ISSN: 0025-5742

THE MATHEMATICS STUDENT

Volume 85, Numbers 1-2, January-June (2016)
(Issued: May, 2016)

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 (not highly technical, but 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.indianmath society.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 85, Numbers 1-2, January-June, (2016)
(Issued: May, 2016)

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. 85, Nos. 1-2, January-June (2016)

CONTENTS

****

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 01-08

MATHEMATICS IN INDIA - PERSONAL PERCEPTIONS

A. M. MATHAI

Dignitaries on the dais, off the dais, my fellow researchers, students and other participants, it is a great pleasure for me to be the President of the oldest scientific society and the truly national mathematical society in India.

1. Proliferation of Scientific/Professional Societies.

In India there are State-wise mathematical societies / associations / forums. Are such associations necessary? Mathematicians in a geographic or linguistic region may want to get together and discuss matters in their own regional language or cultural groups. Thus, such regional associations are needed. When a city has a large number of scientists they may form a local group. If there is a provision then people from outside may also join in such groups and participate in their activities. New York Academy, Gwalior Academy, Calcutta Mathematical Society, etc are such groups. New York Academy has a lot of activities and any researcher with published works from around the world can be a member there and participate in their activities. Such city-based or localized societies are also needed. If there are a sufficient number of people in a subject area in a country then they may launch a society for that subject area such as the Society for Special Functions and their Applications, Indian Society for Probability and Statistics, Applied Mathematics Forum, Semigroup Forum, etc are such subject-wise societies. Such societies are also needed to cater to the special needs. Then some people may like the work of a certain individual and they may be devotees of such a person and they may form a society to perpetuate the person's name. In olden days the followers of Pythagoras created a Pythagorean Society. We have a Ramanujan Society for Mathematics and Mathematical Sciences, Ramanujan Mathematical Society, etc. Such groups are also needed to cater to the special interest. Thus, in the Indian scene also, as in other places in the world, there are a large number of societies / associations catering to various aspects of mathematical sciences, catering to regional needs, catering to local needs, catering to topic-wise needs. Does this mean that we

* The text of the Presidential Address (general) delivered at the 81^{st} Annual Conference of the Indian Mathematical Society held at the Visvesvaraya National Institute of Technology, Nagpur-440 010, Maharashtra, during the period December 27 - 30, 2015.

need a consortium of mathematical societies / associations / forums in India to amalgamate and coordinate the activities? No such consortium is needed because all these regional, local, subject-wise societies are catering to special needs. As national activity, we have the Indian Mathematical Society (IMS), catering to all regions, all localities, all topics, all individuals' works, etc. Hence I would like to request those who are not yet Members of IMS, and interested in any aspect of mathematical sciences in the wider sense, to become the Life Members of IMS.

2. Indias Share of World Contributions in Mathematical Sciences

Nearly one-fifth of the world population is in the Indian subcontinent. Hence, one should naturally expect that around one-fifth of the research contributions in mathematical sciences in the world should be from India. But what is the reality? Not even one percent is from India. There may be some Indian names in the upper cadres but they are the ones settled abroad. Why? What are the real reasons behind this situation even after nearly 70 years of independence? Research in mathematical sciences do not call for expensive laboratories, expensive equipments, expensive infrastructure, etc. Then, why? I have located a few of the problems from my experience in India and in various countries outside India. I shall mention a few of the major ones that I have recognized.

(a). *Concept of learning has to change.*

For over forty years I had taught students at the faculty level (below average), major level (average) and honors level (brilliant) in various topics in Mathematics and Statistics in different countries outside India. During the past 9 years I had run 26 undergraduate mathematics training camps in India (trained over 1000 undergraduates) and run 12 research-orientation programs (trained over 400 all-India participants). In India, especially in Kerala, the concept of learning is to memorize a few mathematical formulae without really knowing the significance or relevance. Some students can recite theorems and even their proofs! But they have no idea why such a theorem is there, what is its importance, its relevance to real-life situations, etc. Most of the students had negative background in the subject matter under consideration. In the name of that subject a lot of nonsense had gone into their brains, and thus the background was negative. I had to de-program them and start from zero. I could bring out most of their latent abilities. Unless our students learn the subject matter, rather than memorizing formulae, they will not be able to make use of whatever they memorize. A real-life problem does not come in the form of a ready-made formula to compute.

I had recruited 17 M.Sc graduates from various colleges and started training them at my Centre (CMSS: Centre for Mathematical and Statistical Sciences, India) for their Ph.Ds. During the past seven years my students had won 19 national level awards for best published paper of the year, best paper presentation, best Ph.D thesis, young scientist award etc. They have published over 150 papers in standard refereed international journals. A number of them have written papers jointly with the top researchers in the world in various topics and over a dozen such papers are published. They could present papers in 8 conferences abroad. 13 out of the 17 received their Ph.Ds from Banaras Hindu University (BHU, 6 of them), Anna University, Chennai (3 or them), MG University, Kerala (4 of them). CMSS is a recognized research centre of these three universities.

(b). *Need for cutting down the number of holidays and extending the work hours.*

I achieved all the above publications and awards through a few simple steps. First, I put a condition that all research scholars must stay within walking distance from CMSS because in Kerala on almost every day some problems will be there on the road, either a bandh declared by a political party, or labor union, or local groups, etc or a bus strike or problems created by students' unions, etc. Then I abolished all the holidays except Sundays, Independence Day and Republic Day. Institution functioned on all days. Individuals are given leaves for personal and religious reasons. In Kerala a Vice-Chancellor of one of the universities said that his greatest ambition was to have 75 working days in a year. Then I put the working hours as 9 am to 5 pm but I used to be present from 7 am until 7 pm. Naturally, most of the research scholars and staff came by 8 am and left only around 6 pm. Then I removed all peons, servants, etc and made a rule that all must clean the premises and all, including me, must share all work in the Centre. Then I put strict discipline. At the first violation of any rule I removed the person from the program. Out of 17 only 3 were dismissed and one went abroad with her husband.

Outside India, no such discipline and enforcement are necessary to achieve the results. There the children are trained at home and in schools at various levels about civilized behavioral patterns, about personal cleanliness, about the use of private and public places, about respect and consideration for others' rights, etc. Hence one has to only mention the rules to be followed and discipline will be there.

If India has to catch up with the developed world then it is highly essential to cut down the number of holidays and extend the work hours. A work culture has to be created. It is possible in India and I have shown that it is possible. Outsiders will laugh at the system if we tell that the teaching and research institutions in India start at 10 am and finish by 3 to 3.30 pm, with two coffee breaks and one lunch break in between, and government offices function from 11 to 11.30 am until 2.30 to 3 pm.

(c). *Periodic work evaluations in all sectors*

Poor performance of the students, poor background and negative knowledge are not due to the fault of the students but these are due to the faults of their teachers. In my department of Mathematics and Statistics at McGill University there is no fixed salary increase every month or every year. Salary increment is based on the work output. Every faculty member is evaluated every year on three criteria of research output, teaching and administration. Then the total merit point is computed based on an open formula. Based on the merit points, the faulty members are put into 3 to 5 categories. Then the total amount available to the department for salary increase is divided among these 3 to 5 categories. Thus, in a certain year the top category may get $2000 increase but in certain other year the amount may be $1000 depending upon the total fund available to the department in that year.

When a faculty member is appointed at Assistant Professor level (starting level) there will be an expert committee set up for that purpose. Usually the members are mathematicians from outside the university. Then at each stage of promotion to Associate Professor, Full Professor, tenure, etc there is strict screening by expert committees, usually from outside. This type of continuous screening will make everyone productive and active in research, teaching and administration. Course-credit system gives the students wide choices of courses and they are free to register into any course provided the student has the required pre-requisite for that course. This can result in a professor not getting students to register in his/her courses. This will result in that professor losing the job. Also the students may make complaints questioning the knowledge of the teacher to teach a particular course. If the complaint is found to have substance, then also the professor loses the job. Then there is a periodic evaluation of every department by outside experts. Usually this is done in every five years. The recommendations of such outside committees are enforced by the university. The recommendations may be to remove certain professors or to bring in youngsters to certain areas or to scrap a certain sector of the department or to scrap the whole department. Such continuous evaluations of every faculty in every department, in every academic institution in India, and periodic evaluations of the departments and universities are necessary to improve the standards of higher education in India. UGC is the right agency in India to enforce such procedures in colleges, universities and institutions in India.

(d). *Overenthusiastic enforcement of local languages*

China has the advantage that the written language is the same all over the country, even though spoken language varies. India has a great disadvantage in this respect. Local languages and dialects develop due to the static nature of the society in olden days. If a group of people are born at a particular place, live and die within a small spread of land then naturally a local language or at least a local dialect will develop over time. But their ability to communicate with others will be nil or limited. If a child growing up in such an atmosphere wants to become a scientist then he / she should at least pick up the technical terms in the current language of science so that the child can easily learn whatever is available on the subject matter. The language of science in the world used to be French for some time. All over Europe, whoever learned French was considered to be learned. Then the language of science became German. After the Second World War, the language of science is English. Hence, any budding scientist will have great advantage if he/she knows English or can communicate in English. Once the symbols and technical terms are picked up in English then it is easy for a child to follow a mathematical statement. When a child is hearing a technical term for the first time, it is equally alien to the child, whatever be the language. But if a technical term is phrased in the mother tongue then it is likely that the child may misinterpret it and connect it to something else in the mother tongue. For example, instead of calling the symbol + as "plus" (English word) if the child is taught to say "adhikam" (Malayalam) then, to the child, "plus" and "adhikam" are both alien and "adhikam" is more difficult to pronounce. In Malayalam, "adhikam" is commonly used for "more", "more than", "greater", "greater than", etc. The child will not understand the precise meaning of the symbol + if he is learning it as "adhikam". In the name of promoting local language, such awkward technical terms are coined in each language in India now. How can a person communicate in science with his next State neighbor, let alone communicating with the outside world?

When I was giving the undergraduate mathematics training at CMSS to college level students, I heard some students saying "chamithi", "chamithi". Where I grew up, this word "chamithi" meant cleaning up after going to the toilet. I never knew that it was a mathematical operation. Later I learnt that it was a mispronounced technical term for some mathematical operation in the name of promoting Malayalam.

Any forward looking country must make sure that the youngsters can talk to each other as well as with outsiders, in science, when they grow up. India is going backward in this respect and destroying opportunities to the vast majority of children, whose parents cannot afford to send them to English medium schools, in the name of "promoting" local language. This is not a promotion but nipping off at the bud prospective scientists of the country.

3. Establishment of Research and Training Centres all Across India

About 30 years back Chinese students from mainland China started coming to North America in large numbers in all subject areas. In Canada the Chinese government representatives came to each Province and made deals so that these Chinese scholars would get fee exemption or fee reduction or fellowships etc. I found the Chinese students coming to our department very good, well motivated and eager to learn everything that they could learn from us. I was wondering how they could all be good students without exception. While talking to them I found that China had established various centres such as my Centre (CMSS) in different parts of China. Then they recruited motivated students from all across China and they were given special training in such centres before sending abroad. What is the net result? When such students come to a department in a university in Canada or USA they create a very good impression. Professors normally help students who show promise. They get their Ph.Ds with flying colors and they will be the best candidates for appointments in universities and they naturally fill all vacancies since the North American system is an open system based on merit.

Instead of making national effort in creating such training centres (such as CMSS) all across India and recruiting motivated students to train in such centres, what is being done in India is to dismantle the only working centre CMSS, by narrow-minded people sitting in decision-making committees. I have given a summary of the achievements of CMSS during the past seven years. Such performance may not be there anywhere in India. The renewal request for further grant to CMSS, to take one more batch of students and train them, went through all the relevant stages except the very last committee. The request was turned down by this last committee, citing age limit as a reason. The maximum age was specified as 70 years and I am older than that. In other countries what is looked into are the following in such a situation: Is the applicant currently active in research? Is he/she active during the past three years? Was he/she consistently productive? Is he/she physically and mentally healthy to carry out the proposed work? Is his/her work up to international standards? If all these criteria are met, then the age limit, if any, is relaxed and funds are released. This requires the presence of people with broad-mind, vision and foresight in the decision-making super committees. If such a step does not come, then India will not be able to catch up with the developed countries and contributions from India in mathematical sciences will stay at the current insignificant level.

4. Need for Appointing Proper People in Decision-making Bodies in India

When top bureaucrats (the government) are looking to appoint persons into decision-making super committees or granting agencies, they seem to listen only to people who sit around and praise each other. Now-a-days bureaucrats do not have to rely on such gossips to evaluate the standing of a person in the field. Google's scholar citations are there. This gives criteria such as h-index, i10-index, etc. This is a criterion showing how others found the person's research output useful. Then the bureaucrats can type in the topic or subject matter. Google will print the rank list of people who are cited most in that topic. This is another criterion to check the person's standing in the topic or subject matter. The bureaucrats can check into the person's publications and impact factor of each journal where the papers appeared. Impact factor is another criterion that can be used to check the relative standing of the journal. Impact factor is computed by using all citations. Mathematicians may claim that the impact factors of mathematics journals are very low. The bureaucrats can compare mathematics journals and check the relative standing of a particular journal among mathematics journals. This is another criterion. Through such internationally accepted criteria the bureaucrats can line up qualified people for appointments to decision-making bodies, rather than relying on self-generated gossips about a person's standing in the field. After lining up qualified people, it is very essential that one should look into the broad-mindedness, balanced views, lack of biases in the form of caste, creed, regionalism, etc and appoint the best among the qualified people into such committees. Then research output from India in mathematical sciences will improve dramatically from year to year.

5. Need for Transparency and Changing the Funding System for Mathematical Sciences in India

If India has to come to the forefront in all topics in mathematical sciences at the international scene then the definition of mathematics used in India, in practice, has to change. Mathematics is not only number theory or differential equations or nonlinear analysis or combinations of these. It has a wide spectrum of areas and in each area a wide spectrum of topics. Unless we give equal respect to all topics in each area, India can never come to the forefront. In any particular topic if India's contribution comes to the top internationally at any time, then give more than average encouragement to this topic, but not by cutting off funds to any other topic. Then topic by topic India will be able to capture the top positions in international scene. If you look into Google's scholar citations, even for small topics, some Chinese names are there in the set of top ten. This is a recent phenomenon, as a result of national efforts in China. A small group sitting around and praising each other will not make India's contribution in any topic great internationally. It is said that if one has a constant diet of starch alone then he gets a disease called beriberi, affecting the brain, thinking capacity and nervous system. This seems to be what has happened to the Indian mathematical scene. Strong and immediate steps are needed to cure the situation.

Before I make any comment on funding agencies in India, I would like to take this opportunity to thank NBHM on behalf of IMS and on my own for its generosity in granting sufficient funds for the 2015 annual conference of IMS. I hope that this healthy shift in NBHM's policy towards IMS and its activities will continue and that future funding will be there at a healthy level.

I had dealt with DST, NBHM and UGC on various occasions from 1985 onward. I found DST the most transparent one. The Department of Atomic Energy (DAE) and National Board for Higher Mathematics (NBHM) handled research funds in mathematical sciences all these years. What is the net result? Research contribution from India in mathematical sciences is miniscule at the international scene. There must be something wrong, either in the philosophy of funding for research in mathematical sciences or in the procedures used. I would like to see a broader-based National Board for Mathematical and Statistical Sciences (NBMSS), covering all subject matters in the wider sense, and reconstituted as an autonomous body within DST, parallel SERB in the mathematical sciences division. There is no use of creating another body or reconstituting existing body unless the administrators are selected as per the criteria suggested in paragraph 4 above.

I would like to make another remark in this context. People involved in central grant giving agencies should be selected by using international criteria. They should be broad-minded without biases of any sort in the name of caste, creed, religion and regionalism etc. This will encourage research output from all sections of people and all regions of India.

A. M. Mathai
Centre for Mathematical and Statistical Sciences
Peechi Campus, KFRI Peechi-680653, Kerala, India
E-mail: *directorcms458@gmail.com, mathai@math.mcgill.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 09-44

A VERSATILE AUTHOR'S CONTRIBUTIONS TO VARIOUS AREAS OF MATHEMATICS, STATISTICS, ASTROPHYSICS, BIOLOGY AND SOCIAL SCIENCES*

A. M. MATHAI

Abstract. An overview of Mathai's work in the following topics is given: Fractional Calculus - real scalar variable case, solutions of fractional differential equations, extensions of fractional integrals to functions of matrix argument, extension to complex domains, establishing a connection between Fractional Calculus and Statistical Distribution Theory, a new general definition for Fractional Calculus; Functions of Matrix Argument - introduction of M-transforms, M-convolutions; Krätzel integrals - Krätzel density, extensions; Pathway Models - scalar and matrix-variate cases, extension to complex matrix-variate cases; Geometrical Probabilities - proof of a conjecture, new conjectures and proofs, random parallelotopes; Astrophysics - reaction rate theory, resonant and non-resonant reactions, depleted and tail cut-off cases, solar and stellar models, gravitational instability and solutions to differential equations; Special Functions - Popularization in Statistics and Physical Sciences, computable representations, G and H-functions; Multivariate Analysis - structural decompositions of λ-criteria, exact null and non-null distributions in the general cases, 11-digit accurate percentage points; Algorithms for non-linear least squares; Characterizations - characterizations of densities, information measure, axiomatic definitions, pseudo analytic functions of matrix argument and characterization of the normal probability law; Mathai's entropy - entropy optimization; Graph Theory - almost cubic maps, minimum number of specifiers, various descriptors; Analysis of Variance - approximate analysis , matrix series; Dispersion Theory; Population Problems and Social Sciences; Integer Programming - optimizing a linear function under quadratic constraints on a grid of positive integers; Quadratic and Bilinear Forms.

1. Mathai's Contributions to Fractional Calculus

Fractional integrals, fractional derivatives and fractional differential equations were available only for real scalar variables. The most popular fractional integrals in the literature are Riemann-Liouville fractional integrals given by the following:

---

* The text of the Presidential address (technical) delivered at the 81^{st} Annual Conference of the Indian Mathematical Society held at the Visvesvaraya National Institute of Technology, Nagpur -440 010, Maharashtra, during the period December 27-30, 2015.

Key words and Phrases: Fractional calculus, functions of matrix argument, Krätzel integral, special functions, geometrical probabilities, astrophysics, multivariate analysis, characterization, algorithms, quadratic and bilinear forms.

$$ {}_aD_x^{-\alpha}f = \frac{1}{\Gamma(\alpha)}\int_a^x (x-t)^{\alpha-1}f(t)\mathrm{d}t, \Re(\alpha) > 0, \tag{1.1} $$

where $\Re(\cdot)$ denotes the real part of $(\cdot)$.

$$ {}_xD_b^{-\alpha}f = \frac{1}{\Gamma(\alpha)}\int_x^b (t-x)^{\alpha-1}f(t)\mathrm{d}t, \Re(\alpha) > 0. \tag{1.2} $$

Here $-\alpha$ in the exponent of D indicates an integral. The D with positive exponent ${}_aD_x^{\alpha}f, {}_xD_b^{\alpha}f$ will be used to denote the corresponding fractional derivatives. Here (1.1) is called Riemann-Liouville left-sided or first kind fractional integral of order α and (2.2) is called Riemann-Liouville fractional integral of order α of the second kind or right-sided. If $a = -\infty$ and $b = \infty$ then (1.1) and (1.2) are called Weyl fractional integrals of order α and of the first kind and second kind respectively or the left-sided and right-sided ones. There are various other fractional integrals in the literature, introduced by various authors from time to time. Mathai was trying to find an interpretation or connection of fractional integrals in terms of statistical densities and random variables. In Mathai (2009), an interpretation is given for Weyl fractional integrals as densities of sum (first kind) and difference (second kind) of independently distributed real positive random variables having special types of densities. Fractional integrals were also given interpretations as fractions of total integrals coming from gamma and type-1 beta random variables. Also Weyl fractional integrals were extended to real matrix-variate cases there. These ideas did not make a proper connection to statistical distribution theory.

1.1. Mellin convolutions of products and ratios

Then while working on Mellin convolutions of products and ratios, Mathai found that a fusion of Fractional Calculus and Statistical Distribution Theory was possible which also opened up ways of extending fractional calculus to real scalar functions of matrix argument, when the argument matrix is real or in the complex domain. Let us consider real scalar variables first. The Mellin convolution of a product of two functions $f_1(x_1)$ and $f_2(x_2)$ says the following: Consider the integral

$$ g_2(u_2) = \int_v \frac{1}{v} f_1(\frac{u}{v}) f_2(v)\mathrm{d}v. \tag{1.3} $$

Then the Mellin transform of $g_2(u_2)$, with Mellin parameter s, is the product of the Mellin transforms of f_1 and f_2. That is

$$ M_{g_2}(s) = M_{f_1}(s)M_{f_2}(s), \tag{1.4} $$

where

$$ M_{f_1}(s) = \int_0^\infty x_1^{s-1}f_1(x_1)\mathrm{d}x_1 \text{ and } \int_0^\infty x_2^{s-1}f_2(x_2)\mathrm{d}x_2 = M_{f_2}(s). $$

It is easy to note that if $g_2(u_2)$ is written as

$$g_2(u_2) = \int_v \frac{1}{v} f_1(v) f_2(\frac{u}{v}) \mathrm{d}v \tag{1.5}$$

then also the formula in (1.4) holds. Thus, the Mellin convolution of a product has the two integral forms in (1.3) and (1.5). But, Mellin convolution of a ratio will have four different representations. Two of these that we will make use of are the following

$$M_{g_1}(u_1) = M_{f_1}(s) M_{f_2}(2-s) \tag{1.6}$$

where

$$g_1(u_1) = \int_v v f_1(uv) f_2(v) \mathrm{d}v \tag{1.7}$$

and

$$M_{g_1}(s) = M_{f_1}(2-s) M_{f_2}(s) \tag{1.8}$$

where

$$g_1(u_1) = \int_v \frac{v}{u_1^2} f_1(\frac{v}{u_1}) f_2(v) \mathrm{d}v. \tag{1.9}$$

1.2. Statistical interpretations of Mellin convolutions

Let x_1 and x_2 be real scalar positive random variables, independently distributed, with densities $f_1(x_1)$ and $f_2(x_2)$ respectively. Let $u_2 = x_1 x_2$ and $u_1 = \frac{x_2}{x_1}$, $v = x_2$. Then the Jacobians are $\frac{1}{v}$ and $-\frac{v}{u_1^2}$ respectively or

$$\mathrm{d}x_1 \wedge \mathrm{d}x_2 = \frac{1}{v} \mathrm{d}u_2 \wedge \mathrm{d}v$$

and

$$\mathrm{d}x_1 \wedge \mathrm{d}x_2 = -\frac{v}{u_1^2} \mathrm{d}u_1 \wedge \mathrm{d}v.$$

The joint density of x_1 and x_2 is $f_1(x_1) f_2(x_2)$ due to statistical independence and then the marginal densities of u_2 and u_1, denoted by $g_2(u_2)$ and $g_1(u_1)$ are given by the following

$$g_2(u_2) = \int_v \frac{1}{v} f_1(\frac{u_2}{v}) f_2(v) \mathrm{d}v \tag{1.10}$$

and

$$g_1(u_1) = \int_v \frac{v}{u_1^2} f_1(\frac{v}{u_1}) f_2(v) \mathrm{d}v. \tag{1.11}$$

In u_2, if x_1 is taken as v then the roles of f_1 and f_2 change in (1.10). If x_1 in u_1 is taken as v then we get (1.7) with the roles of f_1 and f_2 interchanged. Hence $\frac{x_1}{x_2}$ gives two forms and $\frac{x_2}{x_1}$ gives two forms for Mellin convolution of ratios. The Mellin convolutions in (1.4) and (1.8) can be easily interpreted in terms of random variables. $u_2 = x_1 x_2$ gives $E(u_2^{s-1}) = E(x_1^{s-1}) E(x_2^{s-1})$ due to independence where $E(\cdot)$ denotes the expected value. That is,

$$E(u_2^{s-1}) = \int_0^\infty u_2^{s-1} g_2(u_2) \mathrm{d}u_2 = M_{g_2}(s).$$

Similarly

$$E(x_1^{s-1}) = \int_0^\infty x_1^{s-1} f_1(x_1) = M_{f_1}(s) \text{ and } E(x_2^{s-1}) = M_{f_2}(s).$$

This means $M_{g_2}(s) = M_{f_1}(s)M_{f_2}(s)$, which is (1.4) the Mellin convolution of a product. Now, consider $u_1 = \frac{x_2}{x_1}$. Then $E(u_1^{s-1}) = E(x_2^{s-1})E(x_1^{-s+1})$ due to statistical independence. This means

$$E(u_1^{s-1}) = \int_0^\infty u_1^{s-1} g_1(u_1)\mathrm{d}u_1 = M_{g_1}(s),$$

$$E(x_2^{s-1}) = \int_0^\infty x_2^{s-1} f_2(x_2)\mathrm{d}x_2 = M_{f_2}(s)$$

and

$$E(x_1^{-s+1}) = \int_0^\infty x_1^{-s+1} f_1(x_1)\mathrm{d}x_1 = \int_0^\infty x_1^{(2-s)-1} f_1(x_1)\mathrm{d}x_1 = M_{f_1}(2-s).$$

In other words, $M_{g_1}(s) = M_{f_1}(2-s)M_{f_2}(s)$, which is (1.8), one form of Mellin convolution of a ratio. Mellin convolutions of products and ratios make direct connection to product and ratio of real positive random variables.

1.3. Mellin convolutions, statistical densities and fractional integrals
Let $f_1(x_1)$ and $f_2(x_2)$ be statistical densities as in Section 1.2. Let x_1 have a type-1 beta density with parameters $(\gamma+1, \alpha)$ or

$$f_1(x_1) = \frac{\Gamma(\gamma+1+\alpha)}{\Gamma(\gamma+1)\Gamma(\alpha)} x_1^\gamma (1-x_1)^{\alpha-1}, 0 \le x_1 \le 1, \Re(\alpha) > 0, \Re(\gamma) > -1$$

and zero elsewhere. Let $f_2(x_2) = f(x_2)$ an arbitrary density. Then the Mellin convolution of a product is given by

$$\begin{aligned} g_2(u_2) &= \int_v \frac{1}{v} f_1(\frac{u_2}{v}) f(v)\mathrm{d}v \\ &= \frac{\Gamma(\gamma+1+\alpha)}{\Gamma(\gamma+1)} \frac{1}{\Gamma(\alpha)} \int_v (\frac{u_2}{v})^\gamma (1-\frac{u_2}{v})^{\alpha-1} f(v)\mathrm{d}v \\ &= \frac{\Gamma(\gamma+1+\alpha)}{\Gamma(\gamma+1)} K_{2,u_2,\gamma}^{-\alpha} f, \end{aligned} \tag{1.12}$$

where

$$K_{2,u_2,\gamma}^{-\alpha} f = \frac{u_2^\gamma}{\Gamma(\alpha)} \int_{v>u_2} v^{-\gamma-\alpha}(v-u_2)^{\alpha-1} f(v)\mathrm{d}v, \Re(\alpha) > 0, \Re(\gamma) > -1 \tag{1.13}$$

is Kober fractional integral of order α and of the second kind with parameter γ. This is a direct connection among Kober fractional integral of the second kind, Mellin convolution of a product and statistical density of product of two independently distributed real scalar positive random variables where one has a type-1 beta density with parameters $(\gamma+1, \alpha)$ and the other has an arbitrary density. Mathai had illustrated this connection of fractional integrals to statistical distribution theory in 2011-2012 period. Then the idea was extended to fractional calculus of matrix-variate functions and then to complex matrix-variate distributions. These developments were summarized and a series of four articles were put on Cornell University arXiv as Mathai-Haubold papers.

1.4. General definition for fractional integrals

Motivated by this observation, Mathai has given a new definition for fractional integrals of the first and second kinds of order α. This was given in 2011-2012. Let

$$f_1(x_1) = \phi_1(x_1)\frac{1}{\Gamma(\alpha)}(1-x_1)^{\alpha-1}, 0 \le x_1 \le 1, \Re(\alpha) > 0$$

and $f_1(x_1) = 0$ elsewhere, and $f_2(x_2) = \phi_2(x_2)f(x_2)$ where ϕ_1 and ϕ_2 are prefixed functions and $f(x_2)$ is an arbitrary function, where f_1 and f_2 need not be statistical densities. Let us consider the Mellin convolution of a product. Then using the same notation as $g_2(u_2)$ we have

$$\begin{aligned} g_2(u_2) &= \int_v \frac{1}{v} f_1(\frac{u_2}{v}) f_2(v) \mathrm{d}v \\ &= \int_v \frac{1}{v}\phi_1(\frac{u_2}{v})\frac{1}{\Gamma(\alpha)}(1-\frac{u_2}{v})^{\alpha-1}\phi_2(v)f(v)\mathrm{d}v \\ &= \int_{v>u_2} \phi_1(\frac{u_2}{v})\frac{v^{-\alpha}}{\Gamma(\alpha)}(v-u_2)^{\alpha-1}\phi_2(v)f(v)\mathrm{d}v \qquad (1.14) \\ &= \frac{1}{\Gamma(\alpha)}\int_{v>u_2}(v-u_2)^{\alpha-1}f(v)\mathrm{d}v \text{ for } \phi_1 = 1, \phi_2(x_2) = x_2^{\alpha}. \qquad (1.15) \end{aligned}$$

But (1.15) gives Weyl fractional integral of the second kind of order α. If v is bounded above by a constant b then (1.15) is Riemann-Liouville fractional integral of the second kind of order α. Thus, by specifying ϕ_1 and ϕ_2 it can be seen that all fractional integrals of order α of the second kind can be obtained from (1.14). Evidently when $\phi_1(x_1) = x_1^{\gamma}$ and $\phi_2(x_2) = 1$ then we have (1.13) or Kober fractional integral of order α of the second kind with parameter γ.

1.5. First kind fractional integrals

Let $f_1(x_1)$ and $f_2(x_2)$ be as given above in Section 1.4. Let us consider (1.9) the Mellin convolution of a ratio. Let us use the same notation. Then

$$\begin{aligned} g_1(u_1) &= \int_v \frac{v}{u_1^2} f_1(\frac{v}{u_1}) f_2(v)\mathrm{d}v \\ &= \int_v \frac{v}{u_1^2}\phi_1(\frac{v}{u_1})\frac{1}{\Gamma(\alpha)}(1-\frac{v}{u_1})^{\alpha-1}\phi_2(v)f(v)\mathrm{d}v. \qquad (1.16) \end{aligned}$$

Let $\phi_1(x_1) = x_1^{\gamma-1}$ and $\phi_2 = 1$. Then (1.16) reduces to the following form

$$g_1(u_1) = \frac{u_1^{-\gamma-\alpha}}{\Gamma(\alpha)}\int_{v<u_1} v^{\gamma}(u_1-v)^{\alpha-1}f(v)\mathrm{d}v = K_{1,u_1,\gamma}^{-\alpha}f. \qquad (1.17)$$

Then

$$g_1^*(u_1) = \frac{\Gamma(\gamma+\alpha)}{\Gamma(\gamma)}K_{1,u_1,\gamma}^{-\alpha}f, \Re(\alpha) > 0, \Re(\gamma) > 0 \qquad (1.18)$$

is a statistical density when f_1 and f_2 are statistical densities. This is Kober fractional integral operator of the first kind of order α and parameter γ, denoted by $K^{-\alpha}_{1,u_1,\gamma}f$. From (1.16), by specializing ϕ_1 and ϕ_2 one can get all fractional integrals of order α of the first kind in the real scalar case, defined by various authors from time to time.

This formal definition of fractional integrals as Mellin convolutions of ratio and product was introduced formally in Mathai (2013). A geometrical interpretation of fractional integrals as fractions of integral over a simplex in n-space is given in Mathai (2014). Earlier, in Mathai (2009), an interpretation was given as fractionas of total probability in gamma and type-1 beta distributions.

1.6. Extension of fractional integrals to real matrix-variate case

Mathai (2009) introduced fractional integrals in the real matrix-variate case for the first time but they could not be given any physical interpretations. In Mathai (2013) there are interpretations in terms of statistical distribution problem and M-convolutions introduced in Mathai (1997). When M-convolutions were introduced in (1997), the author could not find any physical interpretation. Now, a very meaningful interpretation is given as the densities of product and ratio of matrix-variate random variables. Some details are in the following.

All the matrices appearing here are $p\times p$ real positive definite matrices, unless specified otherwise. The notation $X_j > O$ means the $p\times p$ real symmetric matrix, $X_j = X_j'$, is positive definite, where a prime denotes the transpose. $X_j^{\frac{1}{2}}$ means the positive definite square root of the positive definite matrix X_j. If $X = (x_{ij})$ is $m\times n$ then the wedge product of differentials will be denoted by $\mathrm{d}X$, that is,

$$\mathrm{d}X = \prod_{i=1}^{m}\prod_{j=1}^{m}\wedge \mathrm{d}x_{ij}$$

and it is $\prod_{i\ge j}\wedge \mathrm{d}x_{ij}$ if $m = n = p$ and $X = X'$. Also, $\int_A^B f(X)\mathrm{d}X = \int_{A<X<B} f(X)\mathrm{d}X$ means the integral over all $X > O$ of the real-valued scalar function $f(X)$ of X, such that $A > O, B > O, X - A > O, B - X > O$ where A and B are positive definite constant matrices. We will need some Jacobians of matrix transformations in our discussion. These will be given as lemmas, without proofs. For proofs and for other such Jacobians see Mathai (1997).

Lemma 1.1. *Let $X = (x_{ij})$ be $m\times n$ matrix of distinct real scalar variables x_{ij}'s. Let A be $m\times m$ and B be $n\times n$ nonsingular constant matrices. Then*

$$Y = AXB \Rightarrow \mathrm{d}Y = |A|^n|B|^m\mathrm{d}X \tag{1.19}$$

where $|(\cdot)|$ denotes the determinant of $(\cdot)$.

Lemma 1.2. *Let $X = X'$ be $p\times p$. Let A be a $p\times p$ nonsingular constant matrix. Then*

$$Y = AXA' \Rightarrow \mathrm{d}Y = |A|^{p+1}\mathrm{d}X. \tag{1.20}$$

Lemma 1.3. *Let X be a $p \times p$ nonsingular matrix. Let $Y = X^{-1}$. Then*

$$Y = X^{-1} \Rightarrow \mathrm{d}Y = \begin{cases} |X|^{-2p}\mathrm{d}X, \ \textit{for a general } X \\ |X|^{-(p+1)}\mathrm{d}X \ \textit{for } X = X'. \end{cases} \tag{1.21}$$

Lemma 1.4. *Let $X > O$ be $p \times p$. Let $T = (t_{ij})$ be a lower triangular matrix with positive diagonal elements, that is, $t_{ij} = 0, i < j, t_{jj} > 0, j = 1, ..., p$. Then*

$$X = TT' \Rightarrow \mathrm{d}X = 2^p\{\prod_{j=1}^{p} t_{jj}^{p+1-j}\}\mathrm{d}T. \tag{1.22}$$

With the help of (1.22) we can evaluate a matrix-variate gamma integral and write the result as

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}}\Gamma(\alpha)\Gamma(\alpha - \frac{1}{2})...\Gamma(\alpha - \frac{p-1}{2}), \Re(\alpha) > \frac{p-1}{2} \tag{1.23}$$

where

$$\int_{X>O} |X|^{\alpha - \frac{p+1}{2}} \mathrm{e}^{-\mathrm{tr}(X)}\mathrm{d}X = \Gamma_p(\alpha), \Re(\alpha) > \frac{p-1}{2} \tag{1.24}$$

with $\mathrm{tr}(\cdot)$ denoting the trace of $(\cdot)$. Combining (1.24) and (1.20) we can define a matrix-variate gamma density as

$$h_1(X) = \begin{cases} \frac{|B|^\alpha}{\Gamma_p(\alpha)}|X|^{\alpha - \frac{p+1}{2}}\mathrm{e}^{-\mathrm{tr}(BX)}\mathrm{d}X, X > O, B > O, \Re(\alpha) > \frac{p-1}{2} \\ 0, \ \text{elsewhere}. \end{cases} \tag{1.25}$$

Since the total integral is 1, from (1.25) we have the identity

$$|B|^{-\alpha} \equiv \frac{1}{\Gamma_p(\alpha)}\int_{X>O} |X|^{\alpha - \frac{p+1}{2}}\mathrm{e}^{-\mathrm{tr}(BX)}\mathrm{d}X, B > O. \tag{1.26}$$

This identity will be used to establish fractional derivatives in a class of matrix-variate functions. The real matrix-variate type-1 beta density is defined as

$$h_2(X) = \begin{cases} \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)}|X|^{\alpha - \frac{p+1}{2}}|I - X|^{\beta - \frac{p+1}{2}}, O < X < I, \Re(\alpha) > \frac{p-1}{2}, \Re(\beta) > \frac{p-1}{2} \\ 0, \ \text{elsewhere}. \end{cases} \tag{1.27}$$

There is a corresponding type-2 beta density, which is of the form

$$h_3(X) = \begin{cases} \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)}|X|^{\alpha - \frac{p+1}{2}}|I + X|^{-(\alpha+\beta)}, X > O, \Re(\alpha) > \frac{p-1}{2}, \Re(\beta) > \frac{p-1}{2} \\ 0, \ \text{elsewhere}, \end{cases} \tag{1.28}$$

1.7. Fractional integrals for the real matrix-variate case

With the preliminaries in Section 1.6 we can define fractional integrals in the real matrix-variate case. Let $X_1 > O$ and $X_2 > O$ be $p \times p$ real matrix-variate random variables, independently distributed. Let $U_2 = X_2^{\frac{1}{2}}X_1X_2^{\frac{1}{2}}$ be defined as the product of X_1 and X_2 and let $U_1 = X_2^{\frac{1}{2}}X_1^{-1}X_2^{\frac{1}{2}}$ be defined as the ratio of X_2

over X_1. Let $V = X_2$. Then with the help of the above lemmas we can show that, ignoring sign,

$$\mathrm{d}X_1 \wedge \mathrm{d}X_2 = |V|^{-\frac{p+1}{2}} \mathrm{d}U_2 \wedge \mathrm{d}V = |V|^{\frac{p+1}{2}} |U_1|^{-(p+1)} \mathrm{d}U_1 \wedge \mathrm{d}V. \tag{1.29}$$

Denoting the densities of U_2 and U_1 as $g_2(U_2)$ and $g_1(U_1)$, we can compute these by using the densities $f_1(X_1)$ and $f_2(X_2)$ through transformation of variables and they will be the following

$$g_2(U_2) = \int_V |V|^{-\frac{p+1}{2}} f_1(V^{-\frac{1}{2}} U_2 V^{-\frac{1}{2}}) f_2(V) \mathrm{d}V \tag{1.30}$$

and

$$g_1(U_1) = \int_V |V|^{\frac{p+1}{2}} |U_1|^{-(p+1)} f_1(V^{\frac{1}{2}} U_1^{-1} V^{\frac{1}{2}}) f_2(V) \mathrm{d}V. \tag{1.31}$$

Let $f_1(X_1)$ be a type-1 beta density of the type in (1.27) with parameters $(\gamma + \frac{p+1}{2}, \alpha)$. Note that in (1.27) the parameters are (α, β). Then $g_2(U_2)$ will be of the following form

$$\begin{aligned} g_2(U_2) &= \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})} \frac{|U_2|^{\gamma}}{\Gamma_p(\alpha)} \int_{V>U_2} |V|^{-\alpha-\gamma} |V - U_2|^{\alpha - \frac{p+1}{2}} f(V) \mathrm{d}V \\ &= \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})} K_{2,U_2,\gamma}^{-\alpha} f. \end{aligned} \tag{1.32}$$

This $K_{2,U_2,\gamma}^{-\alpha} f$ in (1.32) for $p = 1$ is Kober fractional integral of the second kind of order α and parameter γ and hence Mathai called the integral as Kober fractional integral of order α and parameter γ of the second kind in the real matrix-variate case. This notation is also due to him. In a similar fashion, Kober fractional integral of order α of the first kind with parameter γ, available from (1.31) by taking $f_1(X_1)$ as a real matrix-variate type-1 beta with parameters (γ, α) is the following

$$K_{1,U_1,\gamma}^{-\alpha} f = \frac{|U_1|^{-\gamma-\alpha}}{\Gamma_p(\alpha)} \int_{V<U_1} |V|^{\gamma} |U_1 - V|^{\alpha - \frac{p+1}{2}} f(V) \mathrm{d}V. \tag{1.33}$$

The density of U_1, again denoted by $g_1(U_1)$, is given by

$$g_1(U_1) = \frac{\Gamma_p(\gamma + \alpha)}{\Gamma_p(\gamma)} K_{1,U_1,\gamma}^{-\alpha} f \tag{1.34}$$

where the first kind Kober fractional integral in the matrix-variate case is given in (1.33).

The above notations as well as a unified notation for fractional integrals and fractional derivatives were introduced by Mathai (2013,2014,2015, *Linear Algebra and its Applications*).

The above results in the real matrix-variate case are extended to complex matrix-variate cases, see Mathai (2013), to many matrix-variate cases, see Mathai (2014) and also the corresponding fractional derivatives in the matrix-variate case are worked out in Mathai (2015). The matrix differential operator introduced in Mathai (2015) is not a universal one, even though it works on some wide classes of functions. The matrix differential operator is introduced through the following symbolic representation. Let D be a differential operator defined for real matrix-variate case. Then D^{α} and $D^{-\alpha}$ represent αth order fractional derivative and fractional integral respectively. Then

$$D^{\alpha}f = D^{n}D^{-(n-\alpha)}f, n = 1, 2, ..., \Re(n-\alpha) > \frac{p-1}{2}, \Re(\alpha) > \frac{p-1}{2}.$$

This is the αth order fractional derivative in Riemann-Liouville sense. Consider

$$D^{\alpha}f = D^{-(n-\alpha)}D^{n}f, n = 1, 2, ..., \Re(\alpha) > \frac{p-1}{2}, \Re(n-\alpha) > \frac{p-1}{2}$$

is the αth order fractional derivative in the Caputo sense. In the Caputo case, D^n operates on f first and then the fractional integral $D^{-(n-\alpha)}$ is taken, whereas in the Riemann-Liouville sense, the $(n-\alpha)$th order fractional integral is taken first and then D^n operates on this. A universal differential operator D in the real as well as complex matrix-variate case is still an open problem.

2. Mathai's Work on Krätzel Integrals

Let x be a real scalar positive variable. Consider the integrals

$$I_1 = \int_0^{\infty} x^{\gamma-1}\mathrm{e}^{-ax^{\delta}-bx^{\rho}}\mathrm{d}x, a > 0, b > 0, \delta > 0, \rho > 0 \tag{2.1}$$

and

$$I_2 = \int_0^{\infty} x^{\gamma}\mathrm{e}^{-ax^{\delta}-bx^{-\rho}}, a > 0, b > 0, \delta > 0, \rho > 0. \tag{2.2}$$

Structures such as the ones in (2.1) and (2.2) appear in many different areas. This (2.2) for $\delta = 1, \rho = 1$ is the basic Krätzel integral, see Krätzel (1979). For $\rho = 1$ and general $\delta > 0$ is the generalized Krätzel integral. An integral transform of the form

$$I_3 = \int_0^{\infty} \mathrm{e}^{-ax-bx^{-1}}f(x)\mathrm{d}x \tag{2.3}$$

where $f(x)$ is arbitrary so that I_3 exists, is known as Krätzel transform. The structures of the integral in (2.2) and (2.1) are very interesting ones. Mathai has investigated various aspects of (2.1) and (2.2) in detail and he has also introduced a statistical density in terms of Krätzel integral. The structures in (2.2) and (2.1) can be generated as Melin convolutions of product and ratio. Consider the real scalar variables x_1 and x_2 and the corresponding functions $f_1(x_1)$ and $f_2(x_2)$. Then it is seen from (1.10) that the Mellin convolution of a product is given by

$$g_2(u_2) = \int_v \frac{1}{v}f_1(\frac{u_2}{v})f_2(v)\mathrm{d}v \text{ and } M_{g_2}(s) = M_{f_1}(s)M_{f_2}(s) \tag{2.4}$$

or $u_2 = x_1x_2, v = x_2$, and the Mellin convolution of a ratio, from (1.7), as

$$g_1(u_1) = \int_v vf_1(u_1v)f_2(v)\mathrm{d}v \text{ and } M_{g_1}(s) = M_{f_1}(s)M_{f_2}(2-s) \tag{2.5}$$

or $u_1 = \frac{x_1}{x_2}, v = x_2$. Let f_1 and f_2 be generalized gamma functions of the form

$$f_j(x_j) = x_j^{\alpha_j} \mathrm{e}^{-a_j x_j^{\beta_j}}, x_j > 0, a_j > 0, \beta_j > 0, j = 1, 2. \tag{2.6}$$

Then $g_1(u_1)$ of (2.5) reduces to the form

$$g_1(u_1) = u_1^{\alpha_1} \int_0^\infty v^{\alpha_1+\alpha_2+1} \mathrm{e}^{-a_1(u_1 v)^{\beta_1} - a_2 v^{\beta_2}} \mathrm{d}v. \tag{2.7}$$

This is the form in (2.1). Now, consider Mellin convolution of a product when f_1 and f_2 are generalized gamma functions in (2.6). Then $g_2(u_2)$ reduces to the form

$$g_2(u_2) = u_2^{\alpha_1} \int_0^\infty v^{\alpha_1-\alpha_2-1} \mathrm{e}^{-a_1(\frac{u_2}{v})^{\beta_1} - a_2 v^{\beta_2}} \mathrm{d}v. \tag{2.8}$$

This is the form in (2.2). Hence (2.1) and (2.2) can be treated as Mellin convolutions of ratio and product when f_1 and f_2 are generalized gamma functions.

Note that if f_1 and f_2 are multiplied by the corresponding normalizing constants c_1 and c_2 then f_1 and f_2 become statistical densities. Let x_1 and x_2 be independently distributed real scalar positive random variables. Let $u_2 = x_1 x_2, u_1 = \frac{x_1}{x_2}, v = x_2$. Then the densities of u_2 and u_1 are given by (2.4) and (2.5) multiplied by the appropriate constants and reduce to the forms in (2.8) and (2.7), multiplied by appropriate constants. In other words, (2.1) and (2.2), multiplied by appropriate constants, can be looked upon as the density of a ratio and product respectively.

The integrand in (2.2) for $\delta = 1, \rho = 1, \gamma = -\frac{3}{2}$ and normalized is the inverse Gaussian density available in stochastic processes. The integral in (2.2) for $\delta = 1, \rho = \frac{1}{2}$ is the basic reaction-rate probability integral, which will be considered later. Mathai (2012) has introduced a Krätzel density associated with (2.1) and (2.2) and it is shown that one has general Bayesian structures in (2.1) and (2.2). For example, let us consider a conditional density of y, given x, in the form

$$h_1(y|x) = \hat{c}_1\, y^{\alpha} \mathrm{e}^{-a(\frac{y}{x})^{\rho}}, y > 0, x > 0, a > 0 \tag{2.9}$$

and $\hat{c}_1$ can act as the normalizing constant. In other words, the conditional density is a generalized gamma density. Let the marginal density of x be given by $h_2(x) = \hat{c}_2 x^{\beta} \mathrm{e}^{-a_1 x^{\delta}}, a > 0, x > 0, \delta > 0$, and $\hat{c}_2$ can act as a normalizing constant, a generalized gamma density. Then the joint density of y and x is given by

$$h_1(y|x) h_2(x) = \hat{c}_1 \hat{c}_2 y^{\alpha} x^{\beta} \mathrm{e}^{-a_1 x^{\delta} - a(\frac{y}{x})^{\rho}}. \tag{2.10}$$

Then the unconditional density of y, $f_y(y)$, is available by integrating out x from this joint density. That is,

$$f_y(y) = \hat{c}_1 \hat{c}_2 y^{\alpha} \int_0^\infty x^{\beta} \mathrm{e}^{-a_1 x^{\delta} - a \frac{y^{\delta}}{x^{\delta}}} \mathrm{d}x. \tag{2.11}$$

Now, compare (2.2) and (2.11). They are one and the same forms. Hence (2.2) can be considered as an unconditional density in a Bayesian structure.

For $\delta = 1, \rho = 1$ in (2.1) and (2.2) one can extend the integrals to the real and complex matrix-variate cases. Mathai has also looked into this problem of Krätzel integrals in the matrix-variate cases. There will be difficulty with the Jacobians if we consider general parameters δ and ρ in the matrix-variate case. The type of difficulties that can arise is described in Mathai (1997) by considering the transformation $Y = X^2$ when $X = X'$. In the real matrix-variate case the scalar quantity x^γ is replaced by the determinant $|X|^\gamma$ and exponent e^{-ax} is replaced by $\mathrm{e}^{-a\mathrm{tr}(X)}$ for $X > O$, where $a > 0$ is a real scalar, or $\mathrm{e}^{-\mathrm{tr}(AX)}$ if a is also replaced by a positive definite constant matrix $A > O$. Mathai has also extended Baysian structures, densities of product and ratio, inverse Gaussian density, Krätzel integral and Krätzel density, to matrix-variate cases. When the matrix is in the complex domain $|X|$ is replaced by $|\det(\tilde{X})|$ = absolute value of the determinant of $\tilde{X}$, where $\tilde{X}$ is a matrix in the complex domain.

3. Pathway Model

In a physical system the stable solution may be exponential or power function or Gaussian. This may be the idealized situation. But in reality the solution may be somewhere nearby the ideal or the stable situation. In order to capture the ideal situation as well as the neighboring unstable situations, a model with a switching mechanism was introduced by Mathai (2005, *Linear Algebra and its Applications*). A form of this was proposed in the 1970's by Mathai in connection with population studies. This was a real scalar variable case. Then the ideas were extended to matrix-variate cases and brought out in 2005. For the real scalar positive variable situation, the model is the following

$$p_1(x) = \tilde{c}_1 x^\gamma [1 - a(1-q)x^\delta]^{\frac{1}{1-q}}, a > 0, \delta > 0, q < 1, x > 0. \tag{3.1}$$

If (3.1) is to be used as a statistical density then $\tilde{c}_1$ is the normalizing constant there. Otherwise $\tilde{c}_1$ is a constant, may be $\tilde{c}_1 = 1$ and then (3.1) will be a mathematical model. For $q > 1$ we can write $1 - q = -(q-1)$ and then (3.1) becomes

$$p_2(x) = \tilde{c}_2 x^\gamma [1 + a(q-1)x^\delta]^{-\frac{1}{q-1}}, a > 0, x > 0, q > 1, \delta > 0. \tag{3.2}$$

When $q \to 1$ then $p_1(x)$ and $p_2(x)$ go to

$$p_3(x) = \tilde{c}_3 x^\gamma \mathrm{e}^{-ax^\delta}, a > 0, x > 0, \delta > 0. \tag{3.3}$$

Note that $p_1(x)$ in (3.1) is in the family of generalized type-1 beta family of functions, whereas $p_2(x)$ is in the family of generalized type-2 beta family of functions and $p_3(x)$ belongs to the generalized gamma family of functions. Thus, when the pathway parameter q, goes from $-\infty$ to 1 we have one family of functions, when q is from 1 to ∞ we have another family of functions and when $q \to 1$ we have a third family of functions. Thus, all the three cases are contained in (3.1), which is the pathway model for the real positive scalar variable case. Replace x by $|x|$, $-\infty < x < \infty$, to extend the families over the real line. In (3.2) and (3.3), δ can be either $\delta > 0$ or $\delta < 0$.

When $p_1(x), p_2(x), p_3(x)$ are statistical densities then (3.1) to (3.3) give a distributional pathway to go to three different families of functions. Mathai has also established a parallel pathway in terms of entropy optimization and in terms of differential equations. These give entropic and differential pathways as well. For example, consider the optimization of Mathai's entropy, namely

$$M_\alpha(f) = \frac{\int_x [f(x)]^{2-\alpha}\mathrm{d}x - 1}{\alpha - 1}, \alpha \neq 1, \alpha < 2 \tag{3.4}$$

where $f(x)$ is a density function of x, and x can be real scalar or vector or matrix variable. A density means that $f(x) \geq 0$ for all x and $\int_x f(x)\mathrm{d}x = 1$. If we take the limit when $\alpha \to 1$ then (3.4), for real scalar x, reduces to

$$M_\alpha(f) \to -\int_x f(x) \ln f(x)\mathrm{d}x = S(f) \tag{3.5}$$

where $S(f)$ is Shannon's entropy or measure of "uncertainty" or the complement of "information". In (3.5), $\int_x f(x) \ln f(x)\mathrm{d}x$ is taken as zero when $f(x) = 0$. Consider the optimization of (3.4) subject to the conditions (a): $\int_x x^{\gamma(1-q)+\delta} f(x)\mathrm{d}x =$ fixed and (b): $\int_x x^{\gamma(1-q)} f(x)\mathrm{d}x =$ fixed. For $\gamma = 0$, condition (b) becomes $\int_x f(x)\mathrm{d}x =$ 1 since the total probability is 1. For $\gamma = 1, q = 0$, (b) means that the first moment is fixed. This can correspond to the physical law of conservation of energy when dealing with energy distribution. If we use Calculus of Variation to optimize (3.4) then the Euler equation there is

$$\frac{\partial}{\partial f}[f^{2-\alpha} - \lambda_1 x^{\gamma(1-q)} f + \lambda_2 x^{\gamma(1-q)+\delta} f] = 0 \tag{3.6}$$

where λ_1 and λ_2 are Lagrangian multipliers. Note that (3.6) gives the structure

$$f^{1-\alpha} = \mu_1 x^{\gamma(1-q)}[1 - \mu_2 x^\delta]$$

for some μ_1 and μ_2, which means

$$f = \gamma_1 x^\gamma [1 - \gamma_2 x^\delta]^{\frac{1}{1-q}} \tag{3.7}$$

for some γ_1 and γ_2. For $\gamma_2 = a(1-q)$ and $\gamma_1 = \tilde{c}_1$ we have the model in (3.1). Thus, for $q < 1, q > 1, q \to 1$ one has an entropic pathway. Similarly we can consider the corresponding differential equations to obtain a differential pathway. The phrases, distributional, entropic and differential pathways, were coined by Hans J. Haubold, co-author of Mathai in various topics in physics, fractional differential equations etc.

The original paper Mathai (2005) deals with rectangular matrix-variate case. Let $X = (x_{ij})$ be $m \times n, m \leq n$ and of rank m be a matrix of distinct real scalar variables x_{ij}'s. Let A be $m \times m$ and B be $n \times n$ constant positive definite matrices. Consider the function

$$P_1(X) = C_1|AXBX'|^\gamma|I - a(1-q)AXBX'|^{\frac{1}{1-q}}, q < 1, a > 0 \tag{3.8}$$

where $a > 0, q < 1$ are scalars, I is a $m \times m$ identity matrix and C_1 is a constant. If (3.8) is to be taken as a density then C_1 is the normalizing constant there and $I - a(1-q)AXBX' > O$. For $q > 1, P_1(X)$ goes to

$$P_2(X) = C_2|AXBX'|^\gamma[I + a(q-1)AXBX'|^{-\frac{1}{q-1}},\ q > 1,\ a > 0 \tag{3.9}$$

and when $q \to 1$, $P_1(X)$ and $P_2(X)$ go to

$$P_3(X) = C_3|AXBX'|^{\gamma}\mathrm{e}^{-a\mathrm{tr}(AXBX')}, \quad a > 0. \tag{3.10}$$

If a location parameter matrix is to be introduced then replace X by $X - M$ where M is a $m \times n$ constant matrix.

Note that the structure $|AXBX'|$ is the structure of the volume content of a parallelotope in n-space. Let us look at the m rows of X. These are $1 \times n$ vectors. These can be taken as m points in n-dimensional Euclidean space. These m vectors, $m \leq n$, are linearly independent when the rank of X is m. These taken in a given order can form a convex hull and a m-parallelotope. The volume of this m-parallelotope is the determinant $|XX'|^{\frac{1}{2}}$. Hence $|AXBX'|^{\frac{1}{2}}$ is the volume content of a generalized m-parallelotope.

Also $AXBX'$ is a generalized quadratic form. For $A = I_m$ and $m = 1$ it is a quadratic form in the $1 \times n$ vector variable. Thus the theory of quadratic form and generalized quadratic form can be extended to a wider class represented by the pathway model (3.8). The current theory of quadratic form and bilinear form in random variables is confined to samples coming from a Gaussian population, see the books Mathai and Provost (1992), Mathai, Provost and Hayakawa (1995). The results on quadratic and bilinear forms can now be extended to the wider class of pathway models. One problem in this direction is discussed in Mathai (2007). The area is still wide open. The matrix-variate pathway model in Mathai (2005) is extended to complex domain in Mathai and Provost (2005, 2006). Some works in the scalar complex variable case, associated with normal or Gaussian population, are available in the literature with lots of applications in sonar, radar, communication and engineering problems. These results are not included in Mathai and Provost (1992) and Mathai, Provost and Hayakawa (1995). This is a deficiency there. These results can be extended to the pathway family, which may produce many useful results in communication, engineering and related areas. These are open problems.

Note that (3.8) for $a = 1, q = 0$ is a matrix-variate type-1 beta density or $AXBX'$ is a type-1 beta matrix. This is the exact form of the matrix appearing in the generalized analysis of variance and design of experiments areas, in the likelihood ratio test involving one or more multivariate normal or Gaussian populations etc, a summary of the contributions of Mathai and his co-workers is available from Mathai and Saxena (1973). The theory available there is based on Gaussian populations. Now, generalized analysis of variance can be examined in a wider pathway family so that the limiting form corresponding to (3.10), will be the Gaussian case. This is wide area, not explored yet. Instead of Gaussian populations, now the likelihood ratio tests in a wider pathway family, corresponding to (3.8) and (3.9), can be examined. This also is a wide area, not explored yet.

While exploring a reliability problem, Mathai (2003) came across a multivariate family of densities, which could be taken as a generalization of type-1 Dirichlet family of densities. Then Mathai and his co-workers introduced several generalizations of type-1 and type-2 Dirichlet densities. Many papers are written in this area, see for example Thomas and Mathai (2009). For the different generalizations of type-1 and type-2 Dirichlet family, a number of characterization results are established showing that these models could also be generated by products of statistically independently type-1 beta distributed real scalar random variables. This is exactly the same structure available in the likelihood ratio criteria in the null cases of testing hypotheses on the parameters of one or more Gaussian populations as well as in the determinant $|AXBX'|$ or in the model (3.8) for $a = 1, q = 0$. Thus, it is already shown that these three areas are connected. This is not yet explored in detail. All the problems here can be set in a general pathway family of functions. These are all open problems.

In (3.1) if we put $\gamma = 0, \delta = 1$ then we get Tsallis statistics in non-extensive statistical mechanics. This is a very popular area. It is said that between 1990 and 2010 more than 5000 papers are published in this area, mostly applications in various problems of physics. Then the pathway model is directly applicable in these problems, which may even produce some significant new theories in physical sciences. Also, (3.2) for $\delta = 1$ as well as for some general $\delta > 0$ is superstatistics. This is also a very hot area recently. Dozens of articles are published in this area also. (3.2) and its limiting form (3.3) are covered in superstatistics but (3.1) is not covered because superstatistics considerations deal with a conditional density of generalized gamma form as well as the marginal density a generalized gamma form then the unconditional density, which is superstatistics in statistical terms from a Bayesian point of view, can only produce a type-2 beta form, namely (3.2) form and not (3.1) form. Thus, superstatistics is also a special case of the pathway model in the real scalar positive variable case.

Mathai's students and others have created a pathway fractional integral operator, a pathway transform or P-transform, pathway Weibull or q-Weibull distribution, q-logistic distribution, q-stochastic process etc, which produced significant results and better-fitting models for many types of real-life data.

In the pathway idea itself there is an open area which is not yet explored. The scalar version of the pathway model in (3.1) to (3.3) can be looked upon as the behavior of a hypergeometric series ${}_1F_0$ (binomial series) going to ${}_0F_0$ (exponential series). That is,

$$ {}_1F_0(-\frac{1}{1-q};\ ;a(1-q)x^{\delta}) = [1-a(1-q)x^{\delta}]^{\frac{1}{1-q}}. \tag{3.11} $$

$$ \lim_{q\to 1_-} {}_1F_0(-\frac{1}{1-q};\ ;a(1-q)x^{\delta}) = \mathrm{e}^{-ax^{\delta}} = {}_0F_0(\ ;\ ;-ax^{\delta}). \tag{3.12} $$

From the point of view of a hypergeometric series, the process (3.11) to (3.12) is the process of a binomial series going to an exponential series. But a Bessel series ${}_0F_1$ can also be sent to an exponential series. For example, consider the Bessel series

$$\lim_{q\to 1_-} {}_0F_1(\ ;\frac{1}{1-q};-\frac{a}{1-q}x^\delta) = \mathrm{e}^{-ax^\delta} = {}_0F_0(\ ;\ ;-ax^\delta). \tag{3.13}$$

Therefore, a generalized form, covering the path towards the exponential form e^{-ax^δ}, is the Bessel form ${}_0F_1(\ ;\frac{1}{1-q};-\frac{a}{1-q}x^\delta)$. Many practical situations are connected to various forms of Bessel functions, the path given by (3.13) can yield rich results. Bessel function with matrix argument is also defined. Hence this area is still there, to be fully explored, as open problems.

4. Mathai's Work on Functions of Matrix Argument

There are not many people working in this area around the world. A multivariate function usually means a function of many scalar variables. This is different from a matrix-variate function of a function of matrix argument. Functions of matrix argument are real-valued scalar functions $f(X)$, where X is a square or rectangular matrix. For example, for a $p\times p$ matrix X, $|X|$ = determinant of X, $\mathrm{tr}(X)$ = trace of X are real-valued scalar functions when X is real. Even for a square $p\times p$ matrix X, the square root cannot be uniquely determined unless further conditions are imposed on X. If we use the definition $A = BB = B^2$ then $B = A^{\frac{1}{2}}$ the square root of A we can have many candidates for B. For example, for a simple matrix like a 2×2 identity matrix $A = I_2$, $B_1, B_2.B_3, ...$ are square roots

$$A = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, B_1 = \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}, B_2 = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, B_3 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

If we restrict A and $A^{\frac{1}{2}}$ to be positive definite matrices then B_3 is the only candidate here. Hence, if X is $p\times p$ real positive definite or Hermitian positive definite then $X^{\frac{1}{2}}$ can be uniquely defined. Therefore, functions of matrix argument are developed mainly when the argument matrix is either real positive definite or Hermitian positive definite. There are three approaches available in the literature for functions of matrix argument, that is, real-valued scalar functions $f(X)$ of matrix argument X. For convenience, all the matrices appearing in this section are $p\times p$ positive definite denoted by $X > O$, real or Hermitian, unless stated otherwise. One definition is through Laplace and inverse Laplace transforms. This development is due to Herz (1955) and others. Here the basic assumption of functional commutativity is used, that is, $f(AB) = f(BA)$ even if $AB \neq BA$. For example, determinant and trace will satisfy this property. When X is real symmetric then there exists an orthonormal matrix Q such that $QQ' = I, Q'Q = I, Q'XQ = D = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ where $\lambda_1, ..., \lambda_p$ are the eigenvalues of X. Then

$$f(X) = f(XI) = f(XQQ') = f(Q'XQ) = f(D) \tag{4.1}$$

or $f(X)$, which is a function of $p(p+1)/2$ real variables x_{ij}'s, when $X = X'$ and real, has become a function of D which is of p real variables $\lambda_1, ..., \lambda_p$, under this assumption of functional commutativity. If $X = (x_{ij}) = X', p \times p$ and $T = (t_{ij}) = T', p \times p$ then

$$\mathrm{tr}(XT) = \sum_{j=1}^{p} x_{jj}t_{jj} + 2\sum_{i<j} x_{ij}t_{ij}. \tag{4.2}$$

Therefore

$$\int_{X>O} \mathrm{e}^{-\mathrm{tr}(TX)} f(X)\mathrm{d}X \neq L_f(T) \tag{4.3}$$

the Laplace transform of $f(X)$ because (4.3) is not consistent with the definition of multivariate Laplace transform. In (4.2) the non-diagonal terms appear twice. In the multivariate Laplace transform, the variables and the corresponding parameters must appear only once each. If we consider a modified parameter matrix $T^* = (t^*_{ij}), t^*_{ij} = t^*_{ji}$ for all i and j, and

$$t^*_{ij} = \begin{cases} t_{ii}, i = j \\ \frac{1}{2}t_{ij}, i \neq j \end{cases}, T = (t_{ij}) = T', \tag{4.4}$$

then

$$\int_{X>O} \mathrm{e}^{-\mathrm{tr}(T^*X)} f(X)\mathrm{d}X = L_f(T^*) \tag{4.5}$$

is the Laplace transform in the real symmetric positive definite matrix-variate case, where T^* is the parameter matrix and $\mathrm{d}X$ stands for the wedge product of the $p(p+1)/2$ differentials $\mathrm{d}x_{ij}$'s or

$$\mathrm{d}X = \prod_{i \geq j} \wedge \mathrm{d}x_{ij}. \tag{4.6}$$

Under this approach, a hypergeometric function of matrix argument, denoted by

$${}_rF_s(a_1, ..., a_r; b_1, ..., b_s; X),$$

where $a_1, ..., a_r$ and $b_1, ..., b_s$ are scalar parameters and X is a $p \times p$ real positive definite matrix, is defined by a Laplace and inverse Laplace pair. Under this definition, explicit forms are available only for ${}_0F_0$ and ${}_1F_0$. Details of the definition and properties may be seen from Herz (1955) and from the book Mathai (1997).

The second approach is through zonal polynomials, developed by James (1961), Constantine (1963) and others. Here also functional commutativity is implicitly assumed, though not stated explicitly. Under this definition, a hypergeometric series is defined as follows

$${}_rF_s(a_1, ..., a_r; b_1, ..., b_s; X) = \sum_{k=0}^{\infty} \sum_{K} \frac{(a_1)_K ... (a_r)_K}{(b_1)_K ... (b_s)_K} \frac{C_K(X)}{k!} \tag{4.7}$$

where $C_K(X)$ are zonal polynomials of order $k, K = (k_1, ..., k_p), k_1 + k_2 + ... + k_p = k$ and

$$(a)_K=\prod_{j=1}^{p}(a-\frac{j-1}{2})_{k_j} \text{ and } (b)_{k_j}=b(b+1)...(b+k_j-1),(b)_0=0,b\neq 0. \quad (4.8)$$

Here $(b)_{k_j}$ is the Pochhammer symbol and $(a)_K$ is the generalized Pochhammer symbol. All terms of the series in (4.7) are explicitly available but since zonal polynomials are complicated to compute, only the first few terms up to $k = 11$ are computed. Details of zonal polynomials may be found, for example from the book Mathai, Provost and Hayakawa (1995). The definition through (4.5) and its inverse Laplace form and the definition through (4.7) are not very powerful in extending results in the univariate case to the corresponding matrix-variate case. When (4.7) is used to extend univariate results to matrix-variate cases the following two basic results will be essential. These will be stated here as lemmas without proofs.

Lemma 4.1. We have

$$\int_{X>O}\mathrm{e}^{-\mathrm{tr}(ZX)}|X|\alpha-\frac{p+1}{2}C_K(XT)\mathrm{d}X=|Z|^{-\alpha}C_K(TZ^{-1})\Gamma_p(\alpha,K), \quad (4.9)$$

where

$$\Gamma_p(\alpha,K)=\pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma(\alpha+k_j-\frac{j-1}{2})=\Gamma_p(\alpha)(\alpha)_K \quad (4.10)$$

with $(\alpha)_K$ *defined as in (4.8).*

Lemma 4.2. We have

$$\int_O^I|X|^{\alpha-\frac{p+1}{2}}|I-X|^{\beta-\frac{p+1}{2}}C_K(TX)\mathrm{d}X=\frac{\Gamma_p(\alpha,K)\Gamma_p(\beta)}{\Gamma_p(\alpha+\beta,K)}C_K(T). \quad (4.11)$$

Starting from 1970, Mathai developed functions of matrix argument through M-transforms and M-convolutions. Under M-transform definition, a hypergeometric function ${}_rF_s$ with $p\times p$ matrix argument $X>O$ is defined as that class of functions $f(X)$ satisfying functional commutativity and the integral equation

$$\int_{X>O}|X|^{\rho-\frac{p+1}{2}}f(-X)\mathrm{d}X=C\Gamma_p(\rho)\frac{\prod_{j=1}^{r}\Gamma_p(a_j-\rho)}{\prod_{j=1}^{s}\Gamma_p(b_j-\rho)},\Re(\alpha)>\frac{p-1}{2} \quad (4.12)$$

where

$$C=\frac{\prod_{j=1}^{s}\Gamma_p(b_j)}{\prod_{j=1}^{r}\Gamma_p(a_j)}.$$

For example

$$\int_{X>O}|X|^{\rho-\frac{p+1}{2}}f(-X)\mathrm{d}X=\Gamma_p(\rho),\Re(\rho)>\frac{p-1}{2}\Rightarrow f(-X)=\mathrm{e}^{-\mathrm{tr}(X)}. \quad (4.13)$$

Since the left side in (4.12) is a function of only one parameter ρ, we cannot normally recover $f(X)=f(D)$ a function of p scalar variables. It is conjectured that when $f(X)$ is analytic in the cone of positive definite matrices X, one has $f(X)$ uniquely recovered from the right side of (4.12). This is not established yet and also an explicit form of an inverse or $f(X)$ through the right-side of (4.12) is not available yet. But (4.12) is the most convenient form to extend univariate results on hypergeometric functions to the corresponding class of matrix-variate cases. In general, when we go from a univariate case, such as a univariate function e^{-x}, to a multivariate case, there is nothing called a unique multivariate analogue. Whatever be the properties of the univariate function that we wish to preserve in the multivariate analogue, we may be able to come up with different functions as multivariate analogues of a univariate function. Hence, a class of multivariate analogues is more appropriate than a single multivariate analogue. Properties of M-transforms and properties of hypergeometric family coming from (4.12) are available in the book Mathai (1997). When Mathai introduced M-convolutions and M-transforms, details in Mathai (1997), no physical meaning could be found. Now, a physical interpretation is available for M-convolutions as densities of products and ratios of matrix random variables, as illustrated in Sections 1.17.

5. Mathai's Work in Geometric Probabilities

The work until 1999 is summarized in the first full textbook in this area Mathai (1999) and later works are available in published papers. The work started as an off-shoot of the work in multivariate statistical analysis. Mathai noted that the moment structure for many types of random geometric configurations was that of product of independently distributed type-1 beta, type-2 beta or gamma random variables. Such structures were already handled by Mathai and his co-workers in connection with problems in multivariate analysis. Earlier contributions of Mathai in this area are available from Chapter 4 of the book Mathai (1999). Then Harold Ruben, a colleague of Mathai at McGill University, one day gave a copy of his paper showing a conjecture in geometrical probabilities, called Miles' conjecture about a re-scaled, relocated random volume, generated by uniformly distributed random points in n-space, as asymptotically normal when $n \to \infty$. The proof was very roundabout. Mathai noted that it could be proved easily with the help of asymptotic expansions of gamma functions. This paper was published in *Annals of Probability*, Mathai (1982). Then Mathai formulated and proved parallel conjectures regarding type-1 beta distributed, type-2 beta distributed points and gamma distributed points and published a series of papers. Then Mathai noted that many European researchers were working on distances between random points, and random areas when the random points are in particular shapes such as triangles, parallelograms, squares, rhombuses etc. As generalizations of all these classes of problems, Mathai generalized Buffon's clean-tile problem, the starting point of geometrical probabilities. He considered placing a ball at random in a pyramid with polygonal base, defining "at random" in terms of kinematic measure, Mathai (1999b). When mixing geometry with probability or measure theory, or in the area of stochastic geometry, the basic axioms of probability are not sufficient, as pointed out by Bertrand's or Russell's paradoxes. We need an additional axiom of invariance under Euclidean motion. Another major contribution of Mathai in this area is Mathai (1999a, *Advances in Applied Probability*) where he has shown that the usual complicated procedures coming from Integral Geometry and Differential Geometry are not necessary for handling certain types of random volumes but only the simple properties of functions of matrix argument and Jacobians of matrix transformation are sufficient. The procedure is illustrated in the distribution of volume content of parallelotope generated by random points in Euclidean n-space. The work on geometrical probabilities is currently progressing in the areas of random sets, image processing etc. The book, Mathai (1999), only deals with distributional aspects of random geometric configurations.

As an application of geometrical probabilities, Mathai and his co-authors looked into a geography problem of city designs of rectangular grid cities as in North America versus circular cities as in Europe, with reference to travel distance, and the associated expense and loss of time, from suburbs to city core, see Mathai (1998), Mathai and Moschopoulos (1999). There are many interesting applications of geometrical probabilities. Nearest neighbor problem, crystal formation, packing and covering problems, antibodies nullifying viruses etc are some of these problems. Some details may be seen from Mathai (1999).

6. Mathai's Work in the Area of Astrophysics

After publishing the books Mathai and Saxena (1973, 1978) many physicists were using various results in special functions in their physics problems. Some of these physicists from Germany were contacting Mathai in Montreal, Canada, to clear some of their doubts in the area of special functions. Then Mathai suggested for the physicists from various parts of Germany to get together, collect all their open problems where help from special functions was needed, and send one person to Montreal, Canada, so that they could sit together and solve all these problems one by one. As a result, Hans J. Haubold came to Montreal, Canada, in 1982 with 12 open problems on reaction-rate theory, solar and stellar models, gravitational instability etc. The idea was to get exact analytical results and analytical models where computations and computer models were available. Mathai figured out that all the problems connected with reaction-rate theory could be solved once the following integral was evaluated explicitly

$$I = \int_0^\infty x^{\gamma-1} \mathrm{e}^{-ax-bx^{-\frac{1}{2}}} \mathrm{d}x, a > 0, b > 0. \tag{6.1}$$

The corresponding general integral is

$$I(\gamma, a, b, \delta, \rho) = \int_0^\infty x^{\gamma-1} \mathrm{e}^{-ax^{\delta}-bx^{-\rho}} \mathrm{d}x, a > 0, b > 0, \delta > 0, \rho > 0. \tag{6.2}$$

In 1982 Mathai could not find any mathematical technique of handling (6.2) or its particular case (6.1). He noted that (6.2) could be written as a product of two integrable functions and thereby as statistical densities by multiplying with appropriate normalizing constants. Then the structure in (6.2) could be converted into the form

$$I(\gamma,a,b,\delta,\rho)=\int_0^\infty \frac{1}{v}f_1(\frac{u}{v})f_2(v)\mathrm{d}v \tag{6.3}$$

and the right side of (6.3) is the density of a product of two real scalar positive independently distributed random variables with densities $f_1(x_1)$ and $f_2(x_2)$ respectively with $u=x_1x_2, v=x_2$. Take

$$f_1(x_1)=c_1\mathrm{e}^{-x_1^\delta}, f_2(x_2)=c_2x_2^\gamma\mathrm{e}^{-ax_2^\rho}, x>0, a>0, \delta>0, \rho>0$$

where c_1,c_2 are normalizing constants. When $u=x_1x_2$ the density of u, denoted by $g(u)$, is given by

$$g(u)=c_1c_2\int_0^\infty \frac{1}{v}v^\gamma\mathrm{e}^{-av^\delta-bv^{-\rho}}\mathrm{d}v, b=u^\rho. \tag{6.4}$$

Now, it is only a matter of evaluating the density $g(u)$ by using some other method. Note that $u=x_1x_2$ where x_1 and x_2 are independently distributed means

$$E(u^{s-1})=E(x_1^{s-1})E(x_2^{s-1})$$

where $E(\cdot)$ denotes the expected value of $(\cdot)$. Then for (6.4)

$$E(x_1^{s-1})=c_1\int_0^\infty x_1^{s-1}\mathrm{e}^{-x_1^\rho}\mathrm{d}x_1=c_1\frac{1}{\rho}\Gamma(\frac{s}{\rho}), \Re(s)>0$$

which also shows that $c_1=\frac{\rho}{\Gamma(\frac{1}{\rho})}$ since $E(x_1^{s-1})$ at $s=1$ is 1. Evaluations of c_1 and c_2 are not necessary for our procedure to hold.

$$E(x_2^{s-1})=c_2\int_0^\infty x_2^{\gamma+s-1}\mathrm{e}^{-ax_2^\delta}\mathrm{d}x_2=\frac{c_2}{\delta}\Gamma(\frac{\gamma+s}{\delta})a^{-(\frac{\gamma+s}{\delta})}, \Re(\gamma+s)>0.$$

Therefore

$$E(u^{s-1})=\frac{c_1c_2}{\delta\rho}a^{-\frac{\gamma}{\delta}}\Gamma(\frac{s}{\rho})\Gamma(\frac{\gamma+s}{\delta})a^{-\frac{s}{\delta}}.$$

Hence $g(u)$ is available from the inverse Mellin transform. That is,

$$g(u)=\frac{c_1c_2}{\delta\rho}a^{-\frac{\gamma}{\delta}}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Gamma(\frac{s}{\rho})\Gamma(\frac{\gamma+s}{\delta})(a^{\frac{1}{\delta}}u)^{-s}\mathrm{d}s, c>0, i=\sqrt{-1}. \tag{6.5}$$

Comparing (6.4) with (6.5) the required integral is given by the following

$$\begin{aligned} I(\gamma,a,b,\delta,\rho)&=\int_0^\infty \frac{1}{v}v^\gamma\mathrm{e}^{-av^\delta-bv^{-\rho}}\mathrm{d}v \\ &=\frac{1}{\delta\rho a^{\frac{\gamma}{\delta}}}H_{0,2}^{2,0}\left[a^{\frac{1}{\delta}}b^{\frac{1}{\rho}}\big|_{(0,\frac{1}{\rho}),(\frac{\gamma}{\delta},\frac{1}{\delta})}\right]. \end{aligned} \tag{6.6}$$

The right side of (6.6) is a H-function. Extensive work in G and H-functions were already done by Mathai and his co-workers earlier.

For the reaction-rate probability integral, $\delta=1$ and $\rho=\frac{1}{2}$. In this case, the H-function in (6.6) reduces to a G-function and explicit computable series forms are also given by Mathai and his co-workers. A series of papers were published in all leading physics journals and some mathematics journals also by Haubold and Mathai. Haubold supplied the physics part and Mathai supplied the mathematics part. Problems considered were resonant reactions, non-resonant reactions, depleted case, high energy tail cut off etc. A summary of the work until 1988 is available in the research monograph Mathai and Haubold (1988). After publishing dozens of papers in physics by using statistical techniques it was realized that the density of a product of independently distributed real positive random variables was nothing but the Mellin convolution of a product. Hence, one could have applied Mellin and inverse Mellin transform techniques there. Mathai and his co-workers are credited with popularizing Mellin and inverse Mellin transform technique in Statistics, Physics and engineering problems. The work in this area of reaction-rate also resulted in two encyclopedia articles, see Haubold and Mathai (1997 on Sun and 1998 on Universe).

6.1. Analytic solar models

Another attempt was to replace the current computer model for the Sun with analytic models. The idea was to assume a basic model for the matter density distribution in the Sun or in main sequence stars which could be treated as a sphere in hydrostatic equilibrium. Let r be an arbitrary distance from the center of the Sun and let $R_\odot$ be the radius of the Sun. Let $x = \frac{r}{R_\odot}$ so that $0 \le x \le 1$. The model for the matter density distribution is taken as

$$f(x) = c(1 - x^\delta)^\gamma \tag{6.7}$$

where c is the central core density when $x = 0$. The parameters δ and γ are selected to agree with observational data. Then by using (6.7), expressions for mass, pressure,temperature and luminosity are computed by using physical laws. Then by using known observations, or comparing with known data on mass, pressure etc the best values for δ and γ are estimated so that close agreement is there with observational values of mass, pressure etc. Some of the results until 1988 are given in the monograph Mathai and Haubold (1988). Later works are available in papers only.

Another area that was looked into was the gravitational instability problem of mixing of cosmic dusts. Russian physicists were working on the problem of mixing two types of cosmic dusts. The problem was brought to the attention of Mathai by his co-worker Haubld. The problem was formulated in the form of complicated differential equations. Mathai tried to change the operator $D = \frac{\mathrm{d}}{\mathrm{d}x}$ to $x\frac{\mathrm{d}}{\mathrm{d}x}$. Then the differential equation got simplified a bit. Then he changed the dependent variable a little bit and found that the differential equation became a particular case of G-function differential equation. Mathai and his co-workers had done a lot of work on G-function. Thus, Mathai could simplify the differential equation and then consider the case of k different dusts mixing. This resulted in the first paper of Mathai in integer order differential equations and it was published in the prestigious MIT journal (Mathai(1989)). Then the results were applied to gravitational instability problem and wrote papers in physics, see for example Haubold and Mathai (1988).

Another area looked into was solar neutrino problem. Haubold and Mathai tried to come up with appropriate models to model the solar neutrino data. Mathai had noted that the graph of the time series data looked similar to the pattern that he had seen when working on modeling of the chemical called Melatonin in human body. Usually what is observed is the residual part of what is produced minus what is consumed or converted or lost. Hence the basic model should be an input-output type model. The necessary theory is available in Mathai (1993b). Mathai had successfully tackled input-output models to study dam capacity, effect of acid rain, grain storage etc. The simplest input-output model is an exponential type input x_1 and an exponential type output x_2 so that the residual part $u = x_1 - x_2$. When x_1 and x_2 are independently and identically exponentially distributed then u has a Laplace distribution. A direct application of Laplace model is in modeling sand dunes created by opposing forces of wind. One model Haubold and Mathai tried was Laplace type random variables over time so that the graph will look like blips at equal or random points on a horizontal line. If the time-lag is shortened then the blips will start joining together. If exponential models of different intensities, that is, in the input-output model $f(x_j) = \theta e^{-\theta x_j}, \theta > 0, x_j > 0.j = 1, 2$, if θ is different for different blips then the pattern can be brought to the pattern seen in nature or the pattern seen from the data. One of the latest models of Haubold and Mathai in 2015 is of this type. This may be close to what is seen in nature. Haubold is the main persuasive force behind Mathai's adventure into astrophysics area.

7. Mathai's Work on Special Functions

Mathai and his co-workers are credited with popularizing special functions, especially G and H-functions, in Statistics and Physics. Major part of the special function work was done with co-worker Saxena and most of the papers were published as Mathai and Saxena papers. They thought that they were the first one to use G and H-functions in statistical literature. But D.G. Kabe pointed out that he had expressed a statistical density in terms of a G-function in 1958. This may be the first paper in Statistics where G or H-function was used. Most probably the use of G and H-function in Physics an engineering areas started after the publication of our books Mathai and Saxena (1973, 1978). We started using G and H-functions in Statistics from 1966 onward. The first work on the fusion of statistical distribution theory and special functions started by creating statistical densities by using generalized special functions. In this connection the most general such density is based on a product of two H-functions, which appeared in the *Annals of Mathematical Statistics* (Mathai and Saxena (1969)). Another area that was looked into was Bayesian structures. The unconditional density in Bayesian analysis is of the form

$$f(x) = \int_y f_1(x|y)g(y)\mathrm{d}y. \tag{7.1}$$

What are the general families of functions for $f_1(x|y)$ and $g(y)$ so that the integral in (7.1) can be evaluated? We have constructed some general mixing families of $f_1(x|y)$ and $g(y)$.

Another family of problems that was looked into were the null and non-null distributions of the likelihood ratio criterion or λ-criterion for testing hypotheses on the parameters of one or more multinormal populations. Consider the $p \times 1$ vector X_j having the density

$$f(X_j) = \frac{1}{(2\pi)^{\frac{p}{2}}|V|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(X_j-\mu)'V^{-1}(X_j-\mu)}, V > O \tag{7.2}$$

where μ is a constant vector, known as the mean-value vector here. For $j = 1, 2, ..., N$ if X_j's are independently distributed with the same density in (7.2) then we say that we have a simple random sample of size N from the p-variate normal or Gaussian population (7.2). Suppose that we want to test a hypothesis $H_o : V =$ is diagonal. This is called the test for independence in the Gaussian case. Then the λ-criterion can be shown to have the structure

$$u = \lambda^{\frac{1}{N}} = \frac{|S_1|}{|S_1 + S_2|} = u_1 u_2 ... u_p \tag{7.3}$$

where S_1 and S_2 are independently distributed matrix-variate gamma variables of (1.25) with the same B. Then the structure in (7.3) is distributed as a product of independently distributed type-1 beta random variables, $u_1, ..., u_p$. Then the density of u can be written as a G-function of the type $G^{p,0}_{p,p}(u)$. The density of λ will go in terms of a H-function. The H-function is more or less the most generalized special function in real scalar variable case and it is defined by the following Mellin-Barnes integral and the following standard notation is used

$$H^{m,n}_{r,s}(z) = H^{m,n}_{r,s}\left[z\Big|^{(a_1,\alpha_1),...,(a_r,\alpha_r)}_{((b_1,\beta_1),...,(b_s,\beta_s)}\right] \tag{7.4}$$

$$= \frac{1}{2\pi i}\int_L \phi(\rho)z^{-\rho}\mathrm{d}\rho, \sqrt{-1} \text{ where} \tag{7.5}$$

$$\phi(\rho) = \frac{\{\prod_{j=1}^m \Gamma(b_j + \beta_j\rho)\}\{\prod_{j=1}^n \Gamma(1 - a_j - \alpha_j\rho)\}}{\{\prod_{j=m+1}^s \Gamma(1 - b_j - \beta_j\rho)\}\{\prod_{j=n+1}^r \Gamma(a_j + \alpha_j\rho)\}} \tag{7.6}$$

where $\alpha_j, j = 1, ..., r; \beta_j, j = 1, ..., s$ are real and positive numbers, a_j's and b_j's are complex numbers. L is a contour separating the poles of $\Gamma(b_j + \beta_j\rho), j = 1, ..., m$ to one side and those of $\Gamma(1 - a_j - \alpha_j\rho), j = 1, ..., n$ to the other side. Existence of the contours and convergence conditions are available from the books Mathai (1993 [309 citations as of August 2015]), Mathai and Saxena (1973, 1978[793 citations as of August 2015]), Mathai and Haubold (2008 [218 citations as of August 2015]), Mathai, Saxena and Haubold (2010 [353 citations as of August 2015]). When $\alpha_1 = 1 = ... = \alpha_r, \beta_1 = 1 = ... = \beta_s$ then the H-function reduces to a G-function denoted as

$$G_{r,s}^{m,n}(z) = G_{r,s}^{m,n}\left[z\Big|_{(b_1,...,b_s}^{a_1,...,a_r}\right]. \tag{7.7}$$

Explicit computable series forms of $G_{0,p}^{p,0}(x), G_{p,p}^{p,0}(x), G_{p,p}^{p,p}(x)$ and for the general $G_{r,s}^{m,n}(x)$, were given by Mathai in a series of papers. The first three forms correspond to product of independently distributed gamma variables, type-1 beta variables and type-2 beta variables respectively. These forms appear in a large number of situations in various fields. The details of the computable representations are available in the book Mathai (1993). This is achieved by developing an operator which can handle poles of all orders. This operator may be seen from Mathai (1971) and its use from Mathai (1993). This is a modification of a procedure developed in Mathai and Rathie (1971) in the *Annals of Mathematical Statistics* paper to handle generalized partial fractions. Let

$$\frac{1}{(x-a_1)^{m_1}(x-a_2)^{m_2}...(x-a_k)^{m_k}} = \sum_{i=1}^{k}\sum_{j=1}^{m_i}\frac{c_{ij}}{(x-a_i)^j} \tag{7.8}$$

where $m_i = 1, 2, ...$ for $i = 1, 2, ..., k$ and the coefficients c_{ij}s are to be evaluated. The technique developed in Mathai and Rathie (1971) enables one to compute c_{ij}'s easily and explicitly.

The G and H-functions are also established by Mathai for the real matrix-variate cases through M-transforms, along with extensions of all special functions of scalar variables to the matrix-variate cases. Also Mathai extended multivariate functions such as Apple functions, Lauricella functions, Kampé de Fériet functions etc to many matrix-variate cases. Some details may be seen from Mathai (1993,1997) and further details from later papers.

By making use of the explicit series forms, MAPLE and MATHEMATICA have produced computer programs for numerical computations of G-functions and MATHEMATICA has a computer program for the evaluation of H-function also. This is why physicists and engineers started using H-function very frequently now. Solutions of fractional differential equations usually end up in terms of Mittag-Leffler function, its generalization as Wright's function and its generalization as H-function. In connection with fractional differential equations for reaction, diffusion, reaction-diffusion problems Haubold, Mathai and Saxena have given solutions for a large number of situations, which may be seen from the joint works of Haubold and Mathai, Haubold, Mathai and Saxena starting from 2000 and the work is still continuing. In all these solutions, either Mittag-Leffler function or Wright's function or H-function appears. Also many other physicists, mathematicians and engineers have tried other fractional partial differential equations where also the solutions are available in terms of H-functions. Mathai and his co-workers have contributed to all aspects of special functions, apart from the five books mentioned above. Also Mathai has extended various results in the scalar variable case to the corresponding matrix-variate case in this area of special functions;

As of December 2015, Google scholar citation counts, that is the counts of others citing the works or making use of the works, gives Mathai **second rank** of most cited in the literature for special functions, **second rank** for applied analysis, **first rank** for statistical distributions, **first rank** for geometrical probabilities and **first rank** for multivariate analysis, globally. All citations = 6101, h-index = 30, i-10 index = 90, Research Gate score = 37.

8. Mathai's Works in Multivariate Statistical Analysis and Statistical Distribution Theory

Mathai has made a large number of very significant contributions in these areas. In the area of multivariate analysis, almost all exact null distributions in the most general cases and a large number of non-null distributions of λ-criteria for testing hypotheses on one or more multivariate Gaussian populations, exponential populations etc are given by Mathai or Mathai with his co-workers Rathie or Saxena or Katiyar or Tan or Tracy or Moschopoulos or Provost. The λ-criterion is explained in (7.3). Null distributions mean the distributions when the null hypotheses are assumed to hold and non-null distributions mean without the restrictions imposed by the hypotheses. In the non-null situations some of the cases are still open problems. In the null cases, u, a one-to-one function of the λ-criterion, has usually the following representations

$$u = u_1 ... u_p; \tag{8.1}$$

$$u = v_1 ... v_p; \tag{8.2}$$

$$u = w_1 ... w_p \tag{8.3}$$

where $u_1, ..., u_p$ are independently distributed real scalar type-1 beta random variables, $v_1, ..., v_p$ are the same type of type-2 beta random variables and $w_1, ..., w_p$ are same type of gamma random variables. The density of u in (8.1) can be written in terms of a $G^{p,o}_{p,p}(u)$, that of (8.2) as a $G^{p,p}_{p,p}(u)$ and that of (8.3) as a $G^{p,0}_{0,p}(u)$. Computational aspects of these forms are already discussed in Section 7 above. In geometrical probabilities also the squares of the volume content of a p-parallelotope can be written as (8.1) when the random points are type-1 beta distributed, as (8.2) when the random points are type-2 beta distributed and as (8.3) when the random points are gamma distributed. There also densities can be evaluated in terms of the three types of G-functions, as explained above.

Also, Mathai and his co-workers have established a connection between λ-criterion in testing of statistical hypotheses, connected with multivariate normal populations, and certain generalizations of type-1 Dirichlet models. Various generalizations of type-1 and type-2 Dirichlet models were introduced by Mathai and his co-workers starting with Mathai (2003). In this area also G and H-functions appear. The forms $G_{p,p}^{p,0})(x)$ and $G_{p,p}^{p,p}(x)$, coming from products of scalar variable type-1 and type-2 beta variables, appear in this area of generalized Dirichlet models.

Exact 11-digit accurate percentage points connected with the null distributions of the λ-criteria were developed by Mathai and Katiyar starting with the *Biometrika* paper Mathai and Katiyar (1979). A series of papers were published by Mathai and Katiyar in this area. As a byproduct, an algorithm for non-linear least squares was also developed by them, see Mathai and Katiyar (1993).

8.1. Work on Mittag-Leffler function and Mittag-Leffler density

Haubold, Mathai and Saxena have solved a large number of fractional differential equations, starting from 2000, where the solutions invariably come in terms of Mittag-Leffler function, Wright's function or H-function. Exponential type solutions of integer order differential equations automatically change to Mittag-Leffler functions when we go from integer order to fractional differential equations. Gorenflo and Mainardi call Mittag-Leffler function the queen function in fractional differential equations. There is also a Mittag-Leffler stochastic process based on a Mittag-Leffler density, which is a non-Gaussian stochastic process. Work in this area is summarized in the review paper Haubold, Mathai and Saxena (2011). Mathai has also introduced a generalized Mittag-Leffler density and shown that it is attracted to heavy-tailed models such as Lévy and Linnik densities, rather than to Gaussian models. Several interesting properties and structural decompositions are also given there.

9. Mathai's Contributions to Characterization Problems

In this area two basic books are Mathai and Rathie (1975) and Mathai and Pederzoli (1977). Characterization is the unique determination through some given properties. Characterization of a density means to show that certain property or properties uniquely determine that density. Unique determination of a concept means to give an axiomatic definition to that concept. That is, to show that the proposed axioms will uniquely determine the concept. The techniques used in this area, to go from the given properties to the density or from the given axioms to the concept such as"uncertainty" or its complement "information" etc, are functional equations, differential equations, Laplace, Mellin, Fourier transforms etc. For example, let us look at the distribution of error. The error ϵ may be the error in measurement in an experiment, the error between observed and predicted values etc. If the factors contributing to the error are known then the experimenter will try to control these factors. Very often the error is contributed by infinitely many unknown factors each factor contributing infinitesimal quantities towards ϵ. Let us put some conditions on this ϵ. Let

$$\epsilon = \epsilon_1 + \epsilon_2 + ..., S_n = \epsilon_1 + ... + \epsilon_n \tag{i}$$

where $\epsilon_1, \epsilon_2, ...$ are assumed to be independently distributed. Let us assume that each $\epsilon_j = +a$ or $-a$ with equal probabilities. That is,

$$Pr\{\epsilon_j = a\} = Pr\{\epsilon_j = -a\} = \frac{1}{2}, j = 1, 2, ... \tag{ii}$$

Let us assume that the total variance of ϵ is finite or

$$\mathrm{Var}(\epsilon) = \sigma^2 < \infty. \tag{iii}$$

Let us check the consequence of these three assumptions. $\mathrm{Var}(S_n) = n\mathrm{Var}(\epsilon_j) = n[\frac{1}{2}a^2 + \frac{1}{2}(-a)^2] = na^2$, where a is fixed and finite. For large n we may take $a = \frac{\sigma^2}{n}$. The moment generating function of ϵ_j is

$$M_{\epsilon_j}(t) = E[\mathrm{e}^{\epsilon_j t}] = \frac{1}{2}[\mathrm{e}^{-at} + \mathrm{e}^{at}] = 1 + \frac{a^2t^2}{2!} + \frac{a^4t^4}{4!} + ... = 1 + \frac{\sigma^2t^2}{2!n} + O(\frac{1}{n^2}).$$

Hence

$$M_{S_n}(t) = [M_{\epsilon_j}(t)]^n = [1 + \frac{\sigma^2t^2}{2!n} + O(\frac{1}{n^2})]^n.$$

That is

$$\ln[M_{S_n}(t)] = n\ln[1 + \frac{\sigma^2t^2}{2!n} + O(\frac{1}{n^2})] = \frac{\sigma^2t^2}{2} + O(\frac{1}{n}) \to \frac{\sigma^2t^2}{2} \text{ as } n \to \infty.$$

Therefore

$$M_\epsilon(t) = \mathrm{e}^{\frac{t^2\sigma^2}{2}}$$

which is the moment generating function of a normal density with mean value zero and variance σ^2 or the density is

$$f(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{\epsilon^2}{2\sigma^2}}, -\infty < \epsilon < \infty, \sigma > 0. \tag{9.1}$$

This is the derivation of the Gaussian or normal density given by Gauss, and hence it is also called the Gaussian density or error curve. The book, Mathai and Pederzoli (1977) contains such characterizations of the normal probability law by using structural properties, regression properties etc. One fundamental idea was introduced in this area by Gordon and Mathai (1972) in their Annals' paper. They tried to come up with pseudo-analytic functions of matrix argument involving rectangular matrices and by using this, characterization theorems were established for multivariate normal densities.

In Mathai and Rathie (1975), axiomatic definitions of information theory measures and basic statistical concepts are given. This is the first book giving axiomatic definitions of information measures. The techniques used are mainly from functional equations, that is , by using the proposed axioms create a functional equation and obtain its unique solution by imposing more conditions, if necessary, thus coming up with a unique definition or characterization of the concept. One such measure there is the one introduced by two Hungarian mathematicians called Havrda-Charvát measure, which for the continuous case is the following

$$H_\alpha(f) = \frac{\int_{-\infty}^{\infty}[f(x)]^\alpha \mathrm{d}x - 1}{2^{1-\alpha} - 1}, \alpha \neq 1 \tag{9.2}$$

where $f(x)$ is a density of the real scalar variable x. There is a corresponding discrete analogue, which is given by

$$H_\alpha(p_1, ..., p_k) = \frac{\sum_{j=1}^{k}[p_j]^\alpha - 1}{2^{1-\alpha} - 1}, \alpha \neq 1 \tag{9.3}$$

where $p_j > 0, j = 1, ..., k; p_1 + ... + p_k = 1$. A modified form of (9.2) and (9.3) is Tsallis entropy given by

$$T_q(f) = \frac{\int_{-\infty}^{\infty}[f(x)]^q \mathrm{d}x - 1}{1 - q}, q \neq 1 \tag{9.4}$$

for the continuous case, with a corresponding discrete analogue. Optimization of (9.4) under the condition that the total energy is preserved or the first moment is fixed, leads to Tsallis statistics of non-extensive statistical mechanics. Tsallis statistics is of the following form

$$p(x) = [1 - (1 - q)x]^{\frac{1}{1-q}} \tag{9.5}$$

which is also a power law in the sense $\frac{\mathrm{d}}{\mathrm{d}x}p(x) = -[p(x)]^q$. Note that a direct optimization of (9.4), under the assumption that the first moment in $f(x)$ is fixed, does not yield (9.5) directly. One has to go through an escort density

$$g(x) = \frac{[p(x)]^q}{\int_{-\infty}^{\infty}[p(x)]^q \mathrm{d}x}$$

and then assume that the first moment is fixed in the escort density $g(x)$, to get the form in (9.5). Mathai's entropy

$$M_\alpha(f) = \frac{\int_x [f(x)]^{2-\alpha} \mathrm{d}x - 1}{\alpha - 1}, \alpha \neq 1, \alpha < 2 \tag{9.6}$$

when optimized under the condition of first moment in $f(x)$ being fixed leads to Tsallis statistics directly. Also the optimization of (9.6) under two moment-type conditions leads to the pathway model, discussed in Section 3, where (9.5) will be a particular case. But (9.4) and (9.5) created the new area of non-extensive statistical mechanics and it is claimed that more than 3000 researcher are working in this area around the globe. After the book Mathai and Rathie (1975) came out, a lot of people in image processing area started to make use of the results. The work of these researchers in image processing resulted in the machines for reading handwriting and automatic sorting of mail, image reading machines at check-out counters of grocery stores etc.

10. Mathai's Work in Biological Modeling

The most significant contribution in this area is the proposal of a theory of growth and form in nature and the explanation of the emergence of beautiful patterns in sunflower, along with explanation for the appearance of Fibonaci sequence and golden ratio there. The proposed theory is only substantiated so far by others and nobody has contradicted yet. The mathematical reconstruction of the sunflower head, with all the features that are seen in nature, is still the cover design of the journal of *Mathematical Biosciences*. The paper of Mathai and Davis appeared in that journal in 1974 and in 1976 the journal adopted the mathematically reconstructed sunflower head of Mathai and Davis (1974) as the journal's cover design. When this paper was sent for publication to this journal, the editor wrote back saying: "enthusiastically accepted for publication" because this was the first time all natural features were explained in full. Several theories were already there, proposed by various people from time to time. As per Davis and Mathai, the programming of the sunflower head is like a point moving along an Archimedes' spiral at a constant speed so that when the point makes an angle θ a second point starts and moves at the same speed. When the second point comes to θ a third point starts, and so on. The rule governing the movement is $f(\theta_1) + f(\theta_2) = f(\theta_1 + \theta_2)$ or $r = k\theta$ where k is a constant, giving Archimedes' spiral. When $\theta \approx 137.5^o$ or $\frac{\theta}{2\pi-\theta} = \frac{\sqrt{5}-1}{2} =$ golden ratio, we have sunflower, coconut tree crown, certain cactus heads and so on. For other values of θ we have other patterns. Such a movement can be generated by a viscous fluid flowing up through a capillary with valves so that when a certain pressure is built up in one chamber the liquid moves up to the next chamber. The continuous flow is made pulse-like at the end. The upward motion can be effected by evaporation process in the leaves, and there is no need for a heart-like mechanism in trees pumping the fluid up.

It is noted that if we take a projection of the growing crown of a coconut tree onto a circle then we get a replica of the sunflower if we look at the center of the leaf attachment to the crown. In a follow-up paper Mathai and Davis (*National Academy of Sciences, India*) have shown that from several mathematical points of view the arrangement of leaves on a coconut tree crown is an ideal arrangement. One more paper in this line was brought out in a FAO publication.

11. Mathai's Contribution to Graph Theory

The first attempt in this area was the paper of Mathai and Rathie (1972) on almost cubic maps. As an application of functional equations, this problem was solved and the paper appeared in the best journal in that area. Mathai had left the area. Then due to the persuasion of Dr B.D. Acharya, Advisor to Government of India in the Department of Science and Technology, Mathai went back to the area and introduced some descriptors based on the norms of adjacency matrices. This was published in the proceedings of an international conference. Then an attempt was made to come up with a criterion called MNS (minimum number of specifiers) so that an underlying pattern can be uniquely determined with the minimum number of 2 or 3 numbers as specifiers. In other words the pattern in the adjacency matrix is uniquely determined by 2 to 3 numbers. Mathai thought that the idea was great and the paper was sent to the best journal in the area but the editor did not appreciate it. Then, since various descriptors were usually used for classifying chemicals, this paper was sent to the journal of Mathematical Chemistry. They did not appreciate it either. Then the work was abandoned. Tom Zaslavsky had seen the draft of this paper. When he heard that Mathai was abandoning the idea of publishing the work, he said that he would make some modifications. Modifications were made and then the paper was published as a joint paper of Mathai and Zaslavsky in 2015.

12. Mathai's Contribution to Design of Experiments and Analysis of Variance

The first paper of Mathai, Mathai (1965), was on an approximate analysis of variance. It was on the analysis of a two-way classification with multiple observations per cell. Here the orthogonality is lost, and when estimating the main effects, one ends up in a singular system of linear equations of the form

$$(I-A)\alpha=b \tag{11.1}$$

where $A=(a_{ij}), a_{ij}=\frac{n_{ij}}{n_{i.}}, n_{i.}=\sum_{j=1}^{m}n_{ij}, n_{ij}>0$ for all i and j, is called the incidence matrix and the sum of the elements in each row is equal to 1. Thus $I-A$ is singular and hence one cannot write it as $\alpha=(I-A)^{-1}b$ where A and b are known and the $m\times 1$ vector α is unknown and is to be estimated. Mathai noted that one could profitably use the conditions in the design and make $I-A$ a nonsingular matrix. One condition in the design is that $\alpha_1+...+\alpha_m=0$ where $\alpha_j, j=1,...,m$ are the elements in α. Let C be a matrix where all elements in the i-th row of C are the median of the i-th row elements in A, namely the median of $a_{i1},...,a_{im}$ for $i=1,...,m$. Then evidently $C\alpha=O$ (null). Then

$$(I-A)\alpha=b\Rightarrow(I-B+C)\alpha=b\Rightarrow(I-B)\alpha=b \tag{12.2}$$

where $B=(b_{ij}), b_{ij}=a_{ij}-c_i$ and c_i is the median of the i-th row elements in A. Then a norm of B is $\|B\|=\max_i\sum_{j=1}^{m}|a_{ij}-c_i|$. But since the mean deviation from the median is the least, $\|B\|$ is the minimum under the circumstances. Therefore, not only that $I-B$ is nonsingular but the series $I+B+B^2+...$ is the fastest converging series for the problem at hand. Then

$$(I-B)\alpha=b\Rightarrow\alpha=(I+B+B^2+...)b.$$

A good approximation for α is available as $\alpha \approx (I + B)b$. This is found to be sufficient for all practical purposes of testing of statistical hypotheses on the components of α. When this paper was sent to the best journal in the area, *Biometrics*, the three referees unanimously stated that the method was an ingenious one. This work was followed up and Mathai worked on the analysis of data under missing values. But the reception by the research community was poor and hence Mathai left the area.

13. Mathai's Work on Dispersion Theory

This was an idea proposed in 1964, papers appeared in 1967-69, showing that the whole area of statistical inference of estimation and testing of hypotheses or decision-making was study of a properly defined "dispersion" or "scatter". An axiomatic definition of "dispersion" was given and it was illustrated that the estimation process was an optimization of a measure of "dispersion", test criteria were the results of optimization of measures of "dispersion", best prediction problem was an optimization of a measure of "dispersion' etc. A few papers were published but since the reception was poor the idea was abandoned. Later, Italinan statisticians claimed that "Dispersion Theory" was "one of the innovations of the 20th century" and persuaded Mathai to return to the area. By then Mathai was involved in other areas and did not return to develop "Dispersion Theory" further.

14. Mathai's Work on Population Problems and Social Sciences

An expert in population studies, Aleyamma George, had assembled demographic data on 38,000 households, under a Ford Foundation survey in Kerala, India. She sought help from Mathai in analyzing this massive data. Various population aspects were looked into and a series of papers were published in the area of demography and social statistics. One problem that was looked into was the study of inter-live-birth interval in cohabiting fecundable women. Between two live births there may be still births, miscarriages and associated sterile periods etc. The standard models used for waiting time in this area are exponential models. In an exponential model, the probability of the woman getting pregnant soon after marriage is the maximum and it decreases as time goes by. This is not the reality. There are cultural and sociological factors inhibiting pregnancy at the time of marriage and from the massive data on hand it was shown that a gamma model was the most appropriate because the maximum chance was away from the starting point or away from the time of marriage. Various models were constructed for the inter-live-birth-intervals and a number of papers were published in various journals including Sankhya.

Another problem that was looked into was how to come up with a measure of "distance" or "closeness" or "affinity" between two sociological groups or how to say that one community is close to another community with respect to a given characteristic. This study was motivated by the need to come up with a scientific way of checking the claims of some politicians that some communities produce more children compared to other communities. Mathai introduced the concepts of "directed divergence", "affinity" etc from Information Theory to social statistics. Let $P = (p_1, ..., p_k), p_j > 0, j = 1, ..., k, p_1 + ... + p_k = 1$ and $Q = (q_1, ..., q_k), q_j > 0, j = 1, ..., k, q_1 + ... + q_k = 1$ be two discrete populations. Consider the representation of P and Q as points on a hypersphere of radius 1, $x_1^2 + ... + x_k^2 = 1$. Then the points are $(\sqrt{p_1}, ..., \sqrt{p_k})$ and $(\sqrt{q_1}, ..., \sqrt{q_k})$. Consider $\cos\theta$ where θ is the angle between these vectors or points on the hypersphere. Note that the lengths of the vectors are

$$\|(\sqrt{p_1}, ..., \sqrt{p_k})\| = 1, \|(\sqrt{q_1}, ..., \sqrt{q_k})\| = 1$$

and hence

$$\cos\theta = \sum_{j=1}^{k} \sqrt{p_j q_j}. \tag{14.1}$$

This is a measure of angular dispersion and it is usually called "affinity" between P and Q or Matusita's measue of affinity between two discrete distributions. George and Mathai computed "affinity" between communities with reference to the characteristic of production of children and found that the politicians' statements did not match with the realities. The paper was published in the *Journal of Biosocial Sciences* published from UK. Several papers were published in *Canadian Studies in Population. Demography of India* etc where different techniques from Information Theory and Statistics were introduced into social sciences.

Another problem that was looked into was the estimation of missing events from information supplied by various agencies. In Kerala, India, birth registers are kept by Registrar's office, villages, churches, temples etc, none of them has a complete list. How to come up with the true estimate of the population? The age-old technique used in this area was based on an approximate likelihood procedure. It was shown by George and Mathai that, in the approximate likelihood, the deleted portion contained the most relevant information and hence the method was not good and they proposed a method based on information measures.

15. Mathai's Work on Integer Programming

An optimization problem such as optimizing a linear function subject to quadratic constraints or a quadratic function subject to linear or quadratic constraints etc are not difficult when the variables are continuous. But if the variables are discrete in the sense if the variables are restricted to integers such as a grid of positive integers only, then the optimization problem becomes quite complicated. The usual results that one may expect can be seen not to hold. Such an optimization problem when the variables are restricted to a grid of integers is called an integer programming problem. A problem of this type of optimizing a linear function subject to a quadratic constraint but the variables are defined on a grid of positive integers, was brought to the attention of Mathai by a Greek statistician and an earlier colleague Stratis Kounias. They looked into the problem and came up with a nice technique of handling such problems and the paper was published in the best journal in the area, *Optimization*, in 1988.

16. Mathai's Work on Quadratic and Bilinear Forms

Major contributions in these areas are summarized in the books Mathai and Provost (1992), and Mathai, Provost and Hayakawa (1995). Mathai-Provost book has 644 citations as of August 2015 as per Google's counts. Almost all available results on quadratic forms in real random variables are included in Mathai and Provost (1992). Complex variable cases are not included there. There are a lot of applications of the theory of quadratic forms in complex random variables, such as applications in radar, sonar etc. Major parts of the two books are contributions from Mathai, Provost, Hayakawa and Mathai and Provost. There is a very important concept in quadratic forms in Gaussian random variables called chisquaredness of quadratic forms. That is, $X'AX \sim \chi_r^2$ iff A is idempotent and of rank r, where X the $p\times 1$ vector having the standard normal distribution $N_p(O.I)$, that is, $X \sim N_p(O,I)$ and χ_r^2 is a chisquare random variable with r degrees of freedom. This result is the most important and fundamental result in the areas of model building, regression analysis, analysis of variance etc. Is there a corresponding concept when dealing with bilinear forms? When the samples come from a bivariate Gaussian or normal population it is not difficult to work out the density of the sample correlation coefficient. But what about the density of the sample covariance, without the scaling factors of the standard deviations? This was a question raised in a course in 1962 when Mathai was a student at the University of Toronto, taking his M.A degree in Mathematics. Both these questions were answered by Mathai (1993b) where Mathai introduced a concept called Laplacianness of bilinear forms ((Mathai 1993a)) and also worked out the density of the covariance structures observing that covariance structure is a bilinear form. The necessary and sufficient conditions for a bilinear form to be noncentral generalized Laplacian are given in Corollary 2.5.2 of Mathai, Provost and Hayakawa (1995). For a noncentral generalized Laplacian the moment generating function is of the form

$$M(t) = (1-\beta^2t^2)^{-\alpha}\exp\{-2\lambda + 2\lambda(1-\beta^2t^2)^{-1}\}$$

where λ is the non-centrality parameter.

17. New Concepts and Procedures Introduced by Mathai

The following are the new concepts, new ideas and new procedures introduced by Mathai: Dispersion Theory - a new theory and a new idea proposed in 1967; Developing a generalized partial fraction technique, with Rathie, in 1971; Developing an operator to evaluate residues when poles of all types of orders occur

(1971); Proposed a theory of growth and forms in nature (1974), the theory still stands, mathematical reconstruction of a sunflower head; Introduction of the concepts of "affinity", "distance" etc in social sciences and creating a procedure to compare sociological groups (1974); Functions of matrix argument through M-transforms and M-convolutions - new ideas introduced in (1990-1997); Non-linear least square algorithm (1993); Solving Miles' conjecture in geometrical probabilities, creating and solving parallel conjectures (1982), introduction of Jacobians of matrix transformations in solving problems of random volumes, replacing the complicated Integral and Differential Geometry procedures (1982); Now meaningful physical interpretations are available for M-convolutions. Unique recovery of $f(X)$ from its M-transform is still a conjecture, extensions of Jacobians from the real case to complex matrix-variate cases in a large number of situations; Concept of Laplacianness of bilinear forms and creating density of covariance structure (1993); Pathway model and pathway idea (2005), particular cases are the most popular Tsallis statistics and superstatistics in the new area of non-extensive statistical mechanics; Extension of Fractional Calculus to real matrix-variate cases (2007); Establishing a connection between Fractional Calculus and Statistical Distribution Theory (2007); Mathai's entropy (2007); Geometrical interpretation and a general definition for fractional integrals (2013-2015); Extension of Fractional Calculus to complex matrix-variate case and complex domain in general (2013); Extension of fractional calculus to many matrix-variate cases (real and complex)(2014); Development of a fractional differential operator in the matrix-variate case (2015).

References

[1] Constantine, A. G., Some noncentral distribution problems in multivariate analysis, *Annals of Mathematical Statistics*, **34** (1963), 1270–1285.

[2] George, A. and Mathai, A. M., Applications of the concepts of affinity and distance to population problems, *Journal of Biosocial Sciences (London)*, **6** (1974), 347–356.

[3] Gordon, F. S. and Mathai, A. M., Characterization of multivariate normal distribtion using regression properties, *Annals of Mathematical Statistics*, **43** (1972), 53–65.

[4] Haubold, H. J. and Mathai, A. M., Gravitational instability in a multicomponent cosmological medium, *Journal of Mathematical Physics*, **29** (9) (1988), 2069–2077.

[5] Haubold, H. J. and Mathai, A. M., Sun, In, *Encyclopaedia of Planetary Sciences*, J. H. Shirley and R. W. Fairbridge, eds, Chapman and hall, London, (1997), 786–794.

[6] Haubold, H. J. and Mathai, A. M., Universe, structure of the, *Encyclopedia of Applied Physics*, **23** (1998), 47–51.

[7] Haubold, H. J., Mathai, A. M. and Saxena, R. K., Mittag-Leffler functions and their applications, *Journal of Applied Mathematics*, (2011), ID 298628, 2011, 51 pp.

[8] Herz, C. S., Bessel functions of matrix argument, *Anals of Mathematics*, **63** (3) (1955), 474–523.

[9] James, A. T., Zonal polynomials of the real positive definite symmetric matrices, *Annals of Mathematics*, **74** (1961), 456–469.

[10] Kounias, S. and Mathai, A. M., Maximizing the sum of integers when their sum of squares is fixed, *Optimization*, **19** (1988), 123–131.

[11] Krätzel, E., Integral transforms the Bessel type, In *Generalized Functions & Operational Calculus, Proc. Conf. Verna, 1975*, Bulg. Acad. Sci., Sofia (1979), 148–165.

[12] Mathai, A. M., An approximate analysis of a two-way layout, *Biometrics*, **21** (1965), 376–385.

[13] Mathai, A. M., On a conjecture in geometric probability regarding asymptotic normality of a random simplex, *Annals of Probability*, **10** (1982), 247–251.

[14] Mathai, A. M., On a system of differential equations connected with the gravitational instability in a multi-component medium in Newtonian cosmology, *Studies in Applied Mathematics*, **80** (1989), 75–93.

[15] Mathai, A. M. *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, (1993), Oxford University Press, Oxford.

[16] Mathai, A. M., On generalized Laplacianness of bilinear forms in normal variables, *Journal of Multivariate Analysis*, **45** (1993a), 239–246.

[17] Mathai, A. M., The residual effect of a growth-decay mechanism and the distribution of covariance structures, *Canadian Journal of Statistics*, **21**(3) (1993b), 277–283.

[18] Mathai, A. M., *Jacobians of Matrix Transformations and Functions of Matrix Argument*, (1997), World Scientific Publishing, New York.

[19] Mathai, A. M., Pollution by vehicular travels from the suburbs to the city core, *Environmetics*, (1998), 617–628.

[20] Mathai, A. M., *An Introduction to Geometrical Probability: Distributional Aspects with Applications*, (1999), Gordon and Breach, Newark.

[21] Mathai, A. M., Random p-parallelotope in Euclidean n-space, *Advances in Applied Probability*, **31**(2) (199a), 343–354.

[22] Mathai, A. M., Buffon type problem in the Euclidean space R3, *Rendiconti del Circolo Matematico di Palermo*, **XLVIII** (1999b), 487–506.

[23] Mathai, A. M., Order statistics from a logistic distribution and applications to survival & reliability analysis, *IEEE Transactions on Reliability*, **52**(2) (2003), 200–206.

[24] Mathai, A. M., (2005): A pathway to matrix-variate gamma and normal densities, *Linear Algebra and its Applications*, **396**, 317-328.

[25] Mathai, A. M., Random volumes under a general matrix-variate model, *Linear Algebra and its Applications*, **425** (2007), 162–170.

[26] Mathai, A. M., Fractional integrals in the matrix-variate cases and connection to statistical distributions, *Integral Transforms and Special Functions*, **20**(12) (2009), 871–882.

[27] Mathai, A. M., Generalized Krätzel integral and associated statistical densities, *International Journal of Mathematical Analysis*, **6**(51) (2012), 2501–2510.

[28] Mathai, A. M., Fractional integral operators in the complex matrix-variate case, *Linear Algebra and its Applications*, **439** (2013), 2901–2913.

[29] Mathai, A. M., Fractional integral operators involving many matrix variables, *Linear Algebra and its Applications*, **446** (2014), 196–215.

[30] Mathai, A. M., Fractional differential operators in the complex matrix-variate case, *Linear Algebra and its Applications*, (2015), (to appear.)

[31] Mathai, A. M. and Davis, T. A., Constructing a sunflower head, *Mathematical Biosciences*, **20** (1974), 117–133.

[32] Mathai, A. M. and Haubold, H. J., *Modern Problems in Nuclear and Neutrino Astrophysics*, (1988), Akademie-Verlag, Berlin.

[33] Mathai, A. M. and Haubold, H. J., Pathway model, superstatistics, Tsallis statistics and a generalized measure of entropy, *Physica A*, **375** (2007), 110–122.

[34] Mathai, A. M. and Haubold, H. J., *Special Functions for Applied Scientists*, (2008), Springer, New York.

[35] Mathai, A. M. and Katiyar, R. S., Exact percentage points for testing independence, *Biometrika*, **66** (1979), 353–356.

[36] Mathai, A. M. and Katiyar, R. S., A new algorithm for nonlinear least squares, *Researches in Mathematical Statistics (Russian)*, **207**(10) (1993), 143–157 [Translation by the American Mathematical Society].

[37] Mathai, A. M. and Moschopoulos, P. G., Pollution by vehicular travels from satellite townships to the city, *Environmetrics*, **10** (1999), 791–802.

[38] Mathai, A. M. and Pederzoli, G., *Characterization of the Normal Probability Law*, (1977), Wiley Eastern, New Delhi and Wiley Halsted, New York.

[39] Mathai, A. M. and Provost, S. B., *Quadratic Forms in Random Variables: Theory and Applications*, (1992), Marcel Dekker, New York.

[40] Mathai, A. M. and Provost, S. B., Some complex matrix-variate statistical distributions on rectangular matrices, *Linear Algebra and its Applications*, **410** (2005), 198–216.

[41] Mathai, A. M. and Provost, S. B., On q-logistic and related distributions, *IEEE Transactions on Reliability*, **55**(2) (2006), 237–244.

[42] Mathai, A. M., Provost, S. B. and Hayakawa, T., *Bilinear Forms and Zonal Polynomials*, (1995), Lecture Notes in Statistics No.102, Springer, New York.

[43] Mathai, A. M. and Rathie, P. N., Exact distribution of Wilks' criterion, *Annals of Mathematical Statistics*, **42** (1971), 1010–1019.

[44] Mathai, A. M. and Rathie, P.N. Enumeration of almost cubic maps, *Journal of Combinatorial Theory*, **13** (1972), 83–90.

[45] Mathai, A. M. and Rathie, P. N., *Basic Concepts in Information Theory and Statistics: Axiomatic Foundations and Application*, (1975), Wiley Halsted, New York and Wiley Eastern, New Delhi.

[46] Mathai, A. M. and Saxena, R. K., Distribution of a product and structural setup of densities, *Annals of Mathematical Statistics*, **40** (1969), 1439–1448.

[47] Mathai, A. M. and Saxena, R. K., *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, (1973), Springer, Lecture Notes No. 348, Heidelberg.

[48] Mathai, A. M. and Saxena, R. K., *The H-function with Applications in Statistics and Other Disciplines*, (1978), Wiley Halsted, New York and Wiley Eastern, New Delhi.

[49] Mathai, A. M., Saxena, R. K. and Haubold, H. J., *The H-function: Theory and Applications*, (2010), Springer, New York.

[50] Thomas, S. and Mathai, A. M., P-content of p-parallelotope and its connection to likelihood ratio statistics, *Sankhya Series A*, **71**(1) (2009), 49-63.

A. M. Mathai
Director, Centre for Mathematical and Statistical Sciences
Peechi Campus, KFRI, Peechi-680653, Kerala, India
and
Emeritus Professor of Mathematics and Statistics
McGill University, Montreal, Canada, H3A 2K6
E-mail: *directorcms458@gmail.com, mathai@math.mcgill.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 45-61

COMPACT OPERATORS AND HILBERT SCALES IN ILL-POSED PROBLEMS*

M. T. NAIR

Abstract. A *Hilbert scale* is a family of Hilbert spaces H_s, $s \in \mathbb{R}$, such that for every $s, t \in \mathbb{R}$ with $s < t$, $H_t \subseteq H_s$ and the inclusion operator is continuous. Given a Hibert space H, we show how to construct a Hilbert scale with $H_0 = H$ using the concept of *Gelfand triple* and give examples of Hilbert scales which are generated by compact operators between Hilbert spces as well as closed densely defined unbounded operators. Citing results from some of the recent work of the author, we discuss the use of Hilbert scales for obtaining improved error estimates for regularized approximate solutions of ill-posed operator equations.

1. Hilbert Scales

The role of Hilbert scales while solving operator equations is similar to the role of Sobolev spaces in partial differential equations.

Definition 1.1. A **Hilbert scale** is a family $\{H_s\}_{s\in\mathbb{R}}$ of Hilbert spaces such that the following hold.

(1) For $s < t$, $H_t \subseteq H_s$ and as a subspace, H_t is a dense subspace of H_s;
(2) As Hilbert spaces, the above inclusion is a continuous embedding, i.e., there exists $c_{s,t} > 0$ such that
$$\|x\|_s \le c_{s,t}\|x\|_t \quad \forall\, x \in H_t.$$

One may have a Hilbert space H and a family $\{H_s\}_{s\geq 0}$ of Hilbert spaces satisfying the properties (1) and (2) in Definition 1.1 with $H_0 = H$. Then the Hilbert spaces H_{-s} for $s > 0$ are defined using the concept of a Gelfand triple.

1.1. **Gelfand triple.** Let H be a Hilbert space and H_1 be a dense subspace of H, which itself is a Hilbert space with respect to a stronger norm $\|\cdot\|_1$, i.e.,
$$\|u\| \le c\|u\|_1 \quad \forall\, u \in H_1$$
for some $c > 0$. For $v \in H$, let

---

* The (modified) text of the 12^{th} Ganesh Prasad Memorial Award lecture delivered at the 81^{st} Annual Conference of the Indian Mathematical Society held at the Visvesvaraya National Institute of Technology, Nagpur - 440 010, Maharashtra, during the period December 27-30, 2015.

2010 Mathematics Subject Classification : 47A52, 65J20, 65J22
Key words and phrases: Hilbert scales, Gelfand triple, Compact operators, Ill-posed problems, Tikhonov regularization.

$$\|v\|_{-1} := \sup\{|\langle u, v\rangle| : u \in H_1, \|u\|_1 \le 1\}.$$

Then, using the fact that H_1 is a dense subspace of H, it can be seen that $v \mapsto \|v\|_{-1}$ is a norm on H and

$$\|v\|_{-1} \le c\|v\| \quad \forall\, v \in H.$$

Let H_{-1} be the completion of H w.r.t. $\|\cdot\|_{-1}$. The inclusions in

$$H_1 \subseteq H \subseteq H_{-1}$$

are continuous embeddings.

Definition 1.2. The triple (H_1, H, H_{-1}) is called a **Gelfand triple**.

We shall show that H_{-1} is linearly isometric with $(H_1)'$ so that H_{-1} is a Hilbert space with respect to the inner product induced from $(H_1)'$. Before that let us give an example of a Gelfand triple.

Example 1.1. Let H be a separable Hilbert space, $\{u_n : n \in \mathbb{N}\}$ be an orthonormal basis of H, and (σ_n) be a sequence of positive real number such that $\sigma_n \to 0$ as $n \to \infty$. Let

$$H_1 := \{x \in H : \sum_{n=1}^{\infty} \frac{|\langle x, u_n\rangle|^2}{\sigma_n^2} < \infty\}.$$

On H_1, define

$$\langle x, y\rangle_1 = \sum_{n=1}^{\infty} \frac{\langle x, u_n\rangle\langle u_n, y\rangle}{\sigma_n^2}, \quad x, y \in H_1.$$

It can be seen that $\langle\cdot,\cdot\rangle_1$ is an inner product on H_1 and H_1 is a Hilbert space with respect to this inner product. Now, we show that for $x \in H$,

$$\|x\|_{-1} = \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n\rangle|^2\Big)^{\frac{1}{2}}.$$

Recall that

$$\|x\|_{-1} := \sup\{|\langle x, u\rangle| : \sum_{n=1}^{\infty} \frac{|\langle u, u_n\rangle|^2}{\sigma_n^2} \le 1\}.$$

Note that for $u \in H_1$ with $\|u\|_1 \le 1$,

$$\begin{aligned}
|\langle x, u\rangle| &= \Big|\sum_{n=1}^{\infty} \sigma_n \langle x, u_n\rangle \frac{\langle u_n, u\rangle}{\sigma_n}\Big| \\
&\le \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n\rangle|^2\Big)^{\frac{1}{2}} \Big(\sum_{n=1}^{\infty} \frac{|\langle u, u_n\rangle|^2}{\sigma_n^2}\Big)^{\frac{1}{2}} \\
&\le \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n\rangle|^2\Big)^{\frac{1}{2}}.
\end{aligned}$$

Thus, $\|x\|_{-1} \le \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n\rangle|^2\Big)^{\frac{1}{2}}$. To show the equality, let

$$u = \sum_{n=1}^{\infty} \frac{\sigma_n^2 \langle x, u_n\rangle}{\alpha} u_n, \quad \alpha := \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n\rangle|^2\Big)^{\frac{1}{2}}.$$

Then $\langle u, u_n \rangle = \frac{\sigma_n^2 \langle x, u_n \rangle}{\alpha}$ so that

$$\sum_{n=1}^{\infty} \frac{|\langle u, u_n \rangle|^2}{\sigma_n^2} = \frac{1}{\alpha^2} \sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n \rangle|^2 = 1$$

and

$$\langle x, u \rangle = \frac{1}{\alpha} \sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n \rangle|^2 = \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n \rangle|^2 \Big)^{\frac{1}{2}}.$$

Therefore, $\|x\|_{-1} = \Big(\sum_{n=1}^{\infty} \sigma_n^2 |\langle x, u_n \rangle|^2 \Big)^{\frac{1}{2}}$.

1.2. **Linear isometry between H_{-1} and $(H_1)'$.** We show H_{-1} is linearly isometric with $(H_1)'$. In particular, H_{-1} is a Hilbert space with respect to the inner product induced from $(H_1)'$.

Theorem 1.1. *Let $\widetilde{H} := H$ with $\|\cdot\|_{-1}$. For $v \in \tilde{H}$, let $f_v(u) = \langle u, v \rangle, \quad u \in H_1$. Then $f_v \in (H_1)'$ and the map $\Phi_0 : \tilde{H} \to (H_1)'$ defined by $\Phi_0(v) := f_v, \quad v \in H_{-1}$, is a linear isometry.*

Proof. Clearly f_v is linear. Moreover

$$|f_v(u)| = |\langle u, v \rangle| \le \|u\| \, \|v\| \le \|u\|_1 \|v\|, \quad u \in H_1.$$

Hence $f_v \in (H_1)'$ and

$$\|f_v\| = \sup\{|f_v(u)| : \|u\|_1 \le 1\} = \|v\|_{-1}.$$

Thus $\Phi_0 : v \mapsto f_v$ is a linear isometry from $\tilde{H}$ to $(H_1)'$. □

We shall extend the linear isometry Φ_0 obtained in the above theorem to the whole of H_{-1} and show that the extended isometry is surjective. First we prove the following.

Theorem 1.2. *The subspace $\{f_v : v \in H\}$ is dense in $(H_1)'$.*

Proof. By Hahn-Banach extension theorem, it is enough to prove that for $\varphi \in (H_1)''$, $\varphi(f_v) = 0$ for all $v \in H$ implies $\varphi = 0$. So, let $\varphi \in (H_1)''$ such that $\varphi(f_v) = 0$ for all $v \in H$. Since H_1 is reflexive, there exists $u \in H_1$ such that

$$\varphi(f) = f(u) \quad \forall\, f \in (H_1)'.$$

Thus, $\langle v, u \rangle = f_v(u) = \varphi(f_v) = 0 \quad \forall\, v \in H$. Hence, $u = 0$. Consequently, $\varphi = 0$. □

Theorem 1.3. *Let $\Phi : H_{-1} \to (H_1)'$ be defined by*

$$\Phi(v) := \lim_{n\to\infty} f_{v_n}, \quad v \in H_{-1},$$

where (v_n) in H is such that $\|v - v_n\|_{-1} \to 0$. Then Φ is a surjective linear isometry.

Proof. Clearly, Φ is linear. Also,

$$\|\Phi(v)\| = \lim_{n\to\infty} \|f_{v_n}\| = \lim_{n\to\infty} \|v_n\|_{-1} = \|v\|_{-1}, \quad v \in H_{-1}.$$

Thus, Φ is a linear isometry. For surjectivity, let $f \in (H_1)'$. Since $\{f_v : v \in H\}$ is dense in $(H_1)'$, there exists (v_n) in H such that $\|f - f_{v_n}\| \to 0$. Then (v_n) is a Cauchy sequence in $\tilde{H}$ and

$$f = \lim_{n\to\infty} f_{v_n} = \Phi(v),$$

where $v \in H_{-1}$ is such that $\|v - v_n\|_{-1} \to 0$. □

1.3. **The Hilbert space H_{-1}.** For $u, v \in H_{-1}$, define

$$\langle u, v\rangle_{-1} := \langle \Phi(u), \Phi(v)\rangle.$$

Recall that, for $f, g \in (H_1)'$,

$$\langle f, g\rangle := \langle u_g, u_f\rangle_{H_1},$$

where u_f is the unique element in H_1 (by Riesz representation theorem) such that

$$\langle v, u_f\rangle_{H_1} = f(v), \quad v \in H_1.$$

Thus, Gelfant triple is a triple of Hilbert spaces. Given $u \in H$, $\Phi(u) := f_u \in (H_1)'$, and there exists a unique $\tilde{u} \in H_1$ such that

$$\langle w, \tilde{u}\rangle_{H_1} = f_u(w) := \langle w, u\rangle \quad \forall\, w \in H_1.$$

Here is an illustration of the above.

Example 1.2. Let Ω be a bounded open subset of $\mathbb{R}^k$. Then the Sobolev space $H_0^1(\Omega)$ is a dense subspace of $L^2(\Omega)$. Given $u \in L^2(\Omega)$, there exists a unique $\tilde{u} \in H_0^1(\Omega)$ such that

$$\langle w, \tilde{u}\rangle_{H_0^1(\Omega)} = f_u(w) := \langle w, u\rangle_{L^2(\Omega)} \quad \forall\, w \in H_0^1(\Omega)$$

i.e.,

$$\int_\Omega \nabla w(x).\nabla \tilde{u}(x)\, dx = \int_\Omega w(x)u(x)\, dx \quad \forall\, w \in H_0^1(\Omega).$$

Thus, $\tilde{u}$ is the *weak solution* of $\ -\Delta\tilde{u} = u\ $ with $\ u_{|\partial\Omega} = 0$ and for $u, v \in H$,

$$\langle u, v\rangle_{-1} := \int_\Omega \nabla\tilde{u}(x).\nabla\tilde{v}(x)\, dx.$$

1.4. **Examples of Hilbert scales.**

Example 1.3. Let H be a separable Hilbert space and $\{u_n : n \in \mathbb{N}\}$ be an orthonormal basis of H. Let (σ_n) be a sequence of positive real numbers with $\sigma_n \to 0$. For $s \geq 0$, let

$$H_s := \{x \in H : \sum_{n=1}^{\infty} \frac{|\langle x, u_n\rangle|^2}{\sigma_n^{2s}} < \infty\}.$$

Then it can be shown that H_s is a Hilbert space with inner product

$$\langle x, y\rangle_s := \sum_{n=1}^{\infty} \frac{\langle x, u_n\rangle\langle u_n, y\rangle}{\sigma_n^{2s}}$$

and the corresponding norm is stronger than the original norm. For $s < 0$, H_s is defined via Gelfand triple. Thus, $\{H_s : s \in \mathbb{R}\}$ is a Hilbert scale. For $x \in H$ and $s \geq 0$,

$$\|x\|_{-s} = \sum_{n=1}^{\infty} \sigma_n^{2s} |\langle x, u_n \rangle|^2.$$

This can be seen as in Example 1.1.

Example 1.4. Let H be a Hilbert space and $L : D(L) \subseteq H \to H$ be a densely defined strictly positive self adjoint operator which is also coercive., i.e.,

$$\langle Lx, x\rangle \geq \gamma \|x\|^2 \quad \forall\, x \in H$$

for some $\gamma > 0$. Let $X = \cap_{k=1}^{\infty} D(L^k)$. For $s > 0$, let

$$\langle u, v\rangle_s := \langle L^s u, v\rangle, \quad u, v \in X,$$

where L^s is defined via spectral representation of L. Thus,

$$\langle L^s u, v\rangle = \int_{\gamma}^{\infty} \lambda^s d\langle E_\lambda u, v\rangle \quad u, v \in X,$$

where $\{E_\lambda : \lambda \geq \gamma\}$ is the *resolution of identity* of L. Clearly, $\langle \cdot, \cdot \rangle_s$ is an inner product on X. Let H_s be the completion of X with respect to $\langle \cdot, \cdot\rangle_s$. The following can be verified.

(1) For $s > 0$, H_s is a dense subspace of H as a vector space.
(2) For $s > 0$, H_s is continuously embedded in H.
(3) For $s > 0$, (H_s, H_0, H_{-s}) is Gelfand triple with $H_0 = H$.
(4) For $0 \leq s \leq t$, $H_t \subseteq H_s$ and the inclusion is continuous.

Thus, $\{H_s : s \in \mathbb{R}\}$ is a Hilbert scale.

Definition 1.3. The Hilbert scale $\{H_s : s \in \mathbb{R}\}$ as in Example 1.4 is called the **Hilbert scale generated by** L.

Traditionally, Hilbert scale is defined as in Example 1.4.

Remark 1.1. The Hilbert scale in Example 1.3 is a special case of Example 1.4: The Hilbert scale is generated by $L : D(L) \subseteq H \to H$,

$$Lx := \sum_{n=1}^{\infty} \frac{\langle x, u_n\rangle}{\sigma_n} u_n, \quad x \in D(L),$$

where

$$D(L) := \{x \in H : \sum_{n=1}^{\infty} \frac{|\langle x, u_n\rangle|^2}{\sigma_n^2} < \infty\}.$$

Note that

$$\langle Lx, x\rangle = \sum_{n=1}^{\infty} \frac{|\langle x, u_n\rangle|^2}{\sigma_n} \geq \frac{1}{\sigma_0}\|x\|^2,$$

where $\sigma_0 := \sup_n \sigma_n$.

We may observe that $A : H \to H$ defined by

$$Ax := \sum_{n=1}^{\infty} \sigma_n \langle x, u_n\rangle u_n, \quad x \in H$$

is an injective compact self adjoint operator and its range is $D(L)$, and its inverse is L.

Remark 1.2. Consider the Hilbert scale as in Example 1.4. Let $s > 0$ and $u \in H$. Then $f_u(x) = \langle x, u \rangle,\ u \in H_1$ and $f_u \in (H_s)'$. By Riesz representation theorem, there exists a unique $\tilde{u} \in H_1$ such that $f_u(x) = \langle x, \tilde{u} \rangle_s,\ x \in H_s$. Thus,

$$\langle x, u \rangle = \langle x, L^s \tilde{u} \rangle \quad \forall\, x \in H,$$

i.e., $\tilde{u}$ is the unique element in H_s such that $L^s \tilde{u} = u$. Note that, even if $u \in H_s$, the element $\tilde{u}$ need not be equal to u.

Example 1.5. Let X and Y be Hilbert spaces and $K : X \to Y$ be a compact operator of infinite rank. Then *singular value representation* of K is given by (see Nair[10])

$$Kx = \sum_{n=1}^{\infty} \sigma_n \langle x, u_n \rangle v_n, \quad x \in X,$$

where (σ_n) is a sequence of positive real numbers which converges to 0, $\{u_n : n \in \mathbb{N}\}$ is an orthonormal basis of $H := N(K)^\perp$, and $\{v_n : n \in \mathbb{N}\}$ is an orthonormal basis of $\overline{R(K)} = N(K^*)^\perp$. In fact, for $n \in \mathbb{N}$, $Ku_n = \sigma_n v_n$ and $K^* v_n = \sigma_n u_n$. Further,

$$(K^*K)^{\frac{1}{2}} x = \sum_{n=1}^{\infty} \sigma_n \langle x, u_n \rangle u_n, \quad x \in X.$$

As observed in Remark 1.1, the operator $L : D(L) \subset H \to H$ defined by

$$Lx := (K^*K)^{-\frac{1}{2}} x = \sum_{n=1}^{\infty} \frac{\langle x, u_n \rangle}{\sigma_n} u_n, \quad x \in D(L),$$

with

$$D(L) := \{x \in H : \sum_{n=1}^{\infty} \frac{|\langle x, u_n \rangle|^2}{\sigma_n^2} < \infty\}$$

is a positive densely defined coercive self-adjoint operator, which generate the Hilbert scale, also called the *Hilbert scale generated by* the compact operator K.

Theorem 1.4. *Let $\{H_s : s \in \mathbb{R}\}$ be as in Example 1.3 and*

$$u_n^{(s)} := \sigma_n^s u_n, \quad s \in \mathbb{R},\, n \in \mathbb{N}.$$

Then the following hold.

(i) $\{u_n^{(s)} : n \in \mathbb{N}\}$ *is an orthonormal basis of H_s.*

(ii) *For $s < t$, the inclusion map $\mathcal{I}_{s,t} : H_t \to H_s$ is a compact embedding.*

Proof. For $x \in H_s$, we have

$$\langle x, u_j^{(s)} \rangle_s = \sum_{n=1}^{\infty} \frac{\langle x, u_n \rangle \langle u_n, u_j^{(s)} \rangle}{\sigma_n^{2s}} = \sum_{n=1}^{\infty} \sigma_j^s \frac{\langle x, u_n \rangle \langle u_n, u_j \rangle}{\sigma_n^{2s}} = \frac{\langle x, u_j \rangle}{\sigma_j^s}.$$

Hence,

$$\langle u_i^{(s)}, u_j^{(s)} \rangle_s = \frac{\langle u_i^{(s)}, u_j \rangle}{\sigma_j^s} = \frac{\sigma_i^s \langle u_i, u_j \rangle}{\sigma_j^s} = \delta_{ij}$$

and for $x \in H_s$,

$$x = \sum_{n=1}^{\infty} \langle x, u_n \rangle u_n = \sum_{n=1}^{\infty} \sigma_n^s \langle x, u_n^{(s)} \rangle_s \, u_n = \sum_{n=1}^{\infty} \langle x, u_n^{(s)} \rangle_s \, u_n^{(s)}.$$

Therefore, $\{u_n^{(s)} : n \in \mathbb{N}\}$ is an orthonormal basis of H_s. Also, for $x \in H_t$,

$$x = \sum_{n=1}^{\infty} \langle x, u_n \rangle u_n = \sum_{n=1}^{\infty} \sigma_n^t \langle x, u_n^{(t)} \rangle_t \, u_n = \sum_{n=1}^{\infty} \sigma_n^{t-s} \langle x, u_n^{(t)} \rangle_t \, u_n^{(s)}$$

Thus,

$$\mathcal{I}_{s,t} x = \sum_{n=1}^{\infty} \sigma_n^{t-s} \langle x, u_n^{(t)} \rangle u_n^{(s)}, \quad x \in H_t.$$

Since $\sigma_n^{t-s} \to 0$, the inclusion map $\mathcal{I}_{s,t}$ is a compact operator, and the above representation is its singular value representation. □

Example 1.6. For $s \geq 0$, recall that

$$H^s(\mathbb{R}^k) := \{f \in L^2(\mathbb{R}^k) : \int_{\mathbb{R}^k} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 d\xi < \infty\}$$

is a Hilbert space with inner product

$$\langle f, g \rangle_s := \int_{\mathbb{R}^k} (1 + |\xi|^2)^s \hat{f}(\xi) \overline{\hat{g}(\xi)} d\xi$$

and the corresponding norm

$$\|f\|_s := \left[\int_{\mathbb{R}^k} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 d\xi \right]^{1/2}.$$

For $s < 0$, $H^s(\mathbb{R}^k)$ is defined via Gelfand triple. Thus, $\{H^s(\mathbb{R}^k) : s \in \mathbb{R}\}$ is a Hilbert scale, called the *Sobolev scale*.

1.5. **Interpolation inequality in Hilbert scales.** In most of the standard Hilbert scales $\{H_s : s \in \mathbb{R}\}$ we have the inequality

$$\|u\|_s \leq \|u\|_r^{1-\lambda} \|u\|_t^{\lambda}, \quad u \in H_t,$$

whenever $r < s < t$ with $\lambda := (t - s)/(t - r)$.
The above inequality is called the **interpolation inequality** in $\{H_s : s \in \mathbb{R}\}$.

Example 1.7. Let $\{H_s\}_{s \in \mathbb{R}}$ be the Hilbert scale as in Example 1.3. Let $r < s < t$ and $x \in H_t$. Then

$$s = (1 - \lambda)r + \lambda t \quad \text{with} \quad \lambda := (t - s)/(t - r).$$

We write

$$\sum_{n=1}^{\infty} \frac{|\langle x, u_n \rangle|^2}{\sigma_n^{2s}} = \sum_{n=1}^{\infty} \Big[\frac{|\langle x, u_n \rangle|^{2(1-\lambda)}}{\sigma_n^{2(1-\lambda)r}} \Big] \Big[\frac{|\langle x, u_n \rangle|^{2\lambda}}{\sigma_n^{2\lambda t}} \Big].$$

Applying Hölder's inequality with $p = \frac{1}{1-\lambda}$, $q = \frac{1}{\lambda}$, we have

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{|\langle x,u_n\rangle|^2}{\sigma_n^{2s}} &= \sum_{n=1}^{\infty}\Big[\frac{|\langle x,u_n\rangle|^{2(1-\lambda)}}{\sigma_n^{2(1-\lambda)r}}\Big]\Big[\frac{|\langle x,u_n\rangle|^{2\lambda}}{\sigma_n^{2\lambda t}}\Big]\\ &\le \left\{\sum_{n=1}^{\infty}\frac{|\langle x,u_n\rangle|^2}{\sigma_n^{2r}}\right\}^{1-\lambda}\left\{\sum_{n=1}^{\infty}\frac{|\langle x,u_n\rangle|^2}{\sigma_n^{2t}}\right\}^{\lambda}\\ &= \left\{\|x\|_r^2\right\}^{1-\lambda}\left\{\|x\|_t^2\right\}^{\lambda}.\end{aligned}$$

Thus, the interpolation inequality $\|x\|_s \le \|x\|_r^{1-\lambda}\|x\|_t^{\lambda}$ holds for all $x \in H_t$.

Example 1.8. Consider the Sobolev scale $\{H^s(\mathbb{R}^k)\}_{s\in\mathbb{R}}$ as in Example 1.6. For $r < s < t$, we have $s = (1-\lambda)r + \lambda t$ with $\lambda := (t-s)/(t-r)$ so that, for $f \in H^t(\mathbb{R}^k)$, by Hölder's inequality, we have

$$\begin{aligned}\|f\|_s^2 &= \int_{\mathbb{R}^k}(1+|\xi|^2)^s|\hat{f}(\xi)|^2 d\xi\\ &= \int_{\mathbb{R}^k}[(1+|\xi|^2)^r|\hat{f}(\xi)|^2]^{(1-\lambda)}\,[(1+|\xi|^2)^t|\hat{f}(\xi)|^2]^{\lambda}d\xi\\ &\le \left[\int_{\mathbb{R}^k}(1+|\xi|^2)^r|\hat{f}(\xi)|^2 d\xi\right]^{1-\lambda}\left[\int_{\mathbb{R}^k}(1+|\xi|^2)^t|\hat{f}(\xi)|^2\right]^{\lambda}d\xi\\ &= \|f\|_r^{2(1-\lambda)}\|f\|_t^{2\lambda}.\end{aligned}$$

Thus, the interpolation inequality $\|f\|_s \le \|f\|_r^{1-\lambda}\|f\|_t^{\lambda}$ holds for all $f \in H^t(\mathbb{R}^k)$.

Example 1.9. Let $\{H_s\}_{s\in\mathbb{R}}$ be the Hilbert scale as in Example 1.4. In this case also, for $s = (1-\lambda)r + \lambda t$ with $\lambda := (t-s)/(t-r)$, we have for $x \in H_t$,

$$\begin{aligned}\|x\|_s^2 &= \langle L^s x, x\rangle = \int_0^{\infty}\lambda^s d\langle E_\lambda x, x\rangle\\ &= \int_0^{\infty}\lambda^{r(1-\lambda)}\lambda^{t\lambda}d\langle E_\lambda x, x\rangle\\ &\le \Big(\int_0^{\infty}\lambda^r d\langle E_\lambda x, x\rangle\Big)^{1-\lambda}\Big(\int_0^{\infty}\lambda^t d\langle E_\lambda x, x\rangle\Big)^{\lambda}\\ &= \|x\|_r^{2(1-\lambda)}\|x\|_t^{2\lambda}.\end{aligned}$$

Thus, the interpolation inequality $\|x\|_s \le \|x\|_r^{1-\lambda}\|x\|_t^{\lambda}$ holds for all $x \in H_t$.

2. Ill-Posed Operator Equations

Let X and Y be Banach spaces. For a given $y \in Y$, consider the problem of finding a solution x of the operator equation

$$F(x) = y, \tag{1}$$

where $F : D(F) \subseteq X \to Y$.

According to Hadamard [4], the above problem is said to be **well-posed** if

(1) for every $y \in Y$ there is a solution x,
(2) the solution x is unique, and
(3) the solution depends continuously on the data (y, F).

Definition 2.1. If the problem is not a well-posed problem, then it is called an **ill-posed** problem.

Operator theoretic formulation of many of the inverse problems that appear in science and engineering are ill-posed. Here are two proto-types of ill-posed problems:

(1) Fredholm integral equations of the first kind (cf. [3, 5])
  (a) Computerized tomography,
  (b) Geophysical prospecting,
  (c) Image reconstruction problems.
(2) Parameter identification problems in PDE (cf. [17, 1, 2]):
  (a) diffraction tomography,
  (b) impedance tomography,
  (c) oil reservoir simulation, and
  (d) under water hydrology.

2.1. **Fredholm integral equations of the first kind.** In this, the problem is to solve the integral equation

$$\int_\Omega k(s,t)x(t)\,dt = y(s), \quad x \in X,\ s \in \Omega,$$

where $k(\cdot,\cdot)$ is a *non-degenerate kernel* in $L^2(\Omega\times\Omega)$ and $y \in L^2(\Omega)$ with Ω a domain in $\mathbb{R}^k$ for some $k \in \mathbb{N}$. We may observe that the operator $T : L^2(\Omega) \to L^2(\Omega)$ defined by

$$(Tx)(s) = \int_\Omega k(s,t)x(t)\,dt, \quad x \in X,\ s \in \Omega,$$

is a *compact operator with non-closed range*. Thus, the problem of solving such integral equations is ill-posed.

Illustration: Let $x, y \in L^2[0,1]$ be such that $Tx = y$. Consider a perturbed data

$$y_n(s) = y(s) + \varepsilon_n(s)$$

with

$$\varepsilon_n(s) := \int_0^1 k(s,t)\sin(n\pi t)dt, \quad s \in [0,1].$$

Then $Tx_n = y_n$ with

$$x_n(t) := x(t) + \sin(n\pi t).$$

We may recall that (φ_n) with $\varphi_n(s) := \sqrt{2}\sin(n\pi s)$ is an orthonormal sequence in $L^2[0,1]$. Hence, for each $s \in [0,1]$,

$$\varepsilon_n(s) = \int_0^1 k(s,t)\sin(n\pi t)dt = \frac{1}{\sqrt{2}}\langle k(s,\cdot), \varphi_n\rangle \to 0,$$

$$|\varepsilon_n(s)|^2 \le \frac{1}{2}\Big(\int_0^1 |k(s,t)|^2 dt\Big),$$

so that

$$\int_0^1 |\varepsilon_n(s)|^2 ds \leq \frac{1}{2}\int_0^1 \Big(\int_0^1 |k(s,t)|^2 dt\Big) ds < \infty.$$

Thus, $|\varepsilon_n(s)|^2 \to 0$ for each $s \in [0,1]$ and $|\varepsilon_n(\cdot)|^2$ is integrable. Hence, by Dominated Convergence Theorem,

$$\|\varepsilon_n(\cdot)\|_2 = \int_0^1 |\varepsilon_n(s)|^2 ds \to 0 \quad \text{as} \quad n \to \infty.$$

Also,

$$\|x_n - x\|_2 = \int_0^1 \sin^2(n\pi t) dt = \frac{1}{2} \quad \forall\, n \in \mathbb{N}.$$

Thus, we arrive at the situation:

$$\|y_n - y\| \to 0 \quad \text{but} \quad \|x_n - x\| \not\to 0.$$

2.2. **Compact operator equations.** Let $T : X \to Y$ be a compact operator between Banach spaces X and Y. We may recall from Functional Analysis that if T is of infinite rank, then T is not bounded below, and hence, for every sequence (λ_n) with $\lambda_n > 0$, there exists (u_n) in X such that

$$\|Tu_n\| \leq \lambda_n \quad \text{and} \quad \|u_n\| = \frac{1}{\lambda_n} \quad \forall\, n \in \mathbb{N}.$$

In particular, if $Tx = y$ and for $n \in \mathbb{N}$ if we take

$$y_n = y + Tu_n, \quad x_n = x + u_n,$$

then $Tx_n = y_n$,

$$\|y - y_n\| \leq \lambda_n \quad \text{and} \quad \|x - x_n\| = \frac{1}{\lambda_n}.$$

In particular, if $\lambda_n \to 0$ as $n \to \infty$, then $(\|y - y_n\|)$ converges to 0 at the rate of (λ_n) whereas $(\|x - x_n\|)$ diverges to ∞ at the rate of $(1/\lambda_n)$.

Illustration: Let X and Y be Hilbert spaces and $T : X \to Y$ be a compact operator of *infinite rank*. Consider its singular value representation

$$Tx = \sum_{n=1}^{\infty} \sigma_n \langle x, u_n\rangle v_n, \quad x \in X.$$

We may recall that (σ_n) is a sequence of positive real numbers such that $\sigma_n \to 0$ as $n \to \infty$, $\{u_n : n \in \mathbb{N}\}$ and $\{v_n : n \in \mathbb{N}\}$ are orthonormal sets in X and Y, respectively.

Let $x \in X$ and $y = Tx$. For $n \in \mathbb{N}$, let

$$y_n = y + \sqrt{\sigma}_n v_n, \quad x_n = x + \frac{1}{\sqrt{\sigma}_n} u_n.$$

Then we have $Tx_n = y_n \quad \forall\, n \in \mathbb{N}$. Note that, as $n \to \infty$,

$$\|y - y_n\| = \sqrt{\sigma}_n \to 0 \quad \text{but} \quad \|x - x_n\| = \frac{1}{\sqrt{\sigma}_n} \to \infty.$$

2.3. **Parameter identification problems in PDE.** Let Ω be a bounded domain in $\mathbb{R}^n$ with *smooth* boundary and $q(\cdot) \in L^\infty(\Omega)$ be such that $q(\cdot) \geq c_0$ a.e. for some $c_0 > 0$. Then for every $f \in L^2(\Omega)$, there exists a unique $u \in H^2(\Omega)$ such that

$$-\nabla.(q(x)\nabla u) = f, \qquad u_{|\partial\Omega} = 0.$$

The map $F : q \mapsto u$ is a nonlinear operator which does not have a continuous inverse (cf. [3]). Thus, the parameter identification problem of determining the parameter functin $q(\cdot)$ from the knowledge of $u(\cdot)$ is a nonlinear ill-posed problem.

Illustration using one-dimensional formulation:

$$\frac{d}{dt}\left[q(t)\frac{du}{dt}\right] = f(t), \quad 0 < t < 1,$$

where $f \in L^2(0,1)$. Note that

$$u(t) = \int_0^t \left[\frac{1}{q(\tau)}\int_0^\tau f(s)ds\right] d\tau.$$

Thus, the problem is same as that of solving the equation

$$F(q)(t) := \int_0^t \left[\frac{1}{q(\tau)}\int_0^\tau f(s)ds\right] d\tau = u(t).$$

Clearly, this equation is nonlinear. Note also that

$$q(t) = \frac{1}{u'(t)}\int_0^t f(s)ds.$$

Suppose $u(t)$ is perturbed to $\tilde{u}(t)$, say

$$\tilde{u}(t) = u(t) + \varepsilon(t).$$

Suppose $\tilde{q}(t)$ is the corresponding solution. Then we have

$$\begin{aligned} q(t) - \tilde{q}(t) &= \left[\frac{1}{u'(t)} - \frac{1}{u'(t)+\varepsilon'(t)}\right]\int_0^t f(s)ds \\ &= \frac{1}{u'(t)}\left[\frac{\varepsilon'(t)}{u'(t)+\varepsilon'(t)}\right]\int_0^t f(s)ds. \end{aligned}$$

Hence,

$$\varepsilon'(t) \approx \infty \quad \Longrightarrow \quad |q(t) - \tilde{q}(t)| \approx |q(t)|.$$

There can be perturbations $\varepsilon(t)$ such that $\varepsilon(t) \approx 0$ but $\varepsilon'(t) \approx \infty$. For example, for large n,

$$\varepsilon_n(t) := (1/n)\sin(n^2 x) \approx 0 \quad \text{but} \quad \varepsilon_n'(t) = n\cos(n^2 x) \approx \infty.$$

Thus, the problem is ill-posed.

3. Regularization

For an ill-posed problem with an approximate data $(\tilde{y}, \tilde{F})$ in place of (y, F), one looks for a family $\{\tilde{x}_\alpha\}_{\alpha>0}$ of approximate solutions such that

(1) each $\tilde{x}_\alpha$ is a solution of a well-posed problem and

(2) $\alpha := \alpha(\tilde{y}, \tilde{F})$ is chosen in such a way that
$$\tilde{x}_\alpha \to x \quad \text{as} \quad (\tilde{y}, \tilde{F}) \to (y, F).$$

The procedure of finding such a stable approximate solution is called a **regularization method**.

3.1. **Tikhonov regularization.** Consider a linear ill-posed problem
$$Tx = y,$$
where $T : X \to Y$ is a bounded linear operator between Hilbert spaces X and Y with non closed range $R(T)$. We may recall (cf. Nair[10] or [11]):

(1) $x_0 \in X$ is an *LRN-solution* if $\quad \|Tx_0 - y\| = \inf_{x\in X} \|Tx - y\|$;

(2) An LRN-solution x_0 exists if and only if $\quad y \in R(T) + R(T)^\perp, \quad$ and in that case $\quad T^*Tx_0 = T^*y$.

(3) If $y \in R(T) + R(T)^\perp$, then there exists a unique LRN-solution $x^\dagger := T^\dagger y$ of minimal norm.

(4) One looks for stable approximations for the the minimum norm LRN-solution $x^\dagger := T^\dagger y, \quad y \in D(T^\dagger) := R(T) + R(T)^\perp$.

In *Tikhonov regularization*, the regularized solution is the unique minimizer of
$$x \mapsto \|Tx - \tilde{y}\|^2 + \alpha\|x\|^2,$$
equivalently, the unique solution of the well-posed operator equation
$$(T^*T + \alpha I)\tilde{x}_\alpha = T^*\tilde{y}.$$

It is known (cf. [11])

(1) If $y \in D(T^\dagger)$, and $\|\tilde{y} - y\| \le \delta$ for some $\delta > 0$, then the best possible error estimate is $\quad \|x^\dagger - \tilde{x}_\alpha\| = O(\delta^{2/3})$.

(2) The above estimate is order optimal for the source set
$$\{x \in X : x = (T^*T)u : \|u\| \le \rho\}$$

(3) It is attained by an a priori choice of α, namely, $\alpha \sim \delta^{2/3}$ or by the a posteriori choice of Arcangeli's method (see [7]) $\quad \|T\tilde{x}_\alpha - \tilde{y}\| = (\delta/\sqrt{\alpha})$.

3.2. **Improvement using Hilbert scales.** In order to improve the error estimate, Natterer [16] suggested a modification using a Hilbert scale $\{H_s : s \in \mathbb{R}\}$ to obtain approximations for the LRN-solution which minimizes the function
$$x \mapsto \|x\|_s.$$

It is assumed that the interpolation inequality
$$\|u\|_s \le \|u\|_r^{1-\lambda}\|u\|_t^\lambda, \quad s = (1-\lambda)r + \lambda t$$
holds. In this modification, the regularized solution is the the minimizer $\tilde{x}_{\alpha,s}$ of
$$x \mapsto \|Tx - \tilde{y}\|^2 + \alpha\|x\|_s^2.$$

Natterer[16] showed the following.

Theorem 3.1. *If T satisfies*

$$\|Tx\| \geq c\|x\|_{-a} \quad \forall\, x \in X$$

for some $a > 0$ and $c > 0$, $\hat{x} \in H_t$ where $0 \leq t \leq 2s + a$, and $\alpha \sim \delta^{\frac{2(a+s)}{a+t}}$, then

$$\|\tilde{x}_{\alpha,s} - \hat{x}\| = O(\delta^{\frac{t}{t+a}}). \tag{2}$$

According to the above theorem, higher smoothness requirement on $\hat{x}$ and with higher level of regularization gives higher order of convergence.

3.3. **Use of unbounded stabilizing operators.** Another approach is to look for an approximation of the LRN-solution which minimizes the function $x \mapsto \|Lx\|$, where $L : D(L) \subseteq X \to X$ is a closed densely defined operator. It is known (cf. [13]) that such an LRN-solution exists whenever

$$y \in R(T_{|_{D(L)}}) + R(T)^{\perp}.$$

Accordingly, one finds the unique minimizer $\tilde{x}_\alpha$ of the function

$$x \mapsto \|Tx - \tilde{y}\|^2 + \alpha\|Lx\|^2, \quad x \in D(L),$$

Existence and uniqueness of the regularized solutions $x_\alpha(\tilde{y})$ are ensured by assuming the completion condition (cf. [13, 6]):

$$\|Tx\|^2 + \|Lx\|^2 \geq \gamma\|x\|^2 \quad \forall\, x \in D(L), \tag{3}$$

The condition (3) is satisfied, if for example, L is bounded below, which is the case for many of the differential operators that appear in applications. The choice $L = I$ corresponds to ordinary Tikhonov regularization.

Also, condition (3) ensures that

(1) $\tilde{x}_\alpha$ is the unique solution of the well-posed equation $(T^*T+\alpha L^*L)x = T^*\tilde{y}$,
(2) $y \in R(T_{|_{D(L)}}) + R(T)^{\perp}$ implies $x_\alpha \to \hat{x}$ as $\alpha \to 0$, where $\hat{x}$ is the unique LRN-solution which minimizes $x \mapsto \|Lx\|$.

Suppose we have the perturbed data y^δ with $\|y - y^\delta\| \leq \delta$, and let x_α^δ be the corresponding regularized solution, i.e., $(T^*T + \alpha L^*L)x_\alpha^\delta = T^*y^\delta$. It is required to choose the regularization parameter $\alpha := \alpha(\delta, y^\delta)$ appropriately so that

$$x_\alpha^\delta \to \hat{x} \quad \text{as} \quad \delta \to 0.$$

In order to obtain error estimates, it is necessary to impose some smoothness assumptions on $\hat{x}$, by requiring it to belong to certain *source set.* This aspect has been considered extensively in the literature in recent years by assuming that the operators T, L are associated with a Hilbert scale $\{X_s\}_{s\in\mathbb{R}}$ in an appropriate manner (cf. [16, 15, 14]).

3.4. **Interplay between Hilbert scales and operator pairs.** Let us consider the following conditions on (T, L) using a Hilbert scale $\{H_s : s \in \mathbb{R}\}$:

(1) There exists $a > 0$, $c > 0$ such that

$$\|Tx\| \geq c\|x\|_{-a} \quad \forall\, x \in X. \tag{4}$$

(2) There exists $b \geq 0$, $d > 0$ such that $D(L) \subseteq X_b$ and

$$\|Lx\| \geq d\|x\|_b \quad \forall x \in D(L). \tag{5}$$

The following result is proved in Nair[8].

Theorem 3.2. *If the Hilbert scale conditions (4) and (5) are satisfied and if $\hat{x}$ belongs to the source set*

$$M_\rho = \{x \in D(L) : \|Lx\| \leq \rho\} \tag{6}$$

for some $\rho > 0$, α is chosen according to the Morozov discrepancy principle

$$c_1\delta \leq \|Tx_\alpha^\delta - y^\delta\| \leq c_0\delta \tag{7}$$

with $c_0, c_1 \geq 1$, then

$$\|\hat{x} - \tilde{x}_\alpha\| \leq 2\Big(\frac{\rho}{d}\Big)^{\frac{a}{a+b}}\Big(\frac{\delta}{c}\Big)^{\frac{b}{a+b}}. \tag{8}$$

Remark 3.1. The estimate in (8) corresponds to the estimate (2) obtained by Natterer [16] for the case $t = s = b$.

In [9], two more source sets are considered, namely,

$$\widetilde{M}_\rho = \{x \in D(L) : \|L^*Lx\| \leq \rho\}, \tag{9}$$

$$M_{\rho,\varphi} := \{x \in D(L^*L) :\; L^*Lx = [\varphi(T^*T)]^{1/2}u,\, \|u\| \leq \rho\}. \tag{10}$$

and obtained improved estimates as given in the following theorem. Here, $\varphi : [0,\infty) \to [0,\infty)$ is a strictly monotonically increasing continuous function such that $\lim_{\lambda\to 0}\varphi(\lambda) = 0$. In regularization theory, such functions are called *index functions*.

Theorem 3.3. *Suppose the Hilbert scale conditions (4) and (5) are satisfied and α is chosen according to the Morozov discrepancy principle (7).*

(i) *If $\hat{x}$ belongs to the source set $\widetilde{M}_\rho$ defined in (9), then*

$$\|\hat{x} - \tilde{x}_\alpha\| \leq 2\Big(\frac{\rho}{d^2}\Big)^{\frac{a}{2a+b}}\Big(\frac{\delta}{c}\Big)^{\frac{2b}{a+2b}}.$$

(ii) *If $\hat{x}$ belongs to the source set $M_{\rho,\varphi}$ defined in (10), then*

$$\|\hat{x} - \tilde{x}_\alpha\| = (1 + c_0)\left(\frac{\rho}{d^2}\right)^{\frac{a}{a+2b}}\left(\frac{\delta}{c}\right)^{\frac{2b}{a+2b}}\left[\psi_p^{-1}\left(\varepsilon_\delta^2\right)\right]^{\frac{a}{2(a+2b)}}$$

where

$$p = \frac{a}{a+2b}, \quad \varepsilon_\delta := c\left(\frac{d^2\delta}{c\rho}\right)^{\frac{a}{a+2b}}, \quad \psi_p(\lambda) := \lambda^{1/p}\varphi^{-1}(\lambda).$$

3.5. **A unified consideration.** In a recent paper [12] a unified approach is adopted by replacing the Hilbert scale conditions (4) and (5) by a single condition involving T and L:

θ-condition: There exist $\eta > 0$ and $0 \le \theta < 1$ such that

$$\eta\|x\| \le \|Tx\|^{\theta}\|Lx\|^{1-\theta} \quad \forall\, x \in D(L). \tag{11}$$

Observe that

(1) $\theta = 0$ corresponds to L being bounded below so that $R(L)$ is closed and $L^{-1} : R(L) \to X$ is a bounded operator. This case also includes the choice $L = I$, the identity operator.

(2) $\theta = 1$ is excluded, as it would imply that T has a continuous inverse.

Theorem 3.4. *The following results hold.*

(1) *The θ-condition (11) implies the completion condition (3) with $\gamma = \eta^2$;*

(2) *Hilbert scales conditions (4) and (5) imply θ-condition (11) with $\theta = \frac{b}{a+b}$ and $\eta = c^{\theta}d^{1-\theta}$.*

Proof. (1). Suppose the θ-condition (11) is satisfied. From the relation $ab \le \frac{a^p}{p} + \frac{b^q}{q}$ with

$$a = \|Tx\|^{2\theta}2, \quad b = \|Lx\|^{2(1-\theta)}, \quad p = \frac{1}{\theta}, \quad q = \frac{1}{1-\theta},$$

we obtain

$$\|Tx\|^{2\theta}\|Lx\|^{2(1-\theta)} \le \theta\|Tx\|^2 + (1-\theta)\|Lx\|^2 \le \|Tx\|^2 + \|Lx\|^2.$$

Thus $\quad \eta^2\|x\|^2 \le \|Tx\|^2 + \|Lx\|^2$.

(2). Suppose the Hilbert scales conditions (4) and (5) are satisfied, i.e.,

$$\|Tx\| \ge c\|x\|_{-a}, \quad \|Lx\| \ge d\|x\|_b.$$

Then $\|x\|_0 \le \|x\|_{-a}^{\theta}\|x\|_b^{1-\theta}$, where θ is such that $0 = (1-\theta)(-a) + \theta b$, that is, $\theta = a/(a+b)$. Thus,

$$\|x\|_0 \le (\|Tx\|/c)^{\theta}(\|Lx\|/d)^{1-\theta},$$

so that $\quad \eta\|x\| \le \|Tx\|^{\theta}\|Lx\|^{1-\theta} \quad$ with $\quad \eta = c^{\theta}d^{1-\theta}$. □

The following theorem in [12] unifies results in the setting of general unbounded stabilizing operator as well as for Hilbert-scale and Hilbert-scale-free settings.

Theorem 3.5. *Under the θ-condition (11) and the discrepancy principle (7), the following hold.*

(i) *If* $\hat{x} \in M_\rho$, *then* $\|\hat{x} - x^{\delta}_{\alpha_\delta}\| \le 2\eta^{-1}\rho^{1-\theta}\delta^{\theta}$.

(ii) *If* $\hat{x} \in \widetilde{M}_\rho$, *then*

$$\|\hat{x} - x^{\delta}_{\alpha_\delta}\| \le (1 + c_0)\left(\frac{1}{\eta^2}\right)^{\frac{1}{1+\theta}} \rho^{\frac{1-\theta}{1+\theta}}\,\delta^{\frac{2\theta}{1+\theta}}.$$

(iii) *If* $\hat{x} \in M_{\rho,\varphi}$ *and* $\delta^2 \le \gamma_1^2 \varphi(1)$ *where* $\gamma_1 = 4\rho/\gamma$, *then*

$$\|\hat{x} - x^\delta_{\alpha_\delta}\| \le (1 + c_0) \left(\frac{1}{\eta^2}\right)^{\frac{1}{1+\theta}} \rho^{\frac{1-\theta}{1+\theta}} \delta^{\frac{2\theta}{1+\theta}} [\psi_p^{-1}(\varepsilon_\delta^2)]^{1/2p},$$

where $p := \frac{1+\theta}{1-\theta}$, $\varepsilon_\delta := \eta^{\frac{2}{1+\theta}} \left(\frac{\delta}{\rho}\right)^{1/p}$.

Remark 3.2. The results in Theorem 3.2 and Theorem 3.3 are recovered from Theorem 3.5, (i) and (ii), respectively, by taking

$$\theta = \frac{b}{a+b}, \quad \eta = c^{\frac{b}{a+b}} d^{\frac{a}{a+b}}.$$

In the ordinary Tikhonov regularization, that is, for the case of $L = I$, equivalently, $\theta = 0$ in part (iii) of Theorem 3.5, we have $p = 1$ and $\varepsilon_\delta = \eta^2 \left(\frac{\delta}{\rho}\right)$. Hence, the estimate reduces to

$$\|\hat{x} - x^\delta_{\alpha_\delta}\| \le (1 + c_0) \left(\frac{1}{\eta^2}\right) \sqrt{\psi_1^{-1}(\eta^2\delta^2/\rho^2)}.$$

Thus, we recover the error estimate for ordinary Tikhonov regularization under the general source condition obtained in [15].

References

[1] Borcea, L., Berryman, J. G. and Papanicolaou, G. C., High-contrast impedance tomography, *Inverse Problems*, **12** (1996).

[2] Devaney, A. J., The limited-view problem in diffraction tomography, *Inverse Problems*, **5**(5)(1989).

[3] Engl, H. W., Hanke, M. and Neubauer, A., *Regularization of Inverse Problems*, Kluwer, Dordrecht, 2003.

[4] Hadamard, J., *Lectures on the Cauchy Problem in Linear partial Differential Equations*, Yale University Press, 1923

[5] Groetsch, C. W., *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Boston-London-Melbourne: Pitman Pub. 1984.

[6] Locker, J. and Prenter, P. M., Regularization with differential operators, *Math. Anal. Appl.*, **74**(1980), 504–529.

[7] Nair, M. T., A generalization of Arcangeli's method for ill-posed problems leading to optimal rates, *Integral Equations and Operator Theory*, **15** (1992), 1042–1046.

[8] Nair, M. T., On Morozov's method Tikhonov regularization as an optimal order yielding algorithm, *Zeit. Analysis und ihre Anwendungen*, **18** (no.1) (1999), 37–46.

[9] Nair, M. T., On improving estimates for Tikhonov regularization using an unbounded operator, *J. Analysis*, **14** (2006), 143–157.

[10] Nair, M. T., *Functional Analysis: A First Course*, PHI Learning, New Delhi 2002 (Fourth Print: 2014).

[11] Nair, M. T., *Linear Operator Equations: Approximation and Regularization*, World Scientific, Singapore, New York, 2009

[12] Nair, M. T., A Unified Treatment for Tikhonov Regularization Using a General Stabilizing Operator, *Analysis and Applications*, 13 (02)(2015), 201–215.

[13] Nair, M. T., Hegland, M. and Anderssen, R. S., The trade-off between regularity and stability in Tikhonov regularization, *Math. Comp.*, **66**(1997), 193–206.

[14] Nair, M. T., Pereverzyev, M. S. and Tautenhahn, U., Regularization in Hilbert scales under general smoothing conditions, *Inverse Problems*, **21**(2003), 1851–1869.

[15] Nair, M. T., Schock, M. E. and Tautenhahn, U., Morozov's discrepancy principle under general source conditions, *Zeit. Analysis und ihre Anwendungen*, **22**(2003), 199–214.

[16] Natterer, F., Error bounds for Tikhonov regularization in Hilbert scales, *Applicable Analysis*, **18**(1984), 29–37.

[17] Parker, R. L., *Geophysical Inverse Theory*, Princeton University Press, Princeton NJ, 1994.

M. T. Nair
Department of Mathematics
I I T Madras, Chennai-600036, India
E-Mail: *mtnair@iitm.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 63-78

LIKELIHOOD, ESTIMATING FUNCTIONS AND MEHOD OF MOMENTS*

UTTARA NAIK NIMBALKAR

Abstract. The likelihood theory is the most central topic in statistical modeling and inference. In this article, we briefly review the method of maximum likelihood estimation attributed to Fisher. We next discuss the theory of estimating functions introduced by Godambe and Durbin in order to overcome the limitations of specifying the functional form of the likelihood while modeling a random phenomenon. The estimating function theory combines the strengths of the least squares approach of Legendre and Gauss and the likelihood approach. We consider the generalized method of moments established by Hansen for applications in econometrics, for which he got, in part, the 2013 Nobel prize in economics. We show the connection between the estimating function and the generalized method of moments approach. We conclude with a discussion of our work on the estimating function approach for 'state-space models' and for Hilbert space valued parameters.

This article is a brief overview of three important methods in statistics and is meant for those not too familiar with statistical inference.

1. Introduction

In statistical applications, one formulates a statistical model for representing the empirical random phenomenon being studied. One identifies the family of probability distributions which can describe the data we observe, that is, based on a sample of observations one formulates a probability distribution for the population. For example, the random variables denoting measurement errors are usually modeled by the normal distribution, the number of accidents at a point by a Poisson distribution, life lengths of units by either a Weibull or a Gamma distribution and so on. If the data are from continuous random variables, that is those that can take any values within a range, a histogram is drawn, and if the data are from discrete random variables, that is those that can take only certain values, finite

* The (modified) text of the 26^{th} V. Ramaswami Aiyar Memorial Award Lecture delivered at the 81^{st} Annual Conference of the Indian Mathematical Society held at the Visvesvaraya National Institute of Technology, Nagpur - 440 010, Maharashtra, during the period December 27-30, 2015.

2010 Mathematics Subject Classification:6201, 62F10
Key words and phrases: Cramér-Rao inequality, estimating equation, maximum likelihood estimation, method of moments

or countable, like the number of runs in a cricket match, a bar chart is formed. Based on the shape of the histogram (or the bar chart), a probability density (a probability mass function) is assumed for the random variables whose values form the observed data. For example, in the continuous case, if the histogram is symmetric about a point and is bell-shaped, the normal distribution is assumed for the random variable, whereas if it is not bell-shaped, then a different distribution has to be assumed. The corresponding cumulative distribution function (cdf) determines the probability that the random variable will assume values in specified intervals. A distribution function usually depends on one or many unknown parameters, which are related to the biological, physical, or other type of system that governs the observed phenomenon or to the domain of application. Based on the observations, these parameters are to be estimated, that is numerical values are obtained for them so that one can obtain numerical values for certain probabilities.

In this article, we consider parametric inference on the basis of the observed data $\{x_1, \cdots, x_n\}$, which are realized values of random variables $\{X_1, \cdots, X_n\}$. It is assumed that a parametric statistical model $F(\underline{x}, \theta)$ specifies the joint distribution of $\{X_1, \cdots, X_n\}$. The interest is in making inference about the parameter θ belonging to some subset Θ of the p dimensional Euclidean space R^p, $p < n$. The subset Θ, is called the parameter space.

Once a parametric statistical model is formulated for the observed data, statistical inference consists mainly of procedures for

(1) obtaining point estimate(s) of the parameter(s) θ, that is, for estimating the 'true' parameter,
(2) obtaining confidence interval estimates for θ with level of confidence $(1 - \alpha)\%$ (= 90%, 95%, 99%). An interval with confidence level 95% means that if the experiment is performed a large number of times, 95% of the interval estimates will contain the true parameter.
(3) Statistical Hypothesis Testing: test and conclude with given level of significance or confidence whether the true parameter belongs to a specified set.
(4) Model selection: test whether the specified statistical model fits the data or some other model would be better.

The likelihood approach, attributed to Fisher, is the most predominant approach in parametric statistical inference and is used for all the four problems stated above. We discuss this approach briefly in the next section but restrict only to the estimation problem. In the section 3, we consider the estimation procedure based on the estimating functions and in the section 4, the generalized method of moments. In the section 5, we describe the application of the estimating function approach to state-space models. In the section 6, we give an extension of the estimating function approach for parameters taking values in an infinite dimensional Hilbert space and conclude with some remarks in the section 7.

2. Likelihood $L(\theta|\underline{x})$

Suppose the random vector $\underline{X}_n = (X_1, \cdots, X_n)$ has the joint probability density function (pdf) (probability mass function (pmf)) $f(\underline{x}_n; \theta)$, where $\underline{x}_n = (x_1, \cdots, x_n)$. Here θ denotes the parameter (which could be a scalar or a vector) and let Θ denote the parameter space. The likelihood is defined as follows.

Definition 2.1. For a given $\underline{x}_n$, the likelihood is a function from Θ to $[0, \infty)$ defined by

$$L(\theta|\underline{x}_n) = f(\underline{x}_n; \theta).$$

If the data $x_1, \cdots, x_n$ are observations on independent and identically distributed (iid) random variables, then the likelihood is

$$L(\theta|\underline{x}_n) = \prod_{i=1}^{n} f(x_i; \theta),$$

where $f(x; \theta)$ is the common pdf (pmf).

Example 2.1: Suppose one tosses a coin 20 times and observes 15 Heads. We want to estimate the probability θ of observing a Head. In this case we can assume that the observations $x_1, x_2, \cdots, x_{20}$, are from a Bernoulli distribution with parameter θ, where $x_i = 1$ if on the $i-$th toss a Head is observed and $x_i = 0$ if a Tail is observed. The likelihood is

$$L(\theta|\underline{x}_{20}) = P_\theta(X = 15) = \begin{pmatrix} 20 \\ 15 \end{pmatrix} \theta^{15}(1-\theta)^5.$$

To obtain an estimate of the parameter θ, one maximizes the likelihood.

Definition 2.2. The maximum likelihood estimate (mle) $\hat{\theta}$ of θ is that value of θ that maximizes the likelihood function $L(\theta|\underline{x}_n)$.

In the above Example 2.1, the mle $\hat{\theta}$ of θ is $\hat{\theta} = 15/20$.

We note that in the discrete set up, if the statistical model is parametrized by a fixed and unknown θ, the likelihood $L(\theta|\underline{x})$ is the probability of the observed data $\underline{x}$ considered as a function of θ. Thus the θ that maximizes the likelihood is the value of θ for which the observed value $\underline{x}$ is most likely.

However in the continuous set up, the likelihood is not a probability. In fact in this case the probability of observing any point x is zero. One may consider the following approximation and think of it as an infinitesimal probability. For small ϵ,

$$L(\theta|x) = f(x; \theta) \approx \int_{x-\epsilon}^{x+\epsilon} f(u; \theta) du = P(x - \epsilon < X < x + \epsilon).$$

Example 2.2: Suppose $x_1, \cdots, x_n$ are observations from iid random variables with the common distribution being normal with mean μ and variance σ^2. Then the likelihood for μ and σ^2 is

$$L(\mu, \sigma^2|\underline{x}_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x_i-\theta)^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2),$$

and the mle of μ and σ^2 are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

In the likelihood approach, the full form of the joint pdf (pmf) is specified. Just for the sake of examples, we give below the functional forms of some common pdfs (pmfs) $f(x;\theta)$.

(1) Poisson(λ): This distribution is used to model some counting random variables such as the number of accidents in a fixed interval, or the number of incoming telephone calls, the number of incoming packages over the internet traffic, etc. The parameter λ denotes the rate of occurrence of the events, i.e., the average number per unit of time.

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!};\ k = 0, 1, \cdots ; \lambda > 0.$$

The parameter space $\Theta = (0, \infty)$.

(2) *Gamma density* ($Gamma(\alpha, \beta)$)

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x};\ x > 0, \alpha > 0, \beta > 0.$$

The parameters α and β are respectively referred to as the shape and the scale parameters and the parameter space $\Theta = (0, \infty) \times (0, \infty)$.

(3) *Weibull density* (used to model the life lengths of items)

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}; x \geq 0, \alpha > 0, \beta > 0.$$

The parameters α and β are respectively referred to as the shape and the scale parameters and the parameter space $\Theta = (0, \infty) \times (0, \infty)$.

We note that an estimate of a parameter θ is some function $\hat{\theta}(\underline{x}_n)$ of the data. For a different set of observed values on the random variables, that is, for another sample, a different value of the estimate is obtained. The corresponding random function $\hat{\theta}(\underline{X}_n)$ is referred to as an 'estimator' of θ. In Example 2.1, $\hat{\theta}(\underline{X}_{20}) = \frac{1}{20}\sum_{i=1}^{20} X_i$ and similarly in Example 2.2, the estimator of the mean μ is $\hat{\mu}(\underline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$, and that of the variance σ^2 is $\sigma^2(\underline{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}(\underline{X}_n))^2$.

Besides the maximum likelihood approach, there are other methods of obtaining estimators of a parameter θ. The quality of the estimator is decided from the properties defined below. Let T_n denote an estimator of θ based on a sample of size n.

Definition 2.3. An estimator T_n is said to be an unbiased estimator of θ if $E_\theta(T_n) = \theta$ for all $\theta \in \Theta$. The bias of an estimator T_n is defined as $Bias_\theta(T_n) = E_\theta(T_n) - \theta$.

Definition 2.4. The mean squared error (MSE) of an estimator T_n of θ is a function of θ defined by $MSE_\theta(T_n) = E_\theta(T_n - \theta)^2$.

The MSE can be expressed as $MSE_\theta(T_n) = Variance_\theta(T_n) + (Bias_\theta(T_n))^2$. A good estimator of θ should have small bias and small variance, or should be unbiased and have minimum variance. An estimator which has the smaller MSE for all $\theta \in \Theta$ is preferable. A good estimate should be very close to the true value. An estimate is a value of the estimator, hence all possible values of the estimator should be close to the true value of θ.

The above measures of performance are considered if the sample size n is small. For large n, the desired properties are of consistency and asymptotic normality.

Definition 2.5. An estimator T_n of θ is said to be consistent if as $n \to \infty$, T_n converges to θ in probability for all $\theta \in \Theta$. That is for all θ, and all $\epsilon > 0$, $lim_{n\to\infty} P_\theta[|T_n - \theta| \leq \epsilon] = 1$

A consistent estimator is close to the true value for large n.

Definition 2.6. An estimator T_n is said to be asymptotically normal if the distribution of $\sqrt{n}(T_n - \theta)$ converges to the distribution of a Normal mean zero random variable as $n \to \infty$.

Under the regularity conditions given below, the maximum Likelihood (ML) estimator is consistent, asymptotically unbiased and asymptotically normal.

The Regularity Conditions:

(1) The parameter space Θ is an open subset of the Euclidean space R (R^p.)
(2) The support $\{x|f(x;\theta) > 0\}$ of the pdf (pmf) does not depend on θ.
(3) The derivatives $\partial f(x;\theta)/\partial\theta$ and $\partial^2 f(x;\theta)/\partial\theta^2$ exist for all $\theta \in \Theta$ and almost all x.
(4) $\int f(x;\theta)dx$ and $\int(\partial f(x;\theta)/\partial\theta) f(x;\theta)dx$ are differentiable under the integral sign.
(5) $E_\theta[(\partial f(X;\theta)/\partial\theta)^2] > 0$, for all $\theta \in \Theta$.
(6) The third order derivative $\partial^3 f(x;\theta)/\partial\theta^3$ exists for every x and is continuous in θ.
(7) For every $\theta_0 \in \Theta$, $\exists$ a positive number c and a function $M(x)$ (both of which may depend on θ_0) such that $|\partial^3 f(x;\theta)/\partial\theta^3| \leq M(x)$ for all x, and $\theta \in (\theta_0 - c, \theta_0 + c)$ with $E_{\theta_0}|M(X)| < \infty$.

Most standard pdfs (pmfs) satisfy these conditions.

The mle is mostly obtained by maximizing the log of the likelihood, that is by solving the equation

$$\frac{\partial}{\partial\theta} logL(\theta|\underline{x}) = 0 \ for \ \theta.$$

Definition 2.7. The function $\frac{\partial}{\partial\theta} logL(\theta|\underline{x})$ is called the score function.

Under the regularity conditions 1, 2, and conditions 3 and 4 for the derivative of order 1, $E_\theta\left[\frac{\partial}{\partial\theta} logL(\theta|\underline{X})\right] = 0$, for all θ.

Under the regularity conditions, the asymptotic variance of the ML estimator is given by the reciprocal of the Fisher information $I(\theta)$, which is defined as

Definition 2.8. The Fisher information $I(\theta) = E_\theta\left[\left(\frac{\partial logL(\theta|\underline{X})}{\partial\theta}\right)^2\right]$.

Under the regularity conditions 1 to 4, $I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2} logL(\theta|\underline{X})\right]$. Thus an estimate of the asymptotic variance of the ML estimator $\hat{\theta}(\underline{X}_n)$, is given by the reciprocal of the observed Fisher information $I(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2} logL(\theta|\underline{x}_n)|_{\theta=\hat{\theta}(\underline{x}_n)}$. The Cramér-Rao inequality gives a lower bound for the variance of an estimator T_n of θ, and holds under the regularity conditions 1 to 4. The inequality states that

$$Variance_\theta(T_n) \geq \frac{\left(\frac{\partial}{\partial\theta} E_\theta[T_n]\right)^2}{I(\theta)},$$

for all θ. An estimator whose variance attains the Cramér-Rao lower bound is said to be an 'optimal' estimator for θ. We note that if the estimator is unbiased for θ, then the numerator of the right hand side of the inequality reduces to 1. The ML estimator is asymptotically unbiased, asymptotically 'optimal', consistent and asymptotically normal (if the regularity conditions hold.) This is the main justification of the ML approach.

One limitation of the ML method is that the full form of the pdf (pmf) has to be specified, only the parameters are unknown. It is sometimes not possible to specify the entire cdf. In some cases, like the stable distributions, the pdf is not analytically expressible except for some parameter values. ML estimation method will not work, if the likelihood function is unbounded. Further, likelihood based inference is not possible if the number of parameters increase with the sample size, like in the 'state-space' models described in the section 5, and if the data is high dimensional like the gene expression data.

We have given only a brief description of the estimation method and mostly for a scalar parameter. For vector valued parameters, for the proofs of the results stated and the likelihood's fundamental role in inference (likelihood ratio tests, likelihood based confidence intervals, etc.) we refer to [19] and [3]. For the historical development of maximum likelihood see [22] and references therein.

3. Estimating Functions

The least squares method of estimation was invented by Gauss in 1795 to solve problems in astronomical studies and later Legendre independently invented the method and published his results in 1806. For this method, assumptions on only the first two moments of the random variable are made. However the resulting estimators are (asymptotically) 'optimal' only when the underlying distribution is normal. Whereas the ML estimators are generally (asymptotically) 'optimal'. The estimating function approach introduced by [5] and [7] combines the strengths of the least squares approach and the likelihood approach.

Similar to the least squares approach, the Estimating Functions approach does not require the full form of the distribution to be specified but only assumptions regarding certain moments like the mean and the variance are made.

Definition 3.1. An **Estimating Function** $g(\underline{x}, \theta)$ is a function of the sample $\underline{x}$ and the parameters θ.

For estimation, an unbiased estimating function is used, which is defined as:

Definition 3.2. An estimating function $g(\underline{x}, \theta)$ is said to be an **Unbiased Estimating Function (EF)** if $E_\theta[g(\underline{X}, \theta)] = 0$, for all $\theta \in \Theta$.

Besides the unbiased property, the EF should satisfy the regularity conditions given below.

Regularity conditions:

(1) For almost all x, $\partial g/\partial\theta$ exists for all $\theta \in \Theta$.

(2) $\int g(x, \theta) f(x, \theta) dx$ is differentiable under the integral sign for all $\theta \in \Theta$.

(3) $E_\theta\left[\frac{\partial}{\partial\theta} g(\underline{X}, \theta)\right] \neq 0$, for all $\theta \in \Theta$.

An estimator of θ is obtained by solving $g(\underline{X}, \theta)] = 0$. The score function $\frac{\partial}{\partial\theta} logL(\theta|\underline{x})$ is an estimating function and under the regularity conditions 1, 2, and conditions 3 and 4 for the derivative of order 1, stated in the section 2, it is an unbiased estimating function.

The EF is usually suggested by the model itself, and one can obtain a class of EFs. From a given class of EFs, one chooses an 'optimal' EF for the estimation purpose. Below we give the definition of an 'Optimal Estimating Function'.

Let $\mathcal{G}$ be a class of EFs.

Definition 3.3. An estimating function g^* in $\mathcal{G}$ is said to be optimal, if

$$\frac{E_\theta[(g^*)^2]}{(E_\theta[\partial g^*/\partial\theta])^2} \leq \frac{E_\theta[g^2]}{(E_\theta[\partial g/\partial\theta])^2}$$

for every $g \in \mathcal{G}$ and for every $\theta \in \Theta$.

In [7] it is shown that if the score function belongs to the class $\mathcal{G}$, then it is the optimal estimating function in that class. The above optimality criterion is equivalent to the following criterion based on the square distances.

An equivalent criterion: An estimating function g^* in $\mathcal{G}$ is optimal if and only if

$$E_\theta[(g^* - \frac{\partial}{\partial\theta} logL(\theta|\underline{X}))^2] \leq E_\theta[(g - \frac{\partial}{\partial\theta} logL(\theta|\underline{X}))^2]$$

for every $g \in \mathcal{G}$ and for every $\theta \in \Theta$.

The following extended Cramér-Rao type inequality was proved in [13] under the regularity conditions. This inequality shows that the score function is 'optimal'.

Cramér-Rao type inequality:

$$\frac{E_\theta[(g)^2]}{(E_\theta[\partial g/\partial\theta])^2} \geq \frac{1}{I(\theta)},$$

for every $g \in \mathcal{G}$ and for every $\theta \in \Theta$.

Example 3.1: Let $X_1, \cdots, X_n$ be independent random variables with $E_\theta[X_i] = \theta$ and $Var_\theta(X_i) = \sigma^2$ for each i.

Consider a class of unbiased estimating functions (for θ),

$$\mathcal{G} = \{g = \sum_{i=1}^{n} b_i(X_i - \theta) | b_i, i = 1, \cdots, n, \text{ are real numbers and } \sum_{i=1}^{n} b_i \neq 0\}.$$

According to the Definition (3.3), the optimal estimating function in this class is $g^* = \sum_{i=1}^n (X_i - \theta)$ and one solves the equation $\sum_{i=1}^n (X_i - \theta) = 0$ to obtain an estimator of θ. We note that this equation is the same as the one obtained while minimizing $\sum_{i=1}^n (X_i - \theta)^2$ to get the least square estimator of θ. The estimator is $\hat{\theta} = \sum_{i=1}^n X_i/n$, which is the sample mean.

We note that the Gauss-Markov theorem states that the sample mean is the linear unbiased minimum variance estimator of θ. The 'unbiased minimum variance' property of the estimator fails if θ is replaced by some non-linear function of θ. For a discussion on this and the failure of the least squares approach, we refer to ([9], Chapter 1).

The EF method is very useful when the observations are dependent or from a stochastic process [8]. Let $\{X_t,\ t = 0, 1, \cdots\}$ be a discrete time stochastic process whose probability distribution P_θ depends on a parameter θ taking values in an open subset Θ of the real line $\mathbf{R}$ (or $\mathbf{R^p}$). We assume that each $(\Omega, \mathcal{F}, P_\theta)$ is a complete probability space. Let $\mathcal{F}_i = \sigma\{X_j,\ j = 0, 1, \cdots, i\}$, the $\sigma-$field generated by the random variables $\{X_j,\ j = 0, 1, \cdots, i\}$. We need not specify the form of P_θ in order to estimate θ. Suppose there exist (measurable) functions $h_i = h_i(X_1, \cdots, X_i; \theta)$ such that the conditional expectations

$$E_\theta[h_i(X_1, \cdots, X_i; \theta)|\mathcal{F}_{i-1}] = 0,\ i = 1, \cdots, T,\ \forall\theta \in \Theta.$$

Consider a class of 'linear' (martingale) estimating functions

$$\mathcal{G} = \{g = \sum_{i=1}^{T} a_{i-1} h_i\}, \tag{1}$$

where the coefficients $a_{i-1}$are $\mathcal{F}_{i-1}$ measurable functions of $X_1, \cdots, X_{i-1}$ and θ, $i = 1, \cdots, T$.

The optimal estimating function in the class $\mathcal{G}$ is given by

$$g_T^* = \sum_{i=1}^{T} \frac{E_\theta[\partial h_i/\partial\theta|\mathcal{F}_{i-1}]}{E_\theta[h_i^2|\mathcal{F}_{i-1}]} h_i. \tag{2}$$

The class $\mathcal{G}$ of EFs is often obtained by combining basic (unbiased) EFs suggested by the model.

Example 3.2 : Autoregressive Process (AR(1)): Let $\{X_t,\ t = 1, 2, ...\}$ be a stochastic process such that

$$X_t = \theta X_{t-1} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated random variables with mean zero and finite variance σ^2. Based on the observations on the X_t's, $t = 1 \cdots, T$, the aim is to estimate the parameter θ. Then the basic estimating functions are $h_i = X_i - \theta X_{i-1}$, $i = 2, \cdots, T$. In this example $E_\theta[\partial h_i/\partial\theta|\mathcal{F}_{i-1}] = X_{i-1}$ and $E_\theta[h_i^2|\mathcal{F}_{i-1}] = \sigma^2$. From (2), the optimal EF in the class defined in (1) is given by

$$g_T^* = \sum_{i=2}^{T} \frac{X_{i-1}}{\sigma^2}(X_i - \theta X_{i-1}),$$

and the estimator $\hat{\theta} = (\sum_{i=2}^{T} X_i X_{i-1})/(\sum_{i=2}^{T} X_{i-1}^2)$.

The above techniques have been extended for inference in continuous time stochastic processes, see [12] and references therein.

Under appropriate regularity conditions and for the proper choice of the class of estimating functions, the estimator obtained by solving the optimal estimating function, is consistent and asymptotically normal. The optimal EF leads to minimum size asymptotic confidence zones. For the proofs of these results and for the use of EFs in hypothesis testing we refer to[12].

The above discussion is given in terms of a scalar parameter but the EF approach holds for a vector valued parameter also. For further details and applications, we refer to [2], [9], [12] and [18].

4. Generalized method of moment (GMM)

Karl Pearson [20] introduced the Method of Moments (MM) to estimate the unknown parameters. In this method the sample moments are equated to the population moments and the equations solved for the unknown parameters to obtain their estimates. The number of equations taken is same as the number of parameters. The GMM was developed in [11] for analyzing models in econometrics. For this method, the full functional form of the distribution function need not be known.

Suppose we have observations on the random variables $\underline{X}_n$ and the parameter θ is a p-dimensional vector. Let $T_n = T_n(\underline{X}_n)$ be a k-dimensional vector of summary statistics, $k \geq p$ and $E_\theta[T_n] = \tau(\theta)$. Let $G_n(\underline{X}_n, \theta) = T_n - \tau(\theta)$, that is, $E_\theta[G(\underline{X}_n, \theta)] = 0$, for all θ. Thus $G_n(\underline{X}_n, \theta)$ is an EF. For the GMM, an estimator of θ is the value of θ that minimizes

$$G_n(\underline{X}_n, \theta)' W_n G_n(\underline{X}_n, \theta), \tag{3}$$

where W_n is a $k \times k$ matrix that may depend on the data (but not on θ) and which converges in probability to a positive definite matrix W, as $n \to \infty$.

The W_n's are called the weighting matrices. The above is like the weighted least squares, which involves minimizing (*w.r.t.* θ) the sum $\{\sum_{i=1}^{n} a_i(x_i - \theta)^2\}$, for known a_i's.

The minimizer of (3) satisfies the estimating equation

$$g(\underline{X}_n, \theta) = (-\frac{\partial \tau}{\partial \theta})' W_n G_n(\underline{X}_n, \theta) = 0,$$

assuming the existence of the derivatives involved. The estimating function will be unbiased for the appropriate choice of W_n. This approach does not necessarily use an optimal EF.

The consistency and asymptotic normality of a GMM estimator hold under certain conditions, see [11].

The choice of W_n affects the properties of the estimators in terms of their variances. An optimal estimator in a class is one that has a minimum asymptotic covariance, in the matrix sense, in that class. The optimal choice for W_n is S^{-1}, where S^{-1} is the inverse of the limiting (as $n \to \infty$) covariance matrix of $G_n(\underline{X}_n, \theta)$. However, in practice S^{-1} is not known and has to be estimated.

The GMM is used for the hypothesis testing problems as well. For a good coverage on this topic we refer to the book by [10].

5. Filtering and Smoothing Via EFs

We discuss the applications of the EF approach in filtering and smoothing from [17]. A State Space Model is specified by two equations, the observational equation and the state equation as follows:

Observational Equation : $$f_{t1}(Y_t, Y_{t-1}, X_t) = \epsilon_t, \tag{4}$$

State Equation : $$f_{t2}(X_t, X_{t-1}) = \delta_t, \tag{5}$$

where $\{\epsilon_t\}$ and $\{\delta_t\}$ are appropriate sequences of random variables. Appropriate initial conditions are assumed.

Only y_t's, realized values of the random variables Y_t's are observed for $t = 1, \cdots, T$ and the problem is to estimate the corresponding states, that is, the x_t's, which are the realized values of the X_t's. The terms filtering and smoothing are defined as follows.

Filtering: Estimate the state x_t at time t, given the observations $y_1, \cdots, y_t$, up to time $t, t = 1, 2, \cdots, T$.

Smoothing: Estimate all the states $(x_1, \cdots x_T)$ based on the observations $y_1, \cdots, y_T$.

The state space models have origins in systems theory and engineering. The state X_t denotes a measurement such as the position or velocity, etc. of a spacecraft at time t. The state equation in this set up is an approximation to physical laws of motion. The observation Y_t is a radar observation (signal) on X_t which is contaminated by noise (error) denoted by ϵ_t. The observational equation is

$Y_t - \alpha_t X_t = \epsilon_t$, and the state equation can be $X_t - \mu - \beta_t(X_{t-1} - \mu) = \delta_t$. The non random quantities $\mu,\ \alpha_t$ and β_t are usually assumed to be known.

In the literature, filtering and smoothing were discussed under the assumption that the $\{\epsilon_t\}$ and $\{\delta_t\}$ are independent sequences of independent random variables and with known distributions, frequently Gaussian. Under these assumptions, the well known *Kalman filter* [14] and *Kalman-Bucy filter* [15] are used to obtain x_t from the observations.

In these models, the quantities to be estimated are random, which was not the case in the 'frequentist' approach described in the earlier sections. It is more in line with a Bayesian set up.

In practice the $\{\epsilon_t\}$ and $\{\delta_t\}$ need not be sequences of independent random variables and the form of the distribution functions may not be fully specified nor be Gaussian. Below we consider two such examples of state-space models.

Example 5.1. Reliability/Life testing: The model given in this example was proposed in [4] for tracking software reliability growth. The observational model is: Y_t given X_t has the $Gamma(\nu, X_t)$ distribution and the state model is: CX_t/X_{t-1} given X_{t-1} has the $Beta(\sigma, \nu)$ distribution. The starting distribution of X_0 is $Gamma(\sigma + \nu, u_0)$ and the nonrandom quantities $C,\ \sigma,\ \nu$ and u_0 are assumed known.

Example 5.2. This model was used in [16] to model rainfall data. The variable Y_t equals 1, if there is an occurrence of rainfall on day t and equals 0 if there is no rainfall on day t. The observational model is specified such that Y_t given X_t is $Bernoulli(\pi_t)$ with $\pi_t = e^{X_t}/(1 + e^{X_t})$. The state equation is $X_t = X_{t-1} + \delta_t$. The aim is to estimate π_t.

The processes $\{Y_t,\ t = 1, 2, ...\}$ and $\{X_t\ t = 0, 1, 2, ...\}$ are defined on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\mathcal{F}_{t1}$ and $\mathcal{F}_{t2}$ be sub-σ fields of $\mathcal{F}$, which increase with t, $t = 1, \cdots, T$.

EFs for smoothing: The state and the observational equations form the basic EFs, which are

$$h_{t1} = f_{t1}(y_t, y_{t-1}, x_t) \text{ and } \mathrm{h_{t2} = f_{t2}(x_t, x_{t-1}),\ t = 1, \cdots, T.}$$

Let

$$a_{tj}^{*(r)} = E[\partial h_{tj}/\partial x_r|\mathcal{F}_{tj}]/var(h_{tj}|\mathcal{F}_{tj}).$$

We make the following assumptions:

Assumption 1: $E[h_{tj}|\mathcal{F}_{tj}] = 0,\ t = 0, 1, \cdots, T; j = 1, 2.$

Assumption 2: $E[a_{tj}^{*(r)} h_{tj} a_{sk}^{*(r)} h_{sk}|\mathcal{F}_{tj}] = 0,\ j, k = 1, 2,$

for all $r,\ s$ and t, $(s, k) \neq (t, j)$. (This condition is known as the 'orthogonality' of the h_{tj}'s.)

Consider the class $\mathcal{G}$ of EFs defined by

$$\mathcal{G} = \{\, g = (g^{(0)}, \cdots, g^{(T)}) \mid\ g^{(r)} = \textstyle\sum_{t=0}^{T} a_{t1}^{(r)} h_{t1} + \sum_{t=0}^{T} a_{t2}^{(r)} h_{t2};$$
$$a_{tj}^{(r)} \text{ is } \mathcal{F}_{tj}\text{-measurable, } r = 0, 1, \cdots, T \text{ and } j = 1, 2 \,\}.$$

The optimal EF $g^* = (g^{*(0)}, \cdots, g^{*(T)})$ in $\mathcal{G}$ is:

$$g^{*(0)} = (x_0 - \mu_0)/var(h_{2,0} + a_{02}^{*(0)} h_{1,2}),$$
$$g^{*(s)} = a_{s1}^{*(s)} h_{s1} + a_{s2}^{*(s)} h_{s2} + a_{(s+1)2}^{*(s)} h_{(s+1)2}, \quad \text{for} \quad s = 1, \cdots, T-1, \quad \text{and}$$
$$g^{*(T)} = a_{T1}^{*(T)} h_{T1} + a_{T2}^{*(T)} h_{T2}.$$

Solution of $g^* = 0$ gives the estimates of the states $(x_1, \cdots x_T)$.

Though the models in Examples 5.1 and 5.2 are not exactly in the form of the model given by ((4), and (5)), the basic EF's for the Example 5.1 can be taken as

$$h_{t1} = y_t - \nu x_t \text{ and } h_{t2} = c x_t / x_{t-1} - \sigma/(\sigma + \nu),$$

$t = 1, \cdots, T$ and for the Example 5.2

$$h_{t1} = y_t - e^{x_t}/(1 + e^{x_t}) \text{ and } h_{t2} = x_t - x_{t-1},$$

$t = 1, \cdots, T.$

EFs for filtering: The basic EFs are

$$h_{t1} = f_{t1}(y_t, y_{t-1}, x_t) \text{ and } \mathrm{h_{t2} = x_t - x_{t|t-1}},$$

where $x_{t|t-1}$ is a function of $y_0, y_1, \cdots, y_{t-1}, \ \ t = 1, 2, \cdots$.
Assume A1 and A2 above. For a fixed t consider

$$\mathcal{G} = \{g | g = a_{t1} h_{t1} + a_{t2} h_{t2} : \ a_{tj} \ is \ \mathcal{F}_{tj} - measurable, \ j = 1, 2\}.$$

The optimal estimating function g^* in $\mathcal{G}$ is $\quad g^* = a_{t1}^* h_{t1} + a_{t2}^* h_{t2}, \quad$ where $\quad a_{tj}^* = E \ [\ \partial h_{tj}/\partial x_t \mid \mathcal{F}_{tj} \ ] \ / \ var(h_{tj} \mid \mathcal{F}_{tj})$.

Different classes of estimating functions are obtained for different choices of $\sigma-$fields $\mathcal{F}_{t1}$ and $\mathcal{F}_{t2}$

Class(i)	$\mathcal{F}_{t1}^{(i)}$	$\mathcal{F}_{t2}^{(i)}$
$\mathcal{G}_1$	$(\Omega, \emptyset)$	$(\Omega, \emptyset)$
$\mathcal{G}_2$	$\sigma\{Y_0, Y_1, \cdots, Y_{t-1}\}$	$(\Omega, \emptyset)$
$\mathcal{G}_3$	$\sigma\{Y_0, Y_1, \cdots, Y_{t-1}\}$	$\sigma\{Y_0, Y_1, \cdots, Y_{t-1}\}$
$\mathcal{G}_4$	$\sigma(\mathbf{Y_{t-1}}; \mathbf{X_t})$	$\sigma(\mathbf{Y_{t-1}}; \mathbf{X_{t-1}})$

where $\mathbf{Y_{t-1}} = (Y_0, \cdots, Y_{t-1})$, and $\mathbf{X_{t-1}} = (X_0, \cdots, X_{t-1})$.
We note that $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{G}_3 \subset \mathcal{G}_4$. Thus the $\mathcal{G}_4$-optimal function g^* is 'better' than the functions 'optimal' in the subclasses.

Within $\mathcal{G}_4$, g^* is closest to the posterior score function and coincides with it in some situations. Computation of the optimal weights in case of $\mathcal{G}_4$ are easier than those for the others. The class $\mathcal{G}_3$ of estimating functions is a subclass of posterior unbiased estimating functions. The optimal estimating function in this class is closest to the posterior score function in the posterior sense (where the posterior distribution is the conditional distribution of the parameters given the data).

The optimal estimating equations may be nonlinear but can be solved using a software package such as the NLSYS module of the software package GAUSS. It may happen that the filtering procedure does not recursively compute a $x_{t|t-1}$ such that the Assumption A1 holds for $X_t - X_{t|t-1}$. In this situation, we carry out the smoothing procedure up to time t for each t, and declare the estimator of x_t obtained through smoothing at t to be the filter for x_t.

As stated earlier we do not make any distributional assumptions on $\{\epsilon_t\}$ and $\{\delta_t\}$, nor do we assume that they are sequences of independent random variables. We do assume the knowledge of the first two conditional moments of $\{\epsilon_t\}$ and $\{\delta_t\}$, namely $E(\epsilon_t|\mathcal{F}_{t1}^{(4)}) = 0,\ E(\delta_t|\mathcal{F}_{t2}^{(4)}) = 0,\ var(\epsilon_t|\mathcal{F}_{t1}^{(4)}) = V_t$, and $var(\delta_t|\mathcal{F}_{t2}^{(4)}) = \sigma_t^2$, where V_t and σ_t^2 are known functions of the conditioning random variables.

If f_{t1} and f_{t2} are linear and σ_t^2 and V_t are constants, then our procedure reduces to the celebrated *Kalman filter and Smoother*.

6. Optimal unbiased estimating functions for Hilbert space valued parameters

Non-parametric problems (form of the probability distribution or probability function is not assumed) or estimation of functions can be treated as parametric ones by taking the parameter space to be an infinite dimensional space. In [23] we consider the parameter space to be a real separable (infinite dimensional) Hilbert space and extend the Cramér-Rao type inequality for the EFs under some conditions, which leads to the definition of an 'optimality' criterion for the estimating function.

Let $\mathcal{X}$ be a sample space with probability measure P_α. Suppose the parameter α belongs to a real separable Hilbert space H. Let $< \cdot, \cdot >$ denote the inner product and $\mathcal{C} = \{e_n, n \geq 1\}$ a complete orthonormal basis in H. Let $f(x,\alpha) = dP_\alpha/d\mu(x)$ denote the probability density w.r.t. some σ-finite measure μ on $\mathcal{X}$.

Definition 6.1. A function $G : \mathcal{X} \times H \to H$ is called an unbiased estimating function (EF) iff $\forall \alpha \in H$ and $\forall e_k \in \mathcal{C}$

$$E_\alpha[< e_k, G(x,\alpha) >] = 0.$$

We assume regularity conditions, analogous to those given in the section 2. Let

$$S_j(x,\alpha) = \{\frac{d}{dt} f(x, \alpha + te_j)|_{t=0}\}/f(x,\alpha),$$

where the quantity in the numerator is the Gateaux derivative of $f(x,\alpha)$ in the direction e_j. We note that the $S_j(x,\alpha)$'s are analogues of the score function $\frac{\partial}{\partial\theta}(log(f(x,\theta)))$. Further, we assume that $\exists$ non-zero real constants $w_j,\ j \geq 1$, such that $E_\alpha[\sum_j |S_j(x,\alpha)w_j|^2] < \infty,\ \forall \alpha \in H$.

For $\beta = \sum_j \beta_j e_j \in H$, let $J_\alpha \beta = \sum_k (\sum_j \beta_j E_\alpha[w_k S_k(x,\alpha) w_j S_j(x,\alpha)]) e_k$. Let $d_i G_k(x,\alpha) = \frac{d}{dt} G_k(x, \alpha + te_i)|_{t=0}$ and $D_{G,\alpha}\beta = \sum_k (\sum_j \beta_j E_\alpha[d_k G_j(x,\alpha)]) w_k e_k$, where $G_k(x,\alpha) =< e_k, G(x,\alpha) >$. Under the appropriate regularity conditions on $f(x,\alpha)$ and on $G(x,\alpha)$ we obtain the following inequality. An EF satisfying the regularity conditions is called a regular EF.

Cramér-Rao type inequality: for all regular $G(x,\alpha)$,

$$Var_\alpha < \beta, G(x,\alpha) > \geq < \beta, D'_{G,\alpha} J_\alpha^{-1} D_{G,\alpha} \beta >, \ \forall \ \alpha \ \& \ \beta \in H, \tag{6}$$

where D' denotes the adjoint of the operator D.

Definition 6.2. : A 'regular' G^* is said to be optimal in the class of 'regular' EFs iff equality holds in (6).

The article [23] has been used in [6], which is on classification of functional time series. The optimality criterion of the definition 6.2 differs from the usual one in the finite dimensional case, which is based on the non-negative definiteness of the difference of the dispersion matrices of standardized regular EFs and which assumes the invertiblity of $D_{G,\alpha}$. Some examples are included in our above cited article. For a comprehensive treatment of Hilbert space methods and estimating functions we refer to [21].

7. Concluding Remarks

In all the three methods of estimation considered in this article, the estimate is obtained by solving an estimating equation. If one knows the functional form of the pdf (pmf) of the data generating process, except up to a parameter, and if the regularity conditions stated in the section 2 hold, then the likelihood approach can be used to obtain an asymptotically optimal estimator. If the functional form can not be specified, but if some moments or conditional moments can be specified, then the EF approach or the GMM are good alternatives. The GMM results in an estimating function, which may not be an 'optimal' estimating function as defined in the section 3. However, the resulting estimators, under certain conditions, are consistent and asymptotically normal but their optimality in terms of minimum variance depends on the weighting matrix used. The skill, in this case, is in choosing a proper weighting matrix. In the EF approach, first the optimal EF is obtained in some specified class of EFs and then it is solved to obtain the estimates. The estimators so obtained are consistent and asymptotically normal under certain conditions. The optimal EF within a class can be used to construct minimum size confidence zones in that class. Moreover, the optimal EF in a class is closest to the score function than all other EFs in that class and coincides with it if the score function belongs to that class. A drawback is that the weights involved in obtaining the optimal EF, may not be easy to compute or may involve the unknown parameters resulting in a complicated estimating equation. For the

EF approach, an appropriate class of EFs should be chosen; as was considered in the section 5.

In some problems, the regularity conditions may not hold, and other methods are considered to obtain 'good' estimators.

References

[1] Bera, A. K. and Biliasb, Y., *The MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis*, Journal of Econometrics, **107** (2002), 51–86.

[2] Bera, A. K., Bilias, Y. and Simlai, P., *Estimating functions and equations: An essay on historical developments with applications to econometrics*, Econometric Theory **1** (2006), 427–476.

[3] Casella, G. and Berger, R. L., *Statistical Inference* edn. 2, Duxbury, Cengage Learning, 2002.

[4] Chen, Y. and Singpurwalla, N. D., *A non-Gaussian Kalman filter model for tracking software reliability*, Statistica Sinica, **4** (1994), 535–548.

[5] Durbin, J., *Estimation of parameters in time-series regression models*, J. R. Stat. Soc. B **22** (1960), 139–153.

[6] Glendinning, R. H. and Fleet, S. L., *Classifying functional time series*, Signal Processing, **87** (2007), 79–100.

[7] Godambe, V. P., *An optimum property of regular maximum likelihood estimation*, Ann. Math. Statist. **31** (1960), 1208–1212.

[8] Godambe, V. P., *The foundations of finite sample estimation in stochastic processes*, Biometrika, **72** (1985), 419–428.

[9] Godambe, V. P. (Ed.), *Estimating Functions*, Oxford Science Publications, Oxford, 1991.

[10] Hall, A. R., *Generalized Method of Moments*, Oxford, University Press, 2005.

[11] Hansen, L., *Large sample properties of generalized method of moments estimators*, Econometrica **50** (1982), 1029–1054.

[12] Heyde, C. C., *Quasi-likelihood and its applications: A general approach to optimal parameter estimation*, Springer, 1997.

[13] Kale, B. K., *An extension of Cramér-Rao inequality for statistical estimation functions*, Skand. Aktur.**45** (1962),60–89.

[14] Kalman, R. E., *A new approach to linear filtering and prediction problems*, Trans. ASME - J. Basic Engineering, **82** (1960), 34–45.

[15] Kalman, R. E. and Bucy, R. S., *New results in linear filtering and prediction theory*, Journal of Basic Engineering, **83** (1961), 95–108.

[16] Kitagawa, G., *Non-Gaussian state-space modeling of non-stationary time series (with discussion)*, Journal of the American Statistical Association, **82** (1987), 1032–1063.

[17] Naik-Nimbalkar U. V. and Rajarshi, M. B., *Filtering and smoothing via estimating functions*, Journal of the American Statistical Association, **90** (1995), 301–306.

[18] Naik-Nimbalkar, U. V., *Estimating functions for stochastic Processes*, In: Stochastic Processes and Statistical Inference, Editors: Prakasa Rao, B. L. S., Bhat, B. R., New Age International Publishers, New Delhi, (1996), 52–72.

[19] Pawitan, Y., *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Clarendon Press, Oxford, 2001.

[20] Pearson, K., *Contribution to the mathematical theory of evolution*, Philosophical Transactions of the Royal Society of London. Series A, **185** (1894), 71–110.

[21] Small, C. G. and McLeish D. L., *Hilbert Space Methods in Probability and Statistics* by John Wiley & Sons, Inc. 1994.

[22] Stigler, S. M., *The Epic Story of Maximum Likelihood*, Statistical Science, **22** (2007), 598–620.

[23] Subramanyam, A. and Naik-Nimbalkar U. V., *Optimal unbiased statistical estimating functions for Hilbert space valued parameters*, Journal of Statistical Planning and Inference, **24** (1990), 95–105.

Uttara Naik-Nimbalkar

Savitribai Phule Pune University & IISER, Pune

Pune-411 007, Maharashtra, India

E-mail: *uvnaik@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 79-96

MATRIX GROUPS OVER RINGS

B. SURY

Abstract. This write-up is in the nature of an exposition of some work dealing with matrix groups over rings and their various decomposition theorems. The topic of matrix groups over rings like the integers is too general and too vast to allow a reasonable survey; so, we give an overview of some topics related to factorization. We discuss two types of related questions on matrix groups over rings here: (i) generating certain matrix groups by abstract subgroups like cyclic groups and implications on the structure of the ambient group; and (ii) '*finite width*' factorization into unipotent subgroups over rings.

1. Introduction.

Groups of matrices are ubiquitous in mathematics via their various avataars: Lie groups - if we work over $\mathbb{R}$ or $\mathbb{C}$, arithmetic subgroups - over integers and other number rings, finite simple groups - over finite fields, representation theory - over any ring. The existence of decompositions/factorizations into special types of pieces (for instance, Iwasawa, Cartan, Bruhat, Langlands,...) have traditionally played key roles. For example, Bruhat decomposition which arose in the theory of linear algebraic groups has proved useful in diverse contexts like numerical stability, and coding theory. In the paper ([19]), it is shown that for certain classes of matrices that have an exponential growth factor when Gaussian elimination with partial pivoting is applied, Bruhat decomposition has at most linear growth. In the paper ([17]), the authors present a new Bruhat decomposition design for constructing full diversity unitary space-time constellations for any number of antennas. The so-called Langlands decomposition of a parabolic subgroup is behind the "philosophy of cusp forms" due to Harish-Chandra (a precursor to Langlands's program) where the discrete groups take the backstage and inducing representations via the Langlands decomposition take center stage. So, generating matrix groups via special kinds of elements is useful. These are trickier

* The (modified) text of the 26^{th} Hansraj Gupta Memorial Award Lecture delivered at the 81^{st} Annual Conference of the Indian Mathematical Society held at the Visvesvaraya National Institute of Technology, Nagpur - 440 010, Maharashtra, during December 27-30, 2015.

2010 Mathematics Subject Classification: 11D45.

Keywords and Phrases: Matrix groups, bounded generation, unitriangularization, finite width, Chevalley groups, rings of stable rank 1.

and more subtle over rings which are not fields. In the next section, we describe a few of the applications where matrix groups over rings play a key role. The examples are chosen for their diversity. Following that, in section 3, we describe the more recent factorization theorems and their proofs.

2. Matrix groups over rings - some old applications

In this section, we briefly describe some of our earlier results on matrix groups over rings which are related to number theoretic and combinatorial group-theoretic questions. These examples are selected purely to demonstrate the diversity of applications that matrix groups over various rings have on other topics.

2.1. **Salem numbers.** A question due to D. H. Lehmer (which is still open from 1933) asks if there is a positive constant $c > 1$ such that for any integer coefficient polynomial, the product of the absolute values of its roots is strictly $> c$ unless the polynomial has only roots of unity as roots. Lehmer's computations revealed that the "worst" polynomials in this respect correspond to reciprocal polynomials with one real root $\tau > 1$ and other roots being $\frac{1}{\tau}$, $\tau_2, \overline{\tau_2}, \cdots, \tau_d, \overline{\tau_d}$ for $|\tau_i| = 1$. Such algebraic integers τ are known as Salem numbers - named after Raphael Salem who studied some of their properties. We can reformulate this question (see [25]) for the above subclass of polynomials in terms of the subgroups of $SL(2,\mathbb{R})$; the question is equivalent to asking if there is a neighbourhood U of the identity matrix such that every arithmetic subgroup Γ with no elements of finite order other than the identity and such that the quotient $SL(2,\mathbb{R})/\Gamma$ is compact, satisfies $\Gamma \cap U = \{I\}$.

2.2. **Generating a family of subgroups.** Here is an example to show how combinatorial-type properties may have bearing on deeper properties of the group. We proved (see [31]) the following theorem.

Theorem 2.1. *For any fixed $n \geq 3$, there is a number $N(n)$ depending only on n so that every group of the form*

$$Ker(SL_n(\mathbb{Z}) \to SL_n(\mathbb{Z}/k\mathbb{Z}))$$

can be generated by $N(n)$ elements for every $k > 1$. One may also wrote out a description of generators for each k.

Recently, Detinko, Flannery and Hulpke used the generators to give an algorithm (see [7]) to decide whether a subgroup of $SL_n(\mathbb{Z})$ (for $n > 2$) has finite index - in general, such problems are undecidable. In the above-mentioned paper, we had also given an example to show that there is no bound like $N(n)$ if we allow all normal subgroups of finite index. Very recently, Mark Shusterman proved a result bounding rank of a group in terms of its index where he elaborates on our example to show that his result is close to optimal.

2.3. **Infinitely presented matrix groups.** The following matrix group over a ring is an example of certain phenomena dealing with factorization, generation and finite presentation (see [26]).

Theorem 2.2. *Let p be a prime. We consider the ring $\mathbb{Z}[1/p]$ of rational numbers whose denominators can only be divisible by powers of p. Let*

$$G = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & p^n & c \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \in \mathbb{Z}[1/p], n \in \mathbb{Z} \right\}$$

has the remarkable properties. Then,
(a) $G = C_1C_2 \cdots C_{12}$ where C_i's are cyclic groups (not necessarily distinct (that is, G has bounded generatios of degree ≤ 12);
(b) the commutator subgroup $[G, G]$ is not finitely generated;
(c) G is not finitely presented.

Indeed, if $x = diag(1, p, 1), y_{12} = I + E_{12}, y_{23} = I + E_{23} \in G$, then

$$\begin{pmatrix} 1 & ap^k & bp^l \\ 0 & p^n & cp^m \\ 0 & 0 & 1 \end{pmatrix} = x^{n-k} y_{12}^a x^{m-n+k} y_{23}^c x^{n-m} x^B y_{12}^A x^{-B} y_{23} x^B y_{12}^{-A} x^{-B}$$

where A, B are defined by $bp^l - acp^{m-n+k} = Ap^{-B}$.

The commutator subgroup of G is the unipotent group $\left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \right\}$ is infinitely generated; indeed, even its abelianization is infinitely generated. The fact that G is not finitely presentable follows from a criterion due to Bieri and Strebel.

2.4. **Matrix groups over finite rings and elementary number theory.** Elementary number-theoretic identities often fall out when one looks at natural actions of matrix groups over *finite rings* (note that finite rings have stable rank 1 - our factorization theorems in the next section deal with rings of stable rank 1). For instance, the identity

$$\sum_{t_1 \in (\mathbb{Z}_n)^*, t_2, \cdots, t_r \in \mathbb{Z}_n} GCD(n, t_1 - 1, t_2, \cdots, t_r) = \phi(n)\sigma_{r-1}(n)$$

can be derived (see [24]) by applying the so-called Cauchy-Frobenius-Burnside lemma to the group

$$G = \{ \begin{pmatrix} t_1 & t_2 & t_3 & \cdots & t_r \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} : t_1 \in (\mathbb{Z}_n)^*, t_i \in \mathbb{Z}_n \ \forall i > 1 \}$$

acting naturally on $(\mathbb{Z}_n)^r$. More generally, the action of the full upper triangular subgroup U_r of $GL(r, \mathbb{Z}_n)$ yields:

$$\sum_{A \in U_r} \prod_{k=1}^{r} d_k = n^{\binom{r}{2}} \phi(n)^r d_r(n)$$

where $A = (a_{ij})$,

$$d_k = GCD\Bigg(n, \frac{na_{1,k}}{(n, a_{1,1}-1, a_{1,2}, \cdots, a_{1,k-1})},$$
$$\frac{na_{2,k}}{(n, a_{22}-1, a_{23}, \cdots, a_{2,k-1})}, \cdots, \frac{na_{k-1,k}}{(n, a_{k-1,k-1}-1)}\Bigg)$$

and $d_1(n) = \sum_{d|n} d$, $d_k = \sum_{d|n} d_{k-1}(d)$.

2.5. **Finite matrix groups as capable groups.** Matrix groups over finite fields provide natural and easy examples of certain phenomena which occur in finite groups. For instance we have the following theorem (see [23]).

Theorem 2.3. *If A is a finite abelian capable group (that is, $A \cong G/Z(G)$ for some group G) where the center $Z(G)$ of G is cyclic, then $A \cong B \times B$ for an abelian group B; in particular, the order of A is a perfect square. Further, this property of A is not necessarily true if $Z(G)$ is not cyclic.*

Thus, it is of interest to find simple examples where $Z(G)$ is not cyclic where $G/Z(G)$ has non-square order. In loc. cit., we constructed the following example.

Example. Let F be a finite field and $E \subset F$ be a proper subfield. Consider the group

$$G = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} : b, c \in F; a \in E \right\}.$$

If we denote a typical element of G by $g(a, b, c)$, then

$$g(a, b, c)g(a', b', c') = g(a + a', b + b', ab' + c + c').$$

Further, $g(a, b, c)^{-1} = g(-a, -b, ab - c)$. Now, note that $g(a, b, c) \in Z(G)$ if and only if $ab' = a'b$ for all $a' \in E, b' \in F$. Thus, some $g(a, 0, c) \in Z(G)$ if and only if $ab' = 0$ for all $b' \in F$; that is, if and only if $a = 0$. On the other hand, if some $g(a, b, c) \in Z(G)$ with $b \neq 0$, then $g(a, b, c)g(1, 0, 0) = g(1, 0, 0)g(a, b, c)$ gives $0 = b$, a contradiction. Thus

$$Z(G) = \{g(0, 0, c) : c \in F\} \quad \text{and} \quad G/Z(G) \cong E \oplus F.$$

Note that the finite, abelian, capable group $G/Z(G)$ can have non-square order - for instance, if E has p elements and F has p^2 elements then $Z(G)$ is not cyclic.

2.6. **Matrix groups as monodromy groups of polynomials.** The problem of finiteness of number of solutions of Diophantine equations of the form $f(x) = g(y)$ where f, g are integer polynomials leads to questions on their monodromy groups which can be fruitfully answered by analyzing certain matrix groups which are isomorphic to finite dihedral groups. Work of Yuri Bilu showed (see [4]) that one may reduce the problem to determining the possible quadratic factors of the polynomial $f(X) - g(Y)$. Over an algebraically closed field K of any characteristic, the latter question is answered in the following manner.

Let x be transcendental over K, and set $t = f(x)$. If $f(X)-g(Y)$ has an irreducible factor of degree 2, then $K(x)$ has a quadratic extension L, the function field of this quadratic factor. Then $L/K(t)$ is Galois. The Galois group is generated by two involutions, hence it is dihedral. The intermediate field $K(x)$ is the fixed field of one of the involutions. By Lüroth's Theorem, KL is a rational field $K(z)$. So $Gal(K(z)/K(t))$ is a subgroup of $Gal(L/K(t))$. Also, the index is at most 2. The group of K-automorphisms of $K(z)$ is $PGL_2(K)$ acting as linear fractional transformations of z. Thus, to determine factors of degree at most 2 of $f(X)-g(Y)$, we have to determine the cyclic and dihedral subgroups of $PGL_2(K)$, and analyze the cases which give pairs f, g such that $f(X) - g(Y)$ has a quadratic factor over K. We may show (see [11]):

Proposition. Let K be an algebraically closed field of characteristic p, and $\rho \in PGL_2(K)$ be an element of finite order n. Then one of the following holds:

(a) p does not divide n, and ρ is conjugate to $\begin{pmatrix} 1 & 0 \\ 0 & \zeta \end{pmatrix}$, where ζ is a primitive n-th root of unity.

(b) $n = p$, and ρ is conjugate to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Using this, we may deduce the following (loc. cit.). First, we introduce two notations. For $u, v \in K[X]$, write $u \sim v$ if and only if there are linear polynomials $L, R \in K[X]$ with $u(X) = L(v(R(X)))$. Also, for $a \in K$, the Dickson polynomial $D_n(X, a)$ is defined by $D_n(z + a/z, a) = z^n + (a/z)^n$. It turns out that we have the following theorem.

Theorem 2.4. *Let $f, g \in K[X]$ be non-constant polynomials over a field K, such that $f(X) - g(Y) \in K[X, Y]$ has a quadratic irreducible factor $q(X, Y)$. If the characteristic p of K is positive, then assume that at least one of the polynomials f, g cannot be written as a polynomial in X^p. Let $\deg f = n$. Then there are $f_1, g_1, \Phi \in K[X]$ with $f = \Phi \circ f_1$, $g = \Phi \circ g_1$ such that $q(X, Y)$ divides $f_1(X)-g_1(Y)$, and one of the following holds*

(a) $\max(\deg f_1, \deg g_1) = 2$ *and* $q(X, Y) = f_1(X) - g_1(Y)$.

(b) *There are $\alpha, \beta, \gamma, \delta \in K$ with $g_1(X) = f_1(\alpha X+\beta)$, and $f_1(X) = h(\gamma X+\delta)$, where $h(X)$ is one of the following polynomials.*

  (i) *p does not divide n, and $h(X) = D_n(X, a)$ for some $a \in K$. If $a \neq 0$, then $\zeta + 1/\zeta \in K$ where ζ is a primitive n-th root of unity.*

  (ii) *$p \geq 3$, and $h(X) = X^p - aX$ for some $a \in K$.*

  (iii) *$p \geq 3$, and $h(X) = (X^p + aX + b)^2$ for some $a, b \in K$.*

  (iv) *$p \geq 3$, and $h(X) = X^p - 2aX^{\frac{p+1}{2}} + a^2X$ for some $a \in K$.*

  (v) *$p = 2$, and $h(X) = X^4 + (1 + a)X^2 + aX$ for some $a \in K$.*

(c) *n is even, p does not divide n, and there are $\alpha, \beta, \gamma, a \in K$ such that $f_1(X) = D_n(X+\beta, a)$, $g_1(X) = -D_n((\alpha X+\gamma)(\xi+1/\xi), a)$. Here ξ denotes a primitive $2n$-th root of unity. Furthermore, if $a \neq 0$, then $\xi^2+1/\xi^2 \in K$.*

(d) *$p \geq 3$, and there are quadratic polynomials $u(X), v(X) \in K[X]$, such that $f_1(X) = h(u(X))$ and $g_1(X) = h(v(X))$ with $h(X) = X^p - 2aX^{\frac{p+1}{2}} + a^2X$ for some $a \in K$.*

The theorem excludes the case that f and g are both polynomials in X^p. The following theorem handles this case; a repeated application of it reduces to the situation of the Theorems

Theorem 2.5. *Let $f, g \in K[X]$ be non-constant polynomials over a field K, such that $f(X) - g(Y) \in K[X, Y]$ has an irreducible factor $q(X, Y)$ of degree at most 2. Suppose that $f(X) = f_0(X^p)$ and $g(X) = g_0(X^p)$, where $p > 0$ is the characteristic of K. Then one of the following holds:*

(a) *$q(X, Y)$ divides $f_0(X) - g_0(Y)$, or*

(b) *$p = 2$, $f(X) = f_0(X^2)$, $g(X) = f_0(aX^2 + b)$ for some $a, b \in K$, and $q(X, Y) = X^2 - aY^2 - b$.*

2.7. **Bounded generation and finite width.** The matrix groups over integers like $SL_n(\mathbb{Z})$ are finitely generated and even have finite presentations. However, a remarkable refinement of the first property came to the fore in the work of A.S.Rapinchuk. This is known as bounded generation. An abstract group G is said to be boundedly generated of degree $\leq n$ if there exists a sequence of (not necessarily distinct) elements $g_1, \cdots, g_n$ such that

$$G =< g_1 >< g_2 > \cdots < g_n >$$

that is,

$$G = \{g_1^{a_1} g_2^{a_2} \cdots g_n^{a_n} : a_i \in \mathbb{Z}\}.$$

A free, non-abelian group (and therefore, $SL_2(\mathbb{Z})$ also) is not boundedly generated. On the other hand, a group like $SL_n(\mathbb{Z})$ for $n \geq 3$, is boundedly generated by elementary matrices (an elementary proof of this can be given using Dirichlet's theorem on primes in arithmetic progressions). It turns out that this difference is an indicator of a deeper attribute called the congruence subgroup property; viz., every subgroup of finite index in $SL_n(\mathbb{Z})$ for $n \geq 3$ contains a subgroup of the form

$$Ker(SL_n(\mathbb{Z}) \to SL_n(\mathbb{Z}/k\mathbb{Z}))$$

This was revealed in the work of V.P.Platonov & A.S.Rapinchuk ([21]) and also in the work of A.Lubotzky ([16]).

Matrix groups which are finitely generated have an abundance of subgroups of finite index. More precisely, they are residually finite - that is, the intersection of all subgroups of finite index is the trivial group. In this case, it is beneficial to define a topology using as a basis the subgroups of finite index - residual finiteness guarantees this is Hausdorff. The completion with respect to this topology is known as the profinite completion; this is a compact group in which the original group embeds.

A finitely generated group for which the normal subgroups which have indices powers of a (fixed) prime p intersect in the identity, is said to be residually-p. The corresponding profinite completion is called the pro-p completion. In general, a profinite group is formed by putting together a tower of finite groups by a limiting process; a classical example is that of Galois group of the algebraic closure. Notions involved in the theory of finite groups and many properties find resonance in the theory of profinite groups and the topology available in the latter theory makes it possible to deduce properties of abstract, discrete groups.

A profinite group G is said to be boundedly generated as a profinite group if there exists a sequence of (not necessarily distinct) elements $g_1, \cdots, g_n$ such that

$$G = \overline{< g_1 >< g_2 >} \cdots \overline{< g_n >}$$

where the 'bar' denotes closure.

It follows from Lazard's deep work on p-adic Lie groups (see [12]) and the solution to the restricted Burnside problem that a pro-p group has bounded generation (as a profinite group) if and only if it is a p-adic compact Lie group; this can be thought of as an analogue of Hilbert's 5th problem for the p-adic case.

If an abstract group has bounded generation, then so do its pro-p completions for each prime p (as does the full profinite completion). Therefore, we have a nice sufficient criterion for an abstract group to have a faithful linear representation - viz., if it has bounded generation and is virtually residually-p. We can use this idea to show that the automorphism group of a free group does not have bounded generation (see [26]).

The question of existence of bounded generation for matrix groups over number-theoretic rings has rather deep connections with other properties. The profinite completion of an arithmetic group is boundedly generated if, and only if, it has the congruence subgroup property - this was proved independently by V. P. Platonov & A. S. Rapinchuk and by A. Lubotzky (see [21]) and [16]). Lubotzky also conjectured that the congruence subgroup property holds for an S-arithmetic group if, and only if, it can be embedded as a closed subgroup of $SL_n(\mathbf{A})$ - a so-called adelic group. This was proved in [22], where we also conjectured that finitely generated closed subgroups of adelic groups have bounded generation. Then M.Liebeck & L.Pyber proved ([15]) that if G is a subgroup of $GL_n(K)$ where K is of characteristic p which is large compared to n, and if G is generated by elements of orders powers of p, then G is a product of 25 Sylow p-subgroups. They used it to prove our conjecture mentioned above.

It is still an intriguing open question as to whether the property of bounded generation for an S-arithmetic group Γ equivalent to bounded generation for its profinite completion (which is, as mentioned earlier, equivalent to the congruence subgroup property holding good for Γ). The answer is perhaps in the negative and certain arithmetic subgroups of $Sp(n,1)$ could provide counter-examples. Moreover, this is a subtle question specific to arithmetic groups and not for more general profinite groups because the group $\prod_{r\geq 1} PSL_n(\mathbf{F}_{2^r})$ is boundedly generated group as a profinite group but none of its discrete subgroups is boundedly generated.

O.Tavgen proved bounded generation of arithmetic groups in rank > 1 groups (see [33]). However, bounded generation for co-compact arithmetic lattices is still an open question in general excepting the case of quadratic forms (see [8]); note here that there are no unipotent elements.

A notion related to but weaker than bounded generation is that of finite width wirh respect to a subset defined as follows.

A group G has finite width with respect to a subset E if there exists a positive integer n such that each element of G can be expressed as $g = e_1e_2\cdots e_r$ with $r \leq n$ and $e_i \in E$.

If we look at rings R that are finitely generated as abelian groups, then $SL(n,R)$ has bounded generation if it has finite width with respect to the set of all elementary matrices $X_{ij}(t)$ with $t \in R$ and $i \neq j$.

More generally, for any commutative ring R, one could look at the question of finite width for elementary group $E_n(R)$ which may be a proper subgroup of $SL(n,R)$. It is not difficult to check that $E_n(R)$ has this property if and only if $K_1(n,R^{\mathbb{N}}) \to K_1(n,R)^{\mathbb{N}}$ is injective, where the K-group K_1 is the quotient of GL by E.

3. Finite unipotent width over stable rank 1 rings

We describe some results on finite width obtained in collaboration with Vavilov and Smolensky. The following problem arises in several independent contexts. It addresses Chevalley groups which we will describe shortly.

Problem. *For a commutative ring R, find the shortest factorization $G = UU^-UU^-\dots U^{\pm}$ of an elementary Chevalley group $E(\Phi,R)$, in terms of the unipotent radical $U = U(\Phi,R)$ of the standard Borel subgroup $B = B(\Phi,R)$, and the unipotent radical $U^- = U^-(\Phi,R)$ of the opposite Borel subgroup $B^- = B^-(\Phi,R)$.*

There are following two problems here.

- first, to establish the *existence* of such factorizations, and
- second, to estimate their *length*.

We can prove the following theorem for rings of stable rank 1 ([28]).

Theorem 3.1. *Let Φ be a reduced irreducible root system and R be a commutative ring such that the stable rank of R is 1. Then the elementary Chevalley group $E(\Phi,R)$ admits a uni-triangular factorisation*

$$E(\Phi,R) = U(\Phi,R)U^-(\Phi,R)U(\Phi,R)U^-(\Phi,R)$$

of length 4*. Further,* 4 *is the minimum possible for such a result to hold good if* R *has a nontrivial unit.*

Here, a commutative ring has **stable rank** 1, if for all $x, y \in R$, which generate R as an ideal, there exists a $z \in R$ such that $x + yz$ is invertible. In this case we write $\mathrm{sr}(R) = 1$. Examples of ring of stable rank 1 are semilocal rings, and the ring of ALL algebraic integers.

The same method allows us to prove (see [29]) the following theorem.

Theorem 3.2. *With* R *as above, the elementary Chevalley group* $E(\Phi, R)$ *admits a Gauss decomposition*

$$E(\Phi, R) = (T(\Phi, R) \cap E(\Phi, R))U(\Phi, R)U^-(\Phi, R)U(\Phi, R).$$

Conversely, if Gauss decomposition holds for some elementary Chevalley group, then $\mathrm{sr}(R) = 1$*.*

Actually, a corollary of this last theorem is the following statement which also shows theorem 3.1 holds at least in the weaker form with length 5.

Corollary 3.3. *Let* Φ *be a reduced irreducible root system and* R *be a commutative ring such that* $\mathrm{sr}(R) = 1$*. Then any element* g *of the elementary Chevalley group* $E(\Phi, R)$ *is conjugate to an element of*

$$U(\Phi, R)H(\Phi, R)U^-(\Phi, R).$$

In the 1960's, N. Iwahori & H. Matsumoto, E. Abe & K. Suzuki, and M. Stein discovered (see [1], [2], [10], [30]) that Chevalley groups $G = G(\Phi, R)$ over a semilocal ring admit the remarkable Gauss decomposition $G = TUU^-U$, where $T = T(\Phi, R)$ is a split maximal torus, whereas $U = U(\Phi, R)$ and $U^- = U^-(\Phi, R)$ are unipotent radicals of two opposite Borel subgroups $B = B(\Phi, R)$ and $B^- = B^-(\Phi, R)$ containing T. It follows from the classical work of Hyman Bass and Michael Stein that for classical groups Gauss decomposition holds under weaker assumptions such as $\mathrm{sr}(R) = 1$ or $\mathrm{asr}(R) = 1$. Later N. Vavilov noticed that condition $\mathrm{sr}(R) = 1$ is necessary for Gauss decomposition to be valid. In our theorems, we show that for the elementary group $E(\Phi, R)$, the condition $\mathrm{sr}(R) = 1$ is also sufficient for Gauss decomposition to hold good. In other words, $E = HUU^-U$, where $H = H(\Phi, R) = T \cap E$. This surprising result pinpoints the fact that stronger conditions on the ground ring, such as being semi-local, $\mathrm{asr}(R) = 1$, $\mathrm{sr}(R, \Lambda) = 1$, etc., were only needed to guarantee that for simply connected groups $G = E$, rather than to verify the Gauss decomposition itself. Our method of proof is an elaboration of a beautiful idea of O. Tavgen ([32]).

Results equivalent to writing matrices in terms of upper and lower triangular matrices have been proved piece-meal in various situations by programmers working on computational linear algebra and others ([20],[13], [5], [3], [9], [27], [34]). So, results such as the above unitriangular factorization admit potential applications in computational linear algebra, wavelet theory, computer graphics and Control theory. For instance, the one-dimensional shears correspond to transvections that are standard in the study of matrix groups over rings.

3.1. **Number rings-finite width.** Number-theoretic rings are usually more complicated. Over the ring $\mathbb{Z}[1/p]$, we prove (see [28]):

Theorem 3.4. *Let p be a prime. The elementary Chevalley group $E\Big(\Phi, \mathbb{Z}\Big[\frac{1}{p}\Big]\Big)$ admits unitriangular factorisation*

$$E\Big(\Phi, \mathbb{Z}\Big[\frac{1}{p}\Big]\Big) = \Big(U\Big(\Phi, \mathbb{Z}\Big[\frac{1}{p}\Big]\Big)U^-\Big(\Phi, \mathbb{Z}\Big[\frac{1}{p}\Big]\Big)\Big)^3$$

of length 6.

The theorem is deduced from the one below for SL_2 which we can prove in the following slightly stronger form.

Lemma 3.5.

$$SL_2\Big(\mathbf{Z}\Big[\frac{1}{p}\Big]\Big) = UU^-UU^-U = U^-UU^-UU^-.$$

From this factorization, we deduce explicitly that $SL_2(\mathbb{Z}[\frac{1}{p}])$ has bounded generation. This is known earlier (see [14]), but the bounded generation was deduced either using generalized Riemann Hypothesis or deep analytic results like Vinogradov's three primes theorem or an indirect model-theoretic proof is given where there was no information on the degree of bounded generation.

Such factorizations can be treated by relating them to division chains (see [6]) in the ring $\mathbb{Z}[\frac{1}{p}]$. Note that expressing a matrix in the group $SL(2, R)$ over a Euclidean ring R as a product of elementary matrices is equivalent to studying continued fractions. Existence of arbitrary long division chains in $\mathbf{Z}$ shows that the group $SL(2, \mathbf{Z})$ cannot have bounded width in elementary generators. If $\begin{pmatrix} A & C \\ B & D \end{pmatrix}$ is in $SL_2(\mathbb{Z}[\frac{1}{p}])$, then we have the following lemma.

Lemma 3.6. $A = Q_1B + R_1, \qquad B = Q_2R_1 + R_2 \qquad R_1 = Q_3R_2 + 1.$

Thus

$$\begin{pmatrix} 1 & 0 \\ -R_2 & 1 \end{pmatrix}\begin{pmatrix} 1 & -Q_3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -Q_2 & 1 \end{pmatrix}\begin{pmatrix} 1 & -Q_1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A & C \\ B & D \end{pmatrix} = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$$

3.2. **Number rings - explicit bounded generation.** Using the above theorem, for the matrices $T = \begin{pmatrix} p^{-1} & 0 \\ 0 & p \end{pmatrix}$, $U_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $V_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, we can deduce unconditionally the following theorem.

Theorem 3.7. *Let p be a prime number. Then $SL_2(\mathbb{Z}[\frac{1}{p}])$ has bounded generation of degree at the most* 11. *In fact, we have*

$$SL_2(\mathbb{Z}[1/p]) = \{T^{a_1}U_1^{b_1}T^{a_2}V_1^{c_1}T^{a_3}U_1^{b_2}T^{a_4}V_1^{c_2}T^{a_5}U_1^{b_3}T^{a_6} : a_i, b_i, c_i \in \mathbb{Z}\}.$$

One should note that this is not completely straightforward from the above unitriangular factorization theorem because one can show that the group $SL_2(\mathbb{Z}[1/p])$ *cannot* have bounded generation with respect to only unipotent matrices!

Let us mention in passing that some matrix groups over rings of polynomials are finitely generated but are not boundedly generated. For instance, the group of 2×2 matrices $\begin{pmatrix} t^m & t^n f(t) \\ 0 & 1 \end{pmatrix}$ where f is any polynomial with integer coefficients and m, n are any integers, is an infinite group (it can be identified with the wreath product of $\mathbb{Z}$ by $\mathbb{Z}$) and it has an infinitely generated abelian subgroup, but is itself generated by just two matrices $\begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. One can prove by combinatorial methods that the above matrix group does not have bounded generation. In fact, we have (see [18]) the following result.

If A and B are groups then $A \wr B$ has bounded generation if and only if A has bounded generation and B is finite.

4. Elementary Chevalley groups over rings

We now proceed to introduce the Chevalley groups and the elementary subgroups occurring in the statements of our theorems 1 and 2.

Let Φ be a reduced irreducible root system of rank l, $W = W(\Phi)$ be its Weyl group and $\mathcal{P}$ be a lattice intermediate between the root lattice $\mathcal{Q}(\Phi)$ and the weight lattice $\mathcal{P}(\Phi)$. Further, we fix an order on Φ and denote by $\Pi = \{\alpha_1, \ldots, \alpha_l\}$, Φ^+ and Φ^- the corresponding sets of fundamental, positive and negative roots, respectively.

It is classically known that with these data one can associate the Chevalley group $G = G_{\mathcal{P}}(\Phi, R)$ for any ring R; this is the group of R-points of an affine groups scheme $G_{\mathcal{P}}(\Phi, -)$ - the Chevalley-Demazure group scheme. The group is said to be simply connected (res. adjoint) if $\mathcal{P}$ is the weight lattice (resp. root lattice). Since our results do not depend on the choice of the lattice $\mathcal{P}$, we will usually assume that $\mathcal{P} = \mathcal{P}(\Phi)$ and omit any reference to $\mathcal{P}$ in the notation. Thus, $G(\Phi, R)$ will denote the simply connected Chevalley group of type Φ over R.

Fix a split maximal torus $T(\Phi, -)$ of the group scheme $G(\Phi, -)$ and set $T = T(\Phi, R)$. Fix isomorphisms $x_\alpha : R \mapsto X_\alpha$. Here, the elements $x_\alpha(\xi)$; $\xi \in R$, $\alpha \in \Phi$ are called root unipotents and, the root groups X_α comprised of these elements when ξ varies in R, are interrelated by the Chevalley commutator formulae. The root subgroups X_α, $\alpha \in \Phi$ generate the elementary subgroup $E(\Phi, R)$ of $G(\Phi, R)$.

Let $\alpha \in \Phi$ and $\varepsilon \in R^*$. Set $h_\alpha(\varepsilon) = w_\alpha(\varepsilon) w_\alpha(1)^{-1}$, where $w_\alpha(\varepsilon) = x_\alpha(\varepsilon) x_{-\alpha}(-\varepsilon^{-1}) x_\alpha(\varepsilon)$. The elements $h_\alpha(\varepsilon)$ are called semisimple root elements. For a simply connected group one has

$$T = T(\Phi, R) = \langle h_\alpha(\varepsilon),\ \alpha \in \Phi,\ \varepsilon \in R^* \rangle.$$

One also defines $H(\Phi, R) = T(\Phi, R) \cap E(\Phi, R)$. Let $N = N(\Phi, R)$ be the algebraic normalizer of the torus $T = T(\Phi, R)$, i. e. the subgroup, generated by $T = T(\Phi, R)$ and all elements $w_\alpha(1)$, $\alpha \in \Phi$. The factor-group N/T is canonically isomorphic to the Weyl group W, and for each $w \in W$ we fix its preimage n_w in N.

The theorems 1 and 2 stated above generalize and strengthen results which were proved piecemeal over finite fields by several authors with a much simpler, uniform proof. Recall the statements again in the following form.

Theorem 4.1. *$E(\Phi, R)$ admits a Gauss decomposition*

$$E(\Phi, R) = H(\Phi, R)U(\Phi, R)U^-(\Phi, R)U(\Phi, R).$$

Conversely, if Gauss decomposition holds for some elementary Chevalley group, then $\mathrm{sr}(R) = 1$.

Corollary 4.2. *(to Theorem 1). We have a unitriangular factorisation*

$$E(\Phi, R) = U(\Phi, R)U^-(\Phi, R)U(\Phi, R)U^-(\Phi, R)$$

of length 4*. Further,* 4 *is the minimum possible for such a result to hold good if R has a nontrivial unit.*

The proofs rely on a beautiful idea of Oleg Tavgen on rank reduction (see [32]); we use the fact that for systems of rank ≥ 2 every fundamental root falls into the subsystem of smaller rank obtained by dropping either the first or the last fundamental root. One needs to study elementary parabolic subgroups then. We just discuss a toy case first.

4.1. **Toy case of theorem 3.1.** The following lemma is this Toy case.

Lemma 4.3. *Let R be a commutative ring of stable rank* 1*. Then*

$$\mathrm{SL}(2, R) = U(2, R)U^-(2, R)U(2, R)U^-(2, R).$$

In particular, $\mathrm{SL}(2, R) = E(2, R)$.

The toy case is the only place where the stability condition on R is invoked. To deduce the general theorem, we use only the theory of linear algebraic groups. Let us prove the toy case above.

proof. Let us trace how many elementary transformations one needs to bring an arbitrary matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, R)$ to the identity. We will not introduce new notation at each step, but rather replace the matrix g by its current value, as is common in computer science. Obviously, its entries a, b, c, d should be also reset to their current values at each step.

Step 1. Multiplication by a single *lower* elementary matrix on the right allows to make the element in the South-West corner invertible.

Indeed, since the rows of the matrix are unimodular, one has $cR + dR = R$ and since $\mathrm{sr}(R) = 1$, there exists such an $z \in R$, that $c + dz \in R^*$. Thus,

$$gt_{21}(z) = \begin{pmatrix} a + bz & b \\ c + dz & d \end{pmatrix}.$$

Step 2. We (can) assume that $c \in R^*$; then, multiplication by a single *upper* elementary matrix on the right allows to make the element in the South-East corner equal to 1. Indeed,

$$gt_{12}(c^{-1}(1-d)) = \begin{pmatrix} a & b + ac^{-1}(1-d) \\ c & 1 \end{pmatrix}.$$

Step 3. We (can) assume that $d = 1$; so, multiplication by a single *lower* elementary matrix on the right allows to make the element in the South-West corner equal to 0. Indeed,

$$gt_{21}(-c) = \begin{pmatrix} a - bc & b \\ 0 & 1 \end{pmatrix}.$$

Since $\det(g) = 1$, the matrix on the right hand side is equal to $t_{12}(b)$. Bringing all elementary factors to the right hand side, we see that any matrix g with determinant 1 can be expressed as a product of the form $t_{12}(*)t_{21}(*)t_{12}(*)t_{21}(*)$, as claimed in the lemma.

4.2. **A concrete case of theorem 3.2.** We discuss a special concrete case. If $N(n, R)$ is the group of monomial matrices over any commutative ring R, then we have the following result.

Proposition 4.4. *Let R be an arbitrary commutative ring. Then one has the following inclusion* $N(n, R) \subseteq U(n, R)U^-(n, R)U(n, R)U^-(n, R)$.

Proof. Let $g = (g_{ij}) \in N(n, R)$. Let us argue by induction on $n \geq 2$.

Case 1. First, let $g_{nn} = 0$. Then, there exists a unique $1 \leq r \leq n-1$ such that $a = g_{rn} \neq 0$ and a unique $1 \leq s \leq n-1$ such that $b = g_{ns} \neq 0$, all other entries in the s-th and the n-th columns are equal to 0. Since g is invertible, automatically $a, b \in R^*$. The matrix $gt_{sn}(b^{-1})$ differs from g only in the position (n, n), where now we have 1 instead of 0. Consecutively multiplying the resulting matrix on the right by $t_{ns}(-b)$ and then by $t_{sn}(b^{-1})$, we get the matrix h, which differs from g only at the intersection of the r-th and the n-the rows with the s-th and the n-th columns, where now instead of $\begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$ one has $\begin{pmatrix} -ab & 0 \\ 0 & 1 \end{pmatrix}$.

Observe, that the determinant of the leading submatrix of order $n-1$ of the matrix h equals 1, and thus we can apply induction hypothesis and obtain for that last matrix the desired factorisation in the group $\mathrm{SL}(n-1, R)$. This factorisation does not affect the last row and the last column. We have shown

$$gt_{sn}(b^{-1})t_{ns}(-b)t_{sn}(b^{-1}) = u_1 u_1^- u_2 u_2^-,$$

where these matrices have no role in the n-th row and column. As they normalize the t_{sn}'s and the t_{ns}'s, the proof can be completed in this case; this is where the general root system requires a carefully proved normalization result stated below as key lemma.

Case 2. Let $b = g_{nn} \neq 0$. Take arbitrary $1 \leq r, s \leq n-1$ for which $a = g_{rs} \neq 0$. Again, automatically $a, b \in R^*$. As in the previous case, let us concentrate on the

r-th and the n-th rows and the s-th and the n-th columns. Since there are no further non-zero entries in these rows and columns, any additions between them do not change other entries of the matrix, and only affect the submatrix at the intersection of the r-th and the n-th rows with the s-th and the n-th columns. Now, multiplying g by $t_{ns}(b^{-1})t_{sn}(1-b)t_{ns}(-1)t_{ns}(-b^{-1}(1-b))$ on the right, we obtain the matrix h, where this submatrix, which was initially equal to $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, will be replaced by $\begin{pmatrix} ab & 0 \\ 0 & 1 \end{pmatrix}$. At this point the proof can be finished in exactly the same way as in the previous case.

4.3. **Towards the proof - elementary parabolics.** The main role in the proofs in general is played by Levi decomposition for elementary parabolic subgroups. Denote by $m_k(\alpha)$ the coefficient of α_k in the expansion of α with respect to the fundamental roots

$$\alpha = \sum m_k(\alpha)\alpha_k, \quad 1 \leq k \leq l.$$

Fix any $r = 1, \dots, l$ - in fact, in the reduction to smaller rank it suffices to employ only terminal parabolic subgroups, even only the ones corresponding to the first and the last fundamental roots, $r = 1, r = l$.

Denote by

$$S = S_r = \{\alpha \in \Phi,\ m_r(\alpha) \geq 0\}$$

the r-th standard parabolic subset in Φ. As usual, the reductive part $\Delta = \Delta_r$ and the special part $\Sigma = \Sigma_r$ of the set $S = S_r$ are defined as

$$\Delta = \{\alpha \in \Phi,\ m_r(\alpha) = 0\}, \quad \Sigma = \{\alpha \in \Phi,\ m_r(\alpha) > 0\}.$$

The opposite parabolic subset and its special part are defined similarly as

$$S^- = S_r^- = \{\alpha \in \Phi,\ m_r(\alpha) \leq 0\}, \quad \Sigma^- = \{\alpha \in \Phi,\ m_r(\alpha) < 0\}.$$

Obviously, the reductive part of S_r^- equals Δ.

Denote by P_r the *elementary* maximal parabolic subgroup of the elementary group $E(\Phi, R)$. By definition

$$P_r = E(S_r, R) = \langle x_\alpha(\xi),\ \alpha \in S_r,\ \xi \in R\rangle.$$

By the Levi decomposition

$$P_r = L_r \leftthreetimes U_r = E(\Delta, R) \leftthreetimes E(\Sigma, R).$$

Recall that

$$L_r = E(\Delta, R) = \langle x_\alpha(\xi), \quad \alpha \in \Delta, \quad \xi \in R\rangle,$$

whereas

$$U_r = E(\Sigma, R) = \langle x_\alpha(\xi),\ \alpha \in \Sigma,\ \xi \in R\rangle.$$

A similar decomposition holds for the opposite parabolic subgroup P_r^-, whereby the Levi subgroup is the same as for P_r, but the unipotent radical U_r is replaced by the opposite unipotent radical $U_r^- = E(-\Sigma, R)$. As a matter of fact, we use Levi decomposition in the following form. It will be convenient to slightly change the notation and write $U(\Sigma, R) = E(\Sigma, R)$ and $U^-(\Sigma, R) = E(-\Sigma, R)$.

Lemma 4.5. *(Key Lemma) The group $\langle U^\sigma(\Delta, R), U^\rho(\Sigma, R)\rangle$, where $\sigma, \rho = \pm 1$, is the semidirect product of its normal subgroup $U^\rho(\Sigma, R)$ and the complementary subgroup $U^\sigma(\Delta, R)$.*

In other words, the subgroup $U^\pm(\Delta, R)$ normalizes each of the groups $U^\pm(\Sigma, R)$ so that, in particular, one has the following four equalities for products

$$U^\pm(\Delta, R)U^\pm(\Sigma, R) = U^\pm(\Sigma, R)U^\pm(\Delta, R).$$

Furthermore, the following four obvious equalities for intersections hold

$$U^\pm(\Delta, R) \cap U^\pm(\Sigma, R) = 1.$$

In particular, one has the following decompositions

$$U(\Phi, R) = U(\Delta, R) \leftthreetimes U(\Sigma, R), \quad U^-(\Phi, R) = U^-(\Delta, R) \leftthreetimes U^-(\Sigma, R).$$

5. Idea of proofs of theorems 1 and 2

Start with the following result which is easy, well known, and very useful.

Lemma 5.1. *The elementary Chevalley group $E(\Phi, R)$ is generated by unipotent root elements $x_\alpha(\xi)$, $\alpha \in \pm\Pi$, $\xi \in R$, corresponding to the fundamental and negative fundamental roots.*

Proof. Indeed, every root is conjugate to a fundamental root by an element of the Weyl group, while the Weyl group itself is generated by the fundamental reflections w_α, $\alpha \in \Pi$. Thus, the elementary group $E(\Phi, R)$ is generated by the root unipotents $x_\alpha(\xi)$, $\alpha \in \Pi$, $\xi \in R$, and the elements $w_\alpha(1)$, $\alpha \in \Pi$. It remains only to observe that $w_\alpha(1) = x_\alpha(1)x_{-\alpha}(-1)x_\alpha(1)$.

Further, let $B = B(\Phi, R)$ and $B^- = B^-(\Phi, R)$ be a pair of opposite Borel subgroups containing $T = T(\Phi, R)$, standard with respect to the given order. Recall that B and B^- are semidirect products $B = T \leftthreetimes U$ and $B^- = T \leftthreetimes U^-$, of the torus T and their unipotent radicals

$$U = U(\Phi, R) = \langle x_\alpha(\xi),\ \alpha \in \Phi^+,\ \xi \in R\rangle,$$

$$U^- = U^-(\Phi, R) = \langle x_\alpha(\xi),\ \alpha \in \Phi^-,\ \xi \in R\rangle.$$

Recall that a subset S in Φ is *closed*, if for any two roots $\alpha, \beta \in S$ whenever $\alpha + \beta \in \Phi$, already $\alpha + \beta \in S$. For closed S, define $E(S) = E(S, R)$ to be the subgroup generated by all elementary root unipotent subgroups X_α, $\alpha \in S$:

$$E(S, R) = \langle x_\alpha(\xi), \quad \alpha \in S, \quad \xi \in R\rangle.$$

In this notation, U and U^- coincide with $E(\Phi^+, R)$ and $E(\Phi^-, R)$, respectively. The groups $E(S, R)$ are particularly important in the case where $S \cap (-S) = \varnothing$. In this case $E(S, R)$ coincides with the *product* of root subgroups X_α, $\alpha \in S$, in some/any fixed order.

Again, let $S \subseteq \Phi$ be a closed set of roots; then S can be decomposed into a disjoint union of its *reductive* = *symmetric* part S^r, consisting of those $\alpha \in S$, for which $-\alpha \in S$, and its *unipotent* part S^u, consisting of those $\alpha \in S$, for which $-\alpha \notin S$. The set S^r is a closed root subsystem, whereas the set S^u is special. Moreover, S^u is an *ideal* of S (i.e., if $\alpha \in S$, $\beta \in S^u$ and $\alpha + \beta \in \Phi$, then $\alpha + \beta \in S^u$).

Levi decomposition shows that the group $E(S, R)$ decomposes into the semidirect product $E(S, R) = E(S^r, R) \leftthreetimes E(S^u, R)$ of its Levi subgroup $E(S^r, R)$ and its unipotent radical $E(S^u, R)$.

5.1. **Reduction to smaller rank.** As mentioned earlier, the proofs depend on the reduction of rank as in the following theorem.

Theorem 5.2. *Let Φ be a reduced irreducible root system of rank $l \geq 2$, and R be a commutative ring.*

(a) Suppose that for subsystems $\Delta = \Delta_1, \Delta_l$ the elementary Chevalley group $E(\Delta, R)$ admits unitriangular factorisation

$$E(\Delta, R) = (U(\Delta, R)U^-(\Delta, R))^L.$$

Then the elementary Chevalley group $E(\Phi, R)$ admits unitriangular factorisation

$$E(\Phi, R) = (U(\Phi, R)U^-(\Phi, R))^L.$$

of the same length $2L$.

(b) Suppose that for subsystems $\Delta = \Delta_1, \Delta_l$ the elementary Chevalley group $E(\Delta, R)$ admits the Gauss decomposition

$$E(\Delta, R) = H(\Delta, R)U(\Delta, R)U^-(\Delta, R)) \cdots U^{\pm}(\Delta, R)$$

of length L. Then, the elementary Chevalley group $E(\Phi, R)$ admits the Gauss decomposition

$$E(\Phi, R) = H(\Phi, R)U(\Phi, R)U^-(\Phi, R)) \cdots U^{\pm}(\Phi, R)$$

of the same length L.

Clearly, Theorem 2 immediately follows from Theorem 5.2 and the rank 1 case; so, it only remains to prove Theorem 5.2.

Observation. If Y is a subset in $E(\Phi, R)$ and if X is a symmetric generating set satisfying $XY \subseteq Y$, then clearly $Y = G$.

Therefore, to prove (a), we will prove $XY \subseteq Y$ with $X = \{x_\alpha(\xi) \mid \alpha \in \pm\Pi,\ \xi \in R\}$ and $Y = (U(\Phi, R)U^-(\Phi, R))^L$. The proof of (b) is similar with the same X and $Y = H(\Phi, R)U(\Phi, R)U^-(\Phi, R) \cdots U^{\pm}(\Phi, R)$.

Proof of Theorem 5.2. As we noted, the group G is generated by the fundamental root elements $X = \{x_\alpha(\xi) \mid \alpha \in \pm\Pi,\ \xi \in R\}$. Thus, to prove (a), it suffices to prove that $XY \subseteq Y$ where $Y = (U(\Phi, R)U^-(\Phi, R))^L$.

Fix a fundamental root unipotent $x_\alpha(\xi)$. Since $\mathrm{rk}(\Phi) \geq 2$, the root α belongs to at least one of the subsystems $\Delta = \Delta_r$, where $r = 1$ or $r = l$, generated by all fundamental roots, except for the first or the last one, respectively. Set $\Sigma = \Sigma_r$ and express $U^{\pm}(\Phi, R)$ in the form

$$U(\Phi, R) = U(\Delta, R)U(\Sigma, R), \quad U^-(\Phi, R) = U^-(\Delta, R)U^-(\Sigma, R).$$

We see that

$$Y = (U(\Delta, R)U^-(\Delta, R))^L (U(\Sigma, R)U^-(\Sigma, R))^L.$$

Since $\alpha \in \Delta$, one has $x_\alpha(\xi) \in E(\Delta, R)$, so that the inclusion $x_\alpha(\xi)Y \subseteq Y$ immediately follows from the assumption. This completes the proof of theorem 5.2 (a). The proof of (b) is entirely similar with

$$Y = H(\Phi, R)U(\Phi, R)U^-(\Phi, R)\dots U^{\pm}(\Phi, R).$$

Remarks. To prove theorems 1 and 2, in theorem 5 above, one needs the decomposition of $E(\Delta, R)$ only for subsystems Δ whose union contains all the fundamental roots. These subsystems do not have to be terminal as in theorem 5.

References

[1] Abe, E., *Chevalley groups over local rings.* — Tôhoku Math. J. **21**, no. 3, (1969), 474–494.

[2] Abe, E. and Suzuki, K., *On normal subgroups of Chevalley groups over commutative rings.* — Tôhoku Math. J. **28**, no. 1, (1976), 185–198.

[3] Baoquan, Chen and Kaufman, A., 3D *volume rotation using shear transformations.* — Graph. Models **62** (2000), 308–322.

[4] Bilu, Y., *Quadratic factors of f(x)-g(y)*, Acta Arith. **90**, no. 4, (1999), 341–355.

[5] Huanyin, Chen and Miaosen, Chen, *On products of three triangular matrices over associative rings.* — Linear Algebra Applic. **387** (2004), 297–311.

[6] Cooke, G. and Weinberger, P. J., *On the construction of division chains in algebraic number rings, with applications to* SL_2, Commun. Algebra, **3** (1975), 481–524.

[7] Detinko, A. S., Flannery, D. L. and Hulpke, A., *Algorithms for arithmetic groups with the congruence subgroup property*, J. Algebra **421** (2015), 234–259.

[8] Erovenko, I. V. and Rapinchuk, A. S., *Bounded generation of some S-arithmetic orthogonal groups.* — C. R. Acad. Sci. **333**, no. 5, (2001), 395–398.

[9] Pengwei, Hao, *Customizable triangular factorizations of matrices*, Linear Algebra Applic., **382** (2004), 135–154.

[10] Iwahori, N. and Matsumoto, H., On some Bruhat decomposition and structure of the Hecke rings of p-adic Chevalley groups, Publ. Math. Inst. Haut. Etudes Sci. **25** (1965), 5–48.

[11] Kulkarni, M., Muller, P. & Sury, B., *Quadratic factors of f(x)-g(y)*, Indagationes Math., **18** (2007), 233–243.

[12] Lazard, M., *Groupes analytiques p-adiques* (French), Inst. Hautes tudes Sci. Publ. Math. **26** (1965), 389–603.

[13] Yang, Lei, Hao, Pengwei and Wu, Dapeng, *Stabilization and optimization of PLUS factorization and its application to image coding.* — J. Visual Communication & Image Representation **22**, no. 1, (2011), 9–22.

[14] Liehl, B., *Beschränkte Wortlänge in* SL_2. — Math. Z. **186** (1984), 509–524.

[15] Liebeck, M. and Pyber, L., *Finite linear groups and bounded generation.* — Duke Math. J. **107** (2001), 159–171.

[16] Lubotzky, A., *Subgroup growth and congruence subgroups*, Invent. Math. **119**, no. 2, (1995), 267–295.

[17] Niyomsataya, T., Miri, A. and Nevins, M., *An Application of the Bruhat Decomposition to the Design of Full Diversity Unitary SpaceTime Codes*, IEEE Transactions on Information Theory, **55**, (2009), 232–244.

[18] Nikolov N. and Sury, B., *Bounded generation of wreath products,* Journal of Group Theory, **18**, no. 6, (2015), 951–959.

[19] Odeh, O. H., Olesky, D. D. and Driessche, P. Van den, *Bruhat decomposition and numerical stability*, SIAM J. Matrix Anal. Appl. **19** (1998), 89–98.

[20] Paeth, A., *A fast algorithm for general raster rotation.* — In: Graphics Gems, Acad. Press (1990), 179–195.

[21] Platonov, V. P and Rapinchuk, A. S., *Abstract characterizations of arithmetic groups with the congruence property*, Soviet Math. Dokl. **44**, no. 1, (1992), 342–347.

[22] Platonov V. P. and Sury, B., *Adelic profinite groups* - Jour. of Algebra, **193** (1997), 757–763.

[23] Sury, B., *An interesting consequence of the Heisenberg construction*, Elemente der Mathematik, **63** (2008), 184–188.

[24] Sury, B., Some number-theoretic identities from group actions, Rendiconti del circolo matematico di Palermo, **58** (2009), 99–108.

[25] Sury, B., *Arithmetic groups and Salem numbers* - Manuscripta Math., **75** (1992), 97–102.

[26] Sury, B., *Bounded generation does not imply finite presentation* – Comm. Alg., **25** (1997), 1673–1683.

[27] Yiyuan, She and Pengwei, Hao, *On the necessity and sufficiency of PLUS factorizations*, Linear Algebra Applic., **400** (2005), 193–202.

[28] Smolensky, A. V., Sury, B. and Vavilov, N. A., Unitriangular factorizations of Chevalley groups, Zapiski Nauchnyh Seminarov POMI, **388** (2011), 17– 47(in Russian).
English translation: J. Math. Sci. (N. Y.) **183**, no. 5, (2012), 584–599. **Read G as E in the main theorems!**

[29] Smolensky, A. V., Sury, B. and Vavilov, N. A., *Gauss decomposition for Chevalley groups - Revisited,* International Journal of Group Theory, **1**, no. 1, (2012), 3–16.

[30] Stein, M. R., *Surjective stability in dimension* 0 *for* K_2 *and related functors*, Trans. Amer. Math. Soc., **178** (1973), 176–191.

[31] Sury, B. and Venkataramana, T. N., *Generators for all principal congruence subgroups of SL(n,Z) with* $n \geq 3$ - Proc. Amer Math. Soc., **122** (1994), 355–358.

[32] Tavgen, O. I., *Finite width of arithmetic subgroups of Chevalley groups of rank* ≥ 2, Soviet Math. Doklady, **41**, no. 1, (1990), 136–140.

[33] Tavgen, O. I., *Bounded generation of normal and twisted Chevalley groups over the rings of* S*-integers*, Contemp. Math., **131**, no. 1, (1992), 409–421.

[34] Toffoli, T., and Quick, J., *Three dimensional rotations by three shears.* — Graphical Models & Image Processing **59** (1997), 89–96.

B. Sury
Statistics & Mathematics Unit, Indian Statistical Institute
8th Mile Mysore Road, Bangalore-560 059, India
E-mail: *sury@ms.isibang.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 97-102

A SIMPLE PROOF OF BURNSIDE'S CRITERION FOR ALL GROUPS OF ORDER n TO BE CYCLIC

SIDDHI PATHAK

(Received : 10 - 09 - 2015 ; Revised : 26 - 01 - 2016)

Abstract. This note gives a simple proof of a famous theorem of Burnside, namely, all groups of order n are cyclic if and only if $(n, \phi(n)) = 1$, where ϕ denotes the Euler totient function.

1. Introduction

The question of determining the number of isomorphism classes of groups of order n has long been of interest to mathematicians. One can ask a more basic question: For what natural numbers n, is there only one isomorphism class of groups of order n? Since we know that there exists a cyclic group of every order, this question reduces to finding natural numbers n such that all groups of order n are cyclic. The answer is given in the following well-known theorem by Burnside [1]. Let ϕ denote the Euler function.

Theorem 1.1. *All groups of order n are cyclic if and only if* $(n,\ \phi(n)) = 1$.

Many different proofs of this fact are available. Practically all of them are inaccessible to the undergraduate student since they use Burnside's transfer theorem and representation theory [2]. Here, we would like to give another proof of this theorem which is elementary and uses only basic Sylow theory. Throughout this note, n denotes a positive integer and C_n denotes the cyclic group of order n.

2. Groups of order pq

Let p and q be two distinct primes, $p < q$. In this section, we investigate the structure of groups of order pq. The two cases to be considered are when $p \mid q-1$ and $p \nmid q-1$.

First, let us suppose that $p \nmid q-1$. In this case, every group of order pq is cyclic. Indeed, let G be a group of order pq. Let n_p be the number of p-Sylow subgroups and n_q be the number of q-Sylow subgroups of G. Then by Sylow's theorem

$$n_q \equiv 1 \bmod q \text{ and } n_q \mid p.$$

Since $p < q$, $n_q = 1$, the q-Sylow subgroup, say Q, is normal in G. Again by Sylow's theorem

---

2010 Mathematics Subject Classification : 20D60, 20E99.
Key words and phrases : Groups of a given order, cyclic groups, Sylow's theorems.

$$n_p \equiv 1 \bmod p \text{ and } n_p \mid q.$$

Since q is prime, either $n_p = 1$ or $n_p = q$. But $p \nmid q-1$. Hence, $n_p = 1$. Thus, the p-Sylow subgroup, say P, is also normal in G. Further, since the order of non-identity elements of P and Q are co-prime, $P \cap Q = \{e\}$. Now, if $a \in P$ and $b \in Q$ then consider the element $c := aba^{-1}b^{-1} \in G$. The normality of Q implies that $aba^{-1} \in Q$ and hence, $c \in Q$. On the other hand, the normality of P implies that $ba^{-1}b^{-1} \in P$ and hence, $c \in P$. Thus, $c \in P \cap Q = \{e\}$. Therefore, the elements of P and Q commute with each other. This gives us a group homomorphism

$$\Psi : P \times Q \to G$$

such that $\Psi(a,b) = ab$. Since, $P \cap Q = \{e\}$, Ψ is injective. $|P \times Q| = |G|$ implies that Ψ is also surjective and hence, an isomorphism. As P and Q are cyclic groups of distinct prime order, $P \times Q$ is cyclic and so is G. Therefore, if $p \nmid q-1$, then all groups of order pq are cyclic.

Now, suppose $p \mid q-1$. We claim that in this case, there exists a group of order pq which is not cyclic.

Note that since $p \mid q-1$, there exists an element in $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ of order p, say α_p. To see this, note that

$$\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z}) \simeq (\mathbb{Z}/q\mathbb{Z})^* \simeq C_{q-1},$$

and a cyclic group of order n contains an element of order d, for every divisor d of n. Thus, we get a group homomorphism, say θ, from C_p to $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ by sending a generator of C_p to α_p. Denote $\theta(u)$ by θ_u. Clearly, θ is a non-trivial map. We define the semi-direct product, $C_p \ltimes_\theta C_q$ as follows.

As a set, $C_p \ltimes_\theta C_q := \{(u,v) : u \in C_p \text{ and } v \in C_q\}$. The group operation on this set is defined as

$$(u,v).(u',v') = (uu', \theta_u(v)v'). \tag{1}$$

One can check that this operation is indeed associative and makes $C_p \ltimes_\theta C_q$ into a group. To see that this group is non-abelian, consider (u,v) and (u',v') in $C_p \ltimes_\theta C_q$. Thus

$$(u',v').(u,v) = (u'u, \theta_{u'}(v')v),$$

which is not equal to $(u,v).(u',v')$ as evaluated in (1) since θ is non-trivial. Thus, if $p \mid q-1$, then there exists a group of order pq which is not abelian, and hence in particular not cyclic.

Remark. It may be observed that given any group G of order pq, one can show that it is either cyclic or isomorphic to the semi-direct product constructed above. Thus, if $p \mid q-1$, there are exactly two isomorphism classes of groups of order pq.

3. Proof of the *only if* part

Suppose all groups of order n are cyclic, i.e, there is only one isomorphism class of groups of order n. Since $\mathbb{Z}/p^2\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ are 2 non-isomorphic groups of order p^2, we see that n is squarefree.

Proof. Let us note that if $n = \prod_{i=1}^{k} p_i$ where, $p_1, \ldots, p_k$ are distinct primes and $p_1 < \cdots < p_k$, then $(n, \phi(n)) = 1 \iff p_i \nmid (p_j - 1)$, for all $1 \le i < j \le k$.

Now, suppose n is squarefree and $(n, \phi(n)) > 1$, i.e, there exists a p_i such that $p_i \mid (p_j - 1)$ for some $1 \le i < j \le k$. As seen in the earlier section, there exists a group, $\mathcal{G}$ of order $p_i p_j$ that is not cyclic. Thus, $\mathcal{G} \times C_{n/p_i p_j}$ is a group of order n and is not cyclic. This contradicts our assumption that all groups of order n are cyclic. Hence, n and $\phi(n)$ must be coprime. □

4. Proof of the *if* part

The condition that $(n, \phi(n)) = 1$ helps us to infer that it is enough to consider only those n that are squarefree.

Our proof hinges upon the following crucial lemma.

Lemma 4.1. *Let G be a finite group such that every proper subgroup of G is abelian. Then either G has prime order, or G has a non-trivial, proper, normal subgroup i.e, G is not simple.*

Proof. Let G be a group of order n. By a maximal subgroup of G, we will mean a nontrivial proper subgroup H of G such that, for any subgroup H' of G that contains H, either $H' = G$ or $H' = H$ itself.

Let M denote a maximal subgroup of G. Let $|M| = m$. Suppose $M = \{e\}$,i.e, G contains no nontrivial proper subgroup. Sylow's first theorem thus implies that the order of G must be prime.

Suppose n is not prime. Hence, $m \ge 2$. Let $N_G(M)$ denote the normalizer of M in G. Recall that

$$N_G(M) = \{g \in G : gMg^{-1} = M\}.$$

If M is normal in G, then clearly G is not simple. Therefore, let us suppose that M is not normal. Hence, $N_G(M) \ne G$. Since $M \subseteq N_G(M)$ and M is maximal, $N_G(M) = M$. Let the number of conjugates of M in G be r, $r > 1$. The number of conjugates of a subgroup in a group is equal to the index of its normalizer. Therefore

$$r = [G : N_G(M)] = [G : M] = \frac{n}{m}.$$

Let $\{M_1, \cdots, M_r\}$ be the set of distinct conjugates of M. Suppose $M_i \cap M_j \ne \{e\}$ for some $1 \le i < j \le r$. Let $K_1 := M_i \cap M_j$. Since M_i and M_j are abelian by hypothesis we have

$$K_1 \lhd M_i \quad , K_1 \lhd M_j. \tag{2}$$

Therefore K_1 is normal in the group generated by M_i and M_j. Since conjugates of maximal subgroups are themselves maximal, the group generated by M_i and M_j is G. Thus, K_1 is normal in G and hence G is not simple.

Therefore, we suppose that all the conjugates of M intersect trivially. Let $V := \cup_{i=1}^{r} M_i$. Then

$$|V| = r(m-1) + 1 = n - \left[\frac{n}{m} - 1\right] < n.$$

Thus, $\exists\ y \in G$, $y \notin V$.

If G is a cyclic group generated by y (of composite order), then the subgroup of G generated by y^k for any $k|n$, $k \neq 1, n$ is a non-trivial normal subgroup. So we can assume that the group generated by y is a proper subgroup of G. Let L be a maximal subgroup containing the subgroup of G generated by y. Since, $y \notin V, L \neq M_i\ \forall\ 1 \leq i \leq r$. If L is normal in G, then G is clearly not simple. Therefore, suppose that L is not normal in G.

Let the number of conjugates of L in G be s, $s > 1$. Let $\{L_1, \cdots, L_s\}$ be the set of distinct conjugates of L in G. If any two distinct conjugates of L or a conjugate of L and a conjugate of M intersect non-trivially, then the corresponding intersection is a normal subgroup of G by an argument similar to the one given above. Thus, G is not simple. Hence, it suffices to assume that

$$M_i \cap M_j = \{e\}, \tag{3}$$

$$M_i \cap L_q = \{e\}, \tag{4}$$

$$L_p \cap L_q = \{e\}, \tag{5}$$

for all $1 \leq i < j \leq r$, for all $1 \leq p < q \leq s$.

Let $|L| = l$, $l \geq 2$. Since L is not normal in G but is maximal, $N_G(L) = L$. Thus, the number of conjugates of L in G is

$$s = [G : N_G(L)] = [G : L] = \frac{n}{l}.$$

Let $W := \cup_{p=1}^{s} L_p$. By (3), (4) and (5),

$$\begin{aligned} |V \cup W| &= r(m-1) + s(l-1) + 1 = n - \frac{n}{m} + n - \frac{n}{l} + 1 \\ &= 2n - n(\frac{1}{m} + \frac{1}{l}) + 1 \geq 2n - n + 1 > n, \end{aligned}$$

since $m, l \geq 2$. But $V \cup W \subseteq G$. Therefore, $|V \cup W| \leq n$. This is a contradiction. Hence, G must have a nontrivial proper normal subgroup. □

We will now prove that if $(n, \phi(n)) = 1$, then all groups of order n are cyclic. As seen earlier, we are reduced to the case when n is squarefree.

Proof. We will proceed by induction on the number of prime factors of n. For the base case, assume that n is prime. Lagrange's theorem implies that any group of prime order is cyclic. Thus, the base case of our induction is true.

Now suppose that the result holds for all n with at most $k-1$ distinct prime factors, for some $k > 1$. Let $n = p_1 \cdots p_k$ for distinct primes $p_1, \cdots, p_k$ and $p_1 < p_2 < \ldots < p_k$. Since $k \geq 2$, Sylow's first theorem implies that G has nontrivial proper subgroups. Let P be a proper subgroup of G. Hence, $|P|$ has fewer prime factors than k. Therefore, by induction hypothesis, P is cyclic and hence abelian. Thus, every proper subgroup of G is abelian. By Lemma 4.1, G has a nontrivial proper normal subgroup, say N. The induction hypothesis implies that G/N is cyclic. Therefore, G/N has a subgroup of index p_i for some $1 \leq i \leq k$. Let this subgroup be denoted by $\mathfrak{H}$. By the correspondence theorem of groups, all subgroups of G/N correspond to subgroups of G containing N. Let the subgroup of G corresponding to $\mathfrak{H}$ via the above correspondence be H, i.e, $\mathfrak{H} = H/N$. Since G/N is abelian, $\mathfrak{H} \triangleleft G/N$ and hence, $H \triangleleft G$. By the third isomorphism theorem of groups

$$G/N \Big/ H/N \simeq G/H.$$

Since, $[G/N : \mathfrak{H}] = p_i$, $[G : H] = p_i$. Thus, G has a normal subgroup of index p_i, namely, H. Note that H is cyclic. In particular

$$H \simeq C_a, \tag{6}$$

where $a = p_1 \cdots p_{i-1}p_{i+1} \cdots p_k$. Let K be a p_i- Sylow subgroup of G. Thus

$$K \simeq C_{p_i}. \tag{7}$$

Consider the map $\Phi : K \to \mathrm{Aut}(H)$ that sends an element $k \in K$ to the automorphism γ_k where, γ_k is conjugation by k. Since $H \triangleleft G$, γ_k is a well-defined map from H to H. Therefore, Φ is a well-defined group homomorphism. Since, $ker(\Phi)$ is a subgroup of K and K has prime order, either $ker(\Phi) = \{e\}$ or $ker(\Phi) = K$. Suppose, $ker(\Phi) = \{e\}$. Then, $\Phi(K)$ is isomorphic to a subgroup of $\mathrm{Aut}(H)$. By the induction hypothesis, H is isomorphic to the cyclic group of order $|H| = p_1 \cdots p_{i-1}p_{i+1} \cdots p_k$. It follows that

$$H \simeq \prod_{j=1, j\neq i}^{k} \mathbb{Z}/p_j\mathbb{Z}.$$

Since $\quad \mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) \simeq (\mathbb{Z}/p\mathbb{Z})^* \quad$ for any prime p, therefore

$$\mathrm{Aut}(H) \simeq \prod_{j=1, j\neq i}^{k} (\mathbb{Z}/p_j\mathbb{Z})^*.$$

Hence $|\mathrm{Aut}(H)| = \prod_{j=1, j\neq i}^{k}(p_j - 1)$. Therefore, by Lagrange's theorem, $|K|$ divides $|\mathrm{Aut}(H)|$, i.e,

$$p_i \Big| \prod_{j=1, j\neq i}^{k} (p_j - 1).$$

Since $(n, \phi(n)) = 1$, we see that $p_i \nmid (p_j - 1)$ for any $1 \leq i, j \leq k$. We thus arrive at a contradiction. Hence, $ker(\Phi) = K$. Let $k \in ker(\Phi)$ i.e, γ_k is the identity homomorphism. Since $ker(\Phi) = K$, $kh = hk$ for all $h \in H$ and for all $k \in K$. We now claim that $G \simeq H \times K$. To prove this claim, consider the map $\Psi : H \times K \to G$ sending a tuple (h, k) to the product hk. Since the elements of H and K commute with each other, Ψ is a group homomorphism. H has no element of order p_i. Thus, $H \cap K = \{e\}$. This implies that Ψ is injective and hence surjective as $|H \times K| = |G|$. Thus Ψ is the desired isomorphism. By (6) and (7) $G \simeq C_n$. It follows that every group of order n is cyclic. $\square$

Acknowledgement: The above proof was a result of a long discussion with Prof. M. Ram Murty. I would like to thank him profusely for his guidance and help in writing the note. I would also like to thank the referee for useful comments on an earlier version of this note.

References

[1] Dixon, J., *Problems in Group Theory*, Dover Books, 1973, pg. **92**, **2.55**.
[2] Scott, W. R., *Group Theory*, Prentice Hall, 1964, pg. **217**.

Siddhi Pathak
Department of Mathematics and Statistics
Queen's University, Kingston, Ontario, Canada, K7L 3N6
E-mail: *siddhi@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 103-111

RATIONAL ANGLED HYPERBOLIC POLYGONS

JACK S. CALCUT

(Received : 16 - 09 - 2015 ; Revised : 18 - 12 - 2015)

Abstract. We prove that every rational angled hyperbolic triangle has transcendental side lengths and that every rational angled hyperbolic quadrilateral has at least one transcendental side length. Thus, there does not exist a rational angled hyperbolic triangle or quadrilateral with algebraic side lengths. We conjecture that there does not exist a rational angled hyperbolic polygon with algebraic side lengths.

1. Introduction

Herein, an angle is *rational* provided its radian measure θ is a rational multiple of π, written $\theta \in \mathbb{Q}\pi$. Rational angles may also be characterized as angles having rational degree measure and as angles commensurable with a straight angle. *Algebraic* means algebraic over $\mathbb{Q}$. Let $\mathbb{A} \subset \mathbb{C}$ denote the subfield of algebraic numbers. If $\theta \in \mathbb{Q}\pi$, then $\cos\theta \in \mathbb{A}$ and $\sin\theta \in \mathbb{A}$.

In Euclidean geometry, there exists a rich collection of rational angled polygons with algebraic side lengths. Besides equilateral triangles, one has the triangles in the Ailles rectangle [Ail71] and in the golden triangle (see Figure 1). Further, let

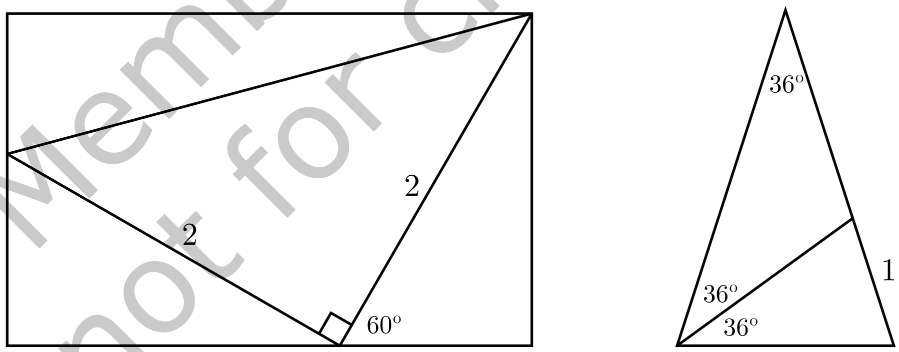


Figure 1. Ailles rectangle and the golden triangle.

2010 Mathematics Subject Classification : Primary 51M10; Secondary 11J81, 11J72.
Key words and phrases : Hyperbolic, triangle, quadrilateral, polygon, rational, angle, transcendental, the Hermite-Lindemann theorem, spherical, Gaussian curvature.

$\sigma(d)$ denote the number of similarity types of rational angled Euclidean triangles containing a representative with side lengths each of degree at most d over $\mathbb{Q}$. Then

(1) $\sigma(d) < \infty$ for each $d \in \mathbb{Z}^+$. A proof of this fact uses three ingredients: (i) the Euclidean law of cosines, (ii) if φ is Euler's totient function, $n > 2$, and k is coprime to n, then the degree over $\mathbb{Q}$ of $\cos(2k\pi/n)$ equals $\varphi(n)/2$, and (iii) $\varphi(n) \geq \sqrt{n/2}$ for each $n \in \mathbb{Z}^+$ (see [Cal10]).

(2) $\sigma(1) = 1$. This class is represented by an equilateral triangle (see [CG06, pp. 228–229] or [Cal10, p. 674]). While Pythagorean triple triangles have integer side lengths, they have irrational acute angles; a simple approach to this result uses unique factorization of Gaussian integers [Cal09]. In fact, the measures of the acute angles in a Pythagorean triple triangle are transcendental in radians and degrees [Cal10].

(3) $\sigma(2) = 14$. Parnami, Agrawal, and Rajwade [PAR82] proved this result using Galois theory. The 14 similarity types are listed in Table 1. It is an exercise to construct representatives for these 14 classes using the triangles above in Figure 1.

60–60–60	45–45–90	30–60–90	15–75–90
30–30–120	30–75–75	15–15–150	30–45–105
45–60–75	15–45–120	15–60–105	15–30–135
36–36–108	36–72–72		

TABLE 1. The 14 similarity types (in degrees) of rational angled Euclidean triangles with side lengths each of degree at most two over $\mathbb{Q}$.

A *polar rational polygon* is a convex rational angled Euclidean polygon with integral length sides. The study of such polygons was initiated by Schoenberg [Sco64] and Mann [Man65] and continues with the recent work of Lam and Leung [LL00].

In sharp contrast, there seem to be no well-known rational angled hyperbolic polygons with algebraic side lengths. Throughout, $\mathbb{H}^2$ denotes the hyperbolic plane of constant Gaussian curvature $K = -1$. We adopt Poincaré's conformal disk model of $\mathbb{H}^2$ for our figures.

Below, we prove that every rational angled hyperbolic triangle has transcendental side lengths and every rational angled hyperbolic quadrilateral (simple, but not necessarily convex) has at least one transcendental side length. Our basic tactic is to obtain relations using hyperbolic trigonometry to which we may apply the following theorem of Hermite and Lindemann.

Theorem 1.1 (Hermite-Lindemann [Niv56, p. 117])**.** *If $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{A}$ are pairwise distinct, then $e^{\alpha_1}, e^{\alpha_2}, \ldots, e^{\alpha_n}$ are linearly independent over $\mathbb{A}$.*

We conjecture that there does not exist a rational angled hyperbolic polygon with algebraic side lengths. We ask whether a rational angled hyperbolic polygon can have any algebraic side lengths whatsoever.

Throughout, polygons are assumed to be simple, nondegenerate (meaning consecutive triples of vertices are noncollinear), and not necessarily convex. In particular, each internal angle of a polygon has radian measure in $(0, \pi) \cup (\pi, 2\pi)$. Our closing section discusses generalized hyperbolic triangles with one ideal vertex, spherical triangles, and the dependency of our results on Gaussian curvature.

2. Hyperbolic Triangles

Consider a triangle $ABC \subset \mathbb{H}^2$ as in Figure 2. Recall the following hyperbolic

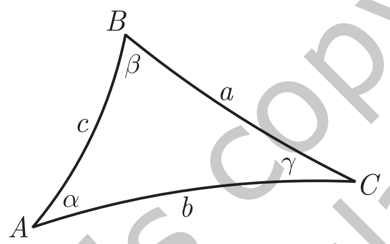


Figure 2. Hyperbolic triangle ABC.

trigonometric identities (see Ratcliffe [Rat06, p. 82]) known respectively as the hyperbolic law of sines, the first hyperbolic law of cosines, and the second hyperbolic law of cosines

$$\frac{\sinh a}{\sin \alpha} = \frac{\sinh b}{\sin \beta} = \frac{\sinh c}{\sin \gamma}, \tag{HLOS}$$

$$\cos \gamma = \frac{\cosh a \cosh b - \cosh c}{\sinh a \sinh b}, \text{ and} \tag{HLOC1}$$

$$\cosh c = \frac{\cos \alpha \cos \beta + \cos \gamma}{\sin \alpha \sin \beta}. \tag{HLOC2}$$

Lemma 2.1. *Let $ABC \subset \mathbb{H}^2$ be a rational angled triangle labelled as in Figure 2. Then, every side length of ABC is transcendental.*

Proof. Define $x := \cosh c > 0$. Then

$$xe^0 = \frac{1}{2}e^c + \frac{1}{2}e^{-c}. \tag{2.1}$$

As $\alpha, \beta, \gamma \in \mathbb{Q}\pi$ and $\mathbb{A}$ is a field, HLOC2 implies that $x \in \mathbb{A}$. As $c > 0$, the Hermite-Lindemann theorem applied to (2.1) implies that $c \notin \mathbb{A}$. The proofs for a and b are similar. $\square$

Corollary 2.2. *Each regular rational angled hyperbolic polygon has transcendental side length, radius, and apothem.*

Proof. Consider the case of a quadrilateral $ABCD \subset \mathbb{H}^2$. By an isometry of $\mathbb{H}^2$, we may assume $ABCD$ appears as in Figure 3 where E is the origin and F is the midpoint of segment AB. Triangle AEF has radian angle measures $\theta/2$, $\pi/4$, and

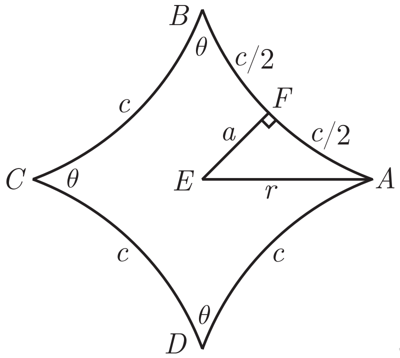


FIGURE 3. Regular hyperbolic quadrilateral $ABCD$.

$\pi/2$. Lemma 2.1 implies that $c/2$, r, and a are transcendental. The cases $n \neq 4$ are proved similarly. $\square$

We will use the following lemma in the next section on quadrilaterals.

Lemma 2.3. *There does not exist a triangle $ABC \subset \mathbb{H}^2$, labelled as in Figure 2, such that $\gamma \in \mathbb{Q}\pi$, $\alpha + \beta \in \mathbb{Q}\pi$, and $a, b \in \mathbb{A}$.*

Proof. Suppose, by way of contradiction, that such a triangle exists. Define

$$\begin{aligned} x &:= \cos(\alpha+\beta) \in (-1,1) \cap \mathbb{A}, \\ y &:= \sin\gamma \in (0,1] \cap \mathbb{A}, \\ z &:= \cos\gamma \in (-1,1) \cap \mathbb{A}, \text{ and} \\ Z &:= \cosh c > 0. \end{aligned} \tag{2.2}$$

Then HLOC1 and HLOS yield

$$\begin{aligned} x = \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\ &= \frac{Z\cosh b - \cosh a}{\sinh b \sinh c} \cdot \frac{Z\cosh a - \cosh b}{\sinh a \sinh c} - y^2 \frac{\sinh a \sinh b}{\sinh^2 c}. \end{aligned}$$

As $\sinh^2 c = Z^2 - 1$, we obtain

$$\begin{aligned} & x\left(Z^2-1\right)\sinh a \sinh b \\ &\qquad = (Z\cosh b - \cosh a)(Z\cosh a - \cosh b) - y^2 \sinh^2 a \sinh^2 b. \end{aligned} \tag{2.3}$$

By HLOC1

$$Z = \cosh a \cosh b - z \sinh a \sinh b. \tag{2.4}$$

Make the substitution (2.4) in (2.3) and then, for $t = a$ and $t = b$, expand hyperbolics into exponentials using the standard identities

$$\cosh t = \frac{1}{2}e^t + \frac{1}{2}e^{-t} \quad \text{and} \quad \sinh t = \frac{1}{2}e^t - \frac{1}{2}e^{-t}.$$

Expand out the resulting equation by multiplication, multiply through by 64, and then subtract the right hand terms to the left side. Now, collect together terms whose exponentials have identical exponents as elements of $\mathbb{Z}[a,b]$. One obtains

$$(x-1)(z-1)^2 e^{3a+3b} + (\text{sum of lower order terms}) = 0. \tag{2.5}$$

Equation (2.5) is our relation. Its left hand side, denoted by R, is the sum of 25 terms each of the form pe^{ma+nb} for some $p \in \mathbb{Z}[x,y,z]$ and some integers m and n in $[-3,3]$. By definition, *lower order* terms of R have $m < 3$ or $n < 3$ (or both). As $a > 0$ and $b > 0$, the Hermite-Lindemann theorem applied to (2.5) implies that $x = 1$ or $z = 1$. This contradicts (2.2) and completes the proof of Lemma 2.3. □

3. Hyperbolic Quadrilaterals

Consider a quadrilateral $ABCD \subset \mathbb{H}^2$. Whether or not $ABCD$ is convex, at least one diagonal AC or BD of $ABCD$ lies inside $ABCD$ (this holds in neutral geometry). Relabelling the vertices of $ABCD$ if necessary, we assume BD lies inside $ABCD$ and that $ABCD$ is labelled as in Figure 4. The radian measures

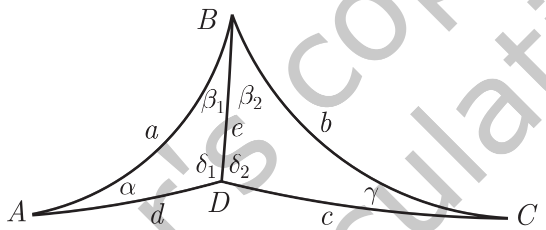


Figure 4. Quadrilateral $ABCD \subset \mathbb{H}^2$ with internal diagonal BD.

of the internal angles of $ABCD$ at B and D are $\beta = \beta_1 + \beta_2$ and $\delta = \delta_1 + \delta_2$ respectively.

Theorem 3.1. *If quadrilateral $ABCD$ is rational angled, then at least one of its side lengths a, b, c, or d is transcendental.*

Proof. We have $\alpha, \beta, \gamma, \delta \in \mathbb{Q}\pi$. Suppose, by way of contradiction, that $a, b, c, d \in \mathbb{A}$. Define

$$\begin{aligned} w &:= \cos\beta \in (-1,1) \cap \mathbb{A}, \\ x &:= \cos\delta \in (-1,1) \cap \mathbb{A}, \\ y &:= \sin\alpha \in [-1,1] \cap \mathbb{A} - \{0\}, \\ z &:= \sin\gamma \in [-1,1] \cap \mathbb{A} - \{0\}, \text{ and} \\ E &:= \cosh e > 0. \end{aligned} \tag{3.1}$$

Then HLOC1 and HLOS yield

$$\begin{aligned} w = \cos\beta = \cos(\beta_1+\beta_2) &= \cos\beta_1\cos\beta_2 - \sin\beta_1\sin\beta_2 \\ &= \frac{E\cosh a - \cosh d}{\sinh a \sinh e} \cdot \frac{E\cosh b - \cosh c}{\sinh b \sinh e} - \frac{\sinh d}{\sinh e}\sin\alpha \frac{\sinh c}{\sinh e}\sin\gamma. \end{aligned}$$

As $\sinh^2 e = E^2 - 1$, we obtain

$$w\left(E^2-1\right)\sinh a \sinh b$$
$$= (E\cosh a - \cosh d)(E\cosh b - \cosh c) - yz\sinh a \sinh b \sinh c \sinh d. \quad (3.2)$$

The analogous calculation beginning with $x = \cos\delta = \cos(\delta_1 + \delta_2)$ yields

$$x\left(E^2-1\right)\sinh c \sinh d$$
$$= (E\cosh d - \cosh a)(E\cosh c - \cosh b) - yz\sinh a \sinh b \sinh c \sinh d. \quad (3.3)$$

In each of the equations (3.2) and (3.3), move all terms to the right hand side and re-group to obtain two quadratic formulas in E given by

$$0 = A_1E^2 + B_1E + C_1 \quad \text{and}$$
$$0 = A_2E^2 + B_2E + C_2.$$

Inspection shows that $B_1 = B_2$. Define $B := B_1 = B_2$. We claim that $A_1 \neq 0$ and $A_2 \neq 0$. Suppose, by way of contradiction, that $A_1 = 0$. Expanding the hyperbolics in A_1 into exponentials and collecting terms by the exponents of their exponentials yields

$$-\frac{1}{4}(w-1)e^{a+b} + (\text{sum of lower order terms}) = 0.$$

The Hermite-Lindemann theorem implies $w = 1$ which contradicts (3.1). Hence $A_1 \neq 0$, and similarly $A_2 \neq 0$.

The quadratic formula yields

$$0 < E = \frac{-B \pm \sqrt{B^2 - 4A_1C_1}}{2A_1} = \frac{-B \pm \sqrt{B^2 - 4A_2C_2}}{2A_2},$$

where the $\pm$ signs are ambiguous and not necessarily equal. We have

$$-A_2B \pm A_2\sqrt{B^2 - 4A_1C_1} = -A_1B \pm A_1\sqrt{B^2 - 4A_2C_2},$$

and hence

$$A_1B - A_2B \pm A_2\sqrt{B^2 - 4A_1C_1} = \pm A_1\sqrt{B^2 - 4A_2C_2}.$$

Squaring both sides yields

$$(A_1B - A_2B)^2 + A_2^2\left(B^2 - 4A_1C_1\right) \pm 2(A_1B - A_2B)A_2\sqrt{B^2 - 4A_1C_1}$$
$$= A_1^2\left(B^2 - 4A_2C_2\right).$$

Therefore

$$(A_1B - A_2B)^2 + A_2^2\left(B^2 - 4A_1C_1\right) - A_1^2\left(B^2 - 4A_2C_2\right)$$
$$= \mp 2(A_1B - A_2B)A_2\sqrt{B^2 - 4A_1C_1}.$$

Again squaring both sides yields

$$\left[(A_1B - A_2B)^2 + A_2^2\left(B^2 - 4A_1C_1\right) - A_1^2\left(B^2 - 4A_2C_2\right)\right]^2$$
$$= 4(A_1B - A_2B)^2 A_2^2\left(B^2 - 4A_1C_1\right). \quad (3.4)$$

Subtract the right hand side of (3.4) to the left side, expand out by multiplication, and cancel the common factor $16A_1^2A_2^2 \neq 0$ to obtain

$$A_1^2C_2^2+A_2^2C_1^2-2A_1A_2C_1C_2+A_1B^2C_1+A_2B^2C_2-A_1B^2C_2-A_2B^2C_1=0. \quad (3.5)$$

At this point, use of a computer algebra system is recommended; we used MAGMA for our computations and Mathematica for independent verification. In (3.5), expand hyperbolics into exponentials and then expand out by multiplication. Multiply through the resulting equation by 4096 and then collect together terms whose exponentials have identical exponents as elements of $\mathbb{Z}[a,b,c,d]$. This yields

$$\begin{aligned}&(w-1)^2y^2z^2e^{(4a+4b+2c+2d)}+(x-1)^2y^2z^2e^{(2a+2b+4c+4d)}\\&+2(w-1)(1-x)y^2z^2e^{(3a+3b+3c+3d)}+(\text{sum of lower order terms})=0. \quad (3.6)\end{aligned}$$

Equation (3.6) is our relation. Its left hand side, denoted by R, is the sum of 1041 terms each of the form $pe^{ka+lb+mc+nd}$ for some $p\in\mathbb{Z}[w,x,y,z]$ and some integers k, l, m, and n in $[-4,4]$. Here, a *lower order* term is defined to be one whose exponential exponent $ka+lb+mc+nd$ is dominated by $4a+4b+2c+2d$, $2a+2b+4c+4d$, or $3a+3b+3c+3d$. This means, by definition, that either

(1) $k\le 4$, $l\le 4$, $m\le 2$, $n\le 2$, and at least one of these inequalities is strict,
(2) $k\le 2$, $l\le 2$, $m\le 4$, $n\le 4$, and at least one of these inequalities is strict, or
(3) $k\le 3$, $l\le 3$, $m\le 3$, $n\le 3$, and at least one of these inequalities is strict.

There are now two cases to consider.

Case 1. One of the sums $4a+4b+2c+2d$, $2a+2b+4c+4d$ and $3a+3b+3c+3d$ is greater than the other two. Then, the Hermite-Lindemann theorem implies that the corresponding polynomial coefficient in equation (3.6) must vanish, that is

$$(w-1)^2y^2z^2=0,\quad (x-1)^2y^2z^2=0,\quad \text{or}\quad 2(w-1)(1-x)y^2z^2=0.$$

By equations (3.1), we have $w\ne 1$, $x\ne 1$, $y\ne 0$, and $z\ne 0$. This contradiction completes the proof of Case 1.

Case 2. Two of the sums $4a+4b+2c+2d$, $2a+2b+4c+4d$, and $3a+3b+3c+3d$ are equal. Then, $a+b=c+d$. Recalling Figure 4, $a+b=c+d$ and two applications of the triangle inequality imply that D cannot lie inside triangle ABC. Similarly, B cannot lie inside triangle ADC. This implies that diagonal AC lies inside quadrilateral $ABCD$. Now, repeat the entire argument of the proof of Theorem 3.1 using diagonal AC in place of diagonal BD. We either obtain a contradiction as in Case 1 or we further obtain $a+d=b+c$. Taken together, $a+b=c+d$ and $a+d=b+c$ imply that $a=c$ and $b=d$. So, opposite sides of quadrilateral $ABCD$ are congruent. By SSS, triangles ABD and CDB are congruent. In particular, $\delta_1=\beta_2$. So, $\beta_1+\delta_1=\beta\in\mathbb{Q}\pi$ and triangle ABD contradicts Lemma 2.3. This completes the proof of Case 2 and of Theorem 3.1. □

4. Concluding Remarks

The results above have analogues for generalized hyperbolic triangles with one ideal vertex and for spherical triangles.

Lemma 4.1. *Let ABC be a generalized hyperbolic triangle with one ideal vertex C as in Figure 5. If $\alpha, \beta \in \mathbb{Q}\pi$, then the finite side length c is transcendental.*

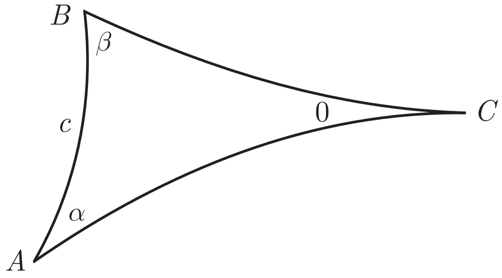


FIGURE 5. Generalized hyperbolic triangle ABC with one ideal vertex C.

Proof. By Ratcliffe [Rat06, p. 88], we have

$$\cosh c = \frac{1 + \cos\alpha \cos\beta}{\sin\alpha \sin\beta}.$$

Thus

$$(\sin\alpha \sin\beta)\, e^{c} + (\sin\alpha \sin\beta)\, e^{-c} = 2\,(1 + \cos\alpha\cos\beta)\, e^{0}. \tag{4.1}$$

As $c > 0$, the Hermite-Lindemann theorem applied to (4.1) implies $c \notin \mathbb{A}$. $\square$

Let $S^2 \subset \mathbb{R}^3$ denote the unit sphere of constant Gaussian curvature $K = +1$.

Lemma 4.2. *Let $ABC \subset S^2$ be a rational angled triangle labelled as in Figure 6. Then, every side length of ABC is transcendental.*

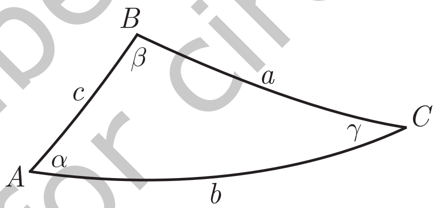


FIGURE 6. Spherical triangle ABC.

Proof. The second spherical law of cosines (see Ratcliffe [Rat06, p. 49]) gives

$$(\sin\alpha \sin\beta)\, e^{ic} + (\sin\alpha \sin\beta)\, e^{-ic} = 2\,(\cos\alpha\cos\beta + \cos\gamma)\, e^{0}, \tag{4.2}$$

where $c > 0$ and $\sin\alpha\sin\beta \in \mathbb{A} - \{0\}$. Whether or not the algebraic number $\cos\alpha\cos\beta + \cos\gamma$ vanishes, the Hermite-Lindemann theorem applied to (4.2) implies $c \notin \mathbb{A}$. The proofs for a and b are similar. $\square$

On the other hand, the results in this note do not hold for arbitrary constant curvature. In a plane of constant Gaussian curvature

$$K = -\left(\cosh^{-1}\left(1 + \sqrt{2}\right)\right)^2 \approx -2.3365\ldots,$$

a regular triangle with angles of radian measure $\pi/4$ has unit side lengths. Similarly, in a sphere of constant Gaussian curvature

$$K = \frac{\pi^2}{4} \approx 2.4674\ldots,$$

a regular right angled triangle has unit side lengths.

References

[Ail71] Ailles, D. S., *Triangles and trigonometry*, Mathematics Teacher **64** (1971), 562.

[Cal09] Calcut, Jack S., *Gaussian integers and arctangent identities for π*, Amer. Math. Monthly **116** (2009), 515–530.

[Cal10] ——, *Grade school triangles*, Amer. Math. Monthly **117** (2010), 673–685.

[CG06] Conway, J. H. and Guy, R. K., *The Book of Numbers*, Copernicus, New York, 2006.

[LL00] Lam, T. Y. and Leung, K. H., *On vanishing sums of roots of unity*, J. Algebra **224** (2000), 91–109.

[Man65] Mann, Henry B., *On linear relations between roots of unity*, Mathematika **12** (1965), 107–117.

[Niv56] Niven, Ivan, *Irrational numbers*, The Mathematical Association of America, John Wiley and Sons, New York, 1956.

[PAR82] Parnami, J. C., Agrawal, M. K. and Rajwade, A. R., *Triangles and cyclic quadrilaterals, with angles that are rational multiples of π and sides at most quadratic over the rationals*, Math. Student **50** (1982), 79–93.

[Rat06] Ratcliffe, John G., *Foundations of hyperbolic manifolds,* second edition, Springer, New York, 2006.

[Sco64] Schoenberg, I. J., *A note on the cyclotomic polynomial,* Mathematika **11** (1964), 131–136.

Jack S. Calcut
Department of Mathematics, Oberlin College
Oberlin, OH 44074, USA
E-mail: *jcalcut@oberlin.edu*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 113-116

METRIZABILITY OF ARITHMETIC PROGRESSION TOPOLOGY

FAHED ZULFEQARR

(Received : 18 - 10 - 2015 ; Revised : 24 - 12 - 2015)

Abstract. Besides giving a way to produce an infinite number of metrics for Furstenberg topology on integers the article rather focuses on a new proof of its metrizability using geometric series.

1. Introduction

Gone are the days when topology was confined to be a branch of pure mathematics. No area of science now escapes from its gravity. For instance, topology is used to understand important concepts of genotype and phenotype in evolutionary biology. Moreover seemingly distant branches of mathematics, number theory and topology have some elegant connections. In this direction, H. Furstenberg in 1955 discovered a strange topology on integers via arithmetic progression and found a beautiful topological proof of infinitude of primes [2]. This article moves a step further and proves metrizability of this topology.

Section 2 introduces Furstenberg topology on integers and contains some useful properties of arithmetic progression. Using convergence of geometric series, a metric on integers is defined in section 3. Last section deals with metrizability of integers with respect to Furstenberg topology.

2. Notations and Preliminaries

The set of all integers is denoted by $\mathbb{Z}$. We let $\mathbb{Z}^* = \mathbb{Z} \setminus \{0\}$. For $m \in \mathbb{Z}$ and $n \in \mathbb{Z}^*$, an arithmetic progression is defined by the set

$$A_{m,n} = \{m + kn \mid k \in \mathbb{Z}\}.$$

We denote the collection of all arithmetic progression by $\mathcal{A}$. The topology generated by $\mathcal{A}$ is the Furstenberg topology on $\mathbb{Z}$ also known as arithmetic progression topology (cf.[1, page 54]). We interchangeably use these terms and let it be denoted by $\tau_{\mathcal{A}}$.

One can easily observe the following useful facts on $\mathbb{Z}$.

Remarks 2.1 (1). $A_{m,n} = A_{m,-n} = A_{r,n}$ for any $r \in A_{m,n}$.

(2). $\mathbb{Z} \setminus A_{m,n} = \bigcup_{i=1}^{i=n-1} A_{m+i,n}$. Hence an arithmetic progression is open as well as closed set in $\tau_{\mathcal{A}}$.

2010 Mathematics Subject Classification : Primary 54-021; Secondary 11B25
Key words and phrases : Arithmetic Progression, Metrizability, Geometric Series

3. Metric on integers

A metric on any non-empty set X is a non-negative real-valued function defined on $X \times X$ taking all its diagonal elements to zero, positive on the complement of diagonal and satisfying symmetric and triangular inequality axioms.

We define an arithmetic function f on $\mathbb{Z}$ using the convergence of geometric series for any fixed real $p > 1$ by

$$f(x) = \sum_{k\in\mathbb{N},\ k\nmid x} \frac{1}{p^k}, \quad x \in \mathbb{Z}^* \quad \text{and} \quad f(0) = 0. \tag{*}$$

The following theorem contains some useful facts about the above arithmetic function.

Theorem 3.1. *The function $f(x)$ defined in (*) satisfies following properties*

(a) $f(x) \geq 0$ *for all* $x \in \mathbb{Z}$.
(b) $f(x) = f(-x)$ *for all* $x \in \mathbb{Z}$.
(c) $f(kx) \leq f(x)$ *for any* $k \in \mathbb{Z}$.
(d) $f(x+y) \leq f(x) + f(y)$ *for all* $x, y \in \mathbb{Z}$.

Proof. (a) follows from the definition of $f(x)$. Divisibility over $\mathbb{Z}$ yields (b). (c) follows from the fact that divisors of x are also the divisors of kx. For (d), if either $x = 0$ or $y = 0$, the equality holds. So we assume that $x \neq 0$ and $y \neq 0$ and observe that

$$\begin{aligned} f(x) &= \sum_{k\nmid x} \frac{1}{p^k} = \sum_{k=1}^{\infty} \frac{1}{p^k} - \sum_{k\mid x} \frac{1}{p^k} \\ \Rightarrow \quad f(x) &= \frac{1}{p-1} - \sum_{k|x,\ k\nmid y} \frac{1}{p^k} - \sum_{k|x,\ k|y} \frac{1}{p^k}. \end{aligned}$$

Similarly

$$\begin{aligned} f(y) &= \frac{1}{p-1} - \sum_{k|y,\ k\nmid x} \frac{1}{p^k} - \sum_{k|x,\ k|y} \frac{1}{p^k} \qquad \text{and} \\ f(x+y) &= \frac{1}{p-1} - \sum_{\substack{k|x+y \\ k\nmid x\ \&\ k\nmid y}} \frac{1}{p^k} - \sum_{k|x,\ k|y} \frac{1}{p^k}. \end{aligned}$$

Therefore we have

$$\begin{aligned} & f(x) + f(y) - f(x+y) \\ & \quad = \frac{1}{p-1} + \sum_{\substack{k|x+y \\ k\nmid x\ \&\ k\nmid y}} \frac{1}{p^k} - \sum_{k|x,\ k\nmid y} \frac{1}{p^k} - \sum_{k|y,\ k\nmid x} \frac{1}{p^k} - \sum_{k|x,\ k|y} \frac{1}{p^k} \\ & \quad \geqslant \sum_{\substack{k|x+y \\ k\nmid x\ \&\ k\nmid y}} \frac{1}{p^k} \end{aligned}$$

which proves (d). ∎

We now define another function d induced by f on $\mathbb{Z}$ by

$$d(m,n) = f(m-n) = \sum_{\substack{k\in\mathbb{N} \\ k\nmid(m-n)}} \frac{1}{p^k}. \tag{**}$$

The next theorem asserts that d actually defines a distance function on $\mathbb{Z}$.

Theorem 3.2. *For $m, n \in \mathbb{Z}$, the function $d(m,n)$ defines a metric on $\mathbb{Z}$.*

Proof. In view of theorem 3.1, all conditions for d to be a metric holds trivially. For instance, triangular inequality follows from (d). In fact, for any $m, l, n \in \mathbb{Z}$
$d(m,n) = f(m-n) = f(m-l+l-n) \leqslant f(m-l) + f(l-n) = d(m,l) + d(l,n).$

□

4. METRIZABILITY OF FURSTENBERG TOPOLOGY

Metrizable spaces have additional structure that can be very useful. For instance, metrizable spaces are Hausdorff so any convergent sequence has a unique limit. Related to metrizability of a topological space, there are two important problems: (1) For a given topological space, construct a metric that induces the topology on the set, (2) Discover a criterion for metrizability of the topological space without specifically finding a metric on it as in the case of Urysohn's metrizable lemma.

Metrizability of Furstenberg topology on $\mathbb{Z}$ is already known to us. Readers are encouraged to see [1] and [3]. We prove this by using the metric $d(m,n)$ on $\mathbb{Z}$ as defined in the previous section.

Theorem 4.1. *The metric $d(m,n)$, $m, n \in \mathbb{Z}$, induces Furstenberg topology $\tau_{\mathcal{A}}$.*

Proof. Let the topology induced by metric d be denoted by τ_d. In the light of Remark 2.1, for the inclusion $\tau_{\mathcal{A}} \subseteq \tau_d$, it suffices to prove that for a given arithmetic progression $A_{m,n}$ there exists some $r > 0$ such that the open ball $B(m,r)$ centered at m with radius r is contained in $A_{m,n}$. For this we choose $r < \frac{1}{p^n}$. Then for any $t \in B(m,r)$ we have

$$\begin{aligned} d(t,m) = \sum_{\substack{k\in\mathbb{N} \\ k\nmid(m-t)}} \frac{1}{p^k} < r < \frac{1}{p^n} & \\ \Rightarrow \quad n \mid (m-t) \qquad & (\text{in fact } \; n! \mid (m-t)) \\ \Rightarrow \quad t = m + kn, \quad & \text{for some } \; k \in \mathbb{Z}. \end{aligned}$$

This implies that $B(m,r) \subseteq A_{m,n}$. For proving the reverse inclusion $\tau_d \subseteq \tau_{\mathcal{A}}$, consider an open ball $B(m,r)$ and choose $n \in \mathbb{Z}^*$ such that $f(n) < r$. Such n indeed exists; in fact, think of sufficiently large $k \in \mathbb{N}$ such that

$$\frac{1}{p^k(p-1)} < r,$$

and put $n = k!$. Then

$$f(n) = \sum_{i\in\mathbb{N},\ i\nmid n} \frac{1}{p^i} \leqslant \sum_{i=k+1}^{\infty} \frac{1}{p^i} = \frac{1}{p^k(p-1)} < r.$$

Let us now take any $t = m + kn \in A_{m,n}$. By using part (c) of Theorem 3.1, we have

$$d(t, m) = f(t - m) = f(kn) \leqslant f(n) < r.$$

Thus we get $A_{m,n} \subseteq B(m, r)$ and the proof is completed. $\square$

References

[1] Adams, C. and Franzosa, R., *Introduction to Topology, Pure and Applied*, Pearson Education, Inc., 2008.

[2] Furstenberg, H., *On infinitude of primes*, Amer. Math. Monthly, **62** (1955), No. 5, 353.

[3] Lovas, R. L. and Mezo, I., *On an exotic topology of integers*, http://arxiv.org /abs /1008.0713.

Fahed Zulfeqarr
Department of Mathematics, Gautam Buddha University
Greater Noida- 201308, U. P., India
E-mail: *fahed@gbu.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 117-133

INTERESTING INFINITE PRODUCTS OF RATIONAL FUNCTIONS MOTIVATED BY EULER

AMRIK SINGH NIMBRAN

(Received : 19 - 10 - 2015 ; Revised : 19 - 12 - 2015)

Abstract. This paper contains a survey of interesting infinite products of rational functions based on Euler's Gamma function and infinite product expansions for $\sin x$ and $\cos x$. Though the index in many products ranges, as usual, over the positive integers, we also treat products where the index ranges over the prime numbers. Historical information concerning the original sources for the gamma function and infinite products is also given.

1. Gamma function and products

The first noteworthy infinite product was discovered by the French mathematician François Viète (1540–1603) in 1593 (Opera, Leyden 1646, p. 400)

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots . \tag{1}$$

The English mathematician John Wallis (1616–1703) gave a neat product for $\frac{4}{\pi}$ (denoted by $\Box$) in Prop. 191 (page 179) of his *Arithmetica infinitorum* (1656) wherein he aimed at finding a general method of quadrature for curved surfaces and interpolated means in the sequence $1, 6, 30, 140, 630, \cdots$ (defined by the recursion $t_{n+1} = \frac{4n+2}{n} \times t_n$, $t_1 = 1$)

$$\Box = \frac{3 \times 3 \times 5 \times 5 \times 7 \times 7 \times etc.}{2 \times 4 \times 4 \times 6 \times 6 \times 8 \times etc.} \tag{2}$$

Leonhard Euler (1707–1783), the versatile Swiss mathematician, did pioneering work on infinite series and products which began when he was 22. An arithmetical function representing the product of the first n natural numbers was known from earlier times though the notation $n!$ was introduced later by a French mathematician Christian Kramp (1760–1826) in his *Eléments d'arithmétique universelle* (1808) while the term 'factorial' was coined by his fellow-countryman Louis François Antoine Arbogast (1759–1803).

Christian Goldbach (1690-1764), who was Euler's colleague at the St. Petersburg Academy but had moved to Moscow soon after Euler's arrival in 1727, and Daniel Bernoulli (1700–1784), with whom Euler stayed then in St. Petersburg,

---

2010 Mathematics Subject Classification : 11Y60, 33B15, 40A20, 41AXX.
Key words and phrases : Infinite products; Gamma function; pi, prime numbers.

having failed to solve the problem of interpolating between the factorials, suggested it to Euler who announced his solution to Goldbach in two letters and soon detailed his ideas in [4].

The first letter dated October 13, 1729 contained the expression

$$\frac{1\cdot 2^m}{1+m}\frac{2^{1-m}\cdot 3^m}{2+m}\frac{3^{1-m}\cdot 4^m}{3+m}\frac{4^{1-m}\cdot 5^m}{4+m}\cdots.$$

Setting $m=\frac{1}{2}$, Euler obtained $\Gamma(\frac{3}{2})=\frac{\sqrt{\pi}}{2}$ and deduced Wallis' product.

In his letter of January 8, 1730 and in [4], Euler took up the series $\sum n!$ and found the general term for the factorial to be $\int(-\ln x)^n\,dx$. Thus, the terms of the series transform into a definite integral $\displaystyle\int_0^1\left(\ln\frac{1}{x}\right)^n\,dx$, which on substituting $x=e^{-t}$, converts into $\int_0^\infty t^n\ e^{-t}\,dt$, the definite integral given later in Art.11 of his paper [8]. Extended to $z\in\mathbb{C}$, it becomes

$$\Gamma(z)=\int_0^\infty t^{z-1}\ e^{-t}\,dt, \qquad \Re(z)>0. \tag{3}$$

One may note that the notation $\Gamma(a)$ was introduced by the French mathematician Adrien-Marie Legendre (1752–1833) who defined it [14, p.477, sec.53](1809) as

$$\Gamma(a)=\int dx\left(\log\frac{1}{x}\right)^{a-1}.$$

In §10 of [8], Euler recorded an infinite product representation

$$y=1\cdot 2\cdot 3\cdots x=\frac{\sqrt{2}\pi(x+n)}{\prod_{k=1}^n(x+k)}\left(\frac{x+n}{e}\right)^{x+n}e^s, \tag{4}$$

where $s=\displaystyle\sum_{m=1}^\infty(-1)^{m-1}\frac{c_{2m}(2m-2)!}{2^{2m-1}(x+n)^{2m-1}}$ and c_{2m} being the coefficient of π^{2m} occurring in Euler's zeta function $\zeta(2m)$.

In §12, Euler noted two more product representations (including one recorded in his first letter to Goldbach)

$$y = \prod_{n=1}^\infty\frac{n}{n+x}\left(\frac{n+1}{n}\right)^x \tag{5}$$

$$= \left(\frac{1+x}{2}\right)^x\prod_{n=1}^\infty\frac{n}{n+x}\left(\frac{2n+1+x}{2n-1+x}\right)^x. \tag{6}$$

In §8 of [4], Euler actually started with another integral $\int x^e\,dx\,(1-x)^n$ which he evaluated by expanding $(1-x)^n$ through the Binomial Theorem

$$\int_0^1 x^e(1-x)^n\,dx=\frac{1\cdot 2\cdots n}{(e+1)(e+2)\cdots(e+n+1)}.$$

His objective now was to isolate $1\cdot 2\cdots n$ from the denominator so as to have an expression for $n!$ as an integral. He thus arrived at the integral $\int(-\ln x)^n\,dx$. The former integral was termed by Legendre the *first Eulerian integral* or the *Beta function* and the latter the *second Eulerian integral* with

$$B(m,n) = \int_0^1 t^{m-1}(1-t)^{n-1}\,dt = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}.$$

Integration of (3) by parts yields the functional equation

$$\Gamma(z+1) = \int_0^\infty t^z\ e^{-t}\,dt = -t^z e^{-t}\Big|_0^\infty + z\int_0^\infty t^{z-1}\ e^{-t}\,dt = z\Gamma(z). \tag{7}$$

Legendre gave the duplication formula [14, page 485]

$$\Gamma(a) = \frac{2^{1-2a}}{\sqrt{\pi}}\cos a\pi\,\Gamma(2a)\,\Gamma\left(\frac{1}{2}-a\right), \quad a < \frac{1}{2}. \tag{8}$$

Carl Friedrich Gauss (1777-1855) studied the Gamma function while dealing with the hypergeometric series [11]. He used the notation Πz, where

$$\Pi z = \lim_{k\to\infty}\frac{k!k^z}{\prod_{j=1}^k (z+j)}. \tag{9}$$

His functional equation (§21, (44)) is slightly different: $\Pi(z+1) = (z+1)\Pi z$ and $\Pi(\frac{-1}{2}) = \sqrt{\pi}$, $\Pi(\frac{1}{2}) = \frac{1}{2}\sqrt{\pi}$. In §26, equation (57), we find

$$\frac{n^{n\,z}\prod_{k=0}^{n-1}\Pi\left(z-\frac{k}{n}\right)}{\Pi nz} = \frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}. \tag{10}$$

Schlömilch [24, p.171, eq(12)](1844) and Newman [16, p.59, eq. (10)] (1848) found

$$\Gamma(1+\alpha) = e^{-C\alpha}\frac{e^{\alpha}}{1+\frac{\alpha}{1}}\cdot\frac{e^{\frac{1}{2}\alpha}}{1+\frac{\alpha}{2}}\cdot\frac{e^{\frac{1}{3}\alpha}}{1+\frac{\alpha}{3}}\cdots, \tag{11}$$

where C is the constant γ introduced by Euler as the constant of integration in [5] (1734) and treated in depth in [9] (1776).

Recognizing the advantages of starting with $\frac{1}{\Gamma(z)}$, Karl Weierstrass (1815-1897) used this form. He called $\frac{1}{\Gamma(u+1)}$ the "*Factorielle* of u", denoted it by $Fc(u)$ [28, p. 7] and gave following representations [28, eq.(10), (15), (46), (47)]

$$\begin{aligned} Fc(u) =& u\prod_{\alpha=1}^{\infty}\left\{\left(\frac{\alpha}{1+\alpha}\right)^u\left(1+\frac{u}{\alpha}\right)\right\} \\ =& \lim_{n\to\infty}\left\{n^{-u}u(1+u)\left(1+\frac{u}{2}\right)\cdots\left(1+\frac{u}{n-1}\right)\right\}. \end{aligned}$$

This can be written as

$$\frac{1}{\Gamma(z)} = ze^{\gamma z}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)e^{-\frac{z}{n}}. \tag{12}$$

Formula (12) shows the roots of $\frac{1}{\Gamma(z)}$. As Davis [3, page 862] remarks, the reciprocal of the gamma function is a much less difficult function to deal with than the gamma function itself.

Since $\ln(\frac{a}{b})=\ln a - \ln b$, we have $\ln(n+1)=\sum_{k=1}^{n}\ln(\frac{k+1}{k})$, and

$$\gamma = \lim_{n\to\infty}\left[\sum_{k=1}^{n}\frac{1}{k} - \sum_{k=1}^{n}\ln\frac{k+1}{k}\right].$$

Then multiplying both sides by $-z$ followed by raising the two sides to the power of e yields

$$e^{-\gamma z} = \lim_{n\to\infty} e^{-\sum_{k=1}^{n}\frac{z}{k}}\, e^{\ln(1+\frac{1}{k})^z} = \lim_{n\to\infty}\prod_{k=1}^{n} e^{-\frac{z}{k}}\left(1+\frac{1}{k}\right)^z.$$

Using Gauss' representation, we get

$$z\Gamma(z) = \lim_{n\to\infty}\prod_{k=1}^{n}\left(1+\frac{1}{k}\right)^z\left(1+\frac{z}{k}\right)^{-1} = \lim_{n\to\infty}\prod_{k=1}^{n}\left(1+\frac{1}{k}\right)^z e^{-\frac{z}{k}}\left(1+\frac{z}{k}\right)^{-1} e^{\frac{z}{k}}$$

which on using the result deduced above yields

$$\Gamma(z) = z^{-1}e^{-\gamma z}\lim_{n\to\infty}\prod_{k=1}^{n}\left(1+\frac{z}{k}\right)^{-1} e^{\frac{z}{k}}$$

and this on being inverted becomes (12).

Euler's reflection/complement formula can be obtained by employing the representation of Weierstrass for the expression $\frac{1}{\Gamma(z)\Gamma(-z)}$ followed by the use of $\Gamma(-z) = -\frac{\Gamma(1-z)}{z}$ and Euler's product expansion for $\sin z$ (to follow soon)

$$\Gamma(1-z)\,\Gamma(z) = \frac{\pi}{\sin(\pi z)} \Rightarrow \Gamma(1-z)\,\Gamma(1+z) = \frac{\pi z}{\sin(\pi z)},\ z \neq 0,\ \pm 1, \cdots. \quad (13)$$

This formula possibly motivated Euler to find the product expansion for $\sin x$. He factorized the function $e^x - e^{-x}$ to obtain [6, Ch. IX, §158]

$$\sin x = x\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{(2\pi)^2}\right)\left(1-\frac{x^2}{(3\pi)^2}\right)\cdots, \quad (14)$$

and

$$\cos x = \left(1-\frac{4x^2}{\pi^2}\right)\left(1-\frac{4x^2}{(3\pi)^2}\right)\left(1-\frac{4x^2}{(5\pi)^2}\right)\cdots. \quad (15)$$

Euler derived Wallis' product in [7, Ch.9, §358] by setting $x = \frac{1}{2}$ in (14) and then shifting the term $\frac{1}{2}$ to the left hand side. We find that it can also be obtained through (15) by transferring the first factor on the right to the left and then taking the limit $x \to \frac{\pi}{2}$. As the limit is of the indeterminate form $\frac{0}{0}$, we apply L' Hospital's rule, differentiating the expression and so getting the desired result.

2. Summary of formulas and special values

We summarize in the following the gamma function identities and values (some noted earlier) that are needed in the rest of the paper

2.1. **Definitions.**

2.1.1. *Euler's integral* : $\Gamma(z) = \int_0^\infty t^{z-1}\, e^{-t}\, dt, \qquad \Re(z) > 0.$

2.1.2. *Euler's product :* $\Gamma(z) = \lim_{n\to\infty} \prod_{k=1}^{n} \frac{(n-1)!}{z(z+1)(z+2)\cdots(n+z-1)} n^z$,

valid for all z except $z \neq 0, -1, -2, \cdots$.

2.1.3. *Weierstrass product :* $\frac{1}{\Gamma(z)} = ze^{\gamma z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) e^{-\frac{z}{n}}$,

valid for all z except $z \neq 0, -1, -2, \cdots$; and $\gamma =_{N\to\infty} \displaystyle\sum_{n=1}^{N} \frac{1}{n} - \ln N$.

2.2. **Important identities.**

2.2.1. *Functional relation :* $\Gamma(z+1) = z\Gamma(z)$.

2.2.2. *Euler's reflection formula :* $\Gamma(z)\,\Gamma(1-z) = \frac{\pi}{\sin(\pi\, z)}$.

2.2.3. *Legendre's duplication formula :* $\Gamma(2z) = \pi^{\frac{-1}{2}}\, 2^{2z-1}\Gamma(z)\,\Gamma\left(z + \frac{1}{2}\right)$.

2.2.4. *Gauss's multiplication formula :*

$$\Gamma(nz) = (2\pi)^{\frac{1-n}{2}}\, n^{nz\frac{1}{2}}\, \Gamma(z)\,\Gamma\left(z + \frac{1}{n}\right)\,\Gamma\left(z + \frac{2}{n}\right)\cdots\Gamma\left(z + \frac{n-1}{n}\right).$$

2.3. **Special values.** (Here n is a nonnegative integer.)

2.3.1. *Factorial :* $\Gamma(n+1) = n!$.

2.3.2. *Half-integer values :* $\Gamma\left(\frac{1}{2}\right) = \pi^{\frac{1}{2}}$; $\Gamma\left(-n + \frac{1}{2}\right) = (-1)^n \frac{2^{2n}\, n!}{(2n)!}\sqrt{\pi}$;

$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\,\Gamma\left(\frac{1}{2}\right) = \frac{(2n)!}{2^{2n}\, n!}\sqrt{\pi}$.

2.3.3. *Quarter-integer values :* $\Gamma\left(\frac{1}{4}\right) = \sqrt{\sqrt{2\pi}\, L}$, $\Gamma\left(n + \frac{1}{4}\right) = \frac{1\cdot 5\cdot 9\cdots(4n-3)}{4^n}\Gamma\left(\frac{1}{4}\right)$, where $L = 2.62205755429212\cdots$ is called the *lemniscate constant* and equals one half the perimeter of the unit lemniscate curve $r^2 = \cos 2\theta$ (a special case of Bernoulli's lemniscate) just as π is one half the circumference of the unit circle [31]. The numbers π and L share several common features. L is also related to Gauss's constant K by $L = \pi\, K$ with

$$K = \frac{2}{\pi}\int_0^1 \frac{dx}{\sqrt[4]{1-x^4}} = \prod_{n=1}^{\infty} \tanh^2 \frac{n\pi}{2} = \frac{\Gamma^2\left(\frac{1}{4}\right)}{(2\pi)^{\frac{3}{2}}}.$$

Note: We are using K for the Gauss constant instead of G used in [30] as we will use G for the Catalan constant later.

2.3.4. *Rational values* $\left(n + \frac{1}{k}\right)$ *:*

$$\Gamma\left(n + \frac{1}{k}\right) = \frac{1\cdot(1+k)\cdot(1+2k)\cdots(1+(n-1)k)}{k^n}\Gamma\left(\frac{1}{k}\right).$$

Euler gave the following formula which is a consequence of the multiplication formula of Gauss

$$\prod_{m=1}^{n-1}\Gamma\left(\frac{m}{n}\right) = \frac{(2\pi)^{\frac{n-1}{2}}}{\sqrt{n}}.$$

Let $\Phi(n) = \{m : 1 \le m < n \wedge \gcd(m,n) = 1\}$. Let $\phi(n) = N(\Phi(n))$ represent Euler's totient function. For any prime p, $\phi(p) = p-1$. If $\gcd(a,b) = 1$, $\phi(a \cdot b) = \phi(a) \times \phi(b)$. And $\phi(m^k) = m^{k-1}\,\phi(m)$. Then we have [23, 15, 17, 2]

$$\prod_{m=1}^{n} \Gamma\left(\frac{m}{n}\right) = \frac{(2\pi)^{\frac{\phi(n)}{2}}}{e^{\frac{\Lambda(n)}{2}}} = \begin{cases} \frac{(2\pi)^{\frac{\phi(n)}{2}}}{\sqrt{p}} & \text{if}\, m = p^r \\ (2\pi)^{\frac{\phi(n)}{2}} & \text{if}\, m \neq p^r \end{cases}$$

where $\Lambda(n)$ is Mangoldt's function defined to be $\log p$ if $n = p^r$; $p-$ prime, $r \in \mathbb{N}$ and 0 otherwise.

3. Products of rational functions of integer index

Probably the most famous product of rational functions with integer index is the Wallis product. We will give a general theorem that explains all such products and gives a closed form for the expansions as a quotient of gamma functions. But before that we touch upon the theory of infinite products.

An infinite product is formed by multiplying, in a given order, the terms of an infinite sequence $\{a_r\}$ of real or complex numbers and is denoted by $\prod_{r=1}^{\infty} a_r$. The product of the first n terms, that is $P_n = \prod_{r=1}^{n} a_r$, is called the n-th *partial product.* If P_n tends to a non-zero finite limit P as $n \to \infty$, then the infinite product is said to converge to (value) P.

To determine the convergence of an infinite product, we write the general term as $a_r = 1 + b_r$ or $a_r = 1 - b_r$. If the infinite product is convergent, P_n and P_{n-1} tend to the same limit as $n \to \infty$. Since $\frac{P_n}{P_{n-1}} = 1 + b_n$, we have

$$\frac{\lim\limits_{n\to\infty} P_n}{\lim\limits_{n\to\infty} P_{n-1}} = 1 \Rightarrow \lim_{n\to\infty}(1+b_n) = 1 \Rightarrow \lim_{n\to\infty} b_n = 0.$$

It is thus a necessary (but not sufficient) condition for $\prod(1+b_n)$ to converge that $b_n \to 0$. A necessary and sufficient condition for the convergence of $\prod(1+b_n)$ and $\prod(1-b_n)$ is the convergence of the series $\sum b_n$.

3.1. Products derived via Gamma function.

We now state and prove the following theorem from Whittaker and Watson [32, page 238].

Theorem 3.1. *Let* $u_n = \dfrac{A\,(n+a_1)(n+a_2)\cdots(n+a_k)}{(n+b_1)(n+b_2)\cdots(n+b_l)}$ *be a general rational function of n. If $A = 1$, $k = l$ and $a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k$ then the infinite product* $\prod\limits_{n=1}^{\infty} u_n$ *converges absolutely; and it has the closed form value*

$$\prod_{n=1}^{\infty} \frac{(n+a_1)(n+a_2)\cdots(n+a_k)}{(n+b_1)(n+b_2)\cdots(n+b_k)} = \frac{\Gamma(1+b_1)\Gamma(1+b_2)\cdots\Gamma(1+b_k)}{\Gamma(1+a_1)\Gamma(1+a_2)\cdots\Gamma(1+a_k)}. \tag{16}$$

Proof. Since we require $\lim\limits_{n\to\infty} u_n = 1$, we must have $k = l$ and $A = 1$. Observe that the general term of the infinite product can be written as

$$u_n = \frac{(1+\frac{a_1}{n})(1+\frac{a_2}{n})\cdots(1+\frac{a_k}{n})}{(1+\frac{b_1}{n})(1+\frac{b_2}{n})\cdots(1+\frac{b_k}{n})}$$

$$= 1 + \frac{a_1 + a_2 + \cdots + a_k - b_1 - b_2 - \cdots - b_k}{n} + O(\frac{1}{n^2}).$$

For the convergence we require that the factors u_n, when expanded in reciprocal powers of n, say as $1 + \frac{\alpha}{n} + \frac{\beta}{n^2} + \cdots$, approach 1 faster than $1 + \frac{\alpha}{n}$, so we must have $\alpha = 0$. Thus we need $a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k$ (See [32, §2.7]).

Further, observe that we can write

$$u_n = \frac{(1+\frac{a_1}{n})e^{-\frac{a_1}{n}}(1+\frac{a_2}{n})e^{-\frac{a_2}{n}}\cdots(1+\frac{a_k}{n})e^{-\frac{a_k}{n}}}{(1+\frac{b_1}{n})e^{-\frac{b_1}{n}}(1+\frac{b_2}{n})e^{-\frac{b_2}{n}}\cdots(1+\frac{b_k}{n})e^{-\frac{b_k}{n}}}$$

by inserting exponential convergence factors in the expression for u_n without altering the value of the product. (Inserting these exponentials causes a factor like $(1+\frac{a}{n})e^{-\frac{a}{n}}$ to approach 1 with order $\frac{1}{n^2}$, thus ensuring absolute convergence of our infinite product). From the Wierstrass definition of the infinite product we have

$$\prod_{n=1}^{\infty}\left\{\left(1+\frac{z}{n}\right)e^{-\frac{z}{n}}\right\} = \frac{1}{z\Gamma(z)e^{\gamma z}} = \frac{1}{\Gamma(z+1)e^{\gamma z}}.$$

The closed form of our infinite product (16) now follows at once. □

As a consequence of this theorem, every infinite product of rational functions has a closed form expansion in terms of a quotient of gamma functions. So the next step would be to find certain values of the constants a and b which reduce the quotient of gamma functions to a more familiar value. With the aid of the known special values of the gamma function, and other identities previously shown, we can obtain an endless number of these products. We also note that other approaches to products of this type, such as the Wallis product, may be found by entirely different (and sometimes simpler) methods. We now illustrate it by deriving a few such infinite products.

Example 1. Wallis product. Taking $a_1 = -\frac{1}{2}$, $a_2 = \frac{1}{2}$, $b_1 = b_2 = 0$, we get

$$\prod_{n=1}^{\infty}\frac{(n-\frac{1}{2})(n+\frac{1}{2})}{n \cdot n} = \frac{\Gamma(1)\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{3}{2}\right)}$$

which immediately simplifies to

$$\prod_{n=1}^{\infty}\frac{(2n-1)(2n+1)}{(2n)^2} = \frac{2}{\pi}.$$

Example 2. Taking $a_1 = -\frac{1}{2}$, $a_2 = 1$ and $b_1 = b_2 = \frac{1}{4}$ we obtain

$$\prod_{n=1}^{\infty}\frac{(n-\frac{1}{2})(n+1)}{(n+\frac{1}{4})(n+\frac{1}{4})} = \frac{\Gamma^2\left(\frac{5}{4}\right)}{\Gamma(\frac{1}{2})\Gamma(2)} = \frac{\Gamma^2\left(\frac{1}{4}\right)}{16\sqrt{\pi}};$$

and since $\Gamma^2\left(\frac{1}{4}\right) = 2\sqrt{2\pi}\,L$ it follows that

$$\prod_{n=1}^{\infty}\frac{(4n-2)(4n+4)}{(4n+1)^2} = \frac{L}{4\sqrt{2}}. \tag{17}$$

Example 3. The values $a_1 = \frac{1}{2} = a_2 = -\frac{1}{4}, b_1 = 0$ and $b_2 = \frac{1}{4}$ yield

$$K = \frac{L}{\pi} = \prod_{n=1}^{\infty} \frac{(2n+1)(4n-1)}{2n(4n+1)}. \tag{18}$$

A number of such products (derived mostly by the author) are listed in the *Appendix.* Most of the products in the *Category C* are based on the relations between particular Gamma values found in [26] and [29]. Two of the products in the *Category B* are motivated by a result occurring in [12].

3.2. **Products derived via the cosine function.** Euler's expansion for $\cos x$ can be expressed as

$$\cos\frac{k\pi}{2m} = \prod_{n=1}^{\infty} \frac{(2mn-(m+k))(2mn-(m-k))}{(2mn-m)^2},\ m,k \in \mathbb{N};\ 1 \le k \le m.$$

Setting $k = 1, 2, 3, \cdots (m-1)$ yields products having factors of the form $2mn - k$. Since $\dfrac{\cos(\pi/4)}{\cos(\pi/3)} = \sqrt{2}$ and 12 is the L.C.M. of 3 and 4, so putting $m = 6$, $k = 3$, 4 in the above identity and then dividing the formula obtained with $k = 3$ by the formula with $k = 4$, we get

$$\sqrt{2} = \prod_{n=1}^{\infty} \frac{(12n-9)(12n-3)}{(12n-10)(12n-2)}.$$

Some results noted in the *Appendix* have been derived by combining various products so obtained using the twenty-four possible values (computed by the author) for $\sin\theta$ and $\cos\theta$ given in *Table 1*.

4. Products involving prime numbers

Beginning with his solution of the famous Basel problem in 1734, Euler evaluated the series of reciprocals of even powers of the natural numbers thus introducing the *Euler zeta function* sometime around 1737. He then discovered in 1744 a beautiful relationship between $\zeta(s)$ and prime numbers

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p-prime} \left(1 - p^{-s}\right)^{-1},\ \Re(s) > 1. \tag{19}$$

Euler obtained (19) by making use of the fundamental theorem of arithmetic (each natural number $n > 1$ can be expressed uniquely as a product of primes) and using the formula for the sum of a geometric series as follows

$$1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right)$$
$$\left(1 + \frac{1}{5^s} + \frac{1}{5^{2s}} + \cdots\right)\cdots = \prod_{p-prime}\left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \prod_{p-prime} \frac{1}{1 - \frac{1}{p^s}}$$

We now show how Euler obtained the following remarkable product [10, §15] that ranges over all odd primes

$$\frac{\pi}{4} = \frac{3}{4}\cdot\frac{5}{4}\cdot\frac{7}{8}\cdot\frac{11}{12}\cdot\frac{13}{12}\cdot\frac{17}{16}\cdot\frac{19}{20}\cdots = \prod_{p-odd\,prime} \frac{p}{p+(-1)^{\frac{p+1}{2}}}. \tag{20}$$

To understand the use of $(-1)^{\frac{p+1}{2}}$ one may observe that all the odd prime numbers can be divided into two classes - those of the form $4n+1$ and those of the form $4n-1$. If $p \equiv -1\ (mod\ 4),\ (-1)^{\frac{p+1}{2}} = 1$ and if $p \equiv 1\ (mod\ 4),\ (-1)^{\frac{p+1}{2}} = -1$. In fact, Euler starts with the Leibniz series (1674)

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots. \tag{21}$$

Multiplying (21) by $\frac{1}{3}$ (3 is a prime of the form $4n-1$) one gets the series expansion

$$\frac{\pi}{4}\cdot\frac{1}{3} = \frac{1}{3} - \frac{1}{9} + \frac{1}{15} - \frac{1}{21} + \frac{1}{27} - \frac{1}{33} + \cdots \tag{22}$$

which when added to (21) will give the following series expansion whose right side is bereft of the multiples of 1/3.

$$\frac{\pi}{4}\left(1+\frac{1}{3}\right) = 1 + \frac{1}{5} - \frac{1}{7} - \frac{1}{11} + \frac{1}{13} + \cdots. \tag{23}$$

Similarly, first multiplying (22) by $\frac{1}{5}$ (5 is prime of the form $4n+1$) one gets the series expansion

$$\frac{\pi}{4}\left(1+\frac{1}{3}\right)\frac{1}{5} = \frac{1}{5} + \frac{1}{25} - \frac{1}{35} - \frac{1}{55} + \frac{1}{65} + \cdots, \tag{24}$$

which when subtracted from (23) will result in to the following series expansion the right side of which is bereft of multiples of 1/5.

$$\frac{\pi}{4}\left(1+\frac{1}{3}\right)\left(1-\frac{1}{5}\right) = 1 - \frac{1}{7} - \frac{1}{11} + \frac{1}{13} + \frac{1}{17} - \frac{1}{19} - \cdots.$$

Continuing in this way, adding when the denominator of the second term on right side of the concerned series expansion is a prime of the form $p = 4n-1$ and subtracting when it is a prime of the form $p = 4n+1$, we can successively eliminate all multiples of $1/7,\ 1/11,\ 1/13, \cdots$ from the right side of the series expansion eventually getting

$$\frac{\pi}{4}\left(1+\frac{1}{3}\right)\left(1-\frac{1}{5}\right)\left(1+\frac{1}{7}\right)\left(1+\frac{1}{11}\right)\left(1-\frac{1}{13}\right)\cdots = 1$$

which can be written as

$$\frac{\pi}{4}\prod_{p-oddprime}\left(1+(-1)^{\frac{p+1}{2}}\frac{1}{p}\right) = \frac{\pi}{4}\prod_{p-oddprime}\left(\frac{p+(-1)^{\frac{p+1}{2}}}{p}\right) = 1$$

from which Euuler's product (20) now follows at once.

Observe that when the index ranges over all odd integers ≥ 3 one obtains

$$\prod_{n\geq 1}\frac{(4n-1)(4n+1)}{(4n)^2} = \prod_{n=0}^{\infty}\frac{\left(n+\frac{3}{4}\right)\left(n+\frac{5}{4}\right)}{(n+1)^2} = \frac{\Gamma^2(1)}{\Gamma(\frac{3}{4})\Gamma(\frac{5}{4})} = \frac{2\sqrt{2}}{\pi}. \tag{25}$$

Dividing (25) by (20) we can deduce the following product involving composite numbers only

$$\prod_{\substack{k-odd>1\\ k\neq p}} \frac{k}{k+(-1)^{\frac{k+1}{2}}} = \frac{9}{8}\cdot\frac{15}{16}\cdot\frac{21}{20}\cdot\frac{25}{20}\cdot\frac{27}{28}\cdot\frac{33}{32}\cdot\frac{35}{36}\cdots = \frac{8\sqrt{2}}{\pi^2}. \tag{26}$$

Euler further observes that the following series of the reciprocals of odd squares

$$1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\frac{1}{9^2}+\frac{1}{11^2}+\frac{1}{13^2}+\cdots$$

whose sum he had shown to be $\pi^2/8$ will turn into

$$\frac{\pi^2}{8} = \frac{3\cdot 3}{2\cdot 4}\cdot\frac{5\cdot 5}{4\cdot 6}\cdot\frac{7\cdot 7}{6\cdot 8}\cdot\frac{11\cdot 11}{10\cdot 12}\cdot\frac{13\cdot 13}{12\cdot 14}\cdots.$$

By squaring the equation (19) and then dividing the result by the preceding product, Euler obtained a product for 2.

Alternatively, consider the following result due to Euler

$$\frac{\pi^2}{6} = \prod_{p-prime}\frac{p^2}{p^2-1} = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{3\cdot 3}{2\cdot 4}\cdot\frac{5\cdot 5}{4\cdot 6}\cdot\frac{7\cdot 7}{6\cdot 8}\cdots. \tag{27}$$

Dividing the two sides of this product by $2^2/(2^2-1)$ and then dividing the result so obtained by (20), we obtain the following product not given by Euler

$$\frac{\pi}{2} = \prod_{p-odd\,prime}\frac{p}{p+(-1)^{\frac{p-1}{2}}} = \frac{3}{2}\cdot\frac{5}{6}\cdot\frac{7}{6}\cdot\frac{11}{10}\cdot\frac{13}{14}\cdot\frac{17}{18}\cdot\frac{19}{18}\cdots. \tag{28}$$

It may be noted here that each denominator leaves remainder 2 when divided by 4, while 4 divides each denominator in the product (20). Using the products (20) and (28), we get the following Euler's product for 2

$$2 = \prod_{p-odd\,prime}\frac{p+(-1)^{\frac{p+1}{2}}}{p+(-1)^{\frac{p-1}{2}}} = \frac{3+1}{3-1}\cdot\frac{5-1}{5+1}\cdot\frac{7+1}{7-1}\cdot\frac{11+1}{11-1}\cdot\frac{13-1}{13+1}\cdots. \tag{29}$$

Since, $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^m} = \prod_{p-prime}\frac{p^m}{p^m-1}$ and $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^4} = \frac{\pi^4}{90}$, we straightway deduce

$$\frac{\zeta(4)}{\zeta(2)} = \frac{\pi^2}{15} = \prod_{p-prime}\frac{p^2}{p^2+1} = \frac{4}{5}\cdot\frac{9}{10}\cdot\frac{25}{26}\cdot\frac{49}{50}\cdots. \tag{30}$$

Using various products noted above, we get the following product not given by Euler

$$3 = \prod_{p-odd\,prime}\frac{p^2+1}{\left(p+(-1)^{\frac{p-1}{2}}\right)^2} = \frac{10}{2^2}\cdot\frac{26}{6^2}\cdot\frac{50}{6^2}\cdot\frac{122}{10^2}\cdot\frac{170}{14^2}\cdots. \tag{31}$$

We have the following product for $G = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2}$, the Catalan constant,

$$G = \prod_{p-odd\,prime} \frac{p^2}{p^2 + (-1)^{\frac{p+1}{2}}}. \tag{32}$$

Using Euler's formulas noted earlier we get

$$\frac{8G}{\pi^2} = \prod_{p-prime\equiv -1\ (mod\ 4)} \frac{p^2-1}{p^2+1}, \tag{33}$$

and we find the formula

$$\frac{8G}{\pi^2} = \frac{2}{\pi}\, {}_3F_2\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; 1, \frac{3}{2}\right) \tag{34}$$

in [1, page 45, Entry 35(i)/35(iv)]. Thus we get

$$\prod_{p-prime\equiv -1\ (mod\ 4)} \frac{p^2-1}{p^2+1} = \frac{2}{\pi}\, {}_3F_2\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; 1, \frac{3}{2}\right). \tag{35}$$

It is easy to deduce from the products noted earlier that

$$\prod_{p-prime} \frac{p^2+1}{p^2-1} = \frac{5}{2} \qquad and \qquad \prod_{p-odd\,prime} \frac{p^2+1}{p^2-1} = \frac{3}{2}. \tag{36}$$

Hence, using the previous product we obtain

$$\frac{12G}{\pi^2} = \prod_{p-prime\equiv 1\ (mod\ 4)} \frac{p^2+1}{p^2-1} \prod_{p-prime\equiv -1\ (mod\ 4)} \frac{p^2-1}{p^2+1}, \tag{37}$$

that is,

$$\prod_{p-prime\equiv 1\ (mod\ 4)} \frac{p^2+1}{p^2-1} = \frac{3}{\pi}\, {}_3F_2\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; 1, \frac{3}{2}\right). \tag{38}$$

Concluding Remarks: We urge the reader to examine the original papers of Euler. They are exciting! They reveal his brilliant intuition and inventive mind. Remember Laplace's advice: "Reader Euler, read Euler. He is the master of us all."

The original derivation of the Wallis product is described in modern terms in [19, 20, 22, 25]. It is very interesting and novel. One can also read the translation of his *The Arithmetic of Infinitesimals* [27]. However, it is a difficult going. Perhaps the simplest derivation of the Wallis product is through the infinite product for the sine function. An interesting twist on this is found in [21].

We also invite the reader to explore new ideas for additional products. There may be hidden patterns in the many products in the *Appendix* that have not yet been revealed. The field is vast and uncharted!!

Acknowledgement: The author is grateful to the anonymous referee for his several useful suggestions which helped in giving a shape to this write up and in improving the presentation.

References

[1] Berndt, B. C., *Ramanujan's Notebooks Part II,* Springer, 1989.

[2] Chamberland, Marc and Straub, Armin, On gamma quotients and infinite products, *Advances in Applied Mathematics,* Elsevier, **51**, Issue 5, Oct. 2013, 546–562. Available at http:// arxiv.org/ abs/ 1309.3455

[3] Davis, P. J., Leonhard Euler's Integral: A Historical Profile of the Gamma Function, *The American Mathematical Monthly,* **66**, No. 10 (Dec. 1959), 849–869.

[4] Euler, L., *De progressionibus transcendentibus seu quarum termini generales algebraice dari nequeunt,* presented to the St. Petersburg Academy on November 28, 1729; originally published in *Commentarii academiae scientiarum Petropolitanae* **5**, 1738, 36–57. Available as E 19 at http://eulerarchive.maa.org/

[5] Euler, L., *De progressionibus harmonicis observationes,* presented to the St. Petersburg Academy on March 11, 1734; originally published in *Commentarii academiae scientiarum Petropolitanae* **7**, 1740, 150–161. E 43.

[6] Euler, L., *Introductio in Analysin Infinitorum* (1748), Book I.

[7] Euler, L., *Institutionum calculi integralis* (1768), Volume 1.

[8] Euler, L., *De curva hypergeometrica hac aequatione expressa* $y = 1.2.3 \cdots x$, presented to the St. Petersburg Academy on December 19, 1765; originally published in *Novi Commentarii academiae scientiarum Petropolitanae* **13**, 1769, 3–66. E 368.

[9] Euler, L., *De numero memorabili in summatione progressionis harmonicae naturalis occurrente,* presented to the St. Petersburg Academy on February 22, 1776; originally published in *Acta Academiae Scientarum Imperialis Petropolitinae* **5**, 1785, 45–75. E583

[10] Euler, L. *De summa seriei ex numeris primis formatae* $\frac{1}{3} - \frac{1}{5} + \frac{1}{7} - \frac{1}{13} - \frac{1}{17} + \frac{1}{19} + \frac{1}{23} - \frac{1}{29} + \frac{1}{31}$, *etc. ubi numeri primi formae 4n-1 habent signum positivum, formae autem 4n+1 signum negativum,* presented to the St. Petersburg Academy on October 2, 1775, originally published in *Opuscula Analytica* **2**, 1785, 240–256. E 596.

[11] Gauss, C. F., *Disquisitiones generales circa seriem infinitam,* Commentationes societatis regiae scientiarum, Gottingen, **II** (1813), 1–46; Werke, Band **III**, 1876, 122–162.

[12] Ben-Ari, I., Hay, D. and Roitershtein, A., *On Wallis-Type Products And Polya's Urn Schemes,* Amer. Math. Monthly **121** (May 2014), 422–432; also available at http:// www.public.iastate.edu / roiterst / papers / gamma8f.pdf

[13] Jensen, J. L. W. V. and (tr.) Gronwall, T. H., *An Elementary Exposition of the Theory of the Gamma Function,* Annals of Mathematics, Second Series, **17**, No. 3 (Mar. 1916), 124–166. Gammafunktionens Theori i elementær Fremstillng. Nyt Tidsskrift for Mathematik, Afdeling B, **2** (1891), 33–35, 57–72, 83–84.

[14] Legendre, A.-M., *Recherches sur diverses sortes d'intégrales définies* (13 November 1809), Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France, 1809, 416–509.

[15] Martin, G., *A Product of Gamma Function Values at Fractions with the Same Denominator,* 2009. http://arxiv.org/abs/0907.4384v2

[16] Newman, F. W., *On* $\Gamma(a)$ *especially when a is negative,* Cambridge and Dublin Mathematical Journal, Barclay and Macmillan, **3** (1848), 57–60.

[17] Nijenhuis, A., Short Gamma Products with Simple Values, *Amer. Math. Monthly,* **117**, No. 8 (Oct. 2010), 733–737. Available at http:// arxiv.org/ abs/ 0907.1689v1

[18] Nimbran, A. S., *Generalized Wallis-Euler Products and New Infinite Products for π*, The Mathematics Student, **83** (2014), 155–164.

[19] Nunn, T. P., *The Arithmetic of Infinites: A School Introduction to the Integral Calculus,* (Part 1), The Mathematical Gazette, **5** (1910), 345–356.

[20] Nunn, T. P., *The Arithmetic of Infinites: A School Introduction to the Integral Calculus,* (Part 2), The Mathematical Gazette, **V**, no. 90 (Jan. 1911), 377–386.

[21] Osler, T.,
emphAn unusual product for sin z and variations of Wallis's product, Math. Gaz., **87** (2003), 134–139.

[22] Osler, T., *The tables of John Wallis and the discovery of his product for pi,* Math. Gaz., **94** (2010), 430–437.

[23] Sándor, J., Harghita, J. and Tóth, L., A remark on the gamma function, *Elemente der Mathematik,* **44** (1989), 73–76. Available at https:// eudml.org/ doc/ 141455

[24] Schlömilch, O., *Einigesüber die Eulerschen Integrale der zweiten Art,* Archiv der Mathematik und Physik. bd.**4** (1844), 167–174. Available at http:// babel.hathitrust.org/ cgi/ pt? id=njp.32101033962992; view=1up; seq=205

[25] Stedall, J. A., *Catching Proteus: The Collaborations of Wallis and Brouncker. I. Squaring the Circle,* Notes and Records of the Royal Society of London, **54**, No. 3 (Sep. 2000), 293–316.

[26] Vidūnas, R., *Expressions for values of the gamma function,* Kyushu J. Math. **59** (2005), 267–283.

[27] Wallis, J., *The Arithmetic of Infinitesimals,* (Translated from Latin by Jacqueline A. Stedall), Springer Verlag, New York, 2004.

[28] Weierstrass, K., *Ueber die Theorie der analytischen Facultäten,* Journ. für Math., **51** (1856), 1–60.

[29] Weisstein, E. W., "Gamma Function." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ GammaFunction.html

[30] Weisstein, E. W., "Gauss's Constant." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ GausssConstant.html

[31] Weisstein, E. W. "Lemniscate Constant." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ LemniscateConstant.html

[32] Whittaker, E. T. and Watson, G. N., *A Course in Modern Analysis,* Cambridge University Press, 4e, 1927.

Amrik Singh Nimbran
Former Director General of Police, Bihar, India
Email: *amrikn622@gmail.com*

Appendix

A. *Products relating to* $2^{\frac{p}{q}}$

$$\sqrt{1+\sqrt{2}} = \prod_{n=1}^{\infty} \frac{(4n-3)(8n-5)}{(4n-2)(8n-7)} = \prod_{n=1}^{\infty} \frac{(4n-2)(8n-3)}{(4n-3)(8n-1)}. \tag{39}$$

$$2^{\frac{1}{m}} = \prod_{n=1}^{\infty} \frac{(2mn-m)(2mn-2)}{(2mn-m-1)(2mn-1)}. \tag{40}$$

$$2^{\frac{1}{m}} = \prod_{n=1}^{\infty} \frac{(2mn-m+1)(2mn+1)}{(2mn-m)(2mn+2)}. \tag{41}$$

$$2^{\frac{m-1}{m}} = \prod_{n=1}^{\infty} \frac{(2mn-2m+2)(2mn-m)}{(2mn-2m+1)(2mn-m+1)},\ m \in \mathbb{N}. \tag{42}$$

$$2^{\frac{m-1}{m}} = \prod_{n=1}^{\infty} \frac{(2mn-1)(2mn+m-1)}{(2mn-m)(2mn+2m-2)},\ m \in \mathbb{N}. \tag{43}$$

$$2^{\frac{m+1}{2m+1}} = \prod_{n=1}^{\infty} \frac{[2(2m+1)n-(2m+2)][2(2m+1)n-(2m+1)]}{[2(2m+1)n-(3m+2)][2(2m+1)n-(m+1)]}. \tag{44}$$

$$2^{\frac{m+1}{2m+1}} = \prod_{n=1}^{\infty} \frac{[2(2m+1)n-m][2(2m+1)n+(m+1)]}{[2(2m+1)n-(2m+1)][2(2m+1)n+(2m+2)]}. \tag{45}$$

B. Products relating to $\sqrt{2k+1}$

$$\sqrt{2k+1} = \prod_{n=0}^{\infty}\prod_{j=1}^{k} \frac{(2k+1)^2(2n+1)^2-(2j-1)^2}{(2k+1)^2(2n+1)^2-(2j)^2}, \quad k \in \mathbb{N}. \tag{46}$$

$$\frac{\sqrt{2k+1}}{2} = \prod_{n=0}^{\infty} \frac{\left(\begin{array}{c} ((4k+2)n+4)((4k+2)n+6)\cdots \\ ((4k+2)n+2k)^2((4k+2)n+2k+2)^2 \\ ((4k+2)n+2k+4)\cdots((4k+2)n+4k-2) \end{array}\right)}{\left(\begin{array}{c} ((4k+2)n+3)((4k+2)n+5)\cdots \\ ((4k+2)n+2k+1)^2 \\ ((4k+2)n+2k+3)\cdots((4k+2)n+4k-3) \end{array}\right)}. \tag{47}$$

$$\sqrt{\frac{2k+1}{2}} = \prod_{n=0}^{\infty}\prod_{j=1}^{k} \frac{(2k)^2(2k+1)^2(2n+1)^2-(2k)^2(2j-1)^2}{(2k)^2(2k+1)^2(2n+1)^2-(2k+1)^2(2j-1)^2}. \tag{48}$$

C. Products relating to $3^{\frac{p}{q}}$ ***and*** $5^{\frac{p}{q}}$

$$\sqrt{3} = \prod_{n=1}^{\infty} \frac{(6n-4)(6n-2)}{(6n-5)(6n-1)}. \tag{49}$$

$$\sqrt{\sqrt{3}} = \prod_{n=1}^{\infty} \frac{(12n-8)^2(12n-7)}{(12n-9)^2(12n-5)}. \tag{50}$$

$$\sqrt{\sqrt{3}} = \prod_{n=1}^{\infty} \frac{(3n-1)(4n-2)^2(12n-5)}{(3n-2)(4n-1)^2(12n-7)}. \tag{51}$$

$$3^{\frac{3}{4}} = \prod_{n=1}^{\infty} \frac{(4n-2)^2(6n-2)^2(12n-5)}{(4n-1)^2(6n-5)(6n-1)(12n-7)}. \tag{52}$$

$$3^{\frac{3}{8}} = \prod_{n=1}^{\infty} \frac{(3n-1)(12n-1)(24n-18)(24n-16)(24n-7)}{(6n-1)(6n-3)(24n-21)(24n-11)(24n-3)}. \tag{53}$$

$$3^{\frac{1}{8}} = \prod_{n=1}^{\infty} \frac{(6n-3)^2(8n-1)(12n-7)(24n-16)^2}{(6n-5)(6n-1)(8n-3)(12n-5)(24n-23)(24n-7)}. \tag{54}$$

$$3^{\frac{3}{10}} = \prod_{n=1}^{\infty} \frac{(5n-4)(30n-25)(30n-5)}{(5n-3)(30n-28)(30n-8)}. \tag{55}$$

$$3^{\frac{1}{5}} = \prod_{n=1}^{\infty} \frac{(5n-2)(15n-13)(15n-8)(30n-25)(30n-5)}{(5n-3)(15n-14)(15n-4)(30n-21)(30n-9)}. \tag{56}$$

$$3^{\frac{1}{10}} = \prod_{n=1}^{\infty} \frac{(12n-9)(12n-3)}{(12n-11)(12n-1)}. \tag{57}$$

$$\frac{3\sqrt{2}+\sqrt{6}}{2} = \prod_{n=1}^{\infty} \frac{(12n-8)(12n-4)}{(12n-11)(12n-1)} = \prod_{n=1}^{\infty} \frac{(24n-19)(24n-17)}{(24n-23)(24n-13)}. \tag{58}$$

$$\sqrt{1+\sqrt{3}} = \prod_{n=1}^{\infty} \frac{(6n-5)(6n-1)(24n-21)(24n-3)}{(3n-2)(12n-1)(24n-23)(24n-7)}. \tag{59}$$

$$\sqrt{\sqrt{2}+\sqrt{3}} = \prod_{n=1}^{\infty} \frac{(3n-2)(12n-1)(24n-7)}{(6n-5)(6n-1)(24n-1)}. \tag{60}$$

$$\sqrt{\left(1+\sqrt{2}\right)\left(1+\sqrt{3}\right)\left(\sqrt{2}+\sqrt{3}\right)} = \prod_{n=1}^{\infty} \frac{(8n-7)(24n-12)(24n-9)}{(8n-6)(24n-1)(24n-23)}. \tag{61}$$

$$\sqrt{5} = \prod_{n=1}^{\infty} \frac{(10n-8)(10n-6)(10n-4)(10n-2)}{(10n-9)(10n-7)(10n-3)(10n-1)}. \tag{62}$$

$$\sqrt{\sqrt{5}} = \prod_{n=1}^{\infty} \frac{(5n-4)(5n-1)(20n-17)(20n-13)}{(5n-3)(5n-2)(20n-19)(20n-11)}. \tag{63}$$

$$\frac{\sqrt{\sqrt{5}}}{2} = \prod_{n=1}^{\infty} \frac{(10n-9)(10n-1)(20n-17)(20n-13)}{(10n-5)^2(20n-19)(20n-11)}. \tag{64}$$

$$2\sqrt{\sqrt{5}} = \prod_{n=1}^{\infty} \frac{(10n-5)^2(12n-17)(20n-13)}{(10n-7)(10n-3)(20n-19)(20n-11)}. \tag{65}$$

$$5^{\frac{1}{6}} = \prod_{n=1}^{\infty} \frac{(3n-2)(30n-25)^2(30n-5)^2}{(3n-1)(30n-28)(30n-22)(30n-16)(30n-4)}. \tag{66}$$

$$\phi = \frac{1+\sqrt{5}}{2} = \prod_{n=1}^{\infty} \frac{(5n-3)(5n-2)}{(5n-4)(5n-1)}. \tag{67}$$

$$\phi^{\frac{1}{2}} = \prod_{n=1}^{\infty} \frac{(10n-6)(10n-4)(20n-19)(20n-11)}{(10n-9)(10n-1)(20n-17)(20n-13)}. \tag{68}$$

$$\phi^{\frac{1}{2}} = \prod_{n=1}^{\infty} \frac{(10n-9)(10n-1)(20n-17)(20n-13)}{(10n-8)(10n-2)(20n-19)(20n-11)}. \tag{69}$$

$$2^{\frac{1}{5}}\,5^{\frac{1}{4}}\sqrt{1+\sqrt{5}} = \prod_{n=1}^{\infty} \frac{(20n-16)(20n-12)}{(20n-19)(20n-9)}. \tag{70}$$

$$2^{\frac{1}{5}}\sqrt{\sqrt{5}-1} = \prod_{n=1}^{\infty} \frac{(20n-16)(20n-12)(20n-11)}{(20n-17)(20n-13)(20n-9)}. \tag{71}$$

$$\sqrt{1+\sqrt{5}} = \prod_{n=1}^{\infty} \frac{(5n-3)(10n-5)^2(20n-17)(20n-13)(20n-9)}{(5n-4)(10n-8)^2(20n-6)^2(20n-11)}. \tag{72}$$

D. Products relating to L

$$\prod_{n=1}^{\infty} \frac{(4n-2)4n}{(4n-3)(4n+1)} = \frac{\left(\Gamma\left(\frac{1}{4}\right)\right)^2}{4\sqrt{\pi}} = \frac{L}{\sqrt{2}}. \tag{73}$$

$$\prod_{n=1}^{\infty} \frac{(4n-1)^2}{(4n-3)(4n+1)} = \frac{\left(\Gamma\left(\frac{1}{4}\right)\right)^4}{8\pi^2} = \frac{L^2}{\pi}. \tag{74}$$

$$\prod_{n=1}^{\infty} \frac{(4n-1)^2(4n+2)}{(4n-2)(4n+1)^2} = \frac{\left(\Gamma\left(\frac{1}{4}\right)\right)^4}{16\pi^2} = \frac{L^2}{2\pi}. \tag{75}$$

$$\prod_{n=1}^{\infty} \frac{(4n-2)(4n-1)(4n+2)}{(4n-3)(4n+1)^2} = \frac{\left(\Gamma\left(\frac{1}{4}\right)\right)^4}{8\sqrt{2}\pi^2} = \frac{L^2}{\sqrt{2}\,\pi}. \tag{76}$$

Table on the next page

TABLE 1. Values of $\sin\theta$ and $\cos\theta$

θ	$\sin\theta$	$\cos\theta$
$\mathbf{0}$	$\mathbf{0}$	$\mathbf{1}$
$\frac{\pi}{24}$	$\frac{\sqrt{4-\sqrt{2}-\sqrt{6}}}{2\sqrt{2}}$	$\frac{\sqrt{4+\sqrt{2}+\sqrt{6}}}{2\sqrt{2}}$
$\frac{\pi}{20}$	$\frac{\sqrt{3+\sqrt{5}}-\sqrt{5-\sqrt{5}}}{4}$	$\frac{\sqrt{3+\sqrt{5}}+\sqrt{5-\sqrt{5}}}{4}$
$\frac{\pi}{16}$	$\frac{\sqrt{2-\sqrt{2+\sqrt{2}}}}{2}$	$\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2}$
$\frac{\pi}{12}$	$\frac{\sqrt{3}-1}{2\sqrt{2}}$	$\frac{\sqrt{3}+1}{2\sqrt{2}}$
$\frac{\pi}{10}$	$\frac{\sqrt{5}-1}{4}$	$\frac{\sqrt{10+2\sqrt{5}}}{4}$
$\frac{\pi}{8}$	$\frac{\sqrt{2-\sqrt{2}}}{2}$	$\frac{\sqrt{2+\sqrt{2}}}{2}$
$\frac{3\pi}{20}$	$\frac{\sqrt{5+\sqrt{5}}-\sqrt{3-\sqrt{5}}}{4}$	$\frac{\sqrt{5+\sqrt{5}}+\sqrt{3-\sqrt{5}}}{4}$
$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$
$\frac{3\pi}{16}$	$\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}-\sqrt{2-\sqrt{2+\sqrt{2}}}}{2\sqrt{2}}$	$\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}+\sqrt{2-\sqrt{2+\sqrt{2}}}}{2\sqrt{2}}$
$\frac{\pi}{5}$	$\frac{\sqrt{10-2\sqrt{5}}}{4}$	$\frac{\sqrt{5}+1}{4}$
$\frac{5\pi}{24}$	$\frac{\sqrt{4+\sqrt{2}-\sqrt{6}}}{2\sqrt{2}}$	$\frac{\sqrt{4-\sqrt{2}+\sqrt{6}}}{2\sqrt{2}}$
$\frac{\pi}{4}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
$\frac{7\pi}{24}$	$\frac{\sqrt{4-\sqrt{2}+\sqrt{6}}}{2\sqrt{2}}$	$\frac{\sqrt{4+\sqrt{2}-\sqrt{6}}}{2\sqrt{2}}$
$\frac{3\pi}{10}$	$\frac{\sqrt{5}+1}{4}$	$\frac{\sqrt{10-2\sqrt{5}}}{4}$
$\frac{5\pi}{16}$	$\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}+\sqrt{2-\sqrt{2+\sqrt{2}}}}{2\sqrt{2}}$	$\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}-\sqrt{2-\sqrt{2+\sqrt{2}}}}{2\sqrt{2}}$
$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
$\frac{7\pi}{20}$	$\frac{\sqrt{5+\sqrt{5}}+\sqrt{3-\sqrt{5}}}{4}$	$\frac{\sqrt{5+\sqrt{5}}-\sqrt{3-\sqrt{5}}}{4}$
$\frac{3\pi}{8}$	$\frac{\sqrt{2+\sqrt{2}}}{2}$	$\frac{\sqrt{2-\sqrt{2}}}{2}$
$\frac{2\pi}{5}$	$\frac{\sqrt{10+2\sqrt{5}}}{4}$	$\frac{\sqrt{5}-1}{4}$
$\frac{5\pi}{12}$	$\frac{\sqrt{3}+1}{2\sqrt{2}}$	$\frac{\sqrt{3}-1}{2\sqrt{2}}$
$\frac{9\pi}{20}$	$\frac{\sqrt{3+\sqrt{5}}+\sqrt{5-\sqrt{5}}}{4}$	$\frac{\sqrt{3+\sqrt{5}}-\sqrt{5-\sqrt{5}}}{4}$
$\frac{11\pi}{24}$	$\frac{\sqrt{4+\sqrt{2}+\sqrt{6}}}{2\sqrt{2}}$	$\frac{\sqrt{4-\sqrt{2}-\sqrt{6}}}{2\sqrt{2}}$
$\frac{\pi}{2}$	$\mathbf{1}$	$\mathbf{0}$

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2015), 135-141

SOME UNIQUE FIXED POINT AND COMMON FIXED POINT THEOREMS IN b-METRIC SPACES USING RATIONAL INEQUALITY

G. S. SALUJA

(Received : 03 - 11 - 2015 ; Revised : 10 - 12 - 2015)

Abstract. The purpose of this paper is to extend and generalize the results of Khan [8] from complete metric spaces to that in the setting of complete b-metric spaces.

1. Introduction and Preliminaries

Fixed point theory plays a very important role in the development of nonlinear analysis. In this area, the first important result was proved by Banach in 1922 for contraction mapping in complete metric space, known as the Banach contraction principle [10]. In [8], Khan studied a fixed point theorem using symmetric rational expression in complete metric spaces. The Banach contraction principle with rational expressions have been expanded and some fixed and common fixed point theorems have been obtained in [5], [6].

In [2], Bakhtin introduced b-metric spaces as a generalization of metric spaces. He proved the contraction mapping principle in b-metric spaces that generalized the famous contraction principle in metric spaces. Czerwik, in [3], used the concept of b-metric space and generalized the renowned Banach fixed point theorem in b-metric spaces (see, also [4]).

In this note, we extend and generalize the results of Khan [8] and establish some unique fixed point and common fixed point theorems using Khan [8] and almost Khan rational expressions in the framework of complete b-metric spaces.

Definition 1.1. ([2]) Let X be a nonempty set and $s \geq 1$ be a given real number. A mapping $d\colon X \times X \to R_+$ is called a b-metric if for all x, y, $z \in X$, the following conditions are satisfied:

$(b1M)$ $d(x, y) = 0$ if and only if $x = y$;

$(b2M)$ $d(x, y) = d(y, x)$;

$(b3M)$ $d(x, y) \leq s[d(x, z) + d(z, y)]$.

The pair (X, d) is called a b-metric space.

2010 Mathematics Subject Classification : 47H10, 54H25.

Key words and phrases : Fixed point, common fixed point, rational inequality, b- metric space.

It is clear from the definition of b-metric that every metric space is a b-metric space for $s = 1$. Therefore, the class of b-metric spaces is larger than the class of metric spaces.

Some known examples of b-metric, which shows that a b-metric space is real generalization of a metric space, are the following.

Example 1.1. ([1]) The set of real numbers together with the functional $d(x, y) = |x - y|^2$ for all $x, y \in \mathbb{R}$ is a b-metric space with constant $s = 2$. Also, we obtain that d is not a metric on $\mathbb{R}$.

Example 1.2. ([7]) Let $X = \ell^p$ with $0 < p < 1$, where $\ell^p = \{\{x_n\} \subset \mathbb{R} : \sum_{n=1}^{\infty} |x_n|^p < \infty\}$. Let $d\colon X \times X \to \mathbb{R}$ defined by $d(x, y) = \Big(\sum_{n=1}^{\infty} |x_n - y_n|^p\Big)^{\frac{1}{p}}$, where $x = \{x_n\}$, $y = \{y_n\} \in \ell^p$. Then (X, d) is a b-metric space with the coefficient $s = 2^{1/p} > 1$, but not a metric space. It is obtained that the above result also holds for the general case $\ell^p(X)$ with $0 < p < 1$, where X is a Banach space.

Example 1.3. ([9]) Let p be a given real number in the interval $(0, 1)$. The space $L_p[0, 1]$ of all real functions $x(t), y(t) \in [0, 1]$ such that $\int_0^1 |x(t)|^p < 1$, together with the functional

$$d(x, y) := \Big(\int_0^1 |x(t) - y(t)|^p \, dt\Big)^{1/p},$$

for each $x, y \in L_p[0, 1]$ is a b-metric space with constant $s = 2^{1/p}$.

Example 1.4. (see [1]) Let $X = \{0, 1, 2\}$. Define $d\colon X \times X \to \mathbb{R}_+$ as follows $d(0, 0) = d(1, 1) = d(2, 2) = 0$, $d(1, 2) = d(2, 1) = d(0, 1) = d(1, 0) = 1$, $d(2, 0) = d(0, 2) = p \geq 2$ for $s = \frac{p}{2}$ where $p \geq 2$, the function defined as above is a b-metric space but not a metric space for $p > 2$.

Example 1.5. ([7]) Let $X = \{1, 2, 3, 4\}$ and $E = \mathbb{R}^2$. Define $d\colon X \times X \to \mathbb{R}$ by

$$d(x, y) = \begin{cases} (|x - y|^{-1}, |x - y|^{-1}) & \text{if } x \neq y, \\ 0, & \text{if } x = y. \end{cases}$$

Then (X, d) is a b-metric space with the coefficient $s = \frac{6}{5} > 1$. But it is not a metric space since the triangle inequality is not satisfied,

$$d(1, 2) > d(1, 4) + d(4, 2), \quad d(3, 4) > d(3, 1) + d(1, 4).$$

Definition 1.2. [([8])] Let (X, d) be a complete metric space. A self mapping $T\colon X \to X$ is called a Khan contraction if it satisfies following condition:

$$d(Tx, Ty) \;\; \leq \;\; \frac{\alpha\,[d(x, Tx)\, d(x, Ty) + d(y, Ty)\, d(y, Tx)]}{d(x, Ty) + d(y, Tx)} \tag{1}$$

for all $x, y \in X$ and $\alpha \in [0, 1)$.

Now, we define the following:

Definition 1.3. Let (X, d) be a complete metric space. A self mapping $T\colon X \to X$ is called an almost Khan contraction if it satisfies the following condition:

$$d(Tx, Ty) \leq \frac{\alpha\,[d(x, Tx)\,d(x, Ty) + d(y, Ty)\,d(y, Tx)]}{d(x, Ty) + d(y, Tx)} + \beta\, d(x, y) \tag{2}$$

for all $x, y \in X$ and $\alpha,\ \beta \in [0, 1)$ with $\alpha + \beta < 1$.

Remark 1.1. If we take $\beta = 0$, then almost Khan contraction (2) reduces to Khan contraction (1).

Theorem 1.1. *([8]) Let (X, d) be a complete metric space and $T\colon X \to X$ satisfies the contractive condition (1), then T has a unique fixed point in X.*

In our main result we will use the following definitions which can be found in [1] and [9].

Definition 1.4. ([1], [9]) Let (X, d) be a b-metric space, $x \in X$ and $\{x_n\}$ be a sequence in X. Then

$(d1)$ $\{x_n\}$ is a Cauchy sequence if for $\varepsilon > 0$ there exists a positive integer N such that for all $n, m \geq N$, we have $d(x_n, x_m) < \varepsilon$;

$(d2)$ $\{x_n\}$ is called convergent if for $\varepsilon > 0$ and $n \geq N$ we have $d(x_n, x) < \varepsilon$, where x is called the limit point of the sequence $\{x_n\}$. We denote this by $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ as $n \to \infty$.

$(d3)$ (X, d) is said to be a complete b-metric space if every Cauchy sequence in X converges to a point in X.

Remark 1.2. In a b-metric space (X, d), the following assertions hold:

(i) a convergent sequence has a unique limit;

(ii) each convergent sequence is Cauchy;

(iii) in general, a b-metric is not continuous.

2. Main Results

In this section we shall prove some fixed point and common fixed point theorems for rational expressions in the framework of b-metric spaces.

Theorem 2.1. *Let (X, d) be a complete b-metric space (CbMS) with the coefficient $s \geq 1$. Suppose that the mappings S, $T\colon X \to X$ satisfy the rational expression:*

$$d(Sx, Ty) \leq \frac{\alpha\,[d(x, Sx)\,d(x, Ty) + d(y, Ty)\,d(y, Sx)]}{d(x, Ty) + d(y, Sx)} \tag{3}$$

for all $x, y \in X$, $\alpha \in [0, 1)$ with $s\alpha < 1$. Then S and T have a unique common fixed point in X.

Proof. Choose $x_0 \in X$. Let $x_1 = S(x_0)$ and $x_2 = T(x_1)$ such that $x_{2n+1} = Sx_{2n}$ and $x_{2n+2} = Tx_{2n+1}$ for all $n \geq 0$. From (3), we have

$$\begin{aligned} d(x_{2n+1}, x_{2n}) &= d(Sx_{2n}, Tx_{2n-1}) \\ &\leq \frac{\alpha\big[d(x_{2n}, Sx_{2n})\,d(x_{2n}, Tx_{2n-1}) + d(x_{2n-1}, Tx_{2n-1})\,d(x_{2n-1}, Sx_{2n})\big]}{d(x_{2n}, Tx_{2n-1}) + d(x_{2n-1}, Sx_{2n})} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\alpha\big[d(x_{2n},x_{2n+1})\,d(x_{2n},x_{2n}) + d(x_{2n-1},x_{2n})\,d(x_{2n-1},x_{2n+1})\big]}{d(x_{2n},x_{2n}) + d(x_{2n-1},x_{2n+1})} \\
&\le \alpha\, d(x_{2n},x_{2n-1}). \qquad (4)
\end{aligned}
$$

By induction, we have

$$d(x_{n+1},x_n) \le \alpha\, d(x_{n-1},x_n) \le \alpha^2\, d(x_{n-2},x_{n-1}) \le \dots \le \alpha^n\, d(x_0,x_1). \qquad (5)$$

Hence for any $m, n \ge 1$ and $m > n$, we have

$$
\begin{aligned}
d(x_n,x_m) &\le s[d(x_n,x_{n+1}) + d(x_{n+1},x_m)] \\
&= sd(x_n,x_{n+1}) + sd(x_{n+1},x_m) \\
&\le sd(x_n,x_{n+1}) + s^2[d(x_{n+1},x_{n+2}) + d(x_{n+2},x_m)] \\
&= sd(x_n,x_{n+1}) + s^2d(x_{n+1},x_{n+2}) + s^2d(x_{n+2},x_m) \\
&\le sd(x_n,x_{n+1}) + s^2d(x_{n+1},x_{n+2}) + s^3d(x_{n+2},x_{n+3}) \\
&\quad + \dots + s^{n+m-1}d(x_{n+m-1},x_m) \\
&\le s\alpha^n d(x_1,x_0) + s^2\alpha^{n+1}d(x_1,x_0) + s^3\alpha^{n+2}d(x_1,x_0) \\
&\quad + \dots + s^m\alpha^{n+m-1}d(x_1,x_0) \\
&= s\alpha^n[1 + s\alpha + s^2\alpha^2 + s^3\alpha^3 + \dots + (s\alpha)^{m-1}]d(x_1,x_0) \\
&\le \Big[\frac{s\alpha^n}{1-s\alpha}\Big]d(x_1,x_0).
\end{aligned}
$$

Since $s\alpha < 1$, therefore taking limit $m, n \to \infty$, we have $\lim_{m,n\to\infty} d(x_n,x_m) = 0$. Hence $\{x_n\}$ is a Cauchy sequence in complete b-metric space X. Since X is complete, so there exists $q \in X$ such that $\lim_{n\to\infty} x_n = q$. Now we have to show that q is a common fixed point of S and T. For this consider

$$
\begin{aligned}
d(x_{2n+1},Tq) &= d(Sx_{2n},Tq)] \\
&\le \frac{\alpha\big[d(x_{2n},Sx_{2n})\,d(x_{2n},Tq) + d(q,Tq)\,d(q,Sx_{2n})\big]}{d(x_{2n},Tq) + d(q,Sx_{2n})} \\
&= \frac{\alpha\big[d(x_{2n},x_{2n+1})\,d(x_{2n},Tq) + d(q,Tq)\,d(q,x_{2n+1})\big]}{d(x_{2n},Tq) + d(q,x_{2n+1})}.
\end{aligned}
$$

Taking limit $n \to \infty$, we have $d(q,Tq) \le 0$.
Hence by definition 1.1($b1M$), we get $Tq = q$. Thus q is a fixed of T.

In an exactly the same fashion we can prove that $Sq = q$. Hence $Sq = Tq = q$. This shows that q is a common fixed point of S and T.

Uniqueness. Let p be another fixed point common to S and T, that is, $Sp = Tp = p$ such that $p \ne q$. Then from (3), we have

$$
\begin{aligned}
d(p,q) &= d(Sp,Tq) \le \frac{\alpha\,[d(p,Sp)\,d(p,Tq) + d(q,Tq)\,d(q,Sp)]}{d(p,Tq) + d(q,Sp)} \\
&= \frac{\alpha\,[d(p,p)\,d(p,q) + d(q,q)\,d(q,p)]}{d(p,q) + d(q,p)} \le 0.
\end{aligned}
$$

Hence by definition 1.1($b1M$), we get $p = q$. This shows that q is a unique common fixed point of S and T. This completes the proof. $\square$

From Theorem 2.1, we obtain the following result as corollary.

Corollary 2.1. *Let (X, d) be a complete b-metric space (CbMS) with the coefficient $s \geq 1$. Suppose that the mapping $T\colon X \to X$ satisfies the rational expression:*

$$d(Tx, Ty) \leq \frac{\alpha\,[d(x, Tx)\,d(x, Ty) + d(y, Ty)\,d(y, Tx)]}{d(x, Ty) + d(y, Tx)} \tag{6}$$

for all $x, y \in X$, $\alpha \in [0, 1)$ with $s\alpha < 1$. Then T has a unique fixed point in X.

Theorem 2.2. *Let (X, d) be a complete b-metric space (CbMS) with the coefficient $s \geq 1$ such that for positive integer n, T^n satisfies the rational expression (1) for all $x, y \in X$, $\alpha \in [0, 1)$ with $s\alpha < 1$. Then T has a unique fixed point in X.*

Proof. If x_0 is the unique fixed point of T^n, then $Tx_0 = x_0$ because $T(T^n x_0) = Tx_0$ or $T^n(Tx_0) = Tx_0$. This shows that x_0 is the unique fixed point of T, and completes the proof. □

Theorem 2.3. *Let (X, d) be a complete b-metric space (CbMS) with the coefficient $s \geq 1$. Suppose that the mapping $T\colon X \to X$ satisfies an almost Khan rational expression (2) for all $x, y \in X$, $\alpha, \beta \in [0, 1)$ with $s\alpha + s\beta < 1$. Then T has a unique fixed point in X.*

Proof. Choose $x_0 \in X$. We construct the iterative sequence $\{x_n\}$, where $x_n = Tx_{n-1}$, $n \geq 1$, that is, $x_{n+1} = Tx_n = T^{n+1}x_0$. From (2), we have

$$\begin{aligned} d(x_n, x_{n+1}) &= d(Tx_{n-1}, Tx_n) \\ &\leq \frac{\alpha\,[d(x_{n-1}, Tx_{n-1})\,d(x_{n-1}, Tx_n) + d(x_n, Tx_n)\,d(x_n, Tx_{n-1})]}{d(x_{n-1}, Tx_n) + d(x_n, Tx_{n-1})} \\ &\quad + \beta\, d(x_{n-1}, x_n) \\ &= \frac{\alpha\,[d(x_{n-1}, x_n)\,d(x_{n-1}, x_{n+1}) + d(x_n, x_{n+1})\,d(x_n, x_n)]}{d(x_{n-1}, x_{n+1}) + d(x_n, x_n)} \\ &\quad + \beta\, d(x_{n-1}, x_n) \\ &\leq (\alpha + \beta)\, d(x_{n-1}, x_n) = \lambda\, d(x_{n-1}, x_n) \end{aligned} \tag{7}$$

where $\lambda = (\alpha + \beta)$, since $s\alpha + s\beta < 1$, it is clear that $0 < \lambda < 1/s$. By induction, we have

$$\begin{aligned} d(x_{n+1}, x_n) &\leq \lambda\, d(x_{n-1}, x_n) \leq \lambda^2\, d(x_{n-2}, x_{n-1}) \leq \dots \\ &\leq \lambda^n\, d(x_0, x_1). \end{aligned} \tag{8}$$

Let $m, n \geq 1$ and $m > n$, we have

$$\begin{aligned} d(x_n, x_m) &\leq s[d(x_n, x_{n+1}) + d(x_{n+1}, x_m)] \\ &= sd(x_n, x_{n+1}) + sd(x_{n+1}, x_m) \\ &\leq sd(x_n, x_{n+1}) + s^2[d(x_{n+1}, x_{n+2}) + d(x_{n+2}, x_m)] \\ &= sd(x_n, x_{n+1}) + s^2 d(x_{n+1}, x_{n+2}) + s^2 d(x_{n+2}, x_m) \end{aligned}$$

$$\begin{aligned}
&\le sd(x_n,x_{n+1})+s^2d(x_{n+1},x_{n+2})+s^3d(x_{n+2},x_{n+3})\\
&\quad +\cdots+s^{n+m-1}d(x_{n+m-1},x_m)\\
&\le s\lambda^n d(x_1,x_0)+s^2\lambda^{n+1}d(x_1,x_0)+s^3\lambda^{n+2}d(x_1,x_0)\\
&\quad +\cdots+s^m\lambda^{n+m-1}d(x_1,x_0)\\
&= s\lambda^n[1+s\lambda+s^2\lambda^2+s^3\lambda^3+\cdots+(s\lambda)^{m-1}]d(x_1,x_0)\\
&\le \Big[\frac{s\lambda^n}{1-s\lambda}\Big]d(x_1,x_0).
\end{aligned}$$

Since $s\lambda < 1$, therefore taking limit $m,n\to\infty$, we have $\lim_{m,n\to\infty} d(x_n,x_m)=0$. Hence $\{x_n\}$ is a Cauchy sequence in complete b-metric space X. Since X is complete, so there exists $z\in X$ such that $\lim_{n\to\infty} x_n = z$. Now we have to show that z is a fixed point of T. For this consider

$$\begin{aligned}
d(x_{n+1},Tz) &= d(Tx_n,Tz)\\
&\le \frac{\alpha\big[d(x_n,Tx_n)\,d(x_n,Tz)+d(z,Tz)\,d(z,Tx_n)\big]}{d(x_n,Tz)+d(z,Tx_n)}+\beta\,d(x_n,z)\\
&= \frac{\alpha\big[d(x_n,x_{n+1})\,d(x_n,Tz)+d(z,Tz)\,d(z,x_{n+1})\big]}{d(x_n,Tz)+d(z,x_{n+1})}+\beta\,d(x_n,z).
\end{aligned}$$

Taking limit $n\to\infty$, we have $d(z,Tz)\le 0$. Hence by definition 1.1($b1M$), we get $Tz=z$. Thus z is a fixed of T.

Uniqueness. Let z_1 be another fixed point T, that is, $Tz_1=z_1$ such that $z\neq z_1$. Then from (2), we have

$$\begin{aligned}
d(z,z_1) &= d(Tz,Tz_1)\\
&\le \frac{\alpha\,[d(z,Tz)\,d(z,Tz_1)+d(z_1,Tz_1)\,d(z_1,Tz)]}{d(z,Tz_1)+d(z_1,Tz)}+\beta\,d(z,z_1)\\
&= \frac{\alpha\,[d(z,z)\,d(z,z_1)+d(z_1,z_1)\,d(z_1,z)]}{d(z,z_1)+d(z_1,z)}+\beta\,d(z,z_1)\le \beta\,d(z,z_1).
\end{aligned}$$

The above inequality is possible only if $d(z,z_1)=0$ and so $z=z_1$. Thus z is a unique fixed point of T, and the proof is completed. □

From Theorem 2.3, we obtain the following result as corollaries.

Corollary 2.2. *Let (X,d) be a complete b-metric space (CbMS) with the coefficient $s\ge 1$. Suppose that the mapping $T\colon X\to X$ satisfies Khan rational expression (1) for all $x,y\in X$ and $\alpha\in[0,1)$ with $s\alpha<1$. Then T has a unique fixed point in X.*

Corollary 2.3. *Let (X,d) be a complete b-metric space (CbMS) with the coefficient $s\ge 1$. Suppose that the mapping $T\colon X\to X$ satisfies the contraction condition $d(Tx,Ty)\le\beta\,d(x,y)$ for all $x,y\in X$ and $\beta\in[0,1)$ with $s\beta<1$. Then T has a unique fixed point in X.*

Remark 2.1. (i) Theorem 2.1 and 2.2 extend and generalize Theorem 1 of Khan [8] from complete metric space to that setting of complete b-metric space considered in this paper.

(ii) Theorem 2.1 also extends and generalizes Theorem 3 of Khan [8] from complete metric space to that setting of b-metric space considered in this paper.

(iii) Corollary 2.3 extends well known Banach [10] contraction principle from complete metric space to that setting of complete b-metric space considered in this paper.

3. Conclusion

In this paper, we establish some unique fixed point and common fixed point theorems using Khan and an almost Khan rational contractions in the setting of b-metric spaces. Our results extend and generalize several results from the existing literature.

Acknowledgements. The author would like to thank the anonymous referee for his careful reading and useful suggestions on the manuscript.

References

[1] Ayadi, H., Bota, M. F., Karapinar, E., and Mitrovic, S., A point theorem for set valued quasi contraction in b-metric spaces, *Fixed Point Theory Appl.* **2012**, Article ID 2012:88, 1-8. doi:10.1186/1687-1812-2012–88.

[2] Bakhtin, I. A., The contraction mapping principle in almost metric spaces, *Funct. Anal. Gos. Ped. Inst. Unianowsk* **30** (1989), 26–37.

[3] Czerwik, S., Contraction mappings in b-metric spaces, *Acta Math. Inf. Univ. Ostraviensis* **1** (1993), 5–11.

[4] Czerwik, S., Nonlinear set-valued contraction mappings in b-metric spaces, *Atti. Semin. Mat. Fis. Univ. Modena* **46** (1998), 263–276.

[5] Fisher, B., Common fixed points and constant mapping satisfying rational inequality, *Math. Sem. Notes* (Univ. Kobe), 1978.

[6] Fisher, B. and Khan, M. S., Fixed points, common fixed points and constant mappings, *Studia Sci. Math. Hungar.* **11** (1978), 467– 470.

[7] H. Huang, H. and Xu, S., Fixed point theorems of contractive mappings in cone b-metric spaces and applications, *Fixed Point Theory Appl.* **2013**, Article ID 2013:112, 1-10. doi: 10.1186 /1687-1812-2013-112.

[8] Khan, M. S., A fixed point theorem for metric spaces, *Rend. Ist. Mat. Univ. Trieste* **8** (1976), 69-72.

[9] Kir, Mehmet and Kizitune, Hukmi, On some well known fixed point theorems in b-metric spaces, *Turkish J. Anal. Number Theory* **1** (2013), 13-16.

[10] Rudin, W., Principles of Mathematical Analysis, Third Edition, International Editions, 1976.

G. S. Saluja
Department of Mathematics, Govt. Nagarjuna P.G. College of Science
Raipur - 492010 (C. G.), India
Email: *saluja1963@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 143-149

LAZY CONTINUED FRACTION EXPANSIONS FOR COMPLEX NUMBERS

S. G. DANI

(Received : 08 - 11 - 2015 ; Revised : 02 - 01 - 2016)

Abstract. We discuss a general notion of continued fraction expansion for complex numbers in terms of partial quotients in various subrings and prove the convergence of the corresponding sequences of convergents.

1. Introduction

For a real number t the classical simple continued fraction expansion is constructed by setting $t_0 = t$, $a_0 = [t]$, the largest integer not exceeding t, and recursively $t_{n+1} = \frac{1}{t_n - a_n}$ for $n \geq 0$, provided $t_n \neq a_n$, and $a_{n+1} = [t_{n+1}]$; we terminate the sequence when $t_n = a_n$ (for the first time), so $t_n - a_n$ can not be inverted. When t is an irrational number the sequence is non-terminating and we have

$$t = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots}},$$

in the sense that the rational numbers $\frac{p_n}{q_n}$, $n \geq 0$, obtained by omitting from the above expression the part after a_n, converge to t; $\frac{p_n}{q_n}$ are called the *convergents* and the expression is called the *simple continued fraction expansion* of t. Here $a_0 \in \mathbb{Z}$ and $a_n \in \mathbb{N}$ for all $n \geq 1$. These expansions have played an important role in various problems in number theory and its applications in other areas since the eighteenth century when they came to the fore in a big way (the reader is referred to [1] and [5] for a historical perspective and [6] for a general introduction to the topic).

If in place of the largest integer not exceeding the given number we opt for the integer nearest to the given number, for picking a_n depending on t_n at each stage $n \geq 0$, and proceed analogously (allowing the t_n's and a_n's to be negative), then again for an irrational number we get a continued fraction expansion in terms of $a_n \in \mathbb{Z}$, with $|a_n| \geq 2$ for $n \geq 1$, known as the nearest integer continued fraction expansion. These expansions have also appeared in various contexts, and were used recently in [2] for studying certain asymptotics of values of quadratic forms at integer points.

2010 Mathematics Subject Classification : 11A55
Key words and phrases : continued fractions, convergence, complex numbers, Gaussian integers, Eisenstein integers

The nearest integer continued fraction expansion has a natural analogue for complex numbers in terms of Gaussian integers $x + iy$ with $x, y \in \mathbb{Z}$. Let $z \in \mathbb{C}$ and define $z_0 = z$, a_0 the Gaussian integer nearest to z (when there are more than one such, a choice may be made by some convention which does not play much role in the theory) and recursively $z_{n+1} = \frac{1}{z_n - a_n}$ for $n \geq 0$, provided $z_n \neq a_n$, and a_{n+1} is the Gaussian integer nearest to z_{n+1} (with the convention as above); the expansion is non-terminating for $z \notin \mathbb{Q}(i)$, the smallest subfield of $\mathbb{C}$ containing i. This expansion was introduced and studied by A. Hurwitz [8] who proved various interesting results about it - including convergence of the corresponding convergents, monotonicity of the size of the denominators of the convergents, an analogue of the classical Lagrange theorem characterising quadratic irrationals, etc. (the reader is referred to [7], [4] and [3] for details in this respect). Hurwitz also discussed analogous issues for expansions in terms of Eisenstein integers, namely complex numbers of the form $x + y\omega$, where ω is a nontrivial cube root of unity and $x, y \in \mathbb{Z}$. Continued fraction expansions have also been considered with partial quotients from some more discrete subrings of $\mathbb{C}$ (see §3 below, and [3] and other references cited there).

Unlike in the classical case of simple continued fractions, the proof of convergence of the convergents in [8] and in later literature (see for example [7]) turns out to be rather intricate. In this note we present a short proof of convergence, in fact in a substantially broader set up, of what is referred to below (and in the title) as "lazy continued fraction expansion"; it should be noted however that Hurwitz's proof brings in, along with the proof of convergence of the convergents, considerable insight also into certain other aspects of behaviour of the continued fraction expansions which does not have a parallel in our case. Our main emphasis is on the generality of the framework and simplicity of the proof. The idea underlying the proof below is also involved in the proof of Proposition 3.6 in [4] in the special case of Gaussian integers; Theorem 2.1 is a generalization of that result, and in the proof an attempt is also made to improve the presentation.

2. The theorem

Let Γ be a discrete subring of $\mathbb{C}$ containing 1 and let K be the quotient field of Γ. We note that K is a countable subfield of $\mathbb{C}$. We shall be interested in continued fraction expansions of z in $\mathbb{C}\backslash K$; the latter correspond to irrational numbers in the general case. We shall assume Γ to be such that for any $z \in \mathbb{C}\backslash K$ there exists an $a \in \Gamma$ such that $|z - a| < 1$; equivalently, the union of the open unit discs with centers at points of Γ covers $\mathbb{C}\backslash K$. A brief discussion on the rings satisfying this condition is included in §3; for the present it may be noted that the ring of Gaussian integers and the ring of Eisenstein integers have this property.

Now let $z \in \mathbb{C}\backslash K$ and let $\{z_n\}$ be a sequence in $\mathbb{C}\backslash K$ constructed as follows: we start by setting $z_0 = z$ and pick *some* element $a_0 \in \Gamma$ such that $|z_0 - a_0| < 1$, and put $z_1 = (z_0 - a_0)^{-1}$; since $z \in \mathbb{C}\backslash K$, $a_0 \neq z_0$, and moreover $z_1 \in \mathbb{C}\backslash K$. Now starting with z_1 we define, following the same steps, an element z_2, and then z_3 etc.. In other words, for $n \geq 0$, we choose $z_{n+1} = \frac{1}{z_n - a_n}$ and a_{n+1} to be *some* element of Γ such that $|z_{n+1} - a_{n+1}| < 1$. We thus get a sequence $\{z_n\}_{n=0}^{\infty}$ such that $z_0 = z$, and $|z_n| > 1$ with $z_{n-1} - z_n^{-1} \in \Gamma$ for all $n \geq 1$, and the associated sequence $\{a_n\}_0^{\infty}$ in Γ such that $a_n = z_n - z_{n+1}^{-1}$ for all $n \geq 0$. The sequence $\{a_n\}_{n=0}^{\infty}$ is analogous to the classical continued fraction expansion sequences, and we shall call it a *lazy continued fraction expansion* for z; "lazy" refers to the fact that we do not put in any effort looking for an element of Γ fulfilling some criterion like being the nearest etc., but just pick our a_n to be any element of Γ such that $|z_n - a_n| < 1$. We call $\{z_n\}$ an *iteration sequence* and $\{a_n\}$ the sequence of *partial quotients* with respect to the given lazy continued fraction expansion; for $n \geq 0$, z_n and a_n are called the nth iterate and the nth partial quotient respectively. Note that any z will have (uncountably!) many lazy continued fraction expansions and corresponding iteration sequences, and sequences of partial quotients. Now let $\{p_n\}_{n=-1}^{\infty}$ and $\{q_n\}_{n=-1}^{\infty}$ be the sequences defined recursively by the relations

$$p_{-1} = 1,\ p_0 = a_0,\ p_{n+1} = a_{n+1}p_n + p_{n-1},\ \text{for all } n \geq 0,$$

$$q_{-1} = 0,\ q_0 = 1,\ q_{n+1} = a_{n+1}q_n + q_{n-1},\ \text{ for all } n \geq 0.$$

There is, a priori, no certainty that $q_n \neq 0$ for all $n \geq 1$, but if this does hold then it can be readily seen, inductively, that

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots_{\cdots + \frac{1}{a_n}}}},$$

for all $n \geq 0$, as in the case of the usual continued fraction expansions. We call (p_n, q_n) the nth convergent for the given lazy continued fraction expansion.

The following theorem, which is our main result in the paper, shows that the procedure as above indeed gives a continued fraction expansion for z.

Theorem 2.1. *Let Γ be a discrete subring of $\mathbb{C}$ such that the open unit discs with centers at elements from Γ cover $\mathbb{C}\backslash K$. Let $z \in \mathbb{C}$ and $\{a_n\}$ be the sequence of partial quotients, in Γ, with respect to a lazy continued fraction expansion of z. Let $\{(p_n, q_n)\}_{n=-1}^{\infty}$ be the corresponding sequence of convergents. Then $q_n \neq 0$ for all n and $\dfrac{p_n}{q_n} \to z$ as $n \to \infty$.*

Before going to the proof of the theorem we note the following.

Proposition 2.1. *Let the notation be as in the theorem and also let $\{z_n\}$ be the associated iteration sequence. Then $q_n z - p_n = (-1)^n (z_1 \cdots z_{n+1})^{-1}$ for all $n \geq 0$. Consequently $q_n \neq 0$ for all $n \geq 0$.*

Proof. At the outset we note that since Γ is a discrete subring of $\mathbb{C}$, for any nonzero $\gamma \in \Gamma$ we have $|\gamma| \geq 1$.

We now proceed with the proof by induction. Since $p_0 = a_0$, $q_0 = 1$ and $z - a_0 = z_1^{-1}$, the desired assertion holds for $n = 0$. Now let $n \geq 1$ and suppose the desired assertion holds for $0, 1, \ldots, n-1$. Then we have

$$\begin{aligned} q_n z - p_n &= (a_n q_{n-1} + q_{n-2})z - (a_n p_{n-1} + p_{n-2}) \\ &= a_n(q_{n-1}z - p_{n-1}) + (q_{n-2}z - p_{n-2}) \\ &= (-1)^{n-1}(z_1 \cdots z_n)^{-1} a_n + (-1)^{n-2}(z_1 \cdots z_{n-1})^{-1} \\ &= (-1)^n (z_1 \cdots z_n)^{-1}(-a_n + z_n) = (-1)^n (z_1 \cdots z_{n+1})^{-1}, \end{aligned}$$

which proves the first part of the proposition. The second part is immediate, since if $q_n = 0$ for some n then we get $|p_n| = |z_1 \cdots z_{n+1}|^{-1} < 1$, which is a contradiction since $p_n \in \Gamma$, and $p_n \neq 0$. □

Proof of the theorem. Let $\{z_n\}$ be the iteration sequence associated with the given lazy continued fraction expansion. Since $q_n \neq 0$ for all n, and $q_n \in \Gamma$, we have $|q_n| \geq 1$ and hence by Proposition 2.1 we get

$$\left| z - \frac{p_n}{q_n} \right| = |q_n|^{-1} |z_1 \cdots z_{n+1}|^{-1} \leq |z_1 \cdots z_{n+1}|^{-1}.$$

Therefore to prove the theorem it suffices to prove that $\limsup_{n\to\infty} |z_n| > 1$. We shall suppose that this is not true, and arrive at a contradiction. Thus we may assume that $|z_n| \to 1$ as $n \to \infty$.

Let C denote the unit circle in $\mathbb{C}$ centered at the origin, and let

$$\Phi = \{\theta \in C \mid |\theta - \gamma| = 1 \text{ for some } \gamma \in \Gamma, \gamma \neq 0\}.$$

Since Γ is discrete it follows that Φ is a finite subset of C. We first show that the limit points of $\{z_n\}$ are all contained in Φ.

Let ζ be any limit point of $\{z_n\}$. Since $|z_n| \to 1$ as $n \to \infty$, it follows that $\zeta \in C$, and that there exists a sequence $\{n_k\}$ in $\mathbb{N}$ such that $z_{n_k} \to \zeta$ as $k \to \infty$. Then $|z_{n_k} - a_{n_k}| = |z_{n_k+1}|^{-1} \to 1$ as $k \to \infty$. Since $z_{n_k} \to \zeta$ as $k \to \infty$ and $a_{n_k} \in \Gamma$, this implies that a_{n_k}, $k \in \mathbb{N}$, belong to a finite subset of Γ. Therefore, passing to a subsequence of $\{n_k\}$ and modifying notation we may assume that a_{n_k} is a constant sequence, say $a_{n_k} = a$ for all k. Then $a \in \Gamma$ and $a \neq 0$, and since $z_{n_k} \to \zeta$ as $k \to \infty$ we further get that $|\zeta - a| = 1$. Thus $\zeta \in \Phi$, as claimed above.

As the set of limit points of $\{z_n\}$ is contained in Φ, which is a finite set, we can fix $\varphi_n \in \Phi$ for each n such that $z_n - \varphi_n \to 0$ as $n \to \infty$. As $z_n - \varphi_n \to 0$ and $|z_n - a_n| = |z_{n+1}|^{-1} \to 1$, as $n \to \infty$, we have $|\varphi_n - a_n| \to 1$, and since Φ and Γ are discrete subsets it follows that $|\varphi_n - a_n| = 1$ for all large n, say all $n \geq n_0$.

For all n let $\rho_n = z_n - \varphi_n$ and $\sigma_n = \varphi_n - a_n$. Then $\rho_n \to 0$ as $n \to \infty$. Since $|z_n| > 1$ and $|\varphi_n| = 1$, we have $\rho_n \neq 0$ for all n. Also, since Φ is finite and Γ is discrete $\{\sigma_n\}$ is a discrete subset, and since $|\sigma_n| = 1$ for all $n \geq n_0$, it is in fact a finite set.

We have $z_n - a_n = \rho_n + \sigma_n$ for all n. Hence for any n,

$$z_{n+1} = \frac{1}{z_n - a_n} = \frac{1}{\rho_n + \sigma_n} = \frac{1}{\sigma_n} + \frac{-\rho_n}{\sigma_n(\rho_n + \sigma_n)} = \overline{\sigma}_n + \frac{-\rho_n}{\sigma_n(\rho_n + \sigma_n)}.$$

We note that

$$\left|\frac{-\rho_n}{\sigma_n(\rho_n + \sigma_n)}\right| \le \frac{|\rho_n|}{1 - |\rho_n|} \to 0, \quad \text{as } n \to \infty.$$

Thus $z_{n+1} - \overline{\sigma}_n \to 0$ as $n \to \infty$. Since we also have $z_{n+1} - \varphi_{n+1} \to 0$ as $n \to \infty$ and $\{\sigma_n\}$ and $\{\varphi_n\}$ are finite subsets of $\mathbb{C}$ this implies that $\varphi_{n+1} = \overline{\sigma}_n$ for all large n, say for all $n \ge n_1$. Thus we have

$$\rho_{n+1} = z_{n+1} - \varphi_{n+1} = z_{n+1} - \overline{\sigma}_n = \frac{-\rho_n}{\sigma_n(\rho_n + \sigma_n)},$$

for all $n \ge n_1$. Hence, for all $n \ge n_1$,

$$|\rho_{n+1}| = \frac{|\rho_n|}{|\rho_n + \sigma_n|} = \frac{|\rho_n|}{|z_n - a_n|} \ge |\rho_n|.$$

But this is a contradiction, since $\rho_n \neq 0$ and $\rho_n \to 0$ as $n \to \infty$. Hence $\limsup_{n\to\infty} |z_n| > 1$ and consequently, as noted above, $|z - \frac{p_n}{q_n}| \to 0$ as $n \to \infty$. This proves the theorem. □

3. Miscellanea

In this section we discuss various points related to the theorem.

3.1. **The rings Γ.** Let Γ be a discrete subring of $\mathbb{C}$ containing 1, other than $\mathbb{Z}$. Then it is of the form $\mathbb{Z}[i\sqrt{k}]$ or $\mathbb{Z}[\frac{1}{2} + \frac{i}{2}\sqrt{4l-1}]$ for some $k, l \in \mathbb{N}$; this can be proved by noting that as a subgroup Γ is generated by 1 and a number $z \in \mathbb{C}\backslash\mathbb{R}$, and z^2 has to be an integral combination of 1 and z; the element z can be adjusted suitably to get one of the above forms. The condition as in §2, that for all $z \in \mathbb{C}\backslash K$ there exists $a \in \Gamma$ such that $|z - a| < 1$, holds for a ring as above if and only if k and l are $1, 2$ or 3; except in the case $k = 3$ the open discs of unit radius centered at points of Γ cover the whole of $\mathbb{C}$; interestingly the five rings for which this holds are precisely the discrete subrings of $\mathbb{C}$ that are Euclidean rings (the author is thankful to Sudesh Kaur Khanduja for pointing this out). In the case of $k = 3$ the open unit discs centered at points of Γ cover all complex numbers except $\frac{1}{2}\gamma$, $\gamma \in \Gamma$, all of which belong to K.

For $k = 1$ the ring Γ as above is the ring of Gaussian integers and for $l = 1$ it is the ring of Eisenstein integers. The three other rings noted above to be Euclidean and fulfilling the desired condition, are also considered in literature with regard to continued fraction expansions; see [3] for some details and references.

3.2. **Expansions in other discrete subrings.** If Γ is a discrete subring for which not every $z \in \mathbb{C}$ is contained in an open unit disc centered at an element of Γ, one can still follow the procedure as above and get an (infinite) iteration sequence and corresponding sequence of partial quotients, for z from a certain subset of $\mathbb{C}$ (for others, at some stage we end up with a z_n for which there is no

a_n with $|z_n - a_n| < 1$). The set for which this holds can be characterised, which we leave to the reader. The argument in Theorem 2.1 applies to these sequences also (corresponding to the points from the special subset), in the sense that if $\{(p_n, q_n)\}$ is the corresponding sequence of convergents (defined analogously) then $\dfrac{p_n}{q_n}$ converges to z, as $n \to \infty$.

3.3. **More general iteration sequences.** Here we have chosen to put the condition on the iteration sequence $\{z_n\}$ that $|z_n| > 1$ for all $n \geq 1$. As far as the general procedure is concerned it is possible to weaken the condition to $|z_n| \geq 1$ for all $n \geq 1$, provided we assume that $z_{n+1} \neq z_n^{-1}$ for all n. This was in fact done in [4] where a variation of Theorem 2.1 was proved in the particular case of the ring of Gaussian integers; with the generalised notion the conclusion is upheld under an additional condition on $\{z_n\}$. In [4] an iteration sequence was referred as *non-degenerate* if it satisfied the desired condition, and *degenerate* otherwise; while we shall not go into the details of these notions here, it may be mentioned that in degenerate sequences $\{z_n\}$ one has $|z_n| = 1$ for all large n. An analogous criterion for exceptional sequences can be worked out also for the other rings Γ to which Theorem 2.1 applies.

3.4. **Restricted iteration sequences and Diophantine approximation.** In constructing the iteration sequences $\{z_n\}$ if we put a stronger condition that there exists $\rho > 1$ such that $|z_n| > \rho$ for all n, then we get a degree of control on the convergence of the convergents as in the theorem, which relates to Diophantine approximation: under this condition the (p_n, q_n)'s satisfy the relation $\left|z - \frac{p_n}{q_n}\right| \leq (\rho - 1)^{-1}|q_n|^{-1}$ for all n; see [3], Proposition 4.2. For which ρ it is possible to generate such iteration sequences will depend on the ring Γ. The lower bound can be readily determined (see [3]); for the ring of Gaussian integers it is possible for $\rho \geq \frac{1}{\sqrt{2}}$, and for the ring of Eisenstein integers $\rho \geq \frac{1}{\sqrt{3}}$ suffices. (It may be mentioned however that with these values the estimate as above is substantially weaker than those for classical continued fraction expansions).

3.5. **Some curious continued fraction expansions.** Let $t \in \mathbb{R}$ be an irrational number and consider the iteration sequence $\{t_n\}$ defined by setting $t_0 = t$, a_0 the farther of the two (real) integers within distance 1 from $t = t_0$, and inductively, $t_{n+1} = \frac{1}{t_n - a_n}$ for $n \geq 0$, and a_{n+1} the farther of the two integers within distance 1 from t_{n+1}; each successive t_n is an irrational number and hence the sequence is defined for all n. Then it is easy to verify inductively that for all $n \geq 1$, a_n is one of the integers ± 1 or ± 2. These choices are among those admissible for lazy iteration sequences, and hence by Theorem 2.1 the corresponding convergents converge to the number t. Thus we get a continued fraction expansion for each irrational, involving only these four integers as partial quotients, beyond a_0.

Now let Γ be a discrete subring of $\mathbb{C}$, with quotient field K, such that for all $z \in \mathbb{C}\backslash K$ there is $a \in \Gamma$ within distance less than 1. Proceeding analogously to the above we get a continued fraction expansion for any $z \in \mathbb{C}\backslash K$, such that the partial quotients beyond the initial one, viz. a_0, are from a fixed finite subset of Γ.

Let us consider specifically the case with Γ the ring of Gaussian integers. Let $z \in \mathbb{C}\backslash K$ and a_0 be the Gaussian integer which is the farthest among those that are within distance less than 1. Then $z - a_0$ is contained in the unit disc, and either the real part or the imaginary part of $z - a_0$ has absolute value at least $\frac{1}{2}$. It follows that $z_1 = (z - a_0)^{-1}$ is contained in the region bounded by the outer semicircles on the sides of the square with vertices at $\pm 1 \pm i$; viz. the region consisting of the square together with the unit discs with centers at ± 1 and $\pm i$. This implies that a_1 is contained in the subset $\{x + iy \in \Gamma \mid x^2 + y^2 \leq 5\}$. As we follow the same procedure with $z_2, z_3, \ldots$, it follows that all later partial quotients $a_2, a_3 \ldots$ belong to $\{x + iy \in \Gamma \mid x^2 + y^2 \leq 5\}$.

References

[1] Brezinski, Claude, *History of continued fractions and Padé approximants*, Springer Series in Computational Mathematics, 12, Springer-Verlag, Berlin, Germany, 1991, vi+551 pp.

[2] Choudhuri, Manoj and Dani, S. G., On values of quadratic forms at integer points, Math. Res. Lett. **22** (2015), no. 4, 1023–1045.

[3] Dani, S. G., Continued fraction expansions for complex numbers - a general approach, Acta Arith. **171** (2015), no. 4, 355–369.

[4] Dani, S. G. and Nogueira, Arnaldo, Complex continued fractions and values of binary quadratic forms, Trans. Amer. Math. Soc. **366** (2014), 3553–3583.

[5] Dickson, Leonard Eugene, *History of the theory of numbers. Vol. II: Diophantine analysis*, Chelsea Publishing Co., New York, USA, 1966, xxv+803 pp.

[6] Hardy, G. H. and Wright, E. M., *An Introduction to the Theory of Numbers*, Sixth edition, Revised by D. R. Heath-Brown and J. H. Silverman, Oxford University Press, Oxford, England, UK, 2008.

[7] Hensley, D., *Continued Fractions*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, USA, 2006.

[8] Hurwitz, A., Über die Entwicklung complexer Grössen in Kettenbrüche (German), Acta Math. **11** (1887), 187–200.

S. G. Dani
Department of Mathematics
Indian Institute of Technology Bombay
Powai, Mumbai-400076, Maharashtra, India
E-mail: *sdani@math.iitb.ac.in, shrigodani@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-june (2016), 151-161

ON HILBERT'S THEOREM IN DIFFERENTIAL GEOMETRY

H. A. GURURAJA

(Received : 20 - 11 - 2015 ; Revised : 03 - 12 - 2015)

Abstract. In 1901, David Hilbert proved that there does not exist any complete embedded surface of constant negative Gaussian curvature in the Euclidean 3-space $\mathbb{E}^3$. Hilbert's proof was later simplified by Bieberbach and Blaschke. In the mean time, E. Holmgren gave an alternative and a simpler proof of Hilbert's theorem in 1902. In this article, we will present Holmgren's proof and also a sketch of Hilbert's original proof. Hilbert's theorem was extended in 1964 by N. V. Efimov to variable negative curvature, provided that the curvature is bounded above by a negative constant. We will end this article with a few comments on Efimov's result.

1. Introduction

We begin by stating that all surfaces considered in this article have empty boundary. It is a well known result in the theory of embedded surfaces in $\mathbb{E}^3$ that a compact surface must have a point where the Gaussian curvature is positive. This result immediately rules out, among other things, the existence of a *compact* surface of constant negative curvature in $\mathbb{E}^3$. However, one can ask if there is a *complete* surface of constant negative curvature in $\mathbb{E}^3$. The answer is negative and this fact was first proved by Hilbert [4] in 1901.

Hilbert's theorem is of fundamental historical significance. It showed for the first time ever that the Hyperbolic geometry must be defined in an abstract setting. It is of interest to note here that some mathematicians (for instance, see [3], page no.119) have speculated that not possessing an example of a complete surface of constant negative curvature in $\mathbb{E}^3$ was probably the reason why Gauss did not make his views regarding non-Euclidean geometry public.

Hilbert's theorem is actually a little more general. It states that there exists no *isometric immersion* of any complete abstract surface of constant negative curvature into $\mathbb{E}^3$. Recall that a map is called an isometric immersion if its differential is injective at each point and, moreover, preserves the inner products. We shall give a sketch of Hilbert's proof as well as a slightly different and simpler proof of the

2010 Mathematics Subject Classification : Primary 43A85; Secondary 22E30.
Key words and phrases : Hyperbolic geometry, complete surface, Hilbert's theorem.

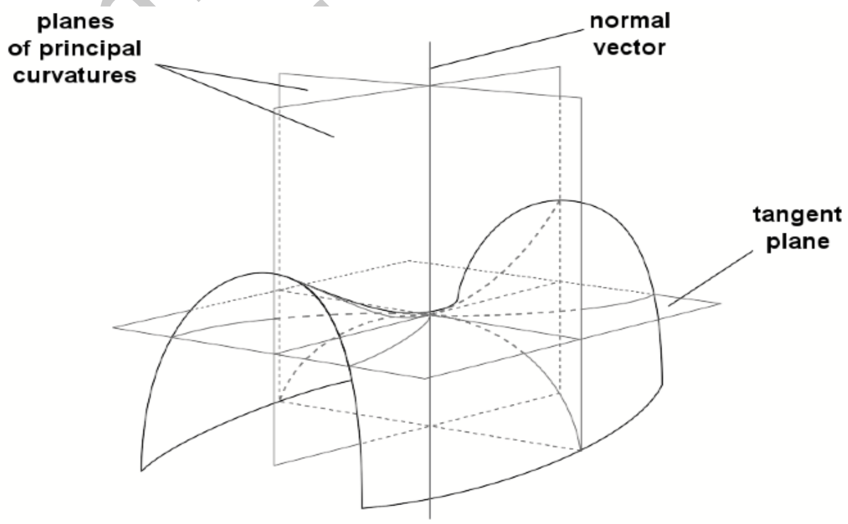
theorem discovered by Holmgren [5] in section 3 below. Efimov's generalization [2] of Hilbert's theorem will be briefly discussed in the final section.

2. PRELIMINARIES FROM SURFACE THEORY

As the proof of Hilbert's theorem involves many concepts from classical surface theory, we will briefly recall them below. For a more detailed discussion of these concepts we refer the reader to any of the standard texts on surface theory, for instance [1].

Eventhough the differential geometry of curves and surfaces in Euclidean space began immediately after the invention of calculus and analytic geometry and that there were a few results about them before, it was Gauss who brought the theory of surfaces into its modern definitive form in his great work on surface theory *Disquisitiones generales circa superficies curvas* in 1827. In this work he made in particular the remarkable discovery, the *Theorema egregium*, that the Gaussian curvature of a surface is an intrinsic invariant of the surface.

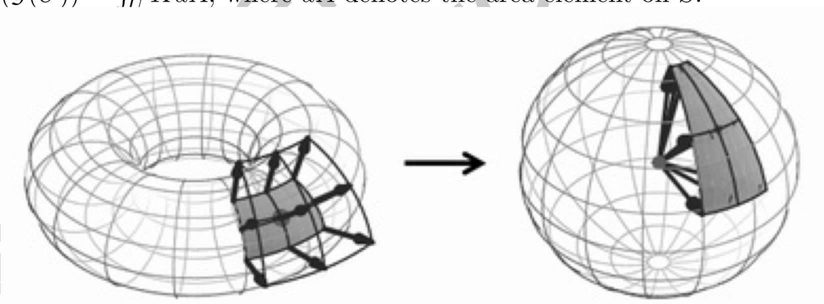
For an embedded surface S in $\mathbb{E}^3$, suppose $\mathbf{N}$ denotes a unit normal vector field to S defined in a neighborhood of some point $p \in S$. For any unit tangent vector v to S at p, the plane Π_v containing the vectors v and $\mathbf{N}(p)$ intersects S along some curve γ_v passing through p, which is called as a *normal section* of S at p. The curvature of the *plane* curve γ_v at p is called the *normal curvature* of S along the vector v at p. The maximum and minimum of the normal curvatures at p are known as the *principle curvatures* of S at p and their product is called the *Gaussian curvature* of S at p. Eventhough the principle curvatures change their signs when the normal field is changed the Gaussian curvature, being their product, will remain the same. From a qualitative point of view, if the Gaussian curvature at p is positive then there is a neighborhood of p in the surface which



lies entirely on one side of the tangent plane to the surface at p and, in the case when the Gaussian curvature is negative, any neighborhood of p lies on either sides of the tangent plane at p.

One can get an explicit expression for the Gaussian curvature of S as follows. For this purpose assume that S is given by a surface patch $\sigma : U \subset \mathbb{R}^2 \to \mathbb{R}^3$. Let $\mathbf{N} = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|}$ denote the *standard unit normal* to σ. Moreover, let $ds^2 = Edu^2 + 2Fdudv + Gdv^2$ and $II = Ldu^2 + 2Mdudv + Ndv^2$ denote the *first and second fundamental forms* of σ, respectively. Then the Gaussian curvature K of σ is given by $K = \frac{LN-M^2}{EG-F^2}$.

Gaussian curvature of a surface S in $\mathbb{R}^3$ can also be described in terms of the *Gauss spherical map* $\mathcal{G} : S \to S^2$. Here S^2 denotes the unit sphere in $\mathbb{R}^3$. In order to define this map we assume that the surface is *orientable,* i. e., S admits a smooth unit normal vector field $\mathcal{N}$. Then the Gauss map is given by $\mathcal{G}(p) = \mathcal{N}(p)$, where we now view $\mathcal{N}(p)$ as a *point* in $\mathbb{R}^3$. The idea is that to measure how curved a surface is at a point we look at how fast a normal field to the surface near the point varies. Since the tangent planes T_pS and $T_{\mathcal{G}(p)}S^2$ are the same we obtain the endomorphism $D\mathcal{G}(p) : T_pS \to T_pS$. Then $K(p) = \det D\mathcal{G}(p)$. In fact, this was how Gauss defined "Gaussian" curvature. We have the following geometric interpretation of the Gaussian curvature: The (oriented) area of the image on S^2 of an infinitesimal region around p in S under the Gauss map is $K(p)$ times the area of this infinitesimal region. Therefore, if $U \subset S$ is a "nice" set then $Area(\mathcal{G}(U)) = \int_U KdA$, where dA denotes the area element on S.



Quite often a problem in differential geometry can be simplified by introducing a suitable parametrization of the surface. A unit tangent vector v at p to S is called an *asymptotic direction at* p if the normal curvature of S at p along v vanishes. In terms of the second fundamental form II of a parametrization this is equivalent to the condition $II(v, v) = 0$. Clearly, if v is an asymptotic direction at p then so is the vector $-v$. If $K(p)$ is positive then both the principle curvatures at p have the same sign and therefore there are no asymptotic directions at p. Now suppose that $K(p) < 0$. If $II = Ldu^2 + 2Mdudv + Ndv^2$, then a unit vector $v = (a, b)$ is a asymptotic direction precisely when $La^2 + 2Mab + Nb^2 = 0$. This last equation may be viewed as a quadratic equation in $\frac{a}{b}$ with *positive* discriminant $M^2 - LN$ and therefore will have two distinct solutions. As a consequence there are (upto sign) exactly two asymptotic directions at p whenever $K(p)$ is negative. A curve on S is called an *asymptotic curve* if at each point on the curve its tangent vector is along an asymptotic direction. One can write down the differential equation for such curves: $\gamma(t) = (u(t), v(t))$ is an asymptotic curve if and only if it satisfies the differential equation $Lu'^2 + 2Mu'v' + Nv'^2 = 0$. If S has negative Gaussian curvature then it follows from the theory of ordinary differential equations (existence, uniqueness and smooth dependence of solutions on the initial data) that through each point of S there pass (upto orientation) exactly two distinct asymptotic curves and that these curves are the parameter curves of some parametrization of a neighborhood of each point in the surface. Clearly, in such a parametrization one has $L = N = 0$.

Another parametrization that can sometimes be useful in certain problems is the so called *Tschebyshef net.* For any *abstract surface* S an *immersion* $\sigma : (a, b) \times (c, d) \to S$ is called a Tchebyshef net if all the parameter curves have unit speed. Equivalently, the first fundamental form of the surface patch σ has the form $ds^2 = du^2 + 2Fdudv + dv^2$. It is not difficult to show that near any point of any surface such a parametrization exists. As we shall see below, the notion of Tschebyshef nets makes a surprising and crucial appearance in the proof of Hilbert's theorem.

We shall also require some global notions. We deal with abstract surfaces in what follows and the word curvature shall mean the *sectional curvature* of the surface. For an embedded surface in $\mathbb{E}^3$ the sectional curvature agrees with the Gaussian curvature.

Recall that a surface is called *complete* if every geodesic on the surface is defined for all time. This corresponds to the familiar axiom for the Euclidean plane geometry that a straight line segment in the plane can be extended to infinity in both directions. A famous result, the Hopf-Rinow theorem, asserts that this notion of completeness of the surface is equivalent to the notion of completeness of the surface as a metric space, where the distance between any two points on the surface is defined to be the infimum of the lengths of all piecewise smooth paths on the surface joining these points.

Another notion that appears in Hilbert's proof is that of a *covering surface.* A covering surface of a surface S is another surface $\widetilde{S}$ together with a surjective map $\Phi : \widetilde{S} \to S$, called the *covering map*, such that for any point $q \in S$, there exists an open neighborhood U of q such that $\Phi^{-1}(U)$ is a disjoint union of open sets in $\widetilde{S}$, each of which is mapped diffeomorphically by Φ onto U. We refer the reader to [1], Chapter 5, for a detailed discussion of this notion. The key point for us will be the fact that any surface admits a (essentially unique) *simply connected* covering surface called the *universal covering surface* of the given surface. Recall that a path connected surface is said to be simply connected if, roughly speaking, every closed curve on the surface can be continuously deformed to a point within the surface. As an illustrative example, for the case of the Torus $T^2 (= S^1 \times S^1)$, the plane $\mathbb{R}^2$ will be the universal covering surface with respect to the mapping $\Phi : \mathbb{R}^2 \to T^2$ given by $\Phi(x, y) = (e^{ix}, e^{iy})$.

The covering map becomes a *local isometry* when the universal covering surface is given the pullback metric by the covering map and, therefore, these surfaces have the same curvatures. Moreover, it is not difficult to see that the universal covering surface is complete precisely when the surface is complete.

The curvature of a surface influences the behavior of geodesics on the surface. Roughly speaking, if the the curvature of the surface is positive then nearby geodesics converge as in the case of a Sphere whereas if the curvature of the surface is negative then they diverge as in the case of the Hyperbolic plane. This last statement is usually made precise by introducing the notion of *exponential map* and *Jacobi field*. Assume that S is complete and $p \in S$. For any $v \in T_pS$ there is a unique geodesic γ such that $\gamma(0) = p$ and $\gamma'(0) = v$. We set $\exp_p(v) = \gamma(1)$, i.e., $\exp_p(v)$ is the point on the surface obtained by following the geodesic γ starting from p and going upto a distance equal to the length of v. The mapping $\exp_p : T_pS \to S$ so obtained is called the exponential map based at p. It is a consequence of the inverse function theorem from calculus that this map is a diffeomorphism of a neighborhood of 0_p in T_pS onto some neighborhood of the point p in the surface. This fact allows us to introduce the analogue of *polar coordinates* around p which turn out to be extremely important in the study of the geometry of the surface. Jacobi fields are variation fields of geodesic variations and relate the behavior of geodesics with the curvature of the surface. By a detailed study of the relationship between geodesics and curvature one can prove the following important global results which will be used later on (Proofs of these theorems can be found in [1], Chapter 5).

Bonnet-Myers theorem. If the curvature of a complete surface S is bounded below by a positive number then S must be compact.

Model spaces of constant curvature. Let S be a complete simply connected surface which has constant curvature K. Then

S must be isometric to the Euclidean plane $\mathbb{E}^2$ if $K = 0$;

S must be isometric to the unit sphere S^2 in $\mathbb{R}^3$ if $K = 1$;

S must be isometric to the Hyperbolic plane $\mathbb{H}^2$ if $K = -1$.

Cartan-Hadamard theorem. Suppose that a complete simply connected surface S has nonpositive curvature. Then S must be diffeomorphic to a plane.

All the three global results extend to higher dimensions as well.

3. Statement and proof of Hilbert's theorem

Theorem (Hilbert, 1901). Let S be a complete abstract surface of constant negative curvature. Then there does not exist any isometric immersion of S into $\mathbb{E}^3$.

Proof. We will simultaneously give Holmgren's and Hilbert's proofs below. The proofs will be divided into many steps. Steps 1 to 3 are common to both the proofs.

Step 1. Let S be a surface in $\mathbb{E}^3$ of constant curvature $K < 0$ and $p \in S$. We can find a parametrization about p which is a Tschebyshef net and such that the parameter curves are asymptotic curves.

Proof of Step 1. Since S has negative curvature, we know that the asymptotic curves give a parametrization near p. Thus we have a parametrization $\sigma : (-\epsilon, \epsilon) \times (-\epsilon, \epsilon) \to S,\ \ \sigma(0,0) = p$, such that the parameter curves are asymptotic curves. Let $ds^2 = Edu^2 + 2Fdudv + Gdv^2$ and $II = Ldu^2 + 2Mdudv + Ndv^2$ denote the first and second fundamental forms of this parametrization, respectively. Without loss of generality, we may assume that the parameter curves passing through p have unit speed. Hence $E(s,0) = 1$ and $G(0,t) = 1$. Using the constancy of the curvature, we show below that *all* parameter curves have unit speed.

Let $\mathbf{N} = (\sigma_u \times \sigma_v / \|\sigma_u \times \sigma_v\|)$ denote the standard unit normal associated to σ and let $D = \sqrt{EG - F^2}$. We know that $\mathbf{N}_u \times \mathbf{N}_v = K\sigma_u \times \sigma_v$. Therefore

$$(\mathbf{N} \times \mathbf{N}_v)_u - (\mathbf{N} \times \mathbf{N}_u)_v = 2\mathbf{N}_u \times \mathbf{N}_v = 2KD\mathbf{N}.$$

Also $\mathbf{N} \times \mathbf{N}_u = \frac{1}{D}\{\langle \sigma_u, \mathbf{N}_u \rangle \sigma_v - \langle \sigma_v, \mathbf{N}_u \rangle \sigma_u\} = \frac{1}{D}(M\sigma_u - L\sigma_v)$ and $\mathbf{N} \times \mathbf{N}_v = \frac{1}{D}(N\sigma_u - M\sigma_v)$. Since $K = \frac{-M^2}{D^2}$ and $L = N = 0$, we conclude that $\mathbf{N} \times \mathbf{N}_u = \pm\sqrt{-K}\sigma_u,\ \ \mathbf{N} \times \mathbf{N}_v = \mp\sqrt{-K}\sigma_v$ so that $2KD\mathbf{N} = \mp\sqrt{-K}\sigma_{uv} \mp \sqrt{-K}\sigma_{vu} = \mp 2\sqrt{-K}\sigma_{uv}$.

Hence σ_{uv} is parallel to $\mathbf{N}$ which implies $E_v = 0 = G_u$. Also $0 = (\langle \sigma_u, \mathbf{N} \rangle)_u = L + \langle \sigma_u,\ N_u \rangle = \langle \sigma_u, \mathbf{N}_u \rangle$. Therefore $E_u = (\langle \sigma_u, \mathbf{N} \rangle)_u = L + \langle \sigma_u, \mathbf{N}_u \rangle = 0$. Similarly $G_v = 0$. Hence E, G are constants and must be equal to 1 throughout. Thus the first fundamental form of σ has the form $ds^2 = du^2 + 2Fdudv + dv^2$ and we are done with the proof of Step 1. □

Step 2. Let S be an abstract surface and let $\sigma : (a,b) \times (c,d) \to S$ be a Tschebyshef net. Define the *angle function* $\omega : (a,b) \times (c,d) \to \mathbb{R}$ as follows: $0 < \omega(s_0, t_0) < \pi$ is the angle between the parameter curves passing through $g(u_0, v_0)$. Then ω satisfies the partial differential equation

$$\frac{\partial^2 \omega}{\partial u \partial v} = -K \sin \omega, \tag{3.1}$$

where K denotes the curvature of S.

Proof of Step 2. Since σ is a Tschebyshef net the first fundamental form of the net has the form $ds^2 = du^2 + 2Fdudv + dv^2$. One form of Gauss's *Theorema Egregium* states that $K = \frac{1}{\sqrt{EG-F^2}}\{(\frac{F_u}{\sqrt{EG-F^2}})_v + (\frac{F_v}{\sqrt{EG-F^2}})_u\}$. Here $E = G = 1, \sqrt{EG - F^2} = \sin\omega$. Therefore the above equation can be written in the form (3.1). □

Step 3. Let S be a complete abstract surface of constant negative curvature. Assume that S can be immersed in $\mathbb{E}^3$. Then there exists an asymptotic Tschebyshef net $F : \mathbb{R}^2 \to S$.

Proof of Step 3. Fix a point $p_0 \in S$ and consider an asymptotic curve $c(u)$ passing through p_0. Since c has unit speed and S is complete, it is not difficult to see that c is defined on all of $\mathbb{R}$. Now choose the other asymptotic curve through p_0 whose velocity at p_0 is V_0. For each $u \in \mathbb{R}$, let $\gamma_u(v)$ denote the unique asymptotic curve passing through $c(u)$ whose velocity vector at $c(u)$ is obtained from continuous extension of V_0. Now define $F : \mathbb{R}^2 \to S$ by $F(u, v) = \gamma_u(v)$.

We have to show that each parameter curve $u \to F(u, v)$ is an asymptotic curve of S. This will be shown using the existence of an asymptotic Tschebyshef net $\sigma : (-\epsilon, \epsilon)\times(-\epsilon, \epsilon) \to S$ about each point of S, as proved in step 1. First observe that if for *some* $v \in (-\epsilon, \epsilon)$ the parameter curve $u \to \sigma(u, v)$ lies along a parameter curve $u \to F(u, \overline{v})$, then *all* parameter curves $u \to \sigma(u, v)$ lie along parameter curves of F. Given $(u_0, v_0) \in \mathbb{R}^2$, we can find a finite number of Tschebyshef nets $\sigma_1, \sigma_2, ..., \sigma_n$ whose images cover the set $\{F(u_0, t) : 0 \leq v \leq v_0.\}$ Arranging the σ_i's so that the images of the consecutive ones overlap and using the fact that the curve $u \to F(u, 0)$ is an asymptotic curve (by definition of F), we see that the curve $u \to f(u, v_0)$ must be an asymptotic curve for u sufficiently close to u_0. □

Holmgren's proof. In view of steps 2 and 3, to prove Hilbert's theorem it is enough to prove the following:

There does not exist a function $\omega : \mathbb{R}^2 \to \mathbb{R}$ satisfying $\frac{\partial^2\omega}{\partial u\partial v} = C\sin\omega$, $0 < \omega < \pi$, for any constant $C > 0$.

Assume that such a solution ω exists. Then $\frac{\partial^2\omega}{\partial u\partial v} > 0$ so that $\frac{\partial\omega}{\partial u}$ is increasing in v and hence $\frac{\partial\omega(u,v)}{\partial u} > \frac{\partial\omega(u,0)}{\partial u}$ for $v > 0$. Integrating this inequality gives $\omega(b, v) - \omega(a, v) > \omega(b, 0) - \omega(a, 0)$ for $v > 0$ and $a < b$. Without loss of generality, we can assume that $\frac{\partial\omega(0,0)}{\partial u} \neq 0$. Further, since the function $(u, v) \to \omega(-u, -v)$ also satisfies the differential equation, we may actually assume that $\frac{\partial\omega(0,0)}{\partial u} > 0$. Now fix three numbers $0 < u_1 < u_2 < u_3$ with

$$\frac{\partial\omega(u, 0)}{\partial u} > 0$$

for $0 \leq u \leq u_3$ and set $\epsilon = min\{\omega(u_3, 0) - \omega(u_2, 0), \omega(u_1, 0) - \omega(0, 0)\}$. Then for $v > 0$ and $u \in [0, u_3]$ we have $\omega(u, v)$ is increasing in $u, \omega(u_1, v) - \omega(0, v) > \epsilon$ and $\omega(u_3, v) - \omega(u_2, v) > \epsilon$. Therefore $\epsilon \leq \omega(u, v) \leq \pi - \epsilon$ for $u \in [u_1, u_2]$ and $v \geq 0$. Thus $\sin(\omega(u, v)) \geq \sin\epsilon$ for $u \in [u_1, u_2]$ and $v \geq 0$.

If we integrate the differential equation $\frac{\partial^2\omega}{\partial u \partial v} = C\sin\omega$ over the rectangle $[u_1, u_2] \times [0, T]$ for any $T > 0$, we obtain $C\int_0^T \int_{u_1}^{u_2} \sin\omega(u,v)dudv = \omega(u_2, T) - \omega(u_1, T) - \omega(u_2, 0) + \omega(u_1, 0)$, so that $\omega(u_2, T) - \omega(u_1, T) = \omega(u_2, 0) - \omega(u_1, 0) + C\int_0^T \int_{u_1}^{u_2} \sin\omega(u,v)dudv \geq \omega(u_2, 0) - \omega(u_1, 0) + CT(u_2 - u_1)\sin\epsilon$. However this gives a contradiction for large enough T, as the left hand side is smaller than π. □

Sketch of Hilbert's proof. The basis of Hilbert's proof is the following lemma.

Lemma 1. (Formula of Hazzidakis). Let S be an abstract surface of constant curvature $K < 0$ and let $\sigma : (a, b) \times (c, d) \to S$ be a Tschebyschef net. Then any quadrilateral formed by parameter curves of σ has area utmost $\frac{2\pi}{-K}$.

Proof. If Q denotes a quadrilateral formed by parameter curves of σ with vortices $(u_1, v_1), (u_2, v_1)$, (u_2, v_2) and (u_1, v_2) with corresponding interior angles $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, then using step2 we see that $Area(Q) = \int_Q \sin\omega du \wedge dv = \frac{1}{-K}\int_{u_1}^{u_2}\int_{v_1}^{v_2} \frac{\partial^2\omega}{\partial u\partial v}dudv = \frac{1}{-K}(\omega(u_2, v_2) - \omega(u_1, v_2) - \omega(u_2, v_1) + \omega(u_1, v_1)) = \frac{1}{-K}([\alpha_3 - (\pi - \alpha_4) - (\pi - \alpha_2) + \alpha_1] = \frac{1}{-K}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - 2\pi) < \frac{2\pi}{(-K)}$. □

Lemma 2. The area of the Hyperbolic plane $\mathbb{H}^2$ is infinite.

Proof. The hyperbolic plane is the upper half plane endowed with the metric $g = \frac{dx^2+dy^2}{y^2}$. Therefore its area is $\int_{-\infty}^{\infty}\int_0^{\infty} \frac{1}{y^2}dxdy = \infty$. □

Now we can complete Hilbert's proof as follows. Without loss of generality we may assume that the constant curvature is -1. If there is an immersion of S into $\mathbb{E}^3$, then by going to the universal covering surface $\tilde{S}$ of S, we can find an isometric immersion of $\tilde{S}$ into $\mathbb{E}^3$. Note that $\tilde{S}$ must be isometric to $\mathbb{H}^2$ and therefore must have infinite area. Now, if we are able to show that the asymptotic Tschebyshef net considered in step 3 above can be extended to form a parametrization of *all* of the *simply connected surface* $\tilde{S}$, then using Lemmas 1 and 2 we obtain a contradiction. Thus establishing the existence of a *global* asymptotic Tschebyshef net is the main (and the harder) part in Hilbert's proof. The reader is referred to [7] or [1] for a complete proof. Also see [8] and [9]. □

4. Efimov's theorem

An example. The Hyperboloid. Consider the surface $x^2 + y^2 = 1 + z^2$ in $\mathbb{R}^3$ which is known as the Hyperboloid of revolution.

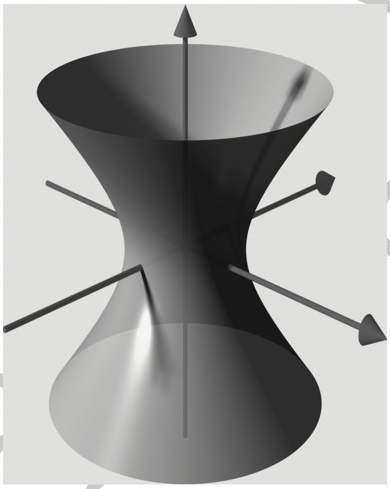
A straightforward calculation shows that the Gaussian curvature K of the Hyperboloid is given by the formula $K(x, y, z) = -\frac{1}{(1+2z^2)^2}$. It follows that the Gaussian curvature of the Hyperboloid is strictly negative everywhere. The fact that the Hyperboloid is complete can be seen, for example, from the following proposition (It may be worth remarking that the converse of the following proposition is false).

Proposition 1. An embedded surface $S \subset \mathbb{R}^3$ which is also a *closed* subset of $\mathbb{R}^3$ must be complete.

Proof. Let γ be any unit speed geodesic of S defined on the interval $[0, b)$. We show that γ can be extended to a geodesic α which is defined on $[0, \infty)$.

Choose a sequence t_n in $[0, b)$ such that $t_n \to b$. Now $|t_n - t_m| \geq d_S(\alpha(t_n), \alpha(t_m)) \geq \|\alpha(t_n) - \alpha(t_m)\|$, since straight line segments realize the Euclidean distance. Thus the sequence $\alpha(t_n)$ is Cauchy and hence converges in $\mathbb{R}^3$ to some point p. Since S is closed we actually have $p \in S$.

We can choose a neighborhood of W of p in S and a number $\delta > 0$ such that for every $q \in W, \exp_q$ is a diffeomorphism on the ball $B_\delta(0)$ and $W \subset \exp_q(B_\delta(0))$ (See [1], page no. 300, Proposition 1). In this case any two points of W are connected by a unique geodesic of length less than δ. If $\alpha(t_n)$ and $\alpha(t_m)$ belong to W with $|t_n - t_m| < \delta$, then the unique geodesic of length less than δ joining $\alpha(t_n)$ and $\alpha(t_m)$ must coincide with γ. Since $\exp_{\alpha(s_n)}$ is a diffeomorphism in $B_\delta(0)$ and $\exp_{\alpha(s_n)}(B_\delta(0)) \subset W$, γ extends α beyond p. Thus α is defined at $t = b$. $\square$



Observe that in the Hyperboloid example the Gaussian curvature tends to 0 as $z \to \pm\infty$. In 1930's, Cohn-Vossen had conjectured that there exists no isometric immersion of a complete abstract surface of negative curvature into $\mathbb{E}^3$, provided that the curvature remains bounded away from 0. This conjecture was settled in [2] by the Russian mathematician Efimov in 1964.

Theorem (Efimov, 1964). Let S be a complete abstract surface whose curvature K is negative and satisfies $K \leq k < 0$ for some constant $k < 0$. Then there exists no isometric immersion from S into $\mathbb{E}^3$.

The proof of Efimov's theorem is too long and delicate to give here in complete detail. We will, therefore, only give a very brief sketch of the key points involved. The interested reader is referred to the paper [6] for the complete proof.

To begin with, we need a generalization of Lemma 2.

Proposition 2. Let S be a complete *simply connected* abstract surface of non-positive curvature. Then the area of S must be infinite.

Proof. Fix any point $p \in S$. The key point here is that the exponential map $exp_p : T_pS \to S$ is a diffeomorphism, by Cartan-Hadamard theorem. This fact allows us to set up a polar co-ordinate system on *all* of S. The metric in polar co-ordinates (r, θ) (with p as the origin) has the form $ds^2 = dr^2 + \phi^2(r,\theta)d\theta^2$, where the function ϕ satisfies $\phi(0,\theta) = 0$ and $\frac{\partial\phi(0,\theta)}{\partial r} = 1$. Since $K(r,\theta) = \frac{-\phi_{rr}(r,\theta)}{\phi(r,\theta)} \leq 0$, we see that $\phi_{rr}(r,\theta) \geq 0$ and hence $r \to \phi(r,\theta)$ is increasing. Since $\frac{\partial\phi(0,\theta)}{\partial r} = 1$ and ϕ satisfies $\phi(0,\theta) = 0$ we conclude that $\phi(r,\theta) \geq r$ for all (r,θ). Therefore the volume of $M = \int_0^\infty \int_0^{2\pi} \phi(r,\theta) d\theta dr \geq \int_0^\infty \int_0^{2\pi} r d\theta dr = \infty$. $\square$

As in the case of Hilbert's proof we may assume that S is simply connected by going to the universal covering surface of S. By Cartan-Hadamard theorem, S must be diffeomorphic to the plane. Let Ω denote the surface S equipped with the metric obtained from pulling back the Riemannian metric of the 2-Sphere S^2 to S via the Gauss map $\mathcal{G} : S \to S^2$. Then Ω has constant positive curvature 1. Note, however, that Ω cannot be complete. Otherwise, it would be compact by the Bonnet-Meyer theorem and, therefore, the surface S must have a point where the curvature must be positive, which is contrary to the hypothesis. Now, Efimov considers the *metric* completion $\overline{\Omega}$ of Ω. The Gauss map $\mathcal{G} : \Omega \to S^2$ extends as a continuous map $\overline{\mathcal{G}} : \overline{\Omega} \to S^2$. By a deep study of the behavior of this map near the "boundary" of Ω, i.e., near the set $\overline{\Omega} \setminus \Omega$, Efimov proves that Ω has *finite* area (in fact, of area atmost 4π). Taking this fact for granted, the proof of Efimov's theorem can be completed as follows

Fix a point $p \in S$. Let $A(r)$ denote the area of the geodesic ball $D_r(p)$ of radius r at p in S and let $A^*(r)$ be the area of this ball in Ω. Since $|K| \geq k > 0$ on S, we have $4\pi \geq A^*(r) = \int_{D_r(p)} |K| dA \geq k \int_{D_r(p)} dA = kA(r)$, which is impossible in view of Proposition 2.

We end this article by remarking that the regularity hypothesis in Efimov's (and Hilbert's) theorem can be weakened. It suffices for the surface to be of class C^2. See [6] for further details.

Acknowledgment. The author wishes to thank Prof. C. S. Aravinda for his careful reading of the original draft of this article and suggesting many changes which have helped in improving the exposition and also for providing the figures appearing in this article. Thanks are also due to the referee for valuable comments on the article.

References

[1] Do Carmo, M., Differential Geometry of Curves and Surfaces, Prentice-Hall, Inc., 1976.

[2] Efimov, N. V., Generation of singularities on surfaces of negative curvature, *Sbornik*, **64** (1964), 286–320.

[3] Gromov, M., Spaces and questions, Visions in Mathematics, *GAFA 2000 special volume*, part 1, 118–161.

[4] Hilbert, D., Über Flächen von konstanter Gausscher Krummung, *Trans. Amer. Math. Society*, **2** (1901), 87–99.

[5] Holmgren, E., Sur les surfaces à courbure constante négative, *C. R. Acad. Sci. Paris*, **134** (1902), 740–743.

[6] Klotz-Milnor, T., Efimov's Theorem About Complete Immersed Surfaces of Negative Curvature, *Adv. in Math.*, **8** 1972, 474–543.

[7] Spivak, M., A Comprehensive Introduction to Differential Geometry, Vol. 3, Publish or Perish.

[8] Stoker, J. J., Differential Geometry, Wiley Classics Library, 1969.

[9] Willmore, T. J., An introduction to differential geometry, Oxford university press, 1959.

H. A. Gururaja
TIFR Centre for Applicable Mathematics, P.O. Box 6503
Sharadanagar, GKVK Post, Bangalore - 560065
E-mail: *gururaj@math.tifrbng.res.in*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June (2016), 163-166

A PAIR OF SIMULTANEOUS DIOPHANTINE EQUATIONS WITH NO SOLUTIONS IN INTEGERS

AJAI CHOUDHRY

(Received : 23 - 11 - 2015)

Abstract. This paper is concerned with the simultaneous diophantine equations $x_1^m + x_2^m = y_1^m + y_2^m$, $x_1^n + x_2^n = y_1^n + y_2^n$, where m, n are distinct positive integers. It is shown that for arbitrary m, n, these equations have no solutions in nonnegative integers; further when m, n are of the same parity, these equations have no solutions in integers, whether positive or negative.

This paper is concerned with the simultaneous diophantine equations

$$x_1^m + x_2^m = y_1^m + y_2^m, \tag{1}$$

$$x_1^n + x_2^n = y_1^n + y_2^n, \tag{2}$$

where m, n are distinct positive integers. It has been shown by Sinha [4] that these two equations do not have a nontrivial solution in integers when $1 \le m \le 4$ and $1 \le n \le 4$. The special case when $(m, n) = (2, 3)$ has also been considered independently by other authors ([1], [2], [3]).

Any solution $(x_1, x_2, y_1, y_2) = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ of eqs. (1) and (2) will be called a trivial solution if the pair of numbers, α_1, α_2, is a permutation of the numbers β_1, β_2. Further, when both m and n are odd integers, any solution of the type $(x_1, x_2, y_1, y_2) = (\alpha, -\alpha, \beta, -\beta)$ will also be considered as a trivial solution. All other solutions will be referred to as nontrivial solutions.

In this paper it is shown that if m, n are arbitrary distinct positive integers, the simultaneous equations (1) and (2) have no nontrivial solutions in nonnegative integers. It is also shown that when the integers m, n are of same parity, the simultaneous equations (1) and (2) have no nontrivial solutions in integers, whether positive or negative. These results are proved in the two theorems that follow.

Theorem 1. If m and n are arbitrary distinct positive integers, the simultaneous equations (1) and (2) have no nontrivial solutions in nonnegative integers.

Proof. Without loss of generality, we assume that $m > n$. Let $(x_1, x_2, y_1, y_2) = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ be a nontrivial solution in nonnegative integers of eqs. (1) and (2) with $\alpha_i \ge 0,\ \beta_i \ge 0,\ i = 1, 2$.

2010 Mathematics Subject Classification : 11D41
Key words and phrases : high degree diophantine equations; symmetric diophantine equations.

We first show that none of the integers α_1, α_2, β_1, β_2 can be 0. It is readily seen that two or more of these four integers cannot be 0. If only one of the four integers, say α_2, is 0 then we have

$$\alpha_1^n = \beta_1^n + \beta_2^n, \tag{3}$$

so that

$$\alpha_1 > \beta_1, \quad \alpha_1 > \beta_2. \tag{4}$$

Multiplying (3) by α_1^{m-n} we get

$$\begin{aligned} \alpha_1^m &= \alpha_1^{m-n}\beta_1^n + \alpha_1^{m-n}\beta_2^n, \\ &> \beta_1^m + \beta_2^m, \end{aligned} \tag{5}$$

in view of the relations (4). This contradicts the assumption that $(\alpha_1, 0, \beta_1, \beta_2)$ is a solution of (1). Similarly, the other integers α_1, β_1, β_2 are seen to be nonzero and henceforth we assume that all the integers α_1, α_2, β_1, β_2 are positive.

We assume without loss of generality that $\alpha_1 \geq \alpha_2$ and $\beta_1 \geq \beta_2$. Let

$$\alpha_1^n + \alpha_2^n = \beta_1^n + \beta_2^n = 2s, \tag{6}$$

so that we may write

$$\alpha_1^n = s + d_1, \quad \alpha_2^n = s - d_1, \quad \beta_1^n = s + d_2, \quad \beta_2^n = s - d_2, \tag{7}$$

where s is a positive rational number and d_1, d_2 are distinct rational numbers such that $0 \leq d_i < s,\ i = 1, 2$ and we may assume without loss of generality that $d_1 > d_2$. It follows from (7) that

$$\alpha_1 = (s + d_1)^{1/n}, \ \ \alpha_2 = (s - d_1)^{1/n}, \ \ \beta_1 = (s + d_2)^{1/n}, \ \ \beta_2 = (s - d_2)^{1/n}. \tag{8}$$

We thus get

$$\alpha_1^m + \alpha_2^m = (s + d_1)^{m/n} + (s - d_1)^{m/n}, \tag{9}$$

$$\beta_1^m + \beta_2^m = (s + d_2)^{m/n} + (s - d_2)^{m/n}. \tag{10}$$

We now consider the following function

$$f(d) = (s + d)^{m/n} + (s - d)^{m/n}, \tag{11}$$

where s is a fixed positive rational number. The derivative of $f(d)$ is given by

$$f'(d) = \frac{m}{n}.\{(s + d)^{(m-n)/n} + (s - d)^{(m-n)/n}\}. \tag{12}$$

When $0 \leq d < s$, both the numbers $(s+d)^{(m-n)/n}$ and $(s-d)^{(m-n)/n}$ are positive and hence their arithmetic mean is not less than their geometric mean. Thus, when $0 \leq d < s$, we get

$$f'(d) \geq \frac{2m}{n}.(s^2 - d^2)^{(m-n)/n} > 0 \tag{13}$$

and hence the function $f(d)$ is monotonically increasing in this interval. Since $0 \leq d_i < s,\ i = 1, 2$, it follows that when $d_1 > d_2$ we have $f(d_1) > f(d_2)$. It now follows from (9), (10) and (11) that

$$\alpha_1^m + \alpha_2^m > \beta_1^m + \beta_2^m. \tag{14}$$

This contradicts the assumption that $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ is a solution of (1). It follows that there cannot exist a solution in nonegative integers of the simultaneous equations (1) and (2). This proves the theorem.

Theorem 2. If m and n are arbitrary distinct positive integers of the same parity then the simultaneous equations (1) and (2) have no nontrivial solutions in integers, whether positive or negative.

Proof. If the integers m and n are both even, any nontrivial solution $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ of eqs. (1) and (2) involving negative integers immediately yields a nontrivial solution $(|\alpha_1|, |\alpha_2|, |\beta_1|, |\beta_2|)$ in nonegative integers of these equations. This contradicts Theorem 1.

If the integers m and n are both odd, we assume without loss of generality that $m > n$. Let $(x_1, x_2, y_1, y_2) = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ be a nontrivial solution of eqs. (1) and (2) so that

$$\alpha_1^m + \alpha_2^m = \beta_1^m + \beta_2^m, \tag{15}$$

$$\alpha_1^n + \alpha_2^n = \beta_1^n + \beta_2^n. \tag{16}$$

In view of Theorem 1 at least one of the four integers $\alpha_1, \alpha_2, \beta_1, \beta_2$ must be negative. We now have the following cases.

Case-1. Only one of the four integers $\alpha_1, \alpha_2, \beta_1, \beta_2$ is negative. Without loss of generality we may take $\alpha_1 > 0,\ \alpha_2 < 0,\ \beta_1 > 0,\ \beta_2 > 0$, so that we obtain, from (15) and (16), the following relations after transposition

$$\alpha_1^m = |\alpha_2|^m + \beta_1^m + \beta_2^m, \tag{17}$$

$$\alpha_1^n = |\alpha_2|^n + \beta_1^n + \beta_2^n. \tag{18}$$

It follows from (18) that $\alpha_1 > |\alpha_2|$, $\alpha_1 > \beta_1$ and $\alpha_1 > \beta_2$. Now, on multiplying (18) by α_1^{m-n} we get

$$\begin{aligned} \alpha_1^m &= \alpha_1^{m-n}|\alpha_2|^n + \alpha_1^{m-n}\beta_1^n + \alpha_1^{m-n}\beta_2^n, \\ \text{or,}\quad \alpha_1^m &> |\alpha_2|^m + \beta_1^m + \beta_2^m, \end{aligned} \tag{19}$$

which contradicts the relation (17).

Case-2. Two of the four integers $\alpha_1, \alpha_2, \beta_1, \beta_2$ are negative. Clearly the two negative integers cannot both be on the same side of eqs. (1) and (2). Thus, one of the two integers α_1, α_2 must be negative, and similarly, one of the two integers β_1, β_2 must be negative. Since both m and n are odd we can obtain from (15) and (16), after suitable transpositions, a nontrivial solution of eqs. (1) and (2) in nonnegative integers in contradiction of Theorem 1.

Case-3. When three of the four integers $\alpha_1, \alpha_2, \beta_1, \beta_2$ are negative, we obtain relations of the type (17) and (18) by suitable transpositions. As already seen earlier such solutions cannot exist. Similarly, it is readily seen that there cannot exist any solution in which all four integers $\alpha_1, \alpha_2, \beta_1, \beta_2$ are negative.

Thus, we see that in all cases when m and n are both even or are both odd, the existence of a nontrivial solution of the simultaneous equations (1) and (2) leads to a contradiction. It follows that when m and n are of the same parity there cannot exist a solution of eqs. (1) and (2) in integers, whether positive or negative. This proves the theorem.

Acknowledgment. I wish to thank the Harish-Chandra Research Institute, Allahabad, for providing me all the necessary facilities that have helped me to pursue my research work in mathematics.

References

[1] Gatteschi, L. and Rosati, L. A., *Risposta ad una questione proposta da A. Moessner*, Boll. Un. Mat. Ital. (3) **5** (1950), 43–48.

[2] Gupta, H., *A system of equations having no nontrivial solutions*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 181–182.

[3] Schmidt, H., *Über einige diophantische Aufgaben 3. und 4. Grades*, Jber. Deutsch. Math.-Verein **67** (1964/65) Abt. 1, 2–13.

[4] Sinha, T. N., *A relation between the coefficients and roots of two equations and its application to diophantine problems*, J. Res. Nat. Bur. Standards Sect. B **74B** (1970), 31–36.

Ajai Choudhry
13/4 A Clay Square, Lucknow - 226001, UP, India
E-mail: *ajaic203@yahoo.com*

The Mathematics Student ISSN: 0025-5742
Vol. 85, Nos. 1-2, January-June, (2016), 167-174

PROBLEM SECTION

Problems of the Mathematics Student Vol. 84, Nos. 1-2, January-June, (2015) (MS, 84, 1-2, 2015 for short): We published Proposer's solution to the *Problem-4* and the correct submitted solution from the floor to the *Problems 5, 9 and 11* in the MS, 84, 3-4, 2015. Till April 30, 2016, no responses are received to the remaining *Problems 2, 3, 6, 7, 8 and 10* and hence we provide in this issue the Proposer's solution to these problems. Problem 1 is an open problem.

Problems of the MS, 84, 3-4, 2015: One submitted solution to the *Problem-18* is received from the floor which is correct and is published in this issue. Readers can try their hand on the remaining *Problems 12, 13, 14, 15, 16 and 17* till October 31, 2016.

In this issue we first present **eight new problems**. Solutions to these problems as also to the remaining *Problems 12, 13, 14, 15, 16 and 17* of MS, 84, 3-4, 2015, received from the floor till October 31, 2016, if approved by the Editorial Board, will be published in the MS, 85, 3-4, 2016.

MS-2016, Nos. 1-2: Problem-01: Proposed by M. Ram Murty.

Let ϕ denote Euler's function. Show that for any positive real number a, there is a constant $C(a)$ such that

$$\sum_{n\leq x}\left(\frac{n}{\phi(n)}\right)^a \leq C(a)x.$$

MS-2016, Nos. 1-2: Problem-02: Proposed by B. Sury

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

$$(a + A)(b + B)(c + C)(d + D) = \pm 1.$$

MS-2016, Nos. 1-2: Problem-03: Proposed by B. Sury .

Show that the number $11\cdots122\cdots25$ where 1 is repeated 2015 times and 2 is repeated 2016 times, is a perfect square.

MS-2016, Nos. 1-2: Problem-04: Proposed by B. Sury .

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

$$f(x)g(x) = a_1x^{p_1} + a_2x^{p_2} + \cdots + a_nx^{p_n}$$

for some integers a_i's.

MS-2016, Nos. 1-2: Problem-05: Proposed by B. Sury .

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.

MS-2016, Nos. 1-2: Problem-06: Proposed by Raja Sridharan, TIFR, Mumbai; submitted through Clare D'Cruz.

Let A be a commutative ring with unity, M a finitely generated A-module, and I be an ideal of A.

(1) Suppose that $IM = M$. Then show that there exists an $x \in I$ such that $(1 + x)M = 0$.
(2) Suppose I is a finitely generated ideal such that $I^2 = I$. Then show that I is generated by an idempotent, i.e., there exists an element x such that $I = (x)$ and $x^2 = x$.
(3) Suppose $x_1, ..., x_n \in I$ and $(x_1, ..., x_n) + I^2 = I$. Then show that there exists $x \in I$ such that $I = (x_1, ..., x_n, x)$.
(4) Let $x \in A$ be an idempotent element and let $y \in A$ be any other element. Then show that the ideal $I = (x, y)$ is principal.
(5) Let $I \subseteq A$ be a finitely generated ideal. Suppose there exists $x_1, ..., x_n \in I$ such that $(x_1, ..., x_n) + I^2 = I$. Then show that for any $y \in A, (I, y)$ can be generated by $n + 1$ elements.
(6) Let (A, m) be a local ring with 1 and $I \subseteq m$ an ideal. Suppose $x_1, ..., x_n \in I$ are such that $(x_1, ..., x_n) + I^2 = I$. Then $I = (x_1, ..., x_n)$.
(7) Is (6) true for any ideal I in a Noetherian ring with 1? What are classes of rings for which (6) is true?

MS-2016, Nos. 1-2: Problem-07: Proposed by Purusottam Rath, CMI, Chennai; submitted through Clare D'Cruz.

Let $M = (a_{ij})$ be an $n \times n$ matrix with $a_{ij} \in \mathbb{R}$ and $|a_{ij}| \leq 1$. Show that the absolute value of determinant of M is at most $n^{n/2}$. When does equality hold?

MS-2016, Nos. 1-2: Problem-08: Proposed by Purusottam Rath, CMI, Chennai; submitted through Clare D'Cruz.

> let A and B be finite subsets of integers. Consider the set $A+B$ given by
> $$A+B := \{a+b : a \in A, b \in B\}.$$
> Show that $|A+B| \geq |A|+|B|-1$.

Solution from the floor: MS-2015, Nos. 3-4: Problem 18: Evaluating the product, prove that

$$\tfrac{\sqrt{2}}{1+\sqrt{2}} \cdot \tfrac{\sqrt{2+\sqrt{2}}}{1+\sqrt{\sqrt{2}}} \cdot \tfrac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{1+\sqrt{\sqrt{\sqrt{2}}}} \cdots = \tfrac{\ln 4}{\pi}.$$

(Solution submitted on 08-01-2016 by **Subhash Chand Bhoria**; Corporal, Air Force Station, Bareilly, Technical Flight, UP-243002, India; *scbhoria@yahoo.com*).

Solution. Put $x_n = \sqrt{2+x_{n-1}}$, $w_n = \frac{x_n}{2}$, for $n \geq 1$, where in $x_0 = 0$. We have

$$\lim_{k\to\infty} \prod_{n=1}^{k} w_n = \frac{2}{\pi}$$

in view of Viéte's formula (see Beckmann, Petr, *A history of* π, Second Edition, The Golden Press, Boulder, Colo. (1971), 94-95). For $y_n = 1+2^{1/2^n}$ for all $n \geq 1$, we then obtain

$$\lim_{k\to\infty} \prod_{n=1}^{k} \frac{x_n}{y_n} = \frac{2}{\pi}\left(\lim_{k\to\infty} \prod_{n=1}^{k} \frac{2}{y_n}\right) = \frac{2}{\pi}\left(\lim_{k\to\infty} \frac{2^k}{\prod\limits_{n=1}^{k}\left(1+2^{1/2^n}\right)}\right).$$

Observe that

$$\begin{aligned}
\frac{1}{\prod_{n=1}^{k}\left(1+2^{1/2^n}\right)} &= \frac{1 \times \left(1-2^{1/2^k}\right)}{\left(\left(1+2^{1/2}\right)\left(1+2^{1/4}\right)\cdots\left(1+2^{1/2^k}\right)\right) \times \left(1-2^{1/2^k}\right)} \\
&= \frac{\left(1-2^{1/2^k}\right)}{\left(1+2^{1/2}\right)\left(1+2^{1/4}\right)\cdots\left(1+2^{\frac{1}{2^{k-1}}}\right)\left(1-2^{\frac{1}{2^{k-1}}}\right)} \\
&= \frac{\left(1-2^{1/2^k}\right)}{\left(1+2^{1/2}\right)\left(1-2^{1/2}\right)} = \left(2^{1/2^k}-1\right).
\end{aligned}$$

Therefore

$$\lim_{k\to\infty} \prod_{n=1}^{k} \frac{x_n}{y_n} = \frac{2}{\pi}\left(\lim_{k\to\infty} 2^k\left(2^{1/2^k}-1\right)\right).$$

Now take $\frac{1}{2^k} = t$. Noticing that t tends to 0 as k tends to infinity and using L'Hôpital's rule, we then obtain

$$\frac{2}{\pi}\left(\lim_{k\to\infty} 2^k\left(2^{1/2^k}-1\right)\right) = \frac{2}{\pi}\left(\lim_{t\to 0} \frac{(2^t-1)}{t}\right) = \frac{2}{\pi}(\ln 2) = \frac{\ln 4}{\pi}$$

which completes the proof.

Solution by the Proposer George E. Andrews: MS-2015, Nos. 1-2:
Problem 2: Prove that, for $|q| < 1$ and b not a negative power of q,

$$\sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)(1-bq^n)} = \sum_{n=1}^{\infty} nq^n \prod_{m=n}^{\infty} \frac{(1-q^{m+1})}{(1-bq^m)}. \tag{0.1}$$

Fokkink, R., Fokkink, W. and Wang, B. (*A relation between partitions and the number of divisors*, Amer. Math. Monthly, **102**(1995), 345-347) proved that the number of divisors of n equals

$$-\sum_{\pi \in \mathcal{D}_n} (-1)^{\#(\pi)} \sigma(\pi),$$

where $\mathcal{D}_n$ is the set of integer partitions of n into distinct parts, $\#(\pi)$ is the number of parts of π and $\sigma(\pi)$ is the smallest part of π.

Ex. When $n = 6$, $\mathcal{D}_6 = \{6, 5+1, 4+2, 3+2+1\}$, so the above sum is $6 - 1 - 2 + 1 = 4$, and there are four divisors of 6, namely 1,2,3, and 6.

Deduce the Fokkink, Fokkink and Wang theorem from (0.1).

Solution. We require the following notation

$$(A;q)_n = \prod_{j=0}^{n-1} (1 - Aq^j) \quad \text{and} \quad (A;q)_\infty = \prod_{j=0}^{\infty} (1 - Aq^j).$$

Also we need the classical q-binomial series

$$\sum_{n=0}^{\infty} \frac{(A;q)_n t^n}{(q;q)_n} = \frac{(At;q)_\infty}{(t;q)_\infty}. \tag{0.2}$$

Then

$$\begin{aligned}
\sum_{n=1}^{\infty} nq^n \prod_{m=n}^{\infty} \frac{(1-q^{m+1})}{(1-bq^m)} &= \frac{d}{dz}\Big|_{z=1} \sum_{n=0}^{\infty} z^n q^n \frac{(q^{n+1};q)_\infty}{(bq^n;q)_\infty} \\
&= \frac{d}{dz}\Big|_{z=1} \frac{(q;q)_\infty}{(b;q)_\infty} \sum_{n=0}^{\infty} \frac{z^n q^n (b;q)_n}{(q,q)_n} \\
&= \frac{d}{dz}\Big|_{z=1} \frac{(q;q)_\infty}{(b;q)_\infty} \frac{(zbq;q)_\infty}{(zq,q)_\infty} \qquad \text{(by (0.2))} \\
&= \frac{(q;q)_\infty}{(b;q)_\infty} \left\{ \frac{(bq;q)_\infty}{(q;q)_\infty} \left(\sum_{n=1}^{\infty} \frac{q^n}{1-q^n} - \sum_{n=1}^{\infty} \frac{bq^n}{1-bq^n} \right) \right\} \\
&\qquad \text{(by the rule of differentiation of products)} \\
&= \frac{1}{1-b} \left(\sum_{n=1}^{\infty} \frac{(1-bq^n)q^n - bq^n(1-q^n)}{(1-q^n)(1-bq^n)} \right) \\
&= \frac{1}{1-b} \sum_{n=1}^{\infty} \frac{q^n(1-b)}{(1-q^n)(1-bq^n)} \\
&= \sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)(1-bq^n)}.
\end{aligned}$$

If one sets $b = 0$, we note that the left hand side of (0.1) becomes

$$\sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)} = \sum_{n=1}^{\infty} d(n)q^n,$$

where $d(n)$ is the number of divisors of n. The right hand side of (0.1) becomes

$$\sum_{n=1}^{\infty} nq^n(1-q^{n+1})(1-q^{n+2})(1-q^{n+3})\cdots$$

and we see that the n^{th} term of this series counts, with weight $\pm n$, the partitions with distinct parts whose smallest part is n. The "$+''$ occurs if the partition in question has an odd number of parts, and "$-''$ occurs for an even number of parts. Thus the coefficient of q^n is

$$-\sum_{\pi\in\mathcal{D}_n} (-1)^{\#(\pi)}\sigma(\pi).$$

As an added bonus, we observe that if $b = q$, we have

$$\sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)(1-q^{n+1})} = \sum_{n=1}^{\infty} nq^n,$$

This implies that the number of partitions of n in which the difference between the largest and smallest parts is at most one equals n.

For example, there are five such partitions of 5:

5; $\quad 3+2$; $\quad 2+2+1$; $\quad 2+1+1+1$; $\quad 1+1+1+1+1$.

Solution by the Proposer Kannappan Sampath: MS-2015, Nos. 1-2: Problem 3: Let $\mathbb{H}$ denote the division algebra of quaternions over the field $\mathbb{R}$ of real numbers. Let n be a positive integer. Suppose that A and B are two $n\times n$ matrices such that $AB = I_{n\times n} = I$, say. Show that $BA = I$.

Solution. Throughout, we shall view of $\mathbb{H}$ as the $\mathbb{R}$-algebra given by

$$\mathbb{H} = \mathbb{R}\oplus\mathbb{R}i\oplus\mathbb{R}j\oplus\mathbb{R}k,$$

where i, j, k are subject to the usual relations $i^2 = j^2 = k^2 = ijk = -1$. We begin by remarking that every real quaternion can be written uniquely as $z_1 + z_2 j$ where $z_1, z_2 \in \mathbb{C}$. In view of this, if A is an $n\times n$ matrix over $\mathbb{H}$, we may write A uniquely as

$$A = A_1 + A_2\mathbf{j} \tag{0.3}$$

where $A_1, A_2 \in M_{n\times n}(\mathbb{C})$ and $\mathbf{j} = \mathrm{diag}(j,\ldots,j) \in M_{n\times n}(\mathbb{H})$. For $a, b \in \mathbb{R}$, we note that $(a+bi)j = j(a-bi)$ so that for a matrix $C \in M_{n\times n}(\mathbb{C})$, we have $C\mathbf{j} = \mathbf{j}\bar{C}$. It is now routine to verify that

$$M_{n\times n}(\mathbb{H}) \to M_{2n\times 2n}(\mathbb{C}) \quad \text{such that } A \mapsto \begin{pmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{pmatrix}$$

is an $\mathbb{R}$-algebra isomorphism. The proof is now straightforward. We have

$$AB = I \iff \begin{pmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{pmatrix}\begin{pmatrix} B_1 & B_2 \\ -\bar{B}_2 & \bar{B}_1 \end{pmatrix} = I$$

$$\iff \begin{pmatrix} B_1 & B_2 \\ -\bar{B}_2 & \bar{B}_1 \end{pmatrix} \begin{pmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{pmatrix} = I$$
$$\iff \qquad BA = I.$$

invoking the fact that $ST = I$ if and only if $TS = I$ for square matrices S and T over a field. This completes the proof.

Remark. Compare with Bourbaki's Algebra, Corollary to Proposition 9 in Chapter II, §7.4, pg. 298.

Solution by the Proposer Amritanshu Prasad: MS-2015, Nos. 1-2: Problem 6: Let X be the set of all subsets of $\{1, \ldots, n\}$ of size 2 (so X has $\binom{n}{2}$ elements). Let V be the complex vector space

$$V = \Big\{ f : X \to \mathbf{C} \mid \sum_{\{s \in X \mid i \in s\}} f(s) = 0 \ \forall\ i \in \{1, 2, \cdots, n\} \Big\}^*.$$

Show that V is a vector space of dimension $n(n-3)/2$ over $\mathbf{C}$.

* The inadvertant typographical error in the definition of V in the statement of this problem that appeared in the Mathematics Student **84**, nos. 1-2, (2015) is regretted.

Solution. Let U be the space of all functions $X \to \mathbf{C}$. For each $i \in \{1, \ldots, n\}$, let L_i be the linear functional $U \to \mathbf{C}$ defined by

$$L_i f = \sum_{\{s \in X \mid i \in s\}} f(s).$$

Then V is the subspace of U on which all the L_i's vanish. We claim that

$$\{L_i \mid 1 \le i \le n\}$$

is a linearly independent set in the dual space of U. Indeed, if $L = \sum_{i=1}^n \alpha_i L_i$ for some $\alpha_1, \ldots, \alpha_n \in \mathbf{C}$, and for each $s \in X$ $\delta_s \in U$ is the function which 1 at s and 0 at all other points of U, then

$$L\Big(\sum_{\{s \in X \mid i \in s\}} \delta_s \Big) = \sum_{j \neq i} (\alpha_i + \alpha_j) = (n-1)\alpha_i + \sum_{j \neq i} \alpha_j.$$

Also

$$L\Big(\sum_{\{s \in X \mid i \notin s\}} \delta_s \Big) = (n-2) \sum_{j \neq i} \alpha_j.$$

Thus

$$\alpha_i = L\Big(\frac{1}{n-1} \sum_{\{s \in X \mid i \in s\}} \delta_s - \frac{1}{(n-1)(n-2)} \sum_{\{s \in X \mid i \notin s\}} \delta_s \Big),$$

showing that α_i can be recovered from L for each $i \in \{1, \ldots, n\}$. It follows that L is linearly independent, and so V, which is the intersection of the kernels of the linear functionals $L_1, \ldots, L_n$ must have dimension

$$\dim U - n = \binom{n}{2} - n = \frac{n(n-3)}{2}, \quad \text{as required.}$$

Solution by the Proposer Ashay Burungale: MS-2015, Nos. 1-2: Problem 7: Consider a kingdom made up of an odd number $2k+1$ of islands. There are bridges connecting some of the islands. Suppose each set S of k islands has the property that there is an island not among S which is connected by a bridge to each island in S. Show that some island of the kingdom has bridges connecting it to all the other islands.

Solution. Any island I is part of some set of k islands and hence, there is another island J connected to it by a bridge. Thus, the set $\{I, J\}$ has the property that each pair of islands in it (there is only one!) is connected by a bridge. So, one may consider a (non-empty) set T of islands in the kingdom which has maximum cardinality such that every pair of islands in T is connected by a bridge. There may be more than one choice for T but fix one. We claim that T has cardinality at least $k+1$. If not, then consider any set T_0 of k islands containing T. By hypothesis, there is an island I not in T_0 which is connected to each island in T_0. But, then $T \cup \{I\}$ is a larger (than T) set of islands in which every pair is connected. This contradicts the choice of T. Therefore, $|T| \geq k+1$. Let $S \subset T$ be any subset with exactly $k+1$ islands. Of course, each pair of islands in S is connected by bridges (as this is true of T). Since the complement of S has exactly k islands, the hypothesis implies there is an island I in S connected to each island in the complement of S. But then I is connected to all islands.

Solution by the Proposer Abhijin Adiga: MS-2015, Nos. 1-2: Problem 8: Consider the set $S = \{1, 2, ..., n\}$. Let S_i be n distinct non-empty subsets of S. Prove that there exists an element $k \in S$, such that $S_i \setminus \{k\}$ are still distinct

Solution. Consider the subsets as the vertices of an n-dimensional hypercube. Now if there are two subsets S_i and S_j such that $S_i - \{k\} = S_j - \{k\}$, then clearly k was in one of them and not in the other. Thus there was an edge in the hypercube between the corresponding two vertices. Mark this edge with the letter k- if there are many such edges, just pick one and mark it. Now if for all $k \in \{1, 2, \ldots, n\}$, such and edge can be found, then the marked n edges define a subgraph of the hypercube on n vertices. Thus there is a cycle in this subgraph, which means more than one of these edges are marked with the same letter k: If we added k to a subset to move away from it, we have removed it to come back. Contradiction.

Solution by the Proposer Mathew Francis: MS-2015, Nos. 1-2: Problem 10: Given a t-regular graph that is properly t-edge coloured with colours $1, 2, \cdots t$. Consider any cut of the graph and let k_i be the number of edges coloured i crossing the cut. Show that $k_1 = k_2 = \cdots = k_t \ mod\ 2$.

Solution. Suppose not. Note that each color class has to be a perfect matching, since each vertex sees all the three colors. Consider a cut $(A, V-A)$. Now if $|A|$ is even then all of k_1, k_2, k_3 should be even. Else suppose k_1 is odd. Consider the matching of k_1 edges colored with 1 going from A to $V-A$. Let A' be the end points of these edges on the A side. Clearly $A' \subset A$ since $|A|$ is an even number and $|A'|$ is odd. Since the set of edges colored by 1 is a perfect matching, we should have a perfect matching on the set of vertices $A - A'$ which is not possible, since $|A - A'|$ is odd.

In a similar way, if $|A|$ were odd, then all of k_1, k_2, k_3 should be odd. Otherwise let k_1 be even, and we will have $|A - A'|$ odd in this case also, which makes it impossible to have a perfect matching on $A - A'$.

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
Signature of the Publisher

Dated: 3rd May 2016

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. 85, Nos. 1-2, January-June, (2016)

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