Remarks

We observe that determining the quadratic Gauss sum in the case $4|n$ is more delicate than the case n odd. The study of the eigenvalues and their multiplicites of the matrix $\mathcal{A}(n, m)$ lies deeper than the law of quadratic reciprocity. The matrix $\mathcal{A}(n, m)$ also appears in the context of the discrete Fourier transform of periodic arithmetical functions. Thus, the study of its eigenvalues and their multiplicities is interesting in its own right. Moreover, the investigation of the eigenvectors of $\mathcal{A}(n, 1)$ is even deeper than the study of its eigenvalues and their multiplicites (for example, see [6]). Surprisingly, the explicit construction of these eigenvectors was first done as late as 1972 in the paper of McClellan and Parks [5].

Acknowledgements. We thank the referee and Anup Dixit for helpful comments on an earlier version of this paper.

References

[1] Borevich, Z. and Shafarevich, I., *Number Theory*, translated by Newcomb Greenleaf, Academic Press, New York, 1966.

[2] Berndt, B. and Evans, R., The determination of Gauss sums, *Bull. Amer. Math. Soc.*, **5**, No. 2 (1981), 107–129.

[3] Ireland, K. and Rosen, M., *A classical introduction to modern number theory*, Graduate Texts in Mathematics, Springer-Verlag, (1990), 66–78.
[4] Lang, S., *Algebraic Number Theory*, Second Edition, Graduate Texts in Mathematics, Springer-Verlag, (1994), 87–90.
[5] McClellan, J. H. and Parks, T. W., *Eigenvalue and eigenvector decomposition of the discrete Fourier transform*, IEEE Transactions Audio Electroacoust., Vol. AU-**20**, (1972), 66–74.
[6] Mehta, M. L., Eigenvalues and eigenvectors of the finite Fourier transform, *Journal of Mathematical Physics*, **28**, (1987), 781–785.
[7] Ram Murty, M., *Problems in Analytic Number Theory*, First Edition, Graduate Texts in Mathematics, Springer-Verlag, (2000), 81 and 323–324.
[8] Ram Murty, M., Quadratic Reciprocity via linear algebra, *Bona Mathematica*, **12**, No. 4 (2001), 75–80.
[9] Schur, I., Über die Gausschen Summen, *Nachrichten von der Königlichen Gessellschaft zu Göttingen, Mathematisch - Physikalische Klass*, (1921), 147–153.
[10] Waterhouse, W. C., The sign of the Gauss sum, *Journal of Number Theory*, **2** (1970), 363.

M. Ram Murty
Department of Mathematics and Statistics
Queen's University, Kingston, Canada, ON K7L 3N6.
E-mail: *murty@mast.queensu.ca*

Siddhi Pathak
Department of Mathematics and Statistics
Queen's University, Kingston, Canada, ON K7L 3N6.
E-mail: *siddhi@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June, (2017), 151-157

PROBLEM SECTION

In the last issue of the Math. Student Vol. **85**, Nos. 3-4, July-December (2016), we had invited solutions from the floor to the remaining problems *1, 2, 4, 5* and *6* of the MS, 85, 1-2, 2016 as well as to the five new problems *9, 10, 11, 12* and *13* presented therein till April 30, 2017.

No solution was received from the floor to the remaining problems *1, 2, 4* and *5* of the MS, 85, 1-2, 2016 and hence we provide in this issue the Proposer's solution to these problems. Problem 6 is a research problem.

We received from the floor **four** correct solutions to the Problem *11* & **two** correct solutions to the Problem *13* of MS, 85, 3-4, 2016 and we publish here a solution received from the floor to the Problem *11* and *13*. Readers can try their hand on the remaining problems *9, 10* and *12* till October 31, 2017.

In this issue we first present **five new problems**. Solutions to these problems as also to the remaining problems *9, 10* and *12* of MS, 85, 3-4, 2016, received from the floor till October 30, 2017, if approved by the Editorial Board, will be published in the MS, 86, 3-4, 2017.

MS-2017, Nos. 1-2: Problem-1:
Proposed by **Zoltán Boros** and **Árpád Száz**, Institute of Mathematics, University of Debrecen, H-4002, Debrecen, Pf. 400, Hungary. zboros@science.unideb.hu, szaz@science.unideb.hu; submitted through B. Sury.

For all $a, b \in \mathbb{R}$ determine

$$F(a,b) = \inf_{n\in\mathbb{N}} \Big(an + \frac{b}{n}\Big),$$

where $\mathbb{N}$ and $\mathbb{R}$ denote the sets of all natural and real numbers respectively.

Problems proposed by B. Sury:

MS-2017, Nos. 1-2: Problem-2:

Let n be a positive integer and $d_1, \cdots, d_r$ be certain divisors of n (not necessarily all and not necessarily distinct). Let $a_1, \cdots, a_r, b$ be arbitrary integers. Determine the number of solutions of the congruence

$$a_1x_1 + \cdots + a_rx_r \equiv b \ mod \ n$$

for integers x_i satisfying $(x_i, n) = d_i$.

MS-2017, Nos. 1-2: Problem-3:

If there is a painting of dimensions 40 inches by 100 inches, what is the smallest square frame which can cover it completely? Answer the same question when the painting has dimensions 40 inches by 90 inches.

MS-2017, Nos. 1-2: Problem-4:

There is a one way street which has n parking spaces numbered 1 to n. Suppose there are n cars $C_1, \cdots, C_n$. Each car C_i has a certain favourite parking slot a_i $(1 \leq a_i \leq n)$. Each car enters the street and, if its favourite slot is empty, it parks there. If not, it proceeds ahead and parks at the next available slot. If no slot is available, the car has to leave the street. A sequence $(a_1, \cdots, a_n)$ is a parking sequence if every car can park. For example, $(1,1), (1,2), (2,1)$ are parking sequences when $n = 2$.
(i) Find the number of parking sequences.
(ii) Find a natural bijection between the parking sequences and the number of regions formed in $\mathbf{R}^n$ by the hyperplanes $x_i - x_j = 0, x_i - x_j = 1$ for $1 \leq i < j \leq n$.

MS-2017, Nos. 1-2: Problem-5:

Prove that $\sum_{k=1}^{n} \frac{1}{k^r} = \sum_I \binom{n}{i_r} \frac{(-1)^{r-1}}{i_1 i_2 \cdots i_r}$ where $I = (i_1, \cdots, i_r)$ runs through r-tuples satisfying $1 \leq i_1 \leq i_2 \leq \cdots \leq i_r \leq n$.

Solution from the floor: MS-2016, Nos. 3-4: Problem 11: If a strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(2t - f(t)) = t = 2f(t) - f(f(t))$ for all t then prove that there exists $c \in \mathbb{R}$ such that $f(t) = t + c$ for all t.
(Solution submitted on 16-02-2017 by **S. U. Weeraratne, K. K. D. S. de Silva, T. R. Ekanayake** (Department of Mathematics, Faculty of Science, University of Colombo, Sri Lanka) and **D. N. Pannipitiya** (IT Unit-2, Faculty of Science, University of Colombo, Sri Lanka.); *suweerainfo@gmail.com, kkdsdesilva@gmail.com, thusaraekanayake@gmail.com, diyathnp@yahoo.com*).
Solution. It can be easily shown that $f^{-1} : \mathbb{R} \longmapsto \mathbb{R}$ exists and $f^{-1} = 2t - f(t)$. Since f increases, so does f^{-1}. Let $n \in \mathbb{N}$. Suppose

$$\underbrace{ff \cdots f}_{n}(t) = nf(t) - (n-1)t$$

Then

$$\underbrace{ff \cdots f}_{n+1}(t) = \underbrace{ff \cdots f}_{n}(f(t)) = nf(f(t)) - (n-1)f(t)$$

$$= n(2f(t) - t) - (n-1)f(t) \;=\; (n+1)f(t) - nt.$$

Thus, with the given fact $t = 2f(t) - f(f(t))$, it follows by the mathematical induction that $\underbrace{ff \cdots f}_{n}(t) = nf(t) - (n-1)t$ for any $n \in \mathbb{N}$.

Replacing t by $f^{-1}(t)$ in $t = 2f(t) - f(f(t))$ we have $f^{-1}(t) = 2t - f(t)$. Suppose

$$\underbrace{f^{-1}f^{-1}\cdots f^{-1}}_{n}(t) = (n+1)t - nf(t).$$

Then

$$\begin{aligned}\underbrace{f^{-1}f^{-1}\cdots f^{-1}}_{n+1}(t) &= (n+1)f^{-1}(t) - nf(f^{-1}(t)) \\ &= (n+1)f^{-1}(t) - nt \\ &= (n+1)(2t - f(t)) - nt \quad = \quad (n+2)t - (n+1)f(t).\end{aligned}$$

Thus, it follows by the mathematical induction that

$$\underbrace{f^{-1}f^{-1}\cdots f^{-1}}_{n}(t) = (n+1)t - nf(t) \quad \text{for any} \quad n \in \mathbb{N}$$

Let $x_1, x_2 \in \mathbb{R}$ be such that $x_1 < x_2$. Then $\underbrace{ff\cdots f}_{n}(x_1) < \underbrace{ff\cdots f}_{n}(x_2)$, that is, $nf(x_1) - (n-1)x_1 < nf(x_2) - (n-1)x_2$ which gives

$$(n-1)/n < (f(x_2) - f(x_1))/(x_2 - x_1)$$

Similarly we can show that

$$(f(x_2) - f(x_1))/(x_2 - x_1) < (n+1)/n$$

by using the fact $\underbrace{f^{-1}f^{-1}\cdots f^{-1}}_{n}(t) = (n+1)t - nf(t)$ and the fact f^{-1} is increasing.

Thus

$$1 - 1/n < (f(x_2) - f(x_1))/(x_2 - x_1) < 1 + 1/n$$

for any $n \in \mathbb{N}$. It follows that $(f(x_2) - f(x_1))/(x_2 - x_1) = 1$, thus proving the claim as x_1 and x_2 are arbitrary.

Correct solution was also received from the floor from:

Aritro Pathak; (Department of Mathematics, Brandeis University, Waltham, Massachusetts, 02453, USA; Eimail: *ap323@brandeis.edu*; received on 08-01-2017)

Dasari Naga Vijay Krishna; (Machilipatnam, Andhra Pradesh -521001; E-mail: *Vijay9290009015@gmail.com*; received on 13-02-2017)

Prajanaswaroopa S.; (Bangalore, India; E-mail: *sntrm4@rediffmail.com*, received on 30-04-2017)

Solution from the floor: MS-2016, Nos. 3-4: Problem 13: If $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function satisfying the conditions $|f'(x)| \leq |f(x)|$ and $f(0) = 0$ then show that $f(x) = 0$ for all $x \in \mathbb{R}$. Give also a proof that uses only the Mean Value Theorem.

(Solution submitted on 13-02-2017 by **Dasari Naga Vijay Krishna** (Machilipatnam, Andhra Pradesh-521001; *Vijay9290009015@gmail.com*).

Solution. Suppose to the contrary, there exists a, real, such that $f(a) = b \neq 0$". Without loss of generality, we may assume that $a, b > 0$ (if $a < 0$, work with $g(x) = f(-x)$; if $b < 0$, work with $g(x) = -f(x)$).

f being continuous, there exists u, v such that $0 \leq u < v < u + 1$ and $f(u) = 0$,

$f(x) > 0$ for all $x \in (u, v]$.
In fact, let $A = \{x \geq 0 : f(t) \leq 0 \ \forall \ t \in [0, x)\}$, and put $u = SupA$. Obviously $u \in [0, a)$. Continuity of f and $f(a) > 0$ imply $f(u) = 0$ and there exists $t > u$ such that $f(x) > 0 \ \forall \ x \in [u, t)$. Choose now $v \in (u, \ min\{u+1, \ t\})$. Observe that existence of a imply existence of u; without existence of a, may be $A = [0, \ \infty)$ and $SupA$ does not exist.
Now, define $g : [u, \ v] \to [0, \ \infty)$ as $g(u) = 0$ and $g(x) = \frac{f(x)}{x-u} \ \forall \ x \in (u, \ v]$. g is continuous over $[u, \ v]$ and so has maximum value at some $t \in (u, \ v]$. Since f is differentiable, there exists $w \in [u, \ t]$ such that

$$f'(w) \ = \ \frac{f(t) - f(u)}{t - u}.$$

Therefore

$$g(w) = \frac{f(w)}{w - u} \ > \ f(w) \ \geq \ |f^{'}(w)| \ = \ \frac{f(t)}{t - u} = \ g(t)$$

in contradiction with the fact that $g(t)$ is the maximum of $g(x)$ over $[u, \ v]$.
Proof-2. Let $[f = 0]$ be the zero set of f. Assume $[f \neq 0]$ is not empty. We may assume without loss of generality that there is $x_0 > 0$ such that $f(x_0) > 0$. Since $[f = 0]$ is closed, we can pick its biggest element y such that $f(y) = 0$ and $y < x_0$. By continuity of f, $f(x) > 0$ on $(y, \ x_0)$. Again by continuity, f takes a maximum value on $[y, \ y + t]$ for $0 < t < min\{x_0, \ 1\}$, say at the point x_m. Now using mean value theorem on $[y, \ x_0]$, we obtain

$$f(x_m) = (x_m - y)f'(\theta) \leq (x_m - y)f(\theta)$$

a contradiction since $(x_m - y) < 1$.
Correct solution was also received from the floor from:
Prajanaswaroopa S.; (Bangalore, India; E-mail: *sntrm4@rediffmail.com*, received on 30-04-2017)

Solution by the Proposer M. Ram Murty: MS-2016, Nos. 1-2: Problem 1: Let ϕ denote Euler's function. Show that for any positive real number a, there is a constant $C(a)$ such that

$$\sum_{n \leq x} \left(\frac{n}{\phi(n)} \right)^a \leq C(a)x.$$

Solution. Writing

$$\left(\frac{n}{\phi(n)} \right)^a = \sum_{d|n} g(d),$$

we easily see that g is multiplicative since $n/\phi(n)$ is multiplicative. It therefore suffices to determine $g(n)$ when n is a prime power. Thus,

$$g(1) = 1, \quad g(p) = \left(1 + \frac{1}{p-1} \right)^a - 1, \quad g(p^\alpha) = 0, \forall \alpha \geq 2,$$

for any prime p. Observe that $g(n)$ is non-negative for all values of n and that $g(p) = O(1/p)$ by a simple approximation in the binomial theorem. Now,

$$\sum_{n\le x}\left(\frac{n}{\phi(n)}\right)^a=\sum_{n\le x}\sum_{d|n}g(d)=\sum_{d\le x}g(d)[x/d],$$

upon interchanging the summation. As g is non-negative, this is

$$\le x\sum_{d\le x}\frac{g(d)}{d}\le x\sum_{d=1}^{\infty}\frac{g(d)}{d}.$$

But g is multiplicative, so the series can be written as an infinite product over the prime numbers:

$$\sum_{d=1}^{\infty}\frac{g(d)}{d}=\prod_p\left(1+\frac{g(p)}{p}\right).$$

Because, $g(p)=O(1/p)$, the product converges absolutely giving us the desired result.

Solution by the Proposer B. Sury: MS-2016, Nos. 1-2: Problem 2: Suppose a,b,c,d are integers such that the last 2016 digits of the number $ab+cd$ are all 9's. Show that there exist integers A,B,C,D each ending in 2016 zeroes such that

$$(a+A)(b+B)(c+C)(d+D)=\pm1.$$

Solution. Let n be a positive integer and suppose $ab+cd+1$ is a multiple of n. We will show that adding integral multiples nA,nB,nC,nD of n, respectively to a,b,c,d, we have that $(a+nA)(b+nB)+(c+nC)(d+nD)=-1$. Applying this to $n=10^{2016}$, we have the assertion of the problem.

Write $ab+cd+1=qn$. Note that GCD $(b,d,n)=1$. First, let us suppose that GCD $(b,d)=1$. Consider $a'=a+un$ and $c'=c+vn$ where u,v are integers to be chosen. Now $a'b+c'd+1=ab+cd+1+(ub+vd)n=(q+ub+vd)n$.

Since GCD $(b,d)=1$, we may choose integers u,v with $q=-ub-vd$; this gives $a'd+b'c+1=0$ and we will be done. So, we only have to prove that we may change b,d modulo n so that they are relatively prime.

Let $p_1,\cdots,p_r$ be the set of all primes which divide b. If each $p_i|n$, then clearly, none of the p_i's divide d since GCD $(b,d,n)=1$. In such a case, evidently GCD $(b,d)=1$.

So, let us suppose that some of the p_i's do not divide n; let $p_1,\cdots,p_k$ be the subset of those prime factors of b which divide n. So GCD $(n,p_1,\cdots p_k)=1$.

By the Chinese remainder theorem, choose an integer $x\equiv d \bmod n$ and $x\equiv 1 \bmod p_1\cdots p_k$. Then, clearly $p_1,\cdots,p_k\not|x$.

Also, writing $x=d+ln$, we have that the prime factors p_i $(k<i\le r)$ of b which divide n, cannot divide $d+ln$ as GCD $(b,d,n)=1$. Hence $(b,d+ln)=1$ and we are done.

Solution by the Proposer B. Sury: MS-2016, Nos. 1-2: Problem 4: Let $f=c_0+c_1X+\cdots+c_nX^n$ be a polynomial with integer coefficients. Prove

that there exist $n+1$ primes $p_0, p_1, p_2, \cdots, p_n$ and a polynomial g with integer coefficients such that

$$f(x)g(x) = a_1x^{p_1} + a_2x^{p_2} + \cdots + a_nx^{p_n}$$

Solution. In the ring $\mathbf{Q}[X]$ of polynomials with rational coefficients, consider the ideal (f) generated by f. The quotient ring $\mathbf{Q}[X]/(f)$ is a vector space of dimension equal to the degree of f (the images of $1, X, X^2, \cdots, X^{deg(f)-1}$ is a basis). So, the dimension is at most n (it equals n if $c_n \neq 0$). Thus, for any set of $n+1$ distinct primes $q_0, q_1, \cdots, q_n$, the images of the $n+1$ polynomials $X^{q_0}, X^{q_1}, \cdots, X^{q_n}$ in $\mathbf{Q}[X]/(f)$ form a linearly independent set over $\mathbf{Q}$. Hence, there exist rational numbers $a_0, a_1, \cdots, a_n$ (not all zero) such that the image of $\sum_{i=0}^{n} a_iX^{q_i}$ is 0 in the quotient ring $\mathbf{Q}[X]/(f)$; in other words, there is a polynomial $q \in \mathbf{Q}[X]$ such that

$$\sum_{i=0}^{n} a_iX^{q_i} = f(X)q(X).$$

We can clear denominators from a_i's as well as from $q(X)$ to get integers $b_0, \cdots, b_n$ and a polynomial $g \in \mathbf{Z}[X]$ such that $\sum_{i=0}^{n} b_iX^{q_i} = f(X)g(X)$.

Solution by the Proposer B. Sury: MS-2016, Nos. 1-2: Problem 5: Let $f = \sum_{i=0}^{n} c_iX^i$ where n is a positive integer and each $c_i = \pm 1$. If all the roots of f are real, determine all the possibilities for f.

Solution. We find all monic polynomials f whose roots are all real. Then, the required possibilities are $\pm f$. Now, write

$$f = X^n + c_{n-1}X^{n-1} + \cdots c_1X + c_0$$

with each $c_i = \pm 1$. Now, if $\alpha_1, \cdots, \alpha_n$ are the roots of f, then

$$\sum_{i=1}^{n} \alpha_i = -c_{n-1}; \qquad \sum_{i<j} \alpha_i\alpha_j = c_{n-2}; \qquad \prod_{i=1}^{n} \alpha_i = (-1)^n c_0.$$

Hence

$$\sum_{i=1}^{n} \alpha_i^2 = c_{n-1}^2 - 2c_{n-2} \quad \text{and} \quad \prod_{i=1}^{n} \alpha_i^2 = c_0^2$$

If all the α_i's are real, the α_i's are positive. Applying AM $\geq$ GM for the numbers α_i^2, we have

$$(c_{n-1} - 2c_{n-2})/n \geq c_0^{2/n}.$$

As c_{n-1}, c_{n-2} are ± 1, we must have $c_{n-2} = 1$. Therefore $3 \geq nc_0^{2/n}$, and hence

$$3^n \geq n^n$$

implying $n \leq 3$. By looking at all the cases, we arrive at the possibilities

$$\pm(X+1), \quad \pm(X-1), \quad \pm(X^2+X-1), \quad \pm(X^2-X-1),$$
$$\pm(X^3+X^2-X-1), \quad \pm(X^3-X^2-X+1).$$

Note. Partial solution to this problem was received from **Vikas Chakraborty**, Department of Mathematics, University of Kalyani, Kalyani-741235, West Bengal,

India; E-mail: *bikashchakraborty.math@yahoo.com ; vikashchakrabortyy@gmail.com*,
way back on 06-06-2016.

FORM IV
(See Rule 8)

1.	Place of Publication:	PUNE
2.	Periodicity of publication:	QUARTERLY
3.	Printer's Name: Nationality: Address:	DINESH BARVE INDIAN PARASURAM PROCESS 38/8, ERANDWANE PUNE-411 004, INDIA
4.	Publisher's Name: Nationality: Address:	N. K. THAKARE INDIAN GENERAL SECRETARY THE INDIAN MATHEMATICAL SOCIETY c/o:CENTER FOR ADVANCED STUDY IN MATHEMATICS, S. P. PUNE UNIVERSITY PUNE-400 007, MAHARASHTRA, INDIA
5.	Editor's Name: Nationality: Address:	J. R. PATADIA INDIAN (DEPARTMENT OF MATHEMATIC, THE M. S.UNIVERSITY OF BARODA) 5 , ARJUN PARK, NEAR PATEL COLONY BEHIND DINESH MILL, SHIVANAND MARG VADODARA - 390 007, GUJARAT, INDIA
6.	Names and addresses of individuals who own the newspaper and partners or shareholders holding more than 1% of the total capital:	THE INDIAN MATHEMATICAL SOCIETY

I, N. K. Thakare, hereby declare that the particulars given above are true to the best of my knowledge and belief.

N. K. THAKARE

Dated: 22^{rd} May 2017 — Signature of the Publisher

Published by Prof. N. K. Thakare for the Indian Mathematical Society, type set by J. R. Patadia at 5, Arjun Park, Near Patel Colony, Behind Dinesh Mill, Shivanand Marg, Vadodara - 390 007 and printed by Dinesh Barve at Parashuram Process, Shed No. 1246/3, S. No. 129/5/2, Dalviwadi Road, Barangani Mala, Wadgaon Dhayari, Pune 411 041 (India). Printed in India

The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June, (2017)

EDITORIAL BOARD

Edited by J. R. Patadia and published by N. K. Thakare
for the Indian Mathematical Society.

Type set by J. R. Patadia at 5, Arjun Park, Near Patel Colony, Behind Dinesh Mill, Shivanand Marg, Vadodara-390 007 and printed by Dinesh Barve at Parashuram Process, Shed No. 1246/3, S. No.129/5/2, Dalviwadi Road, Barangani Mala, Wadgaon Dhayari, Pune – 411 041, Maharashtra, India. Printed in India