ISSN: 0025-5742

THE MATHEMATICS STUDENT

Volume 84, Numbers 3-4, July-December, (2015)
(Issued: November, 2015)

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 84, Numbers 3-4, July-December, (2015)
(Issued: November, 2015)

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. 84, Nos. 3-4, July-December, (2015)

CONTENTS

****

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 01-23

WAVE PROPAGATION IN LOCAL AND NONLOCAL MICROSTRETCH ELASTIC MEDIA*

S. K. TOMAR

Abstract. The present article is concerned with the propagation of plane waves in microstretch solid and fluid media. The article has been divided into two parts. Part-I contains the exploration of the possibility of propagation of plane waves in local microstretch solid and fluid media of infinite extent. It has been found that five basic waves may travel through an infinite microstretch media of either solid or liquid nature consisting of two sets of coupled longitudinal waves, two sets of coupled transverse waves and an uncoupled longitudinal micro-rotational wave. All the waves are found to be dispersive and attenuating in nature when propagating through microstretch fluid, while they are dispersive but not attenuating when propagating through microstretch solid medium. The phase speeds and attenuation coefficients are computed numerically for specific models and depicted graphically. Reflection phenomenon of dilatational wave from stress free boundary of a local microstretch fluid half-space has also been investigated. The nature of dependence of various reflection coefficients against the angle of incidence have also been displayed through graphs. Part-II contains first a brief introduction to nonlocal theory of elasticity and then governing equations and constitutive relations for nonlocal microstretch elastic solids have also been derived. These equations and relations of local microstretch solids medium are fully recovered when nonlocality parameter is neglected. It has also been shown that the waves traveling through nonlocal microstretch solid medium are influenced by nonlocality parameter of the medium and remain dispersive as well.

1. Introduction

Eringen and his coworker (1964a,b) have developed the theory of micro-elastic solids in the later half of the previous century, in which local deformation of material points in any of its volume element has been considered. This means that each tiny particle of a continua is considered as a tiny deformable body and Eringen asserted that this tiny body can undergo deformation similar to entire body. This theory is based upon the following intuitive idea: The mathematical model of

* The text (modified) of the 28^{th} P. L. Bhatnagar Memorial Award Lecture delivered at the 80^{th} Annual Conference of the Indian Mathematical Society held at the Indian School of Mines, Dhanbad, Jharknand, during December 27-30, 2014.

2010 Mathematics Subject Classification: 74B99, 74J10, 74J20.

Keywords and Phrases: Microstretch, nonlocal waves, reflection, phase speed, attenuation.

the continuum theory assumes that the mass is a continuous measure so that a continuous mass density ρ exists in a volume element dV that is infinitesimally small. Molecular theories of matter has shown that when volume $\triangle V$ is less than a limit, say $\triangle V^*$, such is not the case. Therefore, on physical grounds, the mathematical volume element dV is an idealization of a volume $\triangle V \geq \triangle V^*$. If we desire to refine our theory so that we can explain mechanical phenomena contained in $\triangle V < \triangle V^*$, we must consider the deformations of the material contained in the volume $\triangle V$. To this end, we assume that the 'macro-mass element' dM contains continuous mass distributions, so that the total macro-mass dM is the average of all masses in dV. Therefore, a continuous mass distribution is assumed to exist at each point of dV such that the average of local masses over dV gives the total mass dM. This implies that the continuum theory is valid at each point of a macro-element dV, however, we must take some statistical averages to obtain the micro-deformation theory of continuous media [see Eringen and Suhubi (1964a)]. Basically, Eringen and his co-workers assumed the existence of three directors at the center of a tiny particle and said that it can deform and rotate about its center of mass. In micropolar theory of elasticity, these three directors at the center of mass of the tiny particle are assumed to be rigid, so that it cannot deform, but may rotate during the deformation of the body. The rotation of tiny particle about its center of mass enable to add an extra kinematic variable in the theory. Linear theory of micropolar elasticity was presented by Eringen in the year 1966, which is an extension of classical theory of elasticity and a special case of micromorphic theory. In classical theory of elasticity, particle is allowed to undergo translation only, giving rise to three degrees of freedom, while in micropolar theory of elasticity, the particle can under go rotation independently, in addition to translation and thereby giving six degrees of freedom. Hence, micropolar theory of elasticity is an extension of classical theory of elasticity. Moreover, since the directors of the tiny particle of micropolar material are assumed to be rigid and preventing it to deform implies that micropolar elasticity is a special case of micromorphic theory of elasticity. Later, Eringen (1990) relaxed the condition of rigidness of directors and allowed that the directors can stretch (contract or extend) in their directions only, giving birth to the theory of microstretch elasticity. Thus, in microstretch theory of elasticity, there are seven degrees of freedom, namely, three of translation, three of rotation and one of scalar microstretch. Hence, the theory of microstretch elasticity is again an extension of micropolar theory of elasticity and a special case of micromorphic theory. In microstretch theory of elasticity, the surface element is subjected to not only with a *force stress* but with a *couple stress* and a *microstress vector* also. Here in micropolar elasticity, the force stress and couple stress are asymmetric, while in classical elasticity the force stress is symmetric. The concept of microstress vector has been introduced corresponding to microstretch coordinate. His book on *Micro-continuum Field Theories* is worth notable, which is in two volumes for solid and fluent bodies that possess inner structure. This book contains the field theories of micromorphic, microstretch and micropolar continua-known as theory of $3M$ continua.

Enormous problems of waves and vibrations in $3M$ continuum elastic media have been attempted by several researchers in the past. Some notables among them includes Tomar and Kumar (1999), Singh (2001, 2002), Tomar and Singh (2003), Midya (2004), Eremeyev (2005), Tomar and Singh (2006), Hsia et al. (2006), Singh and Tomar (2008), Tomar and Rani (2011). The problems of reflection of waves from boundary surfaces of a continuum are of fundamental interest since long. Keeping in view the fact that elastic waves traveling through an elastic medium carry a lot of information about the medium through which they travel, these problems of reflection and transmission of waves from boundary surfaces are of great help in the exploration of materials on either sides of the boundary surface.

Keeping in view several applications of the problems of reflection of elastic waves, we have considered a problem of reflection of an elastic wave from the free surface of a microstretch fluid half-space. The existence of such a model can be found in geophysics. A set of coupled longitudinal waves propagating through the half-space is assumed to strike obliquely at the free boundary surface. Using Eringen's theory of micro-continuum materials, the phase speeds of possible waves propagating in microstretch fluid and solid media are presented. Governing equations of motion are solved in the relevant media through potential method to find the speeds of propagation of various waves. It is found that the reflection coefficients are functions of angle of incidence and elastic properties of the fluid half space. Numerical calculations have been carried out for a specific model by taking Aluminium matrix with randomly distributed epoxy spheres as microstretch solid medium, while microstretch fluid is taken arbitrarily with suitably chosen values of elastic parameters. The computed results obtained have been depicted graphically. Governing equations and constitutive relations for nonlocal microstretch elastic solid medium have been developed by standard method. The exploration analysis of possible propagation of waves in nonlocal microstretch solid medium has also been performed.

Part-I

2. Governing equations and relations

The constitutive relations and equations of motion without body load and body couple densities in a uniform microstretch continua are given as in the following [see Eringen (1999)].

Constitutive relations.

For Microstretch Solid:

$$t_{kl}^s = (\lambda_{0s}\Upsilon^s + \lambda_s u_{r,r}^s)\delta_{kl} + \mu_s(u_{l,k}^s + u_{k,l}^s) + \kappa_s(u_{l,k}^s + \epsilon_{lkm}\phi_m^s), \tag{1}$$

$$m_{kl}^s = \alpha_s\phi_{r,r}^s\delta_{kl} + \beta_s\phi_{k,l}^s + \gamma_s\phi_{l,k}^s + b_{0s}\epsilon_{mlk}\Upsilon_{,m}^s, \tag{2}$$

$$m_k^s = \alpha_{0s}\Upsilon_{,k}^s + b_{0s}\epsilon_{klm}\phi_{l,m}^s. \tag{3}$$

For Microstretch Fluid:

$$t_{kl}^f = (\lambda_{0f}\Upsilon^f + \lambda_f \dot{u}_{r,r}^f)\delta_{kl} + \mu_f(\dot{u}_{l,k}^f + \dot{u}_{k,l}^f) + \kappa_f(\dot{u}_{l,k}^f + \epsilon_{lkm}\dot{\phi}_m^f), \tag{4}$$

$$m_{kl}^f = \alpha_f\dot{\phi}_{r,r}^f\delta_{kl} + \beta_f\dot{\phi}_{k,l}^f + \gamma_f\dot{\phi}_{l,k}^f + b_{0f}\epsilon_{mlk}\Upsilon_{,m}^f, \tag{5}$$

$$m_k^f = \alpha_{0f}\Upsilon_{,k}^f + b_{0f}\epsilon_{klm}\dot{\phi}_{l,m}^f, \tag{6}$$

Here, t_{kl}^r $(r = s, f)$ is the force stress tensor; m_{kl}^r is the couple stress tensor; m_k^r is the microstress vector; δ_{kl} is the Kronecker delta and ϵ_{klm} is the permutation symbol. The quantities λ_r, μ_r, κ_r, α_r, β_r, γ_r, λ_{0r}, λ_{1r}, α_{0r} and b_{0r}, $(r = f, s)$ are the elastic moduli in the corresponding medium. The $\mathbf{u}^r$ and $\boldsymbol{\phi}^r$ are the displacement and microrotation vectors, respectively, in the relevant medium, Υ^s is scalar microstretch in solid medium and Υ^f corresponds to microstretch rate in the fluid medium. A dot over an entity represents the temporal derivative, while a comma (,) in the subscript represents the spatial derivative, e.g., $\dot{u}_{k,k}^f = \dfrac{\partial^2 u_k}{\partial t\, \partial x_k}$, *etc.*

Equations of motion.

For Microstretch Solid:

$$c_{13s}^2\nabla(\nabla\cdot\mathbf{u}^s) - c_{23s}^2\nabla\times(\nabla\times\mathbf{u}^s) + c_{3s}^2\nabla\times\boldsymbol{\phi}^s + \omega_{0s}^2\nabla\Upsilon^s = \ddot{\mathbf{u}}^s \tag{7}$$

$$c_{45s}^2\nabla(\nabla\cdot\boldsymbol{\phi}^s) - c_{5s}^2\nabla\times(\nabla\times\boldsymbol{\phi}^s) + c_{6s}^2(\nabla\times\mathbf{u}^s - 2\boldsymbol{\phi}^s) = \ddot{\boldsymbol{\phi}}^s, \tag{8}$$

$$c_{7s}^2\nabla^2\Upsilon^s - c_{8s}^2\Upsilon^s - c_{9s}^2\nabla\cdot\mathbf{u}^s = \ddot{\Upsilon}^s. \tag{9}$$

For Microstretch Fluid:

$$c_{13f}^2\nabla\nabla\cdot\dot{\mathbf{u}}^f - c_{23f}^2\nabla\times\nabla\times\dot{\mathbf{u}}^f + c_{3f}^2\nabla\times\dot{\boldsymbol{\phi}}^f + \omega_{0f}^2\nabla\Upsilon^f = \ddot{\mathbf{u}}^f, \tag{10}$$

$$c_{45f}^2\nabla\nabla\cdot\dot{\boldsymbol{\phi}}^f - c_{5f}^2\nabla\times\nabla\times\dot{\boldsymbol{\phi}}^f + c_{6f}^2(\nabla\times\dot{\mathbf{u}}^f - 2\dot{\boldsymbol{\phi}}^f) = \ddot{\boldsymbol{\phi}}^f, \tag{11}$$

$$c_{7f}^2\nabla^2\Upsilon^f - c_{8f}^2\Upsilon^f - c_{9f}^2\nabla\cdot\dot{\mathbf{u}}^f = \dot{\Upsilon}^f, \tag{12}$$

where

$$c_{1s}^2 = \frac{\lambda_s + 2\mu_s}{\rho_s},\quad c_{2s}^2 = \frac{\mu_s}{\rho_s},\quad c_{3s}^2 = \frac{\kappa_s}{\rho_s},\quad c_{4s}^2 = \frac{\alpha_s + \beta_s}{\rho_s j_s},\quad c_{5s}^2 = \frac{\gamma_s}{\rho_s j_s},$$

$$c_{6s}^2 = \frac{\kappa_s}{\rho_s j_s},\quad c_{7s}^2 = \frac{2\alpha_{0s}}{\rho_s j_s},\quad c_{8s}^2 = \frac{2\lambda_{1s}}{\rho_s j_s},\quad c_{9s}^2 = \frac{2\lambda_{0s}}{\rho_s j_s},\quad \omega_{0s}^2 = \frac{\lambda_{0s}}{\rho_s},$$

$$c_{13s}^2 = c_{1s}^2 + c_{3s}^2,\qquad c_{23s}^2 = c_{2s}^2 + c_{3s}^2,\qquad c_{45s}^2 = c_{4s}^2 + c_{5s}^2.$$

$$c_{1f}^2 = \frac{\lambda_f + 2\mu_f}{\rho_f},\quad c_{2f}^2 = \frac{\mu_f}{\rho_f},\quad c_{3f}^2 = \frac{\kappa_f}{\rho_f},\quad c_{4f}^2 = \frac{\alpha_f + \beta_f}{\rho_f j_f},\quad c_{5f}^2 = \frac{\gamma_f}{\rho_f j_f},$$

$$c_{6f}^2 = \frac{\kappa_f}{\rho_f j_f}, \quad c_{7f}^2 = \frac{2\alpha_{0f}}{\rho_f j_f}, \quad c_{8f}^2 = \frac{2\lambda_{1f}}{\rho_f j_f}, \quad c_{9f}^2 = \frac{2\lambda_{0f}}{\rho_f j_f}, \quad \omega_{0f}^2 = \frac{\lambda_{0f}}{\rho_f},$$
$$c_{13f}^2 = c_{1f}^2 + c_{3f}^2, \quad c_{23f}^2 = c_{2f}^2 + c_{3f}^2, \quad c_{45f}^2 = c_{4f}^2 + c_{5f}^2.$$

Here the quantities ρ_r and j_r, $(r = s, f)$ are the density and micro-inertia in the relevant medium. Other symbols have their usual meanings.

Eringen (1999) have shown that the following inequalities among the elastic moduli should be satisfied in order to have the dissipation function to be non-negative

$$3\lambda_r + 2\mu_r + \kappa_r \geq \frac{3\lambda_{0r}^2}{\lambda_{1r}}, \quad 3\alpha_r + \beta_r + \gamma_r \geq 0, \quad 2\mu_r + \kappa_r \geq 0, \quad \alpha_{0r} \geq 0,$$
$$\gamma_r + \beta_r \geq 0, \quad \lambda_{1r} \geq 0, \quad \kappa_r \geq 0, \quad \gamma_r - \beta_r \geq 0, \quad (r = s, f). \tag{13}$$

3. Wave propagation

3.1. **Microstretch Solid.** Decomposing the vectors $\mathbf{u}^s$ and $\boldsymbol{\phi}^s$ into scalar potentials q^s, ξ^s; and vector potentials $\boldsymbol{\Psi}^s, \boldsymbol{\Pi}^s$ through Helmholtz decomposition theorem as

$$\{\mathbf{u}^s, \boldsymbol{\phi}^s\} = \nabla\{q^s, \xi^s\} + \nabla \times \{\boldsymbol{\Psi}^s, \boldsymbol{\Pi}^s\}, \qquad \nabla \cdot \{\boldsymbol{\Psi}^s, \boldsymbol{\Pi}^s\} = 0. \tag{14}$$

Inserting (14) into equations (7)-(9), we obtain

$$(c_{1s}^2 + c_{3s}^2)\nabla^2 q^s + \omega_{0s}^2 \Upsilon^s = \ddot{q}^s, \tag{15}$$
$$(c_{2s}^2 + c_{3s}^2)\nabla^2 \boldsymbol{\Psi}^s + c_{3s}^2 \nabla \times \boldsymbol{\Pi}^s = \ddot{\boldsymbol{\Psi}}^s, \tag{16}$$
$$(c_{4s}^2 + c_{5s}^2)\nabla^2 \xi^s - 2c_{6s}^2 \xi^s = \ddot{\xi}^s, \tag{17}$$
$$c_{5s}^2 \nabla^2 \boldsymbol{\Pi}^s + c_{6s}^2 \nabla \times \boldsymbol{\Psi}^s - 2c_{6s}^2 \boldsymbol{\Pi}^s = \ddot{\boldsymbol{\Pi}}^s, \tag{18}$$
$$c_{7s}^2 \nabla^2 \Upsilon^s - c_{8s}^2 \Upsilon^s - c_{9s}^2 \nabla^2 q^s = \ddot{\Upsilon}^s. \tag{19}$$

We see that equations (15) and (19) are coupled through q^s and Υ^s; equations (16) and (18) are coupled through $\boldsymbol{\Psi}^s$ and $\boldsymbol{\Pi}^s$, while equation (17) is uncoupled in ξ^s.

For a wave traveling in the positive direction of a unit vector $\mathbf{n}$, we consider the form of various quantities as

$$\{q^s, \xi^s, \boldsymbol{\Psi}^s, \boldsymbol{\Pi}^s, \Upsilon^s\} = \{a_s, b_s, \mathbf{A}_s, \mathbf{B}_s, c_s\} \exp\{\iota k_s(\mathbf{n} \cdot \mathbf{r} - v_s t)\}, \tag{20}$$

where $a_s, b_s, \mathbf{A}_s, \mathbf{B}_s, c_s$ are constants, $\mathbf{r} = x\hat{i} + y\hat{j} + z\hat{k}$ is the position vector, v_s is the speed of propagation, k_s is the wavenumber connected with v_s through the relation $\omega = k_s v_s$, ω being the circular frequency and $\iota = \sqrt{-1}$. Using the expressions of q^s and Υ^s from (20) into (15) and (19), we obtain the speeds of waves associated with q^s given by

$$v_{1s,2s}^2 = \frac{1}{2a_{1s}}\left(-a_{2s} \pm \sqrt{a_{2s}^2 - 4a_{1s}a_{3s}}\right), \tag{21}$$

where

$a_{1s} = \omega^2 - c_{8s}^2, \qquad a_{2s} = (c_{1s}^2 + c_{3s}^2)(c_{8s}^2 - \omega^2) - \omega_{0s}^2 c_{9s}^2 - c_{7s}^2 \omega^2$ and

$a_{3s} = (c_{1s}^2 + c_{3s}^2) c_{7s}^2 \omega^2$.

The displacement vector $\mathbf{u}^s$ given by

$$\mathbf{u}^s = \nabla q^s = \iota k_s a_s \mathbf{n} \exp\{\iota k_s(\mathbf{n} \cdot \mathbf{r} - v_s t)\}$$

is parallel to the direction of wave propagation $\mathbf{n}$, therefore the waves associated with the potential q^s are longitudinal in nature. Hence the speeds v_{1s} and v_{2s} correspond to the speeds of set of coupled longitudinal waves. Each set of coupled longitudinal wave will consists of scalar microstretch along with the displacement.

Similarly, inserting the expressions of $\mathbf{\Psi}^s$ and $\mathbf{\Pi}^s$ into (16) and (18), we get the speeds of waves given by

$$v_{3s,4s}^2 = \frac{1}{2A_{1s}} \left(-A_{2s} \pm \sqrt{A_{2s}^2 - 4A_{1s}A_{3s}} \right), \tag{22}$$

where

$A_{1s} = 2c_{6s}^2 - \omega^2, \qquad A_{2s} = \omega^2 c_{5s}^2 - (c_{2s}^2 + c_{3s}^2)(2c_{6s}^2 - \omega^2) + c_{3s}^2 c_{6s}^2,$

$A_{3s} = -c_{5s}^2 \omega^2 (c_{2s}^2 + c_{3s}^2).$

Parfit and Eringen (1969) have already shown that these speeds are the speeds of two sets of coupled transverse waves. Each set of coupled waves is transverse in nature consists of microrotation and displacement perpendicular to it.

Next, inserting the expression of ξ^s from the above relation into equations (17), we obtain the speeds of the wave corresponding to the potential ξ^s in the microstretch solid denoted by v_{5s} given by

$$v_{5s}^2 = \frac{\omega^2 (c_{4s}^2 + c_{5s}^2)}{\omega^2 - 2c_{6s}^2}. \tag{23}$$

The microrotation vector $\mathbf{\Phi}^s$ is given by

$$\mathbf{\Phi}^s = \nabla \xi^s = \iota k_s b_s \mathbf{n} \exp\{\iota k_s(\mathbf{n} \cdot \mathbf{r} - v_s t)\},$$

thus, the wave corresponding to potential ξ^s is longitudinal in nature and comes into existence due to rotation of the micro-elements of the medium. This wave is called as *Longitudinal micro-rotational wave.* It is easy to see at once that the speeds of all the waves of micropolar elasticity can be recovered as a special case of present formulation. If we neglect the presence of microstretch property from the microstretch solid medium by considering $\lambda_{0s} = \lambda_{1s} = \alpha_{0s} = 0$ then we are left with $a_{1s} = \omega^2$, $a_{2s} = -\omega^2(c_{1s}^2 + c_{3s}^2)$, $a_{3s} = 0$. We see from (31)) that $v_{1s}^2 = 0$ and $v_{2s}^2 = c_{1s}^2 + c_{3s}^2$ which is the speed of longitudinal displacement wave of micropolar elasticity. The speeds of coupled transverse waves and longitudinal microrotational wave given by (22) and (23) respectively are exactly the same as that of micropolar elasticity as they are not affected by the microstretch property of the medium at all [see also Parfit and Eringen (1969)].

3.2. **Microstretch Fluid.** On similar lines as in microstretch solid medium, we decompose the vectors $\mathbf{u}^f$ and $\boldsymbol{\phi}^f$ into scalar potentials q^f, ξ^f and vector potentials $\boldsymbol{\Psi}^f, \boldsymbol{\Pi}^f$ as

$$\{\mathbf{u}^f, \boldsymbol{\phi}^f\} = \nabla\{q^f, \xi^f\} + \nabla \times \{\boldsymbol{\Psi}^f, \boldsymbol{\Pi}^f\}, \qquad \nabla \cdot \{\boldsymbol{\Psi}^f, \boldsymbol{\Pi}^f\} = 0. \tag{24}$$

and inserting into equations (10) - (12), we obtain

$$(c_{1f}^2 + c_{3f}^2)\nabla^2 \dot{q}^f + \omega_{0f}^2 \Upsilon^f = \ddot{q}^f, \tag{25}$$

$$(c_{2f}^2 + c_{3f}^2)\nabla^2 \dot{\boldsymbol{\Psi}}^f + c_{3f}^2 \nabla \times \dot{\boldsymbol{\Pi}}^f = \ddot{\boldsymbol{\Psi}}^f, \tag{26}$$

$$(c_{4f}^2 + c_{5f}^2)\nabla^2 \dot{\xi}^f - 2c_{6f}^2 \dot{\xi}^f = \ddot{\xi}^f, \tag{27}$$

$$c_{5f}^2 \nabla^2 \dot{\boldsymbol{\Pi}}^f + c_{6f}^2 \nabla \times \dot{\boldsymbol{\Psi}}^f - 2c_{6f}^2 \dot{\boldsymbol{\Pi}}^f = \ddot{\boldsymbol{\Pi}}^f, \tag{28}$$

$$c_{7f}^2 \nabla^2 \Upsilon^f - c_{8f}^2 \Upsilon^f - c_{9f}^2 \nabla^2 \dot{q}^f = \dot{\Upsilon}^f. \tag{29}$$

We consider

$$\left\{q^f, \xi^f, \boldsymbol{\Psi}^f, \boldsymbol{\Pi}^f, \Upsilon^f\right\} = \{a_f, b_f, \mathbf{A}_f, \mathbf{B}_f, c_f\} \exp\{\iota k_f(\mathbf{n} \cdot \mathbf{r} - v_f t)\}, \tag{30}$$

Inserting the expressions of q^f and Υ^f into (25), (29); $\boldsymbol{\Psi}^f$ and $\boldsymbol{\Pi}^f$ into equations (26), (28) and ξ^f into equations (27), we obtain the speeds of various waves given by

$$v_{1f,2f}^2 = \frac{1}{2a_{1f}}\left(-a_{2f} \pm \sqrt{a_{2f}^2 - 4a_{1f}a_{3f}}\right). \tag{31}$$

$$v_{3f,4f}^2 = \frac{1}{2A_{1f}}\left(-A_{2f} \pm \sqrt{A_{2f}^2 - 4A_{1f}A_{3f}}\right), \tag{32}$$

$$v_{5f}^2 = \frac{-\iota\omega^2(c_{4f}^2 + c_{5f}^2)}{(\omega + 2\iota c_{6f}^2)}, \tag{33}$$

where

$$\begin{aligned}
a_{1f} &= c_{8f}^2 - \iota\omega,\\
a_{2f} &= \omega\left[(c_{1f}^2 + c_{3f}^2)(\omega + \iota c_{8f}^2) - \iota\omega_{0f}^2 c_{9f}^2 + \omega c_{7f}^2\right],\\
a_{3f} &= \iota\omega^3 c_{7f}^2 (c_{1f}^2 + c_{3f}^2), \quad A_{1f} = 2c_{6f}^2 - \iota\omega,\\
A_{2f} &= \omega\left[(c_{2f}^2 + c_{3f}^2)(\omega + 2\iota c_{6f}^2) - \iota c_{3f}^2 c_{6f}^2 + \omega c_{5f}^2\right],\\
A_{3f} &= \iota\omega^3 c_{5f}^2 (c_{2f}^2 + c_{3f}^2).
\end{aligned}$$

It is easy to see that the displacement vector $\mathbf{u}^f$ and the microrotation vector $\boldsymbol{\Phi}^f$ are

$$\{\mathbf{u}^f, \boldsymbol{\Phi}^f\} = \nabla\{q^f, \xi^f\} = \iota k_f\{a_f, b_f\}\mathbf{n} \exp\{\iota k_f(\mathbf{n} \cdot \mathbf{r} - v_f t)\},$$

showing that they are along the direction of propagation of wave. Thus the wave corresponding to the potentials q^f, Υ^f and ξ^f are longitudinal in nature. Clearly, the waves associated with q^f, Υ^f are coupled waves and will be called coupled longitudinal waves, while the wave corresponding to ξ^f is uncoupled and will be called longitudinal micro-rotational wave. The speeds v_{1f} and v_{2f} are the speeds of two sets of coupled longitudinal waves, v_{3f}, v_{4f} are the speeds of two sets of coupled transverse waves and the speed v_{5f} is the speed of longitudinal microrotational wave. Note that all the phase speeds of existing waves are complex valued and depend on frequency, indicating that they are dispersive and attenuating in nature.

In general, for a wave propagating with complex phase speed c, the phase speed V and the corresponding attenuation coefficient Q can be computed from the following formula

$$c^{-1} = V^{-1} + \iota\omega^{-1}Q.$$

From this formula, the phase speed and attenuation coefficient are given by

$$V = \frac{c_R^2 + c_I^2}{c_R}, \qquad Q = -\omega\frac{c_I}{(c_R^2 + c_I^2)}, \tag{34}$$

where c_R is the real part of c and c_I is the imaginary part of c.

4. Reflection of coupled longitudinal waves from flat boundary of a microstretch fluid half-space

Here, we shall investigate the reflection phenomena of a set of coupled longitudinal waves striking obliquely at the free boundary surface of a microstretch fluid medium. Let $M = \{z \geq 0, -\infty < x, y < \infty\}$ be the region occupied by the microstretch fluid. Let $x - y$ be the free surface of the half-space M and z-axis is vertically downward in to the microstretch fluid medium. We consider a two-dimensional problem in $x - z$ plane, so that

$$\mathbf{v} = (v_1(x,z,t), 0, v_3(x,z,t)), \quad \boldsymbol{\phi} = (0, \phi_y(x,z,t), 0), \quad \Upsilon = \Upsilon(x,z,t).$$

Let us consider a set of coupled longitudinal waves propagating with speed v_{1f} having amplitude A_0 and striking the boundary surface $z = 0$ making an angle θ_0 with the z-axis. To satisfy the boundary conditions on the boundary surface $z = 0$, it is necessary to assume the existence of following reflected waves:

(i) a set of coupled longitudinal waves of amplitude A_1 propagating with speed v_{1f} and making an angle θ_1 with the normal.

(ii) a similar set of coupled longitudinal waves of amplitude A_2 propagating with speed v_{2f} and making an angle θ_2 with the normal.

(iii) a set of coupled transverse waves of amplitude A_{3y} propagating with speed v_{3f} and making an angle θ_3 with the normal.

(iv) a similar set of coupled transverse waves of amplitude A_{4y} propagating with speed v_{4f} and making an angle θ_4 with the normal. The schematic diagram of the problem has been depicted in Figure 1.

The full wave structure in the half-space M is given by

$$\begin{aligned} q^f &= A_0 \exp[\iota k_0(x\sin\theta_0 - z\cos\theta_0) - \iota\omega_0 t] \\ &\quad + A_1 \exp[\iota k_1(x\sin\theta_1 + z\cos\theta_1) - \iota\omega_1 t] \\ &\quad + A_2 \exp[\iota k_2(x\sin\theta_2 + z\cos\theta_2) - \iota\omega_2 t], \end{aligned} \tag{35}$$

$$\begin{aligned}\xi^f &= \zeta_0 A_0 \exp[\iota k_0(x\sin\theta_0 - z\cos\theta_0) - \iota\omega_0 t] \\ &\quad + \zeta_1 A_1 \exp[\iota k_1(x\sin\theta_1 + z\cos\theta_1) - \iota\omega_1 t] \\ &\quad + \zeta_2 A_2 \exp[\iota k_2(x\sin\theta_2 + z\cos\theta_2) - \iota\omega_2 t],\end{aligned} \tag{36}$$

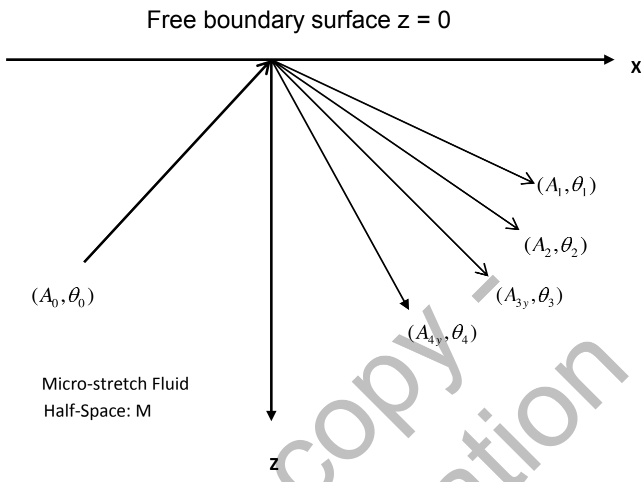


FIGURE 1. Schematic diagram of the problem.

$$\mathbf{\Psi}^f = \sum_{p=3,4} \hat{\mathbf{j}} A_{py} \exp[\iota k_p(x\sin\theta_p + z\cos\theta_p) - \iota\omega_p t], \tag{37}$$

$$\mathbf{\Pi}^f = \sum_{p=3,4} \left(\hat{\mathbf{i}} B_{px} + \hat{\mathbf{k}} B_{pz}\right) \exp[\iota k_p(x\sin\theta_p + z\cos\theta_p) - \iota\omega_p t], \tag{38}$$

The constants $\mathbf{A}$ and $\mathbf{B}$ occurring in (37) and (38) are related to each other through the following relations

$$B_{px} = -\iota k_p \eta_p \cos\theta_p A_{py}, \qquad B_{pz} = \iota k_p \eta_p \sin\theta_p A_{py},$$

$$\eta_p = \frac{c_{6f}^2}{k_p^2 c_{5f}^2 + 2c_{6f}^2 - \iota k_p v_{pf}}. \tag{39}$$

The coupling coefficients ζ_0, ζ_1 and ζ_2 occurring in (36) are given by

$$\zeta_n = c_{9f}^2 k_n^2 \left[k_n^2 c_{7f}^2 + c_{8f}^2 - \iota k_n v_{nf}\right]^{-1}, \qquad (n = 0, 1, 2). \tag{40}$$

4.1. **Boundary conditions.** Since the boundary surface of the microstretch fluid half-space M is free from stresses, therefore the boundary conditions would be the vanishing of stresses, couples stresses and first vector moment at the boundary surface. These boundary conditions can be written in mathematical form as: At the surface $z = 0$, we have

$$t_{zz} = t_{zx} = m_{zy} = m_z = 0. \tag{41}$$

Making use of (4)-(6) and (24), these boundary conditions in potential form can be written as: At the surface $z = 0$

$$\lambda_{0f}\xi^f + \lambda_f \nabla^2 q^f + (2\mu_f + \kappa_f)\left(\frac{\partial^2 q^f}{\partial z^2} + \frac{\partial^2 \Psi_y^f}{\partial x \partial z}\right) = 0, \tag{42}$$

$$(2\mu_f + \kappa_f)\frac{\partial^2 q^f}{\partial x \partial z} + \mu_f \frac{\partial^2 \Psi_y^f}{\partial x^2} - (\mu_f + \kappa_f)\frac{\partial^2 \Psi_y^f}{\partial z^2} + \kappa_f\left(\frac{\partial \Pi_z^f}{\partial x} - \frac{\partial \Pi_x^f}{\partial z}\right) = 0, \tag{43}$$

$$b_{0f}\frac{\partial \xi^f}{\partial x} + \gamma_f\left(\frac{\partial^2 \Pi_x^f}{\partial z^2} - \frac{\partial^2 \Pi_z^f}{\partial z \partial x}\right) = 0, \tag{44}$$

$$\alpha_{0f}\frac{\partial \xi^f}{\partial z} + b_{0f}\left(\frac{\partial^2 \Pi_z^f}{\partial x^2} - \frac{\partial^2 \Pi_x^f}{\partial x \partial z}\right) = 0. \tag{45}$$

Inserting the expressions of various potentials from (35) to (38) into (42) to (45), making use of Snell's law given by

$$k_0 \sin\theta_0 = k_1 \sin\theta_1 = k_2 \sin\theta_2 = k_3 \sin\theta_3 = k_4 \sin\theta_4 \tag{46}$$

and assuming that all frequencies match at the boundary surface, we obtain the following system of equations given by

$$\begin{aligned}
k_0^2 &\left[\lambda_f + (2\mu_f + \kappa_f)\cos^2\theta_0 - \frac{\lambda_{0f}\zeta_0}{k_0^2}\right] A_0 \\
&+ k_1^2 \left[\lambda_f + (2\mu_f + \kappa_f)\cos^2\theta_1 - \frac{\lambda_{0f}\zeta_1}{k_1^2}\right] A_1 \\
&+ k_2^2 \left[\lambda_f + (2\mu_f + \kappa_f)\cos^2\theta_2 - \frac{\lambda_{0f}\zeta_2}{k_2^2}\right] A_2 \\
&+ k_3^2 (2\mu_f + \kappa_f)\sin\theta_3\cos\theta_3 A_{3y} \\
&+ k_4^2 (2\mu_f + \kappa_f)\sin\theta_4\cos\theta_4 A_{4y} = 0,
\end{aligned} \tag{47}$$

$$\begin{aligned}
k_0^2 (2\mu_f + \kappa_f)\sin\theta_0\cos\theta_0 (A_0 - A_1) &- k_2^2 (2\mu_f + \kappa_f)\sin\theta_2\cos\theta_2 A_2 \\
&+ k_3^2 \left[\mu_f \cos 2\theta_3 + \kappa_f \cos^2\theta_3 - \kappa_f \eta_3\right] A_{3y} \\
&+ k_4^2 \left[\mu_f \cos 2\theta_4 + \kappa_f \cos^2\theta_4 - \kappa_f \eta_4\right] A_{4y} = 0,
\end{aligned} \tag{48}$$

$$\begin{aligned}
b_{0f} &\left[k_0 \sin\theta_0 \zeta_0 (A_0 + A_1) + k_2 \sin\theta_2 \zeta_2 A_2\right] \\
&+ \gamma_f \left[k_3^3 \cos\theta_3 \eta_3 A_{3y} + k_4^3 \cos\theta_4 \eta_4 A_{4y}\right] = 0,
\end{aligned} \tag{49}$$

$$\begin{aligned}
\alpha_{0f} &\left[k_0 \cos\theta_0 \zeta_0 (A_1 - A_0) + k_2 \cos\theta_2 \zeta_2 A_2\right] \\
&- b_{0f} \left[k_3^3 \sin\theta_3 \eta_3 A_{3y} + k_4^3 \sin\theta_4 \eta_4 A_{4y}\right] = 0.
\end{aligned} \tag{50}$$

The equations (47) - (50) can be written in matrix form as

$$\mathbf{X}\,\mathbf{Z} = \mathbf{Y}, \tag{51}$$

where $\mathbf{X} = (a_{ij})_{4\times4}$ whose elements in non-dimensional form are given in Appendix-A, $\mathbf{Y} = (-1 \quad +1 \quad -1 \quad +1)^t$, and $\mathbf{Z} = (z_1 \quad z_2 \quad z_3 \quad z_4)^t$. The expressions of z_i are given as

$$z_1 = \frac{A_1}{A_0}, \quad z_2 = \frac{A_2}{A_0}, \quad z_3 = \frac{A_{3y}}{A_0}, \quad z_4 = \frac{A_{4y}}{A_0}.$$

These quantities are known as reflection coefficients corresponding to the set of reflected coupled longitudinal waves with speed v_{1f}, set of reflected coupled longitudinal waves with speed v_{2f}, set of reflected coupled transverse waves with speed v_{3f} and set of reflected coupled transverse wave with speed v_{4f} respectively. On solving the matrix equation (67) using Cramer's rule, one can obtain the reflection coefficients

$$z_j = \frac{\Delta_j}{\Delta}, \qquad (j = 1, 2, 3, 4), \tag{52}$$

where notations Δ and Δ_j are well understood; Δ_j can be obtained by usual method.

Remark: This problem of reflection of coupled longitudinal waves from the free surface of a microstretch fluid half-space has been already investigated by Tomar and Rani (2011), where there is an error in the computation of second boundary condition, namely $t_{zx} = 0$. This has led to enormous error in the numerical computations too. The correct one has been presented here.

Part-II

5. The nonlocal theory of elasticity

Under this theory, the points undergo translational motion as in the classical case. The stress at a point $\mathbf{x}$ depends not only at the strain at that point, but also on the strains at all other points $\mathbf{x}'$ in a region near that point. This observation is in accordance with the atomic theory of lattice dynamics and experimental observation on phonon dispersion. In a limiting case, when the effects of strains at points other than $\mathbf{x}$ are neglected, one recovers classical (local) theory of elasticity.

The local dependence of a physical quantity (the effect r) at a point $\mathbf{x}$ of space at time t on another physical quantity (the cause p) at the same point $\mathbf{x}$ and at the same time t has the general form

$$r(x, y, z, t) = r(p(x, y, z, t))$$

This is local dependence relation. There are three types of nonlocality (1) Spatial nonlocality (2) Temporal nonlocality (3) Mixed nonlocality.
In spatial nonlocality, the effect R at a point $\mathbf{x}$ at a time t depends on the causes at all other points $\mathbf{x}'$ at the same instant of time

$$R(x, y, z, t) = \int_V \alpha(|\mathbf{x} - \mathbf{x}'|, \zeta) r(p(x', y', z', t)) dx'dy'dz'$$

In temporal nonlocality (material with memory), the effect R at a point $\mathbf{x}$ at a time t depends on the history of causes at the point $\mathbf{x}$ over all preceding times and at the present instant

$$R(x,y,z,t) = \int_{-\infty}^{t} \beta(t'-t,\epsilon) r(p(x,y,z,t'))dt'$$

If we consider simultaneously the effects of memory and spatial nonlocalness, we have an effect R at the point $\mathbf{x}$ at time t that depends on the causes at all spatial points $\mathbf{x}'$ and at all times $t' \leq t$

$$R(x,y,z,t) = \int_{-\infty}^{t} \int_{V} \gamma(|\mathbf{x}-\mathbf{x}'|, t-t', \zeta, \varepsilon) r(p(x',y',z',t'))dx'dy'dz'dt'$$

The functions $\alpha(|\mathbf{x}-\mathbf{x}'|,\zeta)$, $\beta(t'-t,\varepsilon)$ and $\gamma(|\mathbf{x}-\mathbf{x}'|, t-t', \zeta, \varepsilon)$ are called the **nonlocal kernels** and describe the influence of nonlocality. Here the quantities $\zeta = \dfrac{L_i}{L_e}$ and $\varepsilon = \dfrac{T_i}{T_e}$, L_i is internal characteristic length (Lattice parameter, size of grain), L_e is the external characteristic length(wave length), T_i is the internal characteristic time (relaxation time, time for a signal to travel between molecules), T_e is the external characteristic time (the time of application of the external action, period of vibration). Certain properties of spatial nonlocal kernel

(a) $\alpha(|\mathbf{x}-\mathbf{x}'|,\zeta)$ has a maximum at $|\mathbf{x}-\mathbf{x}'|$,
(b) $\alpha(|\mathbf{x}-\mathbf{x}'|,\zeta)$ quickly tends to zero as $|\mathbf{x}-\mathbf{x}'|$ increases,
(c) $\alpha(|\mathbf{x}-\mathbf{x}'|,\zeta)$ is continuous function of $|\mathbf{x}-\mathbf{x}'|$,
(d) $\lim_{\zeta\to 0} \alpha(|\mathbf{x}-\mathbf{x}'|,\zeta) = \delta(|\mathbf{x}-\mathbf{x}'|)$,
(e) $\int_V \alpha(|\mathbf{x}-\mathbf{x}'|,\zeta) dV(\mathbf{x}') = 1$.

Eringen (1983) has proposed that

$$\alpha(|\mathbf{x}-\mathbf{x}'|,\zeta) = \frac{1}{8(\pi\eta)^{3/2}} \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{4\eta}\right)$$

where $\eta = \dfrac{a^2}{4e_0^2}$; a is the lattice parameter and e_0 is the corresponding parameter, which can be determined either by experiment or by applying the theory of atomic lattice. For stress - strain relation, if cause is the strain e_{ij} and the effect is the stress τ_{ij}, then

$$\tau_{ij}(x,y,z,t) = c_{ijkm} e_{km}(p(x,y,z,t))$$

$$\tau_{ij}(x,y,z,t) = \int_V \alpha_{ijkm}(|\mathbf{x}-\mathbf{x}'|,\zeta) e_{km}(p(x',y',z',t))dx'dy'dz'$$

$$\tau_{ij}(x,y,z,t) = \int_{-\infty}^{t} \beta_{ijkm}(t'-t,\varepsilon) e_{km}(p(x,y,z,t'))dt'$$

$$\tau_{ij}(x,y,z,t) = \int_{-\infty}^{t} \int_V \gamma_{ijkm}(|\mathbf{x}-\mathbf{x}'|, t-t', \zeta, \varepsilon) e_{km}(p(x',y',z',t'))dx'dy'dz'dt'$$

It can be seen that stress-strain relation of local elasticity can be obtained when $L_e >> L_i$, that is $\zeta \to 0$, since

$$\lim_{\zeta \to 0} \alpha_{ijkm}(|\mathbf{x}-\mathbf{x}'|,\zeta) = c_{ijkm}\delta(|\mathbf{x}-\mathbf{x}'|)$$

and

$$\int f(x)\delta(x-a)dx = f(a).$$

Hence in the long wave limit, the nonlocal theory of elasticity reduces to classical elasticity. Thus we see that nonlocal theory of elasticity is significant in the short wave limit, which agrees with the theory of atomic lattice. It is well known that the micropolar theory of elasticity deals with the granular bodies, e.g., a material having doumb-bell type molecules. The microstructure of such material plays very important role in the problems of waves and vibrations, particularly, in the case of waves having high frequency or short wave length. Thus, in this case $\zeta \neq 0$, rather in this case $\zeta \approx 1$. This is how the micropolar theory of elasticity lies in the domain of nonlocal elasticity. In the subsequent section, the symbols adopted in Part-I have not been borrowed as such, but the symbols followed analogously which are standard and well understood.

6. Waves in nonlocal microstretch solid

The free energy density function F for a linear, nonlocal microstretch solid is as follows [Eringen (2002)]

$$\begin{aligned} 2F = 2F_0 - \frac{2\rho_0}{T_0}CTT' - C_1(T'\psi + T\psi') - D_k(T'\gamma_k + T\gamma_k') \\ -A_{kl}(T'\epsilon_{kl} + T\epsilon_{kl}') - B_{kl}(T'\gamma_{kl} + T\gamma_{kl}') + U, \end{aligned}$$

where U is the strain energy density function given by

$$\begin{aligned} 2U = \frac{\rho_0}{T_0}CTT' + C^S\psi\psi' + C_k^S(\psi'\gamma_k + \psi\gamma_k') + A_{kl}^S(\psi'\epsilon_{kl} + \psi\epsilon_{kl}') \\ +B_{kl}^S(\psi'\gamma_{kl} + \psi\gamma_{kl}') + C_{kl}^S\gamma_k'\gamma_l + A_{klm}^S(\gamma_k'\gamma_{lm} + \gamma_k\gamma_{lm}') \\ +A_{klmn}\epsilon_{kl}'\epsilon_{mn} + B_{klmn}\gamma_{kl}'\gamma_{mn} + C_{klmn}(\epsilon_{kl}'\gamma_{mn} + \epsilon_{kl}\gamma_{mn}'), \end{aligned}$$

where

$$\epsilon_{kl} = u_{l,k} - \epsilon_{klm}\phi_m, \qquad \gamma_{kl} = \phi_{k,l}, \qquad \gamma_k = \psi_{,k}\,. \tag{53}$$

All these constitutive moduli possess the following symmetries

$$\begin{aligned} A_{klmn}(\mathbf{x},\mathbf{x}') = A_{klmn}(\mathbf{x}',\mathbf{x}) = A_{mnkl}(\mathbf{x}',\mathbf{x}), \\ B_{klmn}(\mathbf{x},\mathbf{x}') = B_{klmn}(\mathbf{x}',\mathbf{x}) = B_{mnkl}(\mathbf{x}',\mathbf{x}), \quad etc. \end{aligned}$$

The expressions of requisite tensors are given as [Eringen (2002)]

$$\begin{aligned} t_{kl} = \int_V [-A_{kl}(\mathbf{x},\mathbf{x}')T(\mathbf{x}') + A_{kl}^S(\mathbf{x},\mathbf{x}')\psi(\mathbf{x}') + A_{mkl}^S(\mathbf{x},\mathbf{x}')\psi_{,m}(\mathbf{x}') \\ + A_{klmn}(\mathbf{x},\mathbf{x}')\epsilon_{mn}(\mathbf{x}') + C_{klmn}(\mathbf{x},\mathbf{x}')\gamma_{mn}(\mathbf{x}')]dV(\mathbf{x}'), \end{aligned}$$

$$m_{kl} = \int_V [-B_{lk}(\mathbf{x},\mathbf{x}')T(\mathbf{x}') + B^S_{lk}(\mathbf{x},\mathbf{x}')\psi(\mathbf{x}') + B^S_{mlk}(\mathbf{x},\mathbf{x}')\psi,_m(\mathbf{x}')$$
$$+ C_{mnlk}(\mathbf{x},\mathbf{x}')\epsilon_{mn}(\mathbf{x}') + B_{lkmn}(\mathbf{x},\mathbf{x}')\gamma_{mn}(\mathbf{x}')]dV(\mathbf{x}'),$$
$$m_k = \int_V [-D_k(\mathbf{x},\mathbf{x}')T(\mathbf{x}') + C^S_k(\mathbf{x},\mathbf{x}')\psi(\mathbf{x}') + C^S_{kl}(\mathbf{x},\mathbf{x}')\psi,_l(\mathbf{x}')$$
$$+ A^S_{klm}(\mathbf{x},\mathbf{x}')\epsilon_{lm}(\mathbf{x}') + B^S_{klm}(\mathbf{x},\mathbf{x}')\gamma_{lm}(\mathbf{x}')]dV(\mathbf{x}'),$$
$$s - t = \int_V [-C_1(\mathbf{x},\mathbf{x}')T(\mathbf{x}') + C^S(\mathbf{x},\mathbf{x}')\psi(\mathbf{x}') + C^S_k(\mathbf{x},\mathbf{x}')\psi,_k(\mathbf{x},\mathbf{x}')$$
$$+ A^S_{kl}(\mathbf{x},\mathbf{x}')\epsilon_{kl}(\mathbf{x}') + B^S_{kl}(\mathbf{x},\mathbf{x}')\gamma_{kl}(\mathbf{x}')]dV(\mathbf{x}').$$

For isotropic microstretch solid, the material moduli are given as (see Eringen, 2002)

$$D_k = C^S_k = 0, \quad A_{kl} = \beta_0\delta_{kl}, \quad B_{kl} = B^S_{kl} = 0, \quad C^S = \lambda_1,$$
$$A^S_{kl} = \lambda_0\delta_{kl}, \quad C^S_{kl} = \alpha_0\delta_{kl}, \quad A^S_{klm} = 0, \quad B^S_{klm} = b_0\epsilon_{klm},$$
$$A_{klmn} = \lambda\delta_{kl}\delta_{mn} + (\mu+\kappa)\delta_{km}\delta_{ln} + \mu\delta_{kn}\delta_{lm},$$
$$B_{klmn} = \alpha\delta_{kl}\delta_{mn} + \gamma\delta_{km}\delta_{ln} + \beta\delta_{kn}\delta_{lm} \quad \text{and} \quad C_{klmn} = 0.$$

The material moduli are functions of $|\mathbf{x}' - \mathbf{x}|$. With these values, the above constitutive equations reduce to

$$t_{kl} = \int_V \{[-\beta_0(|\mathbf{x}-\mathbf{x}'|)T(\mathbf{x}') + \lambda_0(|\mathbf{x}-\mathbf{x}'|)\psi(\mathbf{x}') + \lambda(|\mathbf{x}-\mathbf{x}'|)\epsilon_{rr}(\mathbf{x}')]\delta_{kl}$$
$$+[\mu(|\mathbf{x}-\mathbf{x}'|) + \kappa(|\mathbf{x}-\mathbf{x}'|)]\epsilon_{kl}(\mathbf{x}') + \mu(|\mathbf{x}-\mathbf{x}'|)\epsilon_{lk}(\mathbf{x}')\}\, dV(\mathbf{x}'), \tag{54}$$
$$m_{kl} = \int_V [b_0(|\mathbf{x}-\mathbf{x}'|)\epsilon_{mlk}\psi,_m(\mathbf{x}') + \alpha(|\mathbf{x}-\mathbf{x}'|)\gamma_{rr}(\mathbf{x}')\delta_{kl}$$
$$+\beta(|\mathbf{x}-\mathbf{x}'|)\gamma_{kl}(\mathbf{x}') + \gamma(|\mathbf{x}-\mathbf{x}'|)\gamma_{lk}(\mathbf{x}')]dV(\mathbf{x}'), \tag{55}$$
$$m_k = \int_V [\alpha_0(|\mathbf{x}-\mathbf{x}'|)\psi,_k(\mathbf{x}') + b_0(|\mathbf{x}-\mathbf{x}'|)\epsilon_{klm}\gamma_{lm}(\mathbf{x}')]dV(\mathbf{x}'), \tag{56}$$
$$s - t =$$
$$\int_V [-C_1(|\mathbf{x}-\mathbf{x}'|)\mathbf{T}(\mathbf{x}') + \lambda_{\mathbf{1}}(|\mathbf{x}-\mathbf{x}'|)\psi(\mathbf{x}') + \lambda_{\mathbf{0}}(|\mathbf{x}-\mathbf{x}'|)\epsilon_{\mathbf{kk}}(\mathbf{x}')]\mathbf{dV}(\mathbf{x}'). \tag{57}$$

Consider the relation between nonlocal and local elastic moduli

$$\frac{\lambda(|\mathbf{x}-\mathbf{x}'|)}{\lambda} = \frac{\mu(|\mathbf{x}-\mathbf{x}'|)}{\mu} = \frac{\kappa(|\mathbf{x}-\mathbf{x}'|)}{\kappa} = \frac{\lambda_0(|\mathbf{x}-\mathbf{x}'|)}{\lambda_0} = \frac{\beta_0(|\mathbf{x}-\mathbf{x}'|)}{\beta_0}$$
$$= \frac{\alpha_0(|\mathbf{x}-\mathbf{x}'|)}{\alpha_0} = \frac{b_0(|\mathbf{x}-\mathbf{x}'|)}{b_0} = \frac{\lambda_1(|\mathbf{x}-\mathbf{x}'|)}{\lambda_1} = \frac{\alpha(|\mathbf{x}-\mathbf{x}'|)}{\alpha} = \frac{C_1(|\mathbf{x}-\mathbf{x}'|)}{C_1} \tag{58}$$
$$= \frac{\beta(|\mathbf{x}-\mathbf{x}'|)}{\beta} = \frac{\gamma(|\mathbf{x}-\mathbf{x}'|)}{\gamma} = G(|\mathbf{x}-\mathbf{x}'|).$$

Eringen (1984) has already shown that the function G satisfies

$$(1-\epsilon^2\nabla^2)G = \delta(|\mathbf{x}-\mathbf{x}'|), \tag{59}$$

where $\epsilon = e_0 a$. Employing the operator $(1-\epsilon^2\nabla^2)$ on (54)-(57) after using (58) and (59), we obtain

$$(1-\epsilon^2\nabla^2)t_{kl} = t^C_{kl} = (\lambda_0\psi(\mathbf{x}) + \lambda\epsilon_{rr}(\mathbf{x}))\delta_{kl} + (\mu+\kappa)\epsilon_{kl}(\mathbf{x}) + \mu\epsilon_{lk}(\mathbf{x}), \tag{60}$$

$$(1-\epsilon^2\nabla^2)m_{kl} = m^C_{kl} = b_0\epsilon_{mlk}\psi_{,m}(\mathbf{x}) + \alpha\gamma_{rr}(\mathbf{x})\delta_{kl} + \beta\gamma_{kl}(\mathbf{x}) + \gamma\gamma_{lk}(\mathbf{x}), \tag{61}$$

$$(1-\epsilon^2\nabla^2)m_k = m^C_k = \alpha_0\psi_{,k}(\mathbf{x}) + b_0\epsilon_{klm}\gamma_{lm}(\mathbf{x}), \tag{62}$$

$$(1-\epsilon^2\nabla^2)(s-t) = (s-t)^C = \lambda_1\psi(\mathbf{x}) + \lambda_0\epsilon_{kk}(\mathbf{x}), \tag{63}$$

in the absence of temperature field and employing the property $\int f(x)\delta(x-a)dx = f(a)$. For a nonlocal isotropic microstretch solid, the field equations are [Eringen (2002)]

$$t_{kl,k} + \rho(f_l - \ddot{u}_l) = 0, \tag{64}$$

$$m_{kl,k} + \epsilon_{lmn}t_{mn} + \rho(l_l - j\ddot{\phi}_l) = 0, \tag{65}$$

$$m_{k,k} + (t-s) + \rho\left(l - \frac{1}{2}j_0\ddot{\psi}\right) = 0, \tag{66}$$

Now, using the relations (53) and (60)-(63) into the field equations (64)-(66), we obtain

$$\lambda_0\psi_{,l} + (\lambda+\mu)u_{k,kl} + (\mu+\kappa)u_{l,kk} + \kappa\epsilon_{lkm}\phi_{m,k} + \rho(1-\epsilon^2\nabla^2)(f_l - \ddot{u}_l) = 0, \tag{67}$$

$$(\alpha+\beta)\phi_{k,kl} + \gamma\phi_{l,kk} + \kappa\epsilon_{lmn}u_{n,m} - 2\kappa\phi_l + \rho(1-\epsilon^2\nabla^2)(l_l - j\ddot{\phi}_l) = 0, \tag{68}$$

$$\alpha_0\psi_{,kk} - \lambda_1\psi - \lambda_0 u_{k,k} + \rho(1-\epsilon^2\nabla^2)\left(l - \frac{1}{2}j_0\ddot{\psi}\right) = 0. \tag{69}$$

Introducing the scalar potentials (q,ξ) and vector potentials $(\mathbf{U},\mathbf{\Pi})$ through Helmholtz decomposition of vectors as

$$\{\mathbf{u},\boldsymbol{\phi}\} = \nabla\{q,\xi\} + \nabla\times\{\mathbf{U},\mathbf{\Pi}\}; \quad \nabla\cdot\{\mathbf{U},\mathbf{\Pi}\} = 0, \tag{70}$$

and inserting it into (67) to (69), we obtain

$$\lambda_0\psi + (\lambda+2\mu+\kappa)\nabla^2 q = \rho(1-\epsilon^2\nabla^2)\ddot{q}, \tag{71}$$

$$(\mu+\kappa)\nabla^2\mathbf{U} + \kappa\nabla\times\mathbf{\Pi} = \rho(1-\epsilon^2\nabla^2)\ddot{\mathbf{U}}, \tag{72}$$

$$(\alpha+\beta+\gamma)\nabla^2\xi - 2\kappa\xi = \rho j(1-\epsilon^2\nabla^2)\ddot{\xi}, \tag{73}$$

$$\gamma\nabla^2\mathbf{\Pi} + \kappa\nabla\times\mathbf{U} - 2\kappa\mathbf{\Pi} = \rho j(1-\epsilon^2\nabla^2)\ddot{\mathbf{\Pi}}, \tag{74}$$

$$\alpha_0\nabla^2\psi - \lambda_1\psi - \lambda_0\nabla^2 q = \frac{\rho j_0}{2}(1-\epsilon^2\nabla^2)\ddot{\psi}. \tag{75}$$

For plane waves propagating in the positive direction of a unit vector $\mathbf{n}$, we have

$$\{q, \psi, \xi, \mathbf{U}, \mathbf{\Pi}\} = \{a_1, b_1, c_1, \mathbf{A_0}, \mathbf{B_0}\}\exp\{\iota k(\mathbf{n}\cdot\mathbf{r} - Vt)\}. \tag{76}$$

where a_1, b_1, c_1 are scalar constants, $\mathbf{A_0}, \mathbf{B_0}$ are vector constants and V is the phase speed. The circular frequency ω is defined by $\omega = kV$, k being the wavenumber.

Inserting the expressions of q and ψ from (76) into equations (71) and (75), and then eliminating a_1 or b_1, we obtain a quadratic equation in V^2, whose roots are given by

$$V_{1,2}^2 = \frac{1}{2A}\left(-B \pm \sqrt{B^2 - 4AC}\right), \tag{77}$$

where

$$A = \omega^2 - c_7^2, \qquad B = -\omega^2(c_6^2 - \epsilon^2\omega^2) - c_8^2\overline{\lambda}_0 - A(c_1^2 + c_3^2 - \epsilon^2\omega^2),$$
$$C = \omega^2(c_1^2 + c_3^2 - \epsilon^2\omega^2)(c_6^2 - \epsilon^2\omega^2),$$
$$\overline{\lambda}_0 = \frac{\lambda_0}{\rho}, \quad c_1^2 = \frac{\lambda + 2\mu}{\rho}, \quad c_3^2 = \frac{\kappa}{\rho}, \quad c_6^2 = \frac{2\alpha_0}{\rho j_0}, \quad c_7^2 = \frac{2\lambda_1}{\rho j_0}, \quad c_8^2 = \frac{2\lambda_0}{\rho j_0}.$$

It is easy to see that the wave associated with q is longitudinal in nature. We see that q is coupled with ψ, therefore the associated wave will be called as *coupled longitudinal waves*. Thus, there exist two sets of coupled longitudinal waves traveling with speeds V_1 and V_2, each set of which consists of a longitudinal displacement wave associated with microstretch property.

Inserting the expression of ξ from (76) into (73), we obtain

$$V_5^2 = (c_4^2 + c_5^2 - \epsilon^2\omega^2)\left(1 - \frac{2\omega_0^2}{\omega^2}\right)^{-1},$$

where

$$c_4^2 = \frac{\gamma}{\rho j}, \quad c_5^2 = \frac{\alpha + \beta}{\rho j}, \quad \omega_0^2 = \frac{\kappa}{\rho j}.$$

This is the speed of longitudinal micro-rotational wave whose counterpart has been already explored by Parfitt and Eringen (1969) in local micropolar elastic medium.

Next, inserting the expressions of $\mathbf{U}$ and $\mathbf{\Pi}$ from (76) into (72) and (74); then eliminating $\mathbf{A_0}$ or $\mathbf{B_0}$, we obtain a quadratic equation in V^2 whose roots are given by

$$V_{3,4}^2 = \frac{1}{2A^*}\left(-B^* \pm \sqrt{{B^*}^2 - 4A^*C^*}\right),$$

where

$$A^* = \omega^2 - 2\omega_0^2, \quad B^* = -\omega^2(c_4^2 - \epsilon^2\omega^2) - c_3^2\omega_0^2 - A^*(c_2^2 + c_3^2 - \epsilon^2\omega^2),$$
$$C^* = \omega^2(c_2^2 + c_3^2 - \epsilon^2\omega^2)(c_4^2 - \epsilon^2\omega^2), \qquad c_2^2 = \frac{\mu}{\rho}.$$

The quantities $V_{3,4}^2$ are the speeds of sets of coupled transverse waves whose counter part have already been explored by Parfitt and Eringen (1969).

We see that the speeds of coupled longitudinal waves given by $V_{1,2}^2$, coupled transverse waves given by $V_{3,4}^2$ and longitudinal microrotational wave given by V_5^2 depend on frequency and nonlocality parameters. Hence all the waves are dispersive and influenced by the nonlocality of the medium. At a glance, it can be verified that these expressions of speeds reduce to speeds of corresponding waves of micropolar elasticity in the absence of microstretch and nonlocality parameters of the medium (see Parfitt and Eringen, 1969).

7. Computational Results

In order to have better idea of phase speeds and their corresponding attenuations, if any, of the waves propagating through microstretch media, we have computed them for a specific model. For this purpose, the values of various parameters concerning the relevant microstretch media have been given below. For microstretch solid medium, the values have been borrowed from Kiris and Inan (2008), while those of microstretch fluid medium, they have been borrows from Abd-Alla et al. (2011).

For local microstretch solid

Symbol	*Value*	*Unit*	*Symbol*	*Value*	*Unit*
λ_s	7.5869×10^9	N m^{-2}	γ_s	2049.3	N
μ_s	1.896×10^9	N m^{-2}	λ_{0s}	4.8184×10^2	N
κ_s	1.425×10^5	N m^{-2}	λ_{1s}	3.4654×10^4	N
α_s	82.358	N	α_{0s}	1.7309×10^4	N
b_{0s}	1.96×10^4	N	β_s	110.70	N
j_s	1.96×10^{-7}	m^2	ρ_s	2.192×10^3	Kg m^{-3}

For local microstretch fluid

Symbol	*Value*	*Unit*	*Symbol*	*Value*	*Unit*
λ_f	1.5×10^9	N s m^{-2}	γ_f	0.0126×10^8	N s
μ_f	0.7×10^9	N s m^{-2}	λ_{0f}	0.15×10^5	N s
κ_f	0.011×10^9	N s m^{-2}	λ_{1f}	0.2×10^4	N s
α_f	0.0111×10^5	N s	α_{0f}	0.12×10^5	N s
b_{0f}	0.15×10^5	N s	β_f	0.0122×10^4	N s
j_f	0.00140×10^{-6}	m^2	ρ_f	1.0×10^3	Kg m^{-3}

Figures 2, 3 and 4 depict the variation of phase speeds of coupled longitudinal waves, coupled transverse waves and longitudinal micro-rotational wave in local microstretch solid medium with angular frequency varying from 0 to $10^4\pi$.

It can be observed from Figure 2 that phase speeds of coupled longitudinal waves are dispersive in the low frequency range, while they are almost non-dispersive in high frequency rage. The phase speed of second set of coupled longitudinal waves propagating with speed v_{2s} is relatively less than that of the wave

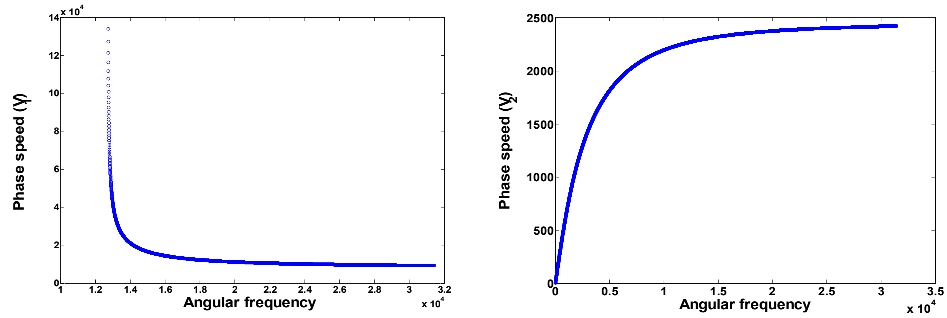


FIGURE 2. Two sets of coupled longitudinal waves in microstretch solid.

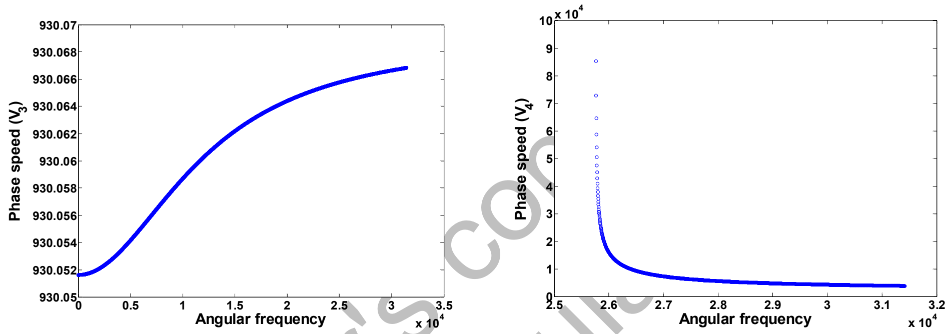


FIGURE 3. Two sets of coupled transverse waves in microstretch solid.

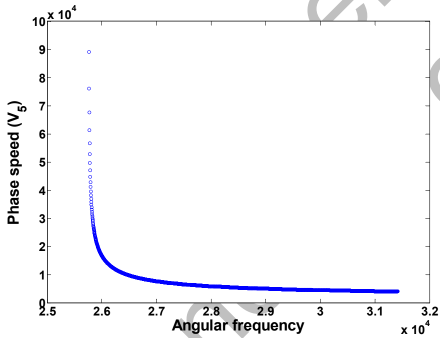


FIGURE 4. Longitudinal microrotational wave in microstretch solid.

propagating with speed v_{1s}. The variation of phase speeds of coupled transverse waves with angular frequency has been shown through Figure 3. One can notice from this figure that the set of coupled transverse waves propagating with speed v_{3s} is relatively less than that of the other set of coupled transverse waves at each value of frequency considered. Both these waves are also highly dispersive in the low frequency range, while less dispersive in high frequency range. It can be seen from Figure 4 that longitudinal micro-rotational wave propagating with speed v_{5s} is also highly dispersive in low frequency range. The pattern of variation of phase speed of waves propagating with speeds v_{1s}, v_{4s} and v_{5s} with frequency ω is similar.

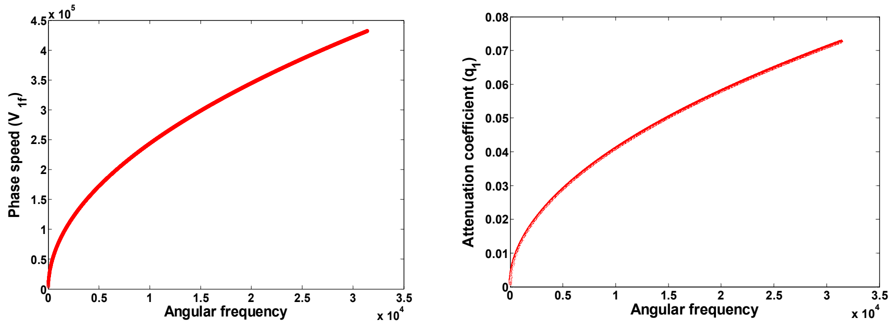


FIGURE 5. First set of coupled longitudinal waves in microstretch fluid

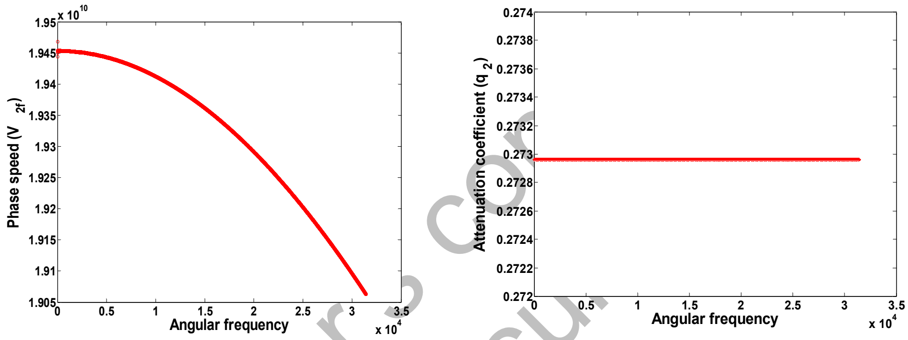


FIGURE 6. Second set of coupled longitudinal waves in microstretch fluid.

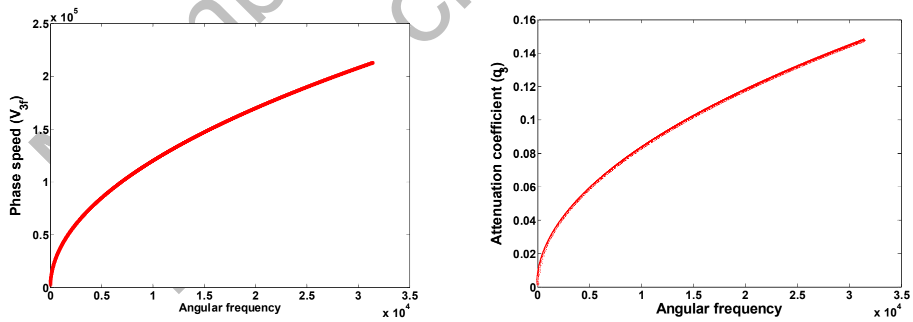


FIGURE 7. First set of coupled transverse waves in microstretch fluid.

Figures 5-9 depict the variation of phase speeds and corresponding attenuation coefficients of coupled longitudinal waves, coupled transverse waves and longitudinal micro-rotation wave with angular frequency varying from 0 to $10^4\pi$ in local microstretch fluid medium. It can be seen that both the sets of coupled longitudinal waves propagating with speed v_{1f} and v_{2f} are greatly depend on frequency and hence highly dispersive, while their attenuation pattern is different. The attenuation coefficient of coupled longitudinal waves propagating with speed v_{1f} depends on frequency, while that of the coupled dilatational waves propagating

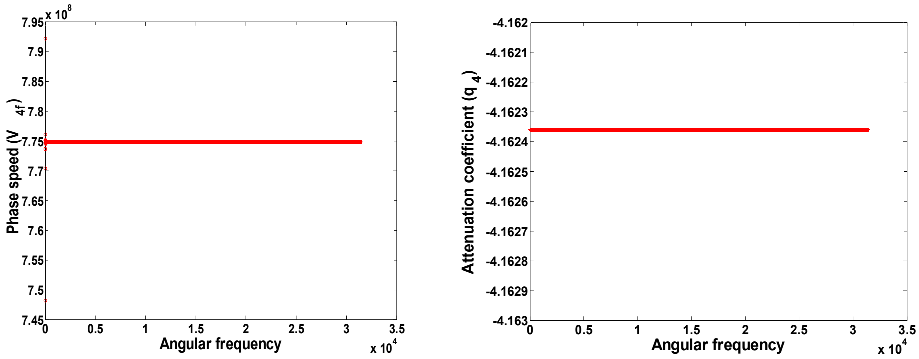


FIGURE 8. Second set of coupled transverse waves in microstretch fluid.

with speed v_{2f} hardly depends on frequency. From Figures 7 and 8, it is interesting to note that one of the sets of coupled transverse waves propagating with speed v_{3f} is dispersive and its corresponding attenuation coefficient depends on frequency. However, the coupled transverse waves propagating with speed v_{4f} is not dispersive at all and its corresponding attenuation coefficients is also independent of frequency. The similar behavior is also observed about the longitudinal micro-rotational wave propagating with speed v_{5f} as shown in Figure 9.

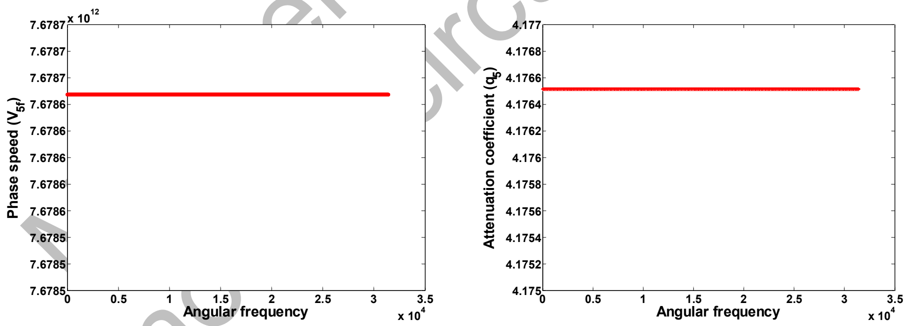


FIGURE 9. Longitudinal microrotational wave in microstretch fluid.

Figure 10 depicts the nature of dependence of various reflection coefficients due to incidence of a set of coupled longitudinal waves propagating with speed v_{1f} against the angle of incidence at fixed value of angular frequency $\omega = 10\pi\ Hz$. It can be easily observed from this figure that the nature of dependence of different reflection coefficient is different at each angle of incidence. But all the reflection coefficients vanish at normal incidence, except that of the reflected wave propagating with speed as of the incident wave. This shows that the set of incident coupled longitudinal waves comes back and does not give rise to any other reflected waves at the boundary. This reflected coupled waves remains dominant even when the angle of incidence runs through 0^0 to 90^0. At grazing incidence, the reflection

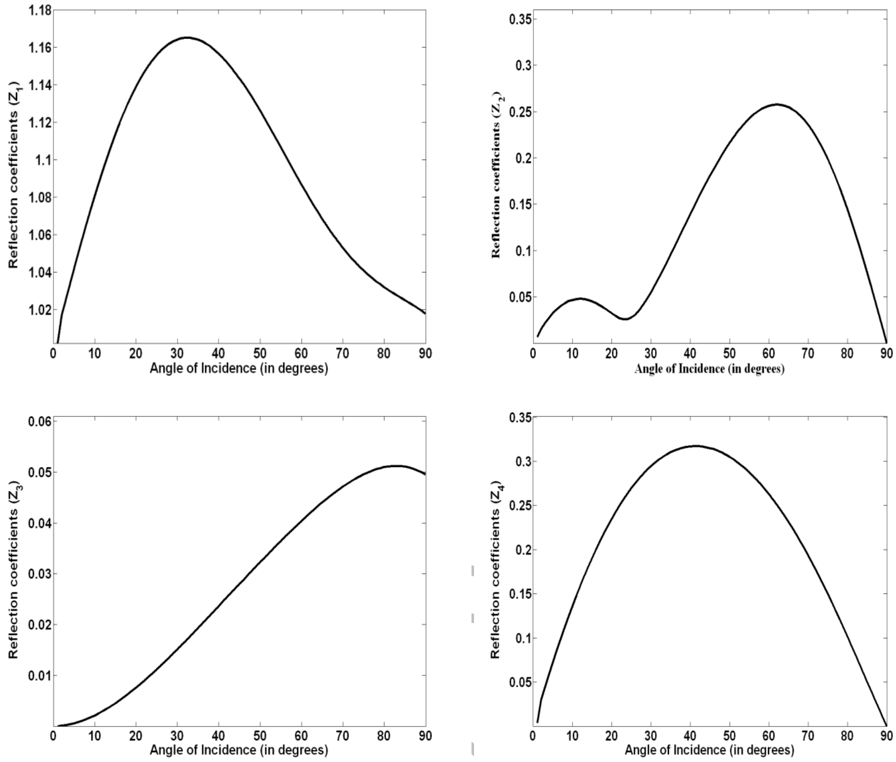


FIGURE 10. Variation of reflection coefficients with angle of incidence at $\omega = 10\pi Hz$.

coefficients z_1 and z_3 corresponding to reflected coupled dilatational waves and reflected coupled transverse waves respectively are appearing and remaining ones are absent showing that reflection phenomena takes place even at grazing incidence.

8. Conclusions

(a) The possibility of propagation of plane waves has been explored in (i) Local microstretch elastic solid, (ii) Local microstretch fluid and (iii) Non-local microstretch solid.

(b) Five basic waves may travel through an infinite microstretch media of either solid or liquid. These waves are two sets of coupled longitudinal waves, two sets of coupled transverse waves and an independent longitudinal micro-rotational wave.

(c) These waves are found to be dispersive and attenuating in nature when propagating through microstretch fluid, while they are dispersive but not attenuating in microstretch solid (local or non-local).

(d) All the waves propagating through the nonlocal microstretch solid medium are found to be influenced by the nonlocality parameter of the medium.

(e) The phenomenon of reflection of a set of coupled longitudinal waves from the free boundary of a microstretch fluid half-space has been investigated. The various reflection coefficients are found to depend on the angle of incidence and material properties of the microstretch fluid half-space.

Appendix-A

$$a_{11} = a_{21} = a_{31} = a_{41} = 1,$$
$$a_{12} = \frac{1}{D_1}\left[\lambda_f + (2\mu_f + \kappa_f)\cos^2\theta_2 - \frac{\lambda_{0f}\zeta_2}{k_2^2}\right]\left(\frac{k_2}{k_0}\right)^2,$$
$$a_{13} = \frac{1}{D_1}(2\mu_f + \kappa_f)\sin\theta_3\cos\theta_3\left(\frac{k_3}{k_0}\right)^2,$$
$$a_{14} = \frac{1}{D_1}(2\mu_f + \kappa_f)\sin\theta_4\cos\theta_4\left(\frac{k_4}{k_0}\right)^2,$$
$$a_{22} = \frac{1}{D_2}(2\mu_f + \kappa_f)\sin\theta_2\cos\theta_2\left(\frac{k_2}{k_0}\right)^2,$$
$$a_{23} = -\frac{1}{D_2}\left[\kappa_f(\cos^2\theta_3 - \eta_3) + \mu_f\cos 2\theta_3\right]\left(\frac{k_3}{k_0}\right)^2,$$
$$a_{24} = -\frac{1}{D_2}\left[\kappa_f(\cos^2\theta_4 - \eta_4) + \mu_f\cos 2\theta_4\right]\left(\frac{k_4}{k_0}\right)^2,$$
$$a_{32} = \frac{1}{D_3}b_{0f}\sin\theta_2\frac{\zeta_2}{k_2^2}\left(\frac{k_2}{k_0}\right)^3,$$
$$a_{33} = \frac{1}{D_3}\gamma_f\eta_3\cos\theta_3\left(\frac{k_3}{k_0}\right)^3,$$
$$a_{34} = \frac{1}{D_3}\gamma_f\eta_4\cos\theta_4\left(\frac{k_4}{k_0}\right)^3,$$

References

[1] Abd-Alla, A. M., Afifi, N. A. S., Mahmoud, S. R. and AL-Sheikh, M. A., *Effect of the Rotation on Waves in a Cylindrical Borehole Filled with Micropolar Fluid,* Int. Math. Forum, **6**(55) (2011), 2711 - 2728.

[2] Chen, Y., Lee, J. D. and Eskandarian, A., *Atomistic view point of the applicability of micro-continuum theories,* Int. J. Solid. Struct. **41** (2004), 2085-2097.

[3] Eremeyev, V. A., *Acceleration waves in micropolar elastic media,* Doklady Phys. **50**(4) (2005), 204-206.

[4] Eringen, A. C., *Nonlocal polar elastic continua,* Int. J. Engng. Sci., **10** (1972a), 1-16.

[5] Eringen, A. C., *Linear theory of nonlocal elasticity and dispersion of plane waves,* Int. J. Engng. Sci. **10** (1972b), 425-435.

[6] Eringen, A. C., *Nonlocal Continuum Field Theories,* Springer Verlag, New York, Chapter 6, 71-176 (2002).

[7] Eringen, A. C., *Microcontinuum Field Theories I: Foundations and Solids, II: Fluent Media,* Springer Verlag, New York, (1999).

[8] Eringen, A. C., *Plane waves in nonlocal micropolar elasticity,* Int. J. Engng. Sci. **22** (1984), 1113-1121.

[9] Eringen, A. C. and Edelen, D. G. B., *On nonlocal elasticity,* Int. J. Engng. Sci. **10** (1972), 233-248.

[10] Eringen, A. C. and Suhubi, E. S., *Non-linear theory of simple micro-elastic solids-I,* Int. J. Engng. Sci. **2**(2) (1964a), 189-203.

[11] Hsia, S.-Y., Chiu, S.-M. and Cheng J.-W., *Wave propagation at the human muscle-compact bone interface,* Theor. Appl. Mech. **33**(3) (2006), 223-243.

[12] Khurana, A. and Tomar, S. K., *Reflection of plane longitudinal waves from the stress-free boundary of a nonlocal, micropolar solid half-space,* J. Mech. Mat. and Struct. **8**(1) (2013), 95-107.

[13] Kiris, A. and Inan, E., *On the identification of microstretch elastic moduli of materials by using vibration data of plates,* Int. J. Engng. Sci. **46** (2008), 585-597.

[14] Lazar, M. and Kirchner, H. O. K., *The Eshelby tensor in nonlocal elasticity and in nonlocal micropolar elasticity,* J. Mech. Mat. Struct. **1** (2006), 325-337.

[15] Midya, G. K., *On Love-type surface waves in homogeneous micropolar elastic media,* Int. J. Engng. Sci. **42**(1112): (2004), 1275-1288.

[16] Parfitt, V. R. and Eringen, A. C., *Reflection of plane waves from a flat boundary of a micropolar elastic half-space,* J. Acoust. Soc. Am. **45** (1969), 1258.

[17] Singh, B., *Reflection and refraction of plane waves at a liquid/thermo-microstretch elastic solid interface,* Int. J. Engng. Sci. **39**(5) (2001), 583-598.

[18] Singh, B., *Reflection of plane waves from free surface of a microstretch elastic solid,* J. Earth Sys. Sci. **111**(1) (2002), 29-37.

[19] Singh, D. and Tomar, S. K., *Longitudinal waves at a micropolar fluid/solid interface,* Int. J. Solid. Struc. **45** (2008), 225-244.

[20] Suhubi, E. S. and Eringen, A. C., *Non-linear theory of simple micro-elastic solids-II,* Int. J. Engng. Sci. **2**(4) (1964b), 389-404.

[21] Tomar, S. K. and Kumar, R., *Wave propagation at liquid/micropolar elastic solid interface,* J. Sound Vib. 222(5), (1999), 858-869.

[22] Tomar, S. K. and Singh, H., *Radial vibrations due to a spherical cavity contained in an unbounded micropolar elastic medium,* Indian J. Pure Appl. Math. **34**(2) (2003), 1785-1796.

[23] Tomar, S. K. and Singh, D., *Propagation of Stoneley waves at an interface between two microstretch elastic half-spaces,* J. Vib. Cont. **12**(9) (2006), 995-1009 [Erratum: ibid **13**(12) (2007), 1835-1836].

[24] Tomar, S. K. and Rani, N., *Propagation of plane waves in microstretch fluid,* Appl. Math. Model. **35**(12) (2011), 5751-5765.

[25] Zeng, X., Chen, Y. and Lee, J. D., *Determining material constants in nonlocal micromorphic theory through phonon dispersion relations,* Int. J. Engng. Sci. **44** (2006), 1334-1345.

S. K. Tomar
Department of Mathematics, Panjab University
Chandigarh-160 014 (Union Territory), India
E-mail: *sktomar@pu.ac.in* (SKT).

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 25-38

ENHANCING THE CONVERGENCE OF RAMANUJAN'S SERIES AND DERIVING SIMILAR SERIES FOR CATALAN'S CONSTANT AND $\pi\sqrt{3}$

AMRIK SINGH NIMBRAN

(Received : 13 - 05 - 2014 ; Revised : 20 - 05 - 2015)

Abstract. This paper seeks to enhance the convergence of Ramanujan's famous series for Catalan's constant by providing extra factors to the denominator. An elementary method is illustrated to derive two infinite classes of similar series with enhanced rate of convergence by way of increasing the number of factors in the denominator. In addition, three infinite classes of series for $\pi\sqrt{3}$ are also introduced in the paper.

1. Introduction

Catalan's constant was introduced in 1865 by the Franco-Belgian mathematician Eugéne Charles Catalan (1814-94). He used the symbol G to denote the constant and defined it on page 20 of [4] by means of an alternating series and on page 33 by means of an integral

$$G = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} = \int_0^1 \frac{\arctan x}{x}\, dx.$$

He investigated G again in 1883 [5] and gave a frustratingly sluggish series

$$G = \sum_{n=0}^{\infty} \frac{2^{2n-1}}{\binom{2n}{n}(2n+1)^2}. \tag{1}$$

Glaisher briefly discusses Catalan constant in his papers [8, 9] without using the symbol G and defines it by means of the integral

$$\frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{x}{\sin x} dx.$$

He records on page 190 of [9] this value of the constant: .915965594177219015 05 . . .

Later mathematicians looked for more rapid series that converge to G faster. To explain what I mean by a faster series, I give here a pair of series with second series converging to $\frac{1}{\pi}$ at a quicker pace than the first

2010 Mathematics Subject Classification : 11Y60, 1106, 65B10, 33F05.
Key words and phrases : Catalan's constant, Ramanujan's series, transformation of series, Convergence acceleration.

$$\sum_{n=0}^{\infty} \frac{\binom{2n}{n}^2}{2^{4n}(2n-1)} = -\frac{2}{\pi},$$
$$\sum_{n=0}^{\infty} \frac{\binom{2n}{n}^2}{2^{4n}(2n-1)^2} = \frac{4}{\pi}.$$

Ramanujan discovered a couple of rapid series for G and first gave the formula [2, 12]

$$\sum_{n=0}^{\infty} \frac{\binom{2n}{n}^2}{2^{4n}(2n+1)} = \frac{4G}{\pi}. \tag{2}$$

A whole class of series like (2) are derived in my recent paper [11].

However, these series involve computation of π followed by multiplication with an approximate value of the infinite sum on the left hand side.

We are going to discuss here Ramanujan's other faster series [1, 13, 14, 15]

$$G = \frac{1}{8}\left[\pi \ln(2+\sqrt{3}) + 3\sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)^2}\right] \tag{3}$$

and derive many more similar series.

This series is faster than Ramanujan's earlier series as it involves no factor in the numerator and has a square factor in the denominator as compared to a single linear factor in the denominator of the previous series. However, it has a comparative disadvantage as it involves computation of a complicated quantity, namely $\pi \ln(2+\sqrt{3})$, which then has to be added to the approximate value of the infinite sum in the second term.

Greg Fee [7] computed G to 300000 digits on September 29, 1996 using (3). Robert J. Setti is reported to have computed G to 100,000,000,000 digits on April 6, 2013, but the series used for the computation is not specified.

Professor Alexandru Lupas (1942-2007), a Romanian mathematician, derived a very fast converging series for G in 2000[10]

$$G = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} 2^{8n}(40n^2-24n+3)}{(4n)^3\binom{4n}{2n}^2\binom{2n}{n}(2n-1)}. \tag{4}$$

Bradley [3] develops some acceleration formulae based on transformations of the log tangent integral and also treats Ramanujan series (3).

In Section 2, we consider transformations of the defining series by Catalan and give our own results.

2. Transformation of G Series

In the first part of his paper [4], Catalan employed a method of transformation of series to derive two efficient series including the following series (see p. 20)

$$G = \frac{2909}{3150} - 768\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)^2(2n+3)^2(2n+5)^2(2n+7)}. \tag{5}$$

The series (5) transforms into the following series having only positive terms

$$G = \frac{2909}{3150} - 7680 \sum_{n=1}^{\infty} \frac{1}{(4n-3)(4n-1)^2(4n+1)^2(4n+3)^2(4n+5)^2(4n+7)}. \quad (6)$$

Catalan used series (5) to compute the first ten decimal digits recorded on page 23.

Let me now explain a simple method of transformation to derive Catalan-type efficient series.

Leibniz discovered in 1674 his celebrated series for computing π which can be deduced from the expansion of the arctan function

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1}. \quad (7)$$

The following recurrence relation with $k \in \mathbb{N}$ is not difficult to establish

$$\begin{aligned} &\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)(2n+3)\cdots(2n+2k-3)(2n+2k-1)} \\ &= \frac{1}{k} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)(2n+3)\cdots(2n+2k-5)(2n+2k-3)} \\ &\quad - \frac{1}{2k \cdot 1 \cdot 3 \cdot 5 \cdots (2k-3)(2k-1)}. \end{aligned}$$

So by using the above relation, we can deduce an infinite class of series with more and more factors in the denominator.

Note that shifting the index in the series for G, we obtain new series. For example, the sum with denominator $(2n-1)^2$ is G and that with $(2n+1)^2$ is $1-G$. Now

$$\begin{aligned} \frac{1}{(2n-1)^2(2n+1)^2} &= \frac{1}{4}\left(\frac{1}{2n-1} - \frac{1}{2n+1}\right)^2 \\ &= \frac{1}{4}\left[\frac{1}{(2k-1)^2} + \frac{1}{(2k+1)^2} - \frac{2}{(2n-1)(2n+1)}\right]. \end{aligned}$$

On taking the sum from 1 to ∞, the two terms involving square factors that contain G cancel each other, leaving a constant and the term involving π. Thus

$$\pi = 4 - 8 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2(2n+1)^2}. \quad (8)$$

We can similarly deduce

$$G = \frac{11}{18} + 8 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2(2n+3)^2}. \quad (9)$$

We may combine the two classes of series to obtain some series for G. For example, combining the sums $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)}$ and $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n+1)^2}$ we obtain

$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)^2}$. Similarly, upon combination, $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n+1)(2n+3)}$ and $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n+1)^2}$ yield $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n+1)^2(2n+3)}$. The combination of the two sums in turn leads to

$$G = \frac{5}{6} + 4\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)(2n+1)^2(2n+3)}. \tag{10}$$

We shall now use this result for deriving a series with three consecutive squares.

$$\frac{1}{(2n-1)^2(2n+1)^2(2n+1)^2} = \frac{1}{16}\left(\frac{1}{(2n-1)(2n+1)} - \frac{1}{(2n+1)(2n+3)}\right)^2$$
$$= \frac{1}{16}\left[\frac{1}{(2k-1)^2(2n+1)^2} + \frac{1}{(2k+1)^2(2n+3)^2} - \frac{2}{(2n-1)(2n+1)^2(2n+3)}\right].$$

Putting the values deduced earlier, we straightway obtain

$$G = \frac{19}{18} - 32\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2(2n+1)^2(2n+3)^2}. \tag{11}$$

We discern that when the number of consecutive square factors are odd we obtain series for G, and when it is even we get series for π. Further, excluding the middle factor in the odd factors yields a series with even factors for G.

$$G = \frac{21131}{22050} - 1536\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2(2n+1)^2(2n+5)^2(2n+7)^2}. \tag{12}$$

$$G = \frac{3919}{4410} + 24576\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^2(2n+1)^2(2n+3)^2(2n+5)^2(2n+7)^2}. \tag{13}$$

This series has more factors in the denominator than (5) has and is therefore relatively faster series. This way, we can keep increasing the number of factors in denominator indefinitely.

3. Two Infinite Classes of Ramanujan-type G Series

Ramanujan's formula has an odd square factor in the denominator. We will derive two types of series here - those which have only odd linear factors and those which have either consecutive linear factors or odd square factors. Though it is easier to treat the first type of series, I would first take up the second type.

3.1. **Modifications of Ramanujan's G Series.** It is easy to see that

$$\sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)^2} = \sum_{n=1}^{\infty} \frac{1}{\binom{2n-2}{n-1}(2n-1)^2} = \sum_{n=1}^{\infty} \frac{2}{\binom{2n}{n}n(2n-1)}.$$

Hence, Ramanujan's series (3) can be written as

$$G = \frac{1}{8}\left[\pi \ln(2+\sqrt{3}) + 6\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n(2n-1)}\right]. \tag{14}$$

By the Binomial Theorem

$$\frac{1}{\sqrt{1-x^2}} = 1 + \frac{1}{2}x^2 + \frac{1\cdot 3}{2\cdot 4}x^4 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}x^6 + \cdots \tag{15}$$

which, upon term-by-term integration, yields

$$\arcsin x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{x^7}{7} + \cdots . \tag{16}$$

Multiplying these two series one gets the expansion given by Euler [6]

$$\frac{\arcsin x}{\sqrt{1-x^2}} = x + \frac{2}{3}x^3 + \frac{2\cdot 4}{3\cdot 5}x^5 + \frac{2\cdot 4\cdot 6}{3\cdot 5\cdot 7}x^7 + \cdots \tag{17}$$

$$= \sum_{n=1}^{\infty} \frac{2^{2n-1}}{\binom{2n}{n}n}x^{2n-1}. \tag{18}$$

Putting $x = \frac{1}{2}$ in this expansion one obtains

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n} = \frac{\pi\sqrt{3}}{9}. \tag{19}$$

Observe that

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n} = \sum_{n=0}^{\infty} \frac{1}{\binom{2n+2}{n+1}(n+1)}$$
$$= \sum_{n=0}^{\infty} \frac{(n+1)^2}{\binom{2n}{n}(2n+2)(2n+1)(n+1)}$$
$$= \sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}2(2n+1)},$$

and hence one has

$$\sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)} = 2\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n} = \frac{2\pi\sqrt{3}}{9}. \tag{20}$$

Combination of the series (19) and (20) gives

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n(2n+1)} = 2 - \frac{\pi\sqrt{3}}{3} \tag{21}$$

which, in turn, when combined with (14) yields

$$G = \frac{3}{2} - \frac{\pi\sqrt{3}}{4} + \frac{\pi}{8}\ln(2+\sqrt{3}) + 3\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n-1)2n(2n+1)}. \tag{22}$$

Observe that the series (22) has three linear factors in the denominator while Ramanujan's series has only a square of one linear factor, but it contains an extra term involving $\pi\sqrt{3}$ and hence requires extra effort to compute it.

The following elegant series is due to Euler

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}n^2} = \frac{\pi^2}{18} \tag{23}$$

which can be rewritten as

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}n^2} &= \sum_{n=0}^{\infty}\frac{1}{\binom{2n+2}{n+1}(n+1)^2}\\ &= \sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)(2n+2)}\\ &= \sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2)}.\end{aligned}$$

Using Euler's series and equation (20), we can deduce that

$$\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(n+1)(2n+1)}=\frac{\pi^2}{9} \tag{24}$$

and

$$\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(n+1)}=\frac{4\pi\sqrt{3}}{9}-\frac{\pi^2}{9}. \tag{25}$$

We may combine equations (21) and (24) to deduce

$$\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}n(n+1)(2n+1)}=3-\frac{\pi\sqrt{3}}{3}-\frac{\pi^2}{9} \tag{26}$$

which, in turn, when combined with (22) yields

$$G=\frac{15}{4}-\frac{\pi\sqrt{3}}{2}-\frac{\pi^2}{12}+\frac{\pi}{8}\ln(2+\sqrt{3})+9\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n-1)2n(2n+1)(2n+2)}. \tag{27}$$

This series has more factors in the denominator than Ramanujan's series but contains two additional terms - one involving $\pi\sqrt{3}$ and the other involving π^2.

We can successfully employ the method of shifting the index to deduce

$$\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+3)},\qquad \sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+4)},\qquad \sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+5)},\qquad \dots$$

and then combine them, thereby getting more and more factors in the denominator.

Let me explain this through a recurrence relation. Observe that

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k-1)} &= \sum_{n=0}^{\infty}\frac{(n+1)^2}{\binom{2n}{n}(2n+1)(2n+2)(2n+2k+1)}\\ &= \frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)}+\frac{1}{8k}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)}\right]\\ &= \frac{2k-1}{8k}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)}+\frac{1}{8k}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}.\end{aligned}$$

This gives a useful recurrence relation

$$\begin{aligned}&\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)}\\&=\frac{8k}{2k-1}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k-1)}-\frac{1}{2k-1}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}.\end{aligned}\tag{28}$$

We can similarly derive recurrence relation for the even factors. We can then evaluate series with linear odd and even factors and combine the two. However, the disadvantage of such series with more than three factors in the denominator is the occurrence of the term involving π^2.

Let us now proceed to have only square factors in the denominator of the series occurring in the formula. This will avoid the term with π^2. We need another recurrence relation for this purpose

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)^2}&=\sum_{n=0}^{\infty}\frac{(n+1)^2}{\binom{2n}{n}(2n+1)(2n+2)(2n+2k+3)^2}\\&=\frac{1}{4}\sum_{n=0}^{\infty}\frac{2n+2}{\binom{2n}{n}(2n+1)(2n+2k+3)^2}\\&=\frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}+\frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)(2n+2k+3)^2}\\&=\frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}\\&\quad+\frac{1}{4(2k+2)}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)(2n+2k+3)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}\right]\\&=\frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}-\frac{1}{4(2k+2)}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}\\&\quad+\frac{1}{4(2k+2)^2}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)}\right]\\&=\frac{2k+1}{4(2k+2)}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}\\&\quad+\frac{1}{4(2k+2)^2}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)}\right].\end{aligned}$$

Hence, we get

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)^2}&=\frac{4(2k+2)}{2k+1}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+1)^2}\\&\quad-\frac{1}{(2k+1)(2k+2)}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+2k+3)}\right].\end{aligned}$$

Putting $k=0$ this gives

$$\begin{aligned}\sum_{n=0}^{\infty} & \frac{1}{\binom{2n}{n}(2n+3)^2} \\ &= 8\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)^2}+\frac{1}{2}\left[\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}-\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+3)}\right].\end{aligned} \tag{29}$$

Using relevant relations/values we can now compute

$$\begin{aligned}&\sum_{n=1}^{\infty}\frac{4}{\binom{2n}{n}(2n+1)^2(2n+3)^2} \\ &=\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)^2}+\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+3)^2}-\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)}+\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+3)}.\end{aligned}$$

This yields a series having two square factors in the denominator as against one in Ramanujan's series though it involves extra term with $\pi\sqrt{3}$

$$G=\frac{23}{27}-\frac{\pi\sqrt{3}}{12}+\frac{1}{8}\pi\ln(2+\sqrt{3})+\frac{1}{6}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)^2(2n+3)^2}. \tag{30}$$

We can go on like this but it would involve formidable computations.

3.2. **Another Class of Ramanujan-type G Series.** We may take an alternative route to derive Ramanujan-type series which will have only linear and odd terms in the denominator. We first obtain the series with one factor in the denominator by splitting the right hand side of series (14) as

$$G=\frac{1}{8}\left[\pi\ln(2+\sqrt{3})+6\sum_{n=1}^{\infty}\frac{2}{\binom{2n}{n}(2n-1)}-6\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}n}\right]. \tag{31}$$

We now eliminate the second series on the right hand side with the help of (19) and thereby get

$$G=-\frac{\pi\sqrt{3}}{12}+\frac{\pi}{8}\ln(2+\sqrt{3})+\frac{3}{2}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n-1)}. \tag{32}$$

This series is the first in an unending hierarchy of series, with ever-increasing number of factors in the denominators, that we can construct. I shall now illustrate my method.

Rearranging the terms of (32) and then combining with (20) gives

$$G=-\frac{3}{2}+\frac{\pi\sqrt{3}}{4}+\frac{\pi}{8}\ln(2+\sqrt{3})+3\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n-1)(2n+1)}. \tag{33}$$

As mentioned earlier, we get this from the equation (28)

$$\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n+3)}=\frac{14\pi\sqrt{3}}{9}-\frac{25}{3}. \tag{34}$$

Combining the series (20) and (34), we get

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)(2n+3)} = \frac{11}{3} - \frac{2\pi\sqrt{3}}{3}. \tag{35}$$

Combining the results (33) and (35), we obtain

$$G = \frac{19}{2} - \frac{7\pi\sqrt{3}}{4} + \frac{\pi}{8}\ln(2+\sqrt{3}) + 12\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n-1)(2n+1)(2n+3)}. \tag{36}$$

Further, we derive from the recurrence relation (28) the following series

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n+5)} = \frac{74\pi\sqrt{3}}{9} - \frac{2009}{45}. \tag{37}$$

This series can be combined with previous results to obtain

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n+3)(2n+5)} = \frac{817}{45} - \frac{10\pi\sqrt{3}}{3}. \tag{38}$$

Furthermore, we get

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)(2n+3)(2n+5)} = \frac{2\pi\sqrt{3}}{3} - \frac{163}{45}. \tag{39}$$

This result combined with (36) yields

$$\begin{aligned} G = &-\frac{1019}{30} + \frac{25\pi\sqrt{3}}{4} + \frac{\pi}{8}\ln(2+\sqrt{3}) \\ &+ 72\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n-1)(2n+1)(2n+3)(2n+5)}. \end{aligned} \tag{40}$$

We can go on indefinitely like this to get G series with more and more factors in the denominator and thus having better rate of convergence than that of Ramanujan's series. Further, we can eliminate the term containing $\pi\sqrt{3}$ by using two consecutive formulas. For example, add the results after multiplying (36) by 25 and (40) by 7 to obtain

$$G = \frac{1}{8}\left[-\frac{1}{15} + \pi\ln(2+\sqrt{3}) + \sum_{n=1}^{\infty} \frac{150n+501}{\binom{2n}{n}(2n-1)(2n+1)(2n+3)(2n+5)}\right]. \tag{41}$$

We may also eliminate the term containing $\pi\sqrt{3}$ by using the equations (22) and (36) thereby getting

$$G = \frac{1}{6} + \frac{1}{8}\pi\ln(2+\sqrt{3}) + \frac{3}{2}\sum_{n=1}^{\infty} \frac{2n+7}{\binom{2n}{n}(2n-1)2n(2n+1)(2n+3)}. \tag{42}$$

We can now combine the equations (41) and (42) to obtain a faster series

$$G = \frac{1}{5} + \frac{1}{8}\pi\ln(2+\sqrt{3}) + \frac{1}{2}\sum_{n=1}^{\infty} \frac{38n+125}{\binom{2n}{n}(2n-1)2n(2n+1)(2n+3)(2n+5)}. \tag{43}$$

Further more, we may combine the equations (40) and (30) to deduce two series – with and without the term involving $\pi\sqrt{3}$

$$\begin{aligned} G = {} & \frac{12059}{12930} - \frac{169}{1724}\pi\sqrt{3} + \frac{1}{8}\pi\ln(2+\sqrt{3}) \\ & - \frac{576}{431}\sum_{n=1}^{\infty}\frac{1}{\binom{2n}{n}(2n-1)(2n+1)^2(2n+3)^2(2n+5)}. \end{aligned} \tag{44}$$

$$\begin{aligned} G = {} & \frac{2693}{6840} + \frac{1}{8}\pi\ln(2+\sqrt{3}) \\ & + \frac{1}{152}\sum_{n=1}^{\infty}\frac{676n^2+1352n+307}{\binom{2n}{n}(2n-1)(2n+1)^2(2n+3)^2(2n+5)}. \end{aligned} \tag{45}$$

The equations (43) and (45) have the same rate of convergence because both have effectively four linear factors in the denominator. They can be combined to obtain the following efficient series without any term involving $\pi\sqrt{3}$

$$\begin{aligned} \mathbf{G} = {} & \mathbf{\frac{385}{918}} + \mathbf{\frac{1}{8}}\pi\ln(\mathbf{2}+\sqrt{\mathbf{3}}) \\ & - \mathbf{\frac{1}{102}}\sum_{\mathbf{n=1}}^{\infty}\frac{\mathbf{3380n^2+7064n+2535}}{\binom{\mathbf{2n}}{\mathbf{n}}\mathbf{(2n-1)2n(2n+1)^2(2n+3)^2(2n+5)}}. \end{aligned} \tag{46}$$

This series is morphologically same as Ramanujan's series. It has seven linear factors in the denominator with two in the numerator and so effectively five factors as against two linear factors in Ramanujan's series and thus converges to G more speedily.

4. Two Efficient Classes of Series for $\pi\sqrt{3}$

While deriving series for G in the previous section, we developed parallel class of series for $\pi\sqrt{3}$ whose denominator had $\binom{2n}{n}$. It means that the multiple in the denominator varied – it began with 1 and rose to approach 4, given by $3+\frac{n-1}{n+1}, n \geq 0$. It would be appropriate to introduce now two classes of series with a greater common ratio $= 3^2$.

Setting $x = \frac{1}{\sqrt{3}}$ in the expansion $\arctan x = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^{2n-1}}{(2n-1)}$ yields the following series attributed to Abraham Sharp who in 1699 calculated π to 72 decimal places

$$\frac{\pi}{6} = \frac{1}{\sqrt{3}}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{3^{n-1}(2n-1)}. \tag{47}$$

Sharp's alternating series transforms easily into one with only positive terms

$$\sum_{n=1}^{\infty}\frac{4n}{3^{2n}(4n-3)(4n-1)} = \frac{\pi\sqrt{3}}{36}. \tag{48}$$

We shall now explain a method to derive similar series with more factors.

It is easy to see that

$$\sum_{n=1}^{\infty}\left(\frac{1}{9^{n-1}(4n-3)} - \frac{1}{9^{n}(4n+1)}\right) = 1. \tag{49}$$

This can be rewritten as

$$\sum_{n=1}^{\infty} \frac{8n+3}{3^{2n}(4n-3)(4n+1)} = \frac{1}{4}. \tag{50}$$

This series combined with (48) yields

$$\sum_{n=1}^{\infty} \frac{4n+3}{3^{2n}(4n-3)(4n-1)(4n+1)} = \frac{\pi\sqrt{3}}{18} - \frac{1}{4}. \tag{51}$$

Now consider the integral

$$\int \frac{dx}{1+x^2} = \frac{x}{1+x^2} + 2\int \frac{dx}{1+x^2} - 2\int \frac{dx}{(1+x^2)^2}.$$

That is,

$$\int \frac{dx}{(1+x^2)^2} = \frac{x}{2(1+x^2)} + \frac{1}{2}\arctan x + C.$$

Using the Binomial Theorem to expand $\frac{1}{(1+x^2)^2}$ and integrating the result term-by-term lead to

$$\int \frac{dx}{(1+x^2)^2} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{nx^{2n-1}}{2n-1} + C. \tag{52}$$

Evaluating the integral from 0 to $\frac{1}{\sqrt{3}}$ and simplifying the result, we get

$$\frac{2\pi\sqrt{3}}{3} = -3 + 8\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n}{3^{n-1}(2n-1)}. \tag{53}$$

This result can be alternatively deduced from Sharp's series by writing it as

$$\begin{aligned}
\frac{\pi}{6} &= \frac{1}{\sqrt{3}} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{2n-(2n-1)}{3^{n-1}(2n-1)} \\
&= \frac{1}{\sqrt{3}} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{2n}{3^{n-1}(2n-1)} - \frac{1}{\sqrt{3}} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{3^{n-1}} \\
&= \frac{1}{\sqrt{3}} \left(\sum_{n=1}^{\infty} (-1)^{n-1} \frac{2n}{3^{n-1}(2n-1)} - \frac{3}{4} \right).
\end{aligned}$$

We may write the series on the right hand side of (53) as

$$\begin{aligned}
\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n}{3^{n-1}(2n-1)} &= \sum_{n=0}^{\infty} (-1)^{n} \frac{n+1}{3^{n}(2n+1)} \\
&= 1 - \sum_{n=1}^{\infty} (-1)^{n+1} \frac{n+1}{3^{n}(2n+1)}
\end{aligned}$$

$$\begin{aligned}
&= 1 - \sum_{n=1}^{\infty} \frac{2n}{3^{2n-1}(4n-1)} + \sum_{n=1}^{\infty} \frac{2n+1}{3^{2n}(4n+1)} \\
&= 1 - \sum_{n=1}^{\infty} \frac{16n^2+4n+1}{3^{2n}(4n-1)(4n+1)} \\
&= 1 - \sum_{n=1}^{\infty} \frac{1}{9^n} - \sum_{n=1}^{\infty} \frac{4n+2}{3^{2n}(4n-1)(4n+1)} \\
&= 1 - \frac{1}{8} - \sum_{n=1}^{\infty} \frac{4n+2}{3^{2n}(4n-1)(4n+1)}.
\end{aligned}$$

By using this result, one derives from (53)

$$\sum_{n=1}^{\infty} \frac{4n+2}{3^{2n}(4n-1)(4n+1)} = \frac{1}{2} - \frac{\pi\sqrt{3}}{12}. \tag{54}$$

Subtracting (54) from (48) yields (51). We may rewrite (48) as

$$\sum_{n=1}^{\infty} \frac{4n+4}{3^{2n}(4n+1)(4n+3)} = \frac{\pi\sqrt{3}}{9} - \frac{4}{3}. \tag{55}$$

Subtracting (55) from (54), we obtain

$$\sum_{n=1}^{\infty} \frac{4n+5}{3^{2n}(4n-1)(4n+1)(4n+3)} = \frac{11}{12} - \frac{\pi\sqrt{3}}{6}. \tag{56}$$

The equation (51) can be rewritten as

$$\sum_{n=1}^{\infty} \frac{4n+7}{3^{2n}(4n+1)(4n+3)(4n+5)} = \frac{\pi\sqrt{3}}{2} - \frac{163}{60}. \tag{57}$$

On subtracting (57) from (56), we get a series with four factors in the denominator

$$\sum_{n=1}^{\infty} \frac{4n+8}{3^{2n}(4n-1)(4n+1)(4n+3)(4n+5)} = \frac{109}{120} - \frac{\pi\sqrt{3}}{6}. \tag{58}$$

We deduce the associated series by subtracting (56) from (51)

$$\sum_{n=1}^{\infty} \frac{4n+6}{3^{2n}(4n-3)(4n-1)(4n+1)(4n+3)} = \frac{\pi\sqrt{3}}{18} - \frac{7}{24}. \tag{59}$$

We can go on this way to derive further series in the two inter-connected classes of series. These two relations, holding true for $m = 0, 1, 2, \cdots$, generate

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{4n+3m}{3^{2n}\prod_{k=0}^{m+1}(4n+2k-3)} - \sum_{n=1}^{\infty} \frac{4n+3m+2}{3^{2n}\prod_{k=0}^{m+1}(4n+2k-1)} \\
= \sum_{n=1}^{\infty} \frac{(2m+2)(4n+3m+3)}{3^{2n}\prod_{k=0}^{m+2}(4n+2k-3)}
\end{aligned} \tag{60}$$

and

$$\sum_{n=1}^{\infty} \frac{4n+3m+2}{3^{2n}\prod_{k=0}^{m+1}(4n+2k-1)} - \sum_{n=1}^{\infty} \frac{4n+3m+4}{3^{2n}\prod_{k=0}^{m+1}(4n+2k+1)} = \sum_{n=1}^{\infty} \frac{(2m+2)(4n+3m+5)}{3^{2n}\prod_{k=0}^{m+2}(4n+2k-1)}. \tag{61}$$

I have derived these general formulas for the two classes of series:

$$\sum_{\mathbf{n=1}}^{\infty} \frac{\mathbf{4n+3m}}{\mathbf{3^{2n}}\prod_{\mathbf{k=0}}^{\mathbf{m+1}}(\mathbf{4n+2k-3})} = \frac{\mathbf{2^{m-2}}\pi\sqrt{\mathbf{3}}}{\mathbf{m!\cdot 9}} - \mathbf{a_m}, \tag{62}$$

where $a_0 = 0$ and $a_{m+1} = \frac{2}{m+1}\, a_m + \frac{1}{4(m+1)1\cdot 3\cdot 5\cdots(2m+1)}$, for $m = 0, 1, 2, \ldots$
and

$$\sum_{\mathbf{n=1}}^{\infty} \frac{\mathbf{4n+3m+2}}{\mathbf{3^{2n}}\prod_{\mathbf{k=0}}^{\mathbf{m+1}}(\mathbf{4n+2k-1})} = \mathbf{b_m} - \frac{\mathbf{2^{m-2}}\pi\sqrt{\mathbf{3}}}{\mathbf{m!\cdot 3}}, \tag{63}$$

where $b_0 = \frac{1}{2}$ and $b_{m+1} = \frac{2}{m+1}\, b_m - \frac{1}{4(m+1)3\cdot 5\cdot 7\cdots(2m+3)}$, for $m = 0, 1, 2, \ldots$
We also have following direct formulas for a_m and b_m :

$$a_m = \frac{2^{m-2}}{m!}\sum_{k=1}^{m}\frac{1}{\binom{2k}{k}k}; \quad b_m = \frac{2^{m-1}}{m!}\Big(1 - \sum_{k=1}^{m}\frac{1}{\binom{2k}{k}2k(2k+1)}\Big). \tag{64}$$

These relations/formulas yield infinitely many pairs of series such as

$$\sum_{n=1}^{\infty} \frac{4n+9}{3^{2n}(4n-3)(4n-1)(4n+1)(4n+3)(4n+5)} = \frac{\pi\sqrt{3}}{27} - \frac{1}{5}. \tag{65}$$

Or equivalently,

$$\sum_{\mathbf{n=0}}^{\infty} \frac{\mathbf{4n+9}}{\mathbf{3^{2n-3}}(\mathbf{4n-3})(\mathbf{4n-1})(\mathbf{4n+1})(\mathbf{4n+3})(\mathbf{4n+5})} = \pi\sqrt{\mathbf{3}};$$

$$\sum_{n=1}^{\infty} \frac{4n+11}{3^{2n}(4n-1)(4n+1)(4n+3)(4n+5)(4n+7)} = \frac{127}{210} - \frac{\pi\sqrt{3}}{9}. \tag{66}$$

We can thus obtain infinitely many increasingly rapid series for $\pi\sqrt{3}$.
Acknowledgement: The author would like to acknowledge the discovery of some $\pi\sqrt{3}$ formulas by Prof. S. K. Lucas of James Madison University, Harrisonburg, Virginia (USA) that motivated the results presented in Section 4. The anonymous referee deserves thanks for pointing to the latest computation of G by Robert J. Setti and for bringing to author's notice a fast converging series derived by A. Lupas in 2000. The author also thanks Prof. Paul Levrie for his useful suggestions.

References

[1] Berndt, Bruce C., *Ramanujan's Notebooks: Part I,* Springer-Verlag, 1985, p. 289.

[2] Berndt, Bruce C., *Ramanujan's Notebooks, Part II,* Springer-Verlag, 1989, p. 40.

[3] Bradley, David M., *A Class of Series Acceleration Formulae for Catalan's Constant,* The Ramanujan Journal, Vol. 3, Issue 2, June 1999, 159–173.

[4] Catalan, E., *Sur la transformation des séries, et sur quelques intégrales définies,* Mémoires in 4 de l'Academie royale de Belgique, 1865.

[5] Catalan, E., *Recherches sur la constante G, et sur les integrales euleriennes,* Mémoires de l'Academie imperiale des sciences de Saint-Pétersbourg, Ser. 7, 31, 1883, p. 41.

[6] Euler, L., *Institutiones calculi differentialis cum eius usu in analysi finitorum ac doctrina serierum,* 1755, Part II, chapter I, p.242. E212 sec2 ch1 at `http://www.eulerarchive.org/`

[7] Fee, Greg, *Project Gutenberg EBook of Catalan's Constant [Ramanujan's Formula],* `http://www.gutenberg.org/6/8/682/`

[8] Glaisher, J. W. L., *On a numerical continued product,* Messenger of Mathematics, vol.VI (September, 1876) 71–76, p. 76, (xiv), (xv).

[9] Glaisher, J. W. L., *Further note on certain numerical continued products,* Messenger of Mathematics, vol.VI (1877) 189–192, p. 190 footnote.

[10] Lupas, Alexandru, *Formulae for some classical constants,* (to appear in Proceedings of ROGER-2000)

[11] Nimbran, A. S., *Deriving Forsyth-Glaisher type series for* $\frac{1}{\pi}$ *and Catalan's constant by an elementary method,* Mathematics Student,Vol. 84, Nos. 1–2, January–June (2015), 69-86.

[12] Ramanujan, S., *Modular equations and approximations to* $\frac{1}{\pi}$, Quarterly Journal of Pure and Applied Mathematics, XLV (1914), 350–372.

[13] Ramanujan, S., *On the integral* $\int_0^x \frac{arctant}{t}\, dt$, Journal of the Indian Mathematical Society, VII (1915), 93–96.

[14] Ramanujan, S., *Manuscript Book 1 of Srinivasa Ramanujan,* Tata Institute of Fundamental Research, chap. XI, p. 85, entry 32, Examples, 3 `http://www.math.tifr.res.in/publ/nsrBook1.pdf`

[15] Ramanujan, S., *Manuscript Book 2 of Srinivasa Ramanujan,* Tata Institute of Fundamental Research, Chap. IX, p. 111, entry 32, Examples, (iii). `http://www.math.tifr.res.in/publ/nsrBook2.pdf`

Amrik Singh Nimbran

Former Director General of Police, Bihar, Patna.

Email: *simnimas@yahoo.co.in ; amrikn622@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 39-52

TAYLOR SERIES FOR ARCTAN AND BBP-TYPE FORMULAS FOR π

AMRIK SINGH NIMBRAN*

(Received : 03 - 12 - 2014 ; Revised : 24 - 08 - 2015)

Abstract. We give here a neat Taylor series for the arctan function and use that to deduce a dozen BBP-type formulas for π. In addition, an alternative approach for deriving more formulas is also explained.

1. Introduction

In 1995 Bailey, Borwein, and Plouffe [3] discovered a lovely formula for π:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k}\left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right).$$

They found it by way of "inspired guessing and extensive searching" using Ferguson's PSLQ integer relation finding algorithm. The "proof" followed the discovery. Earlier, it was believed that if one wanted to determine the n-th digit of π, one had to generate the entire sequence of the first n digits. These authors also found an algorithm for computing individual hexadecimal or binary digits of π. The algorithm is explained by Bailey in his note. [4]

The above-noted *base-2* or binary formula allows us to compute the *nth* hexadecimal or binary digit of π, without computing any of the previous digits. Now a formula of this sort is called as BBP-type formula after the (initials of) three co-discoverers. It may be pointed out that the new algorithm "is not fundamentally faster than best-known schemes for computing all digits of π up to some position."

This notation was introduced later by D. H. Bailey and R. E. Crandall.

$$P(s, b, n, A) = \sum_{k=0}^{\infty} \frac{1}{b^k} \sum_{j=1}^{n} \frac{a_j}{(kn+j)^s},$$

where s, b and n are integers, and $A = (a_1, a_2, \ldots, a_n)$ is a vector of integers. Then the first formula becomes $\pi = P(1, 16, 8, (4, 0, 0, -2, -1, -1, 0, 0))$.

---

* Dedicated to the memory of late Prof. Ratan Prakash Agarwal (1925–2008), past President of IMS and past Editor of both IMS periodicals, who encouraged me to carry on with my work in mathematics.

2010 AMS Subject Classification: 26A24, 30D10, 33B10, 40A25.

Key words and phrases: Taylor series, expansion of arctan, BBP-type formulas, Pi

Adamchik and Wagon [1] employed *Mathematica* to derive an alternating BBP-type formula. We intend to derive here some such formulas (including that of Adamchik and Wagon) *mathematically* from a beautiful Taylor series expansion of the arctan.

2. Taylor's Theorem and A Series for Arctan

As is commonly known, a Taylor series is a power series representation of a function as an infinite sum of terms which are evaluated from the function's derivatives at a single point in its domain. The Scottish mathematician James Gregory (1638–75) communicated to John Collins (1625–1683), English mathematician and librarian of the Royal Society, in 1670–71 a number of results on infinite series expansions of various trigonometric functions, including what is now known as Gregory's series for the arctan function. His 1667 book *Vera Circuli et Hyperbolae Quadratura,* reprinted in 1668 with an appendix, *Geometriae Pars,* contained "the earliest enunciation"[6] of the expansions in series of $\sin x$, $\cos x$, $\arcsin x$ and $\arccos x$.

However, it is now an established fact [15, 17] that an enunciation of the inverse tangent series is found in Sanskrit verses attributed to an Indian mathematician Madhava (1340–1425) of Sangamagrama (near Kochi, Kerala) and quoted in the 16th century commentary *Kriyakramkari* on *Lilavati* of Bhaskaracharya (1114-1185). The series for *arctan* x and other trigonometric functions are given in Sanskrit verse in *Tantrasangraha* of Nilakantha (1450–1550, Kerala) and a commentary on this work called *Tantrasangraha-vyakhya* of unknown authorship. *Yuktibhasa* of Jyesthadeva (1500–1610), a commentary in Malalyalam on the *Tantrasangraha* contains a proof of the arctan series. In his *Aryabhatiya-bhasya*, a commentary on Aryabhata's work on astronomy, Nilakantha attributes the sine series to Madhava.[22]

Brook Taylor (1685–1731), an English mathematician, provided the general method for the formation of such series. His method, now known as Taylor's theorem, appears merely as a corollary (Corollary II of Proposition VII, Theorem III) to the corresponding theorem in Finite Differences in his 1715 work *Methodus Incrementorum Directa & Inversa*[24]. But, as Duncan Gregory (1813-1844) notes, he makes no application of it, or remark on its importance.[14]

What Taylor's theorem in effect does is to develop a function of the algebraic sum of two quantities into a series arranged according to the ascending power of one of these quantities, with coefficients depending on the other. While the Binomial Theorem expands $(x+h)^n$ in a series of powers of h, Taylor's Theorem expands *any infinitely differentiable function* of $(x+h)$ in a similar series. The Taylor series of an infinitely differentiable function f is given by: $f(x+h) = \sum_{n=0}^{\infty} \frac{h^n f^{(n)}(x)}{n!}$, where $f^{(n)}(x)$ denotes the nth derivative of f.

If we put in this $x = 0$ and change h into x, we get a series named after Colin

Maclaurin (1698–1746), who published it in his *A Treatise of Fluxions*[18]. It appears in §751 of Volume II. In the next section, he refers to similar result in Bernoulli's Act. Erud. Lips., 1694. In a letter of Dec. 7, 1728 to the Scottish mathematician James Stirling (1692–1770), Maclaurin refers to "a Theorem in y[our](?) book where a Quantity is expressed by a series whose coefficients are first, second, third fluxions, etc." Stirling had established it on pp.102–103 of his *Methodus Differentialis* [23], but he had used it earlier in a paper entitled 'Methodus Differentialis Newtoniana Illustrata', published in the Philosophical Transactions (1719) of the Royal Society[26]. Augustus De Morgan (1806–71) rightly commented: "...both Maclaurin and Stirling would have been astonished that a particular case of Taylor's theorem would be called by either of their names."[19]

We shall now derive a Taylor series representation for $\arctan(x)$ with simple closed-form coefficients.

Let $f(x) = \arctan(x)$ and $x = \cot\theta = \tan(\frac{\pi}{2} - \theta)$, that is, $\arctan x = \frac{\pi}{2} - \theta = \phi$, meaning that θ and ϕ are complementary angles. Here $|x| < 1$ and $0 < \theta < \frac{\pi}{2}$. So using a right-angled triangle with opposite side 1 and adjacent side x, we immediately get $\frac{1}{\sqrt{1+x^2}} = \sin\theta$.

The equation $1 + x^2 = (\sin\theta)^{-2}$ leads to $2x\frac{dx}{d\theta} = -2(\sin\theta)^{-3}\cos\theta$. That is, $\frac{dx}{d\theta} = -\frac{(\sin\theta)^{-3}\cos\theta}{\cot\theta} = -(\sin\theta)^{-2}$. Hence, $\frac{d\theta}{dx} = -\sin\theta\sin\theta$. Differentiating this, we obtain

$$\frac{d^2\theta}{dx^2} = -(2\sin\theta\cos\theta)\frac{d\theta}{dx} = -(\sin 2\theta)(-\sin^2\theta) = \sin^2\theta\sin 2\theta.$$

Further,

$$\begin{aligned}
\frac{d}{dx}(\sin^n\theta\,\sin(n\theta)) &= n\sin^{n-1}\theta\cos\theta\frac{d\theta}{dx}\sin(n\theta) + n\sin^n\theta\cos(n\theta)\frac{d\theta}{dx} \\
&= n\sin^{n-1}\theta\frac{d\theta}{dx}\left[\sin(n\theta)\cos\theta + cos(n\theta)\sin\theta\right] \\
&= n\sin^{n-1}\theta(-\sin^2\theta)\left[\sin(n\theta + \theta)\right] \\
&= -n\sin^{n+1}\theta\,\sin(n+1)\theta.
\end{aligned}$$

Applying Taylor' theorem to arctan, we obtain

$\arctan(x+h) = \arctan x + h\frac{d}{dx}\arctan x + \frac{h^2}{2!}\frac{d^2}{dx^2}\arctan x + \ldots$

Hence, we get

$$\begin{aligned}
\arctan(x+h) &= \left(\frac{\pi}{2} - \theta\right) + h\frac{d}{dx}\left(\frac{\pi}{2} - \theta\right) + \frac{h^2}{2}\frac{d^2}{dx^2}\left(\frac{\pi}{2} - \theta\right) + \ldots \\
&= \left(\frac{\pi}{2} - \theta\right) - h\frac{d\theta}{dx} - \frac{h^2}{2}\frac{d^2\theta}{dx^2} - \frac{h^3}{3!}\frac{d^3\theta}{dx^3} - \frac{h^4}{4!}\frac{d^4\theta}{dx^4}\ldots
\end{aligned}$$

Now using all the results that we obtained earlier, we obtain

$$\arctan(x+h) = \left(\frac{\pi}{2} - \theta\right) + h\sin\theta\sin\theta - \frac{h^2}{2}\sin^2\theta\sin 2\theta + \frac{h^3}{3}\sin^3\theta\sin 3\theta - \frac{h^4}{4}\sin^4\theta\sin 4\theta + - \dots \quad (2.1)$$

We may point out here that inexplicably this *arctan* expansion does not appear in modern calculus textbooks, or indeed in the numerous Taylor series tables that exist on the web. My extensive search led me eventually to an 1841 calculus text[14] by Gregory and [16, 25] where this series appears.

3. Series for deriving formulas for π

We now derive some relevant series.

Putting $x = 0$ in (2.1), that is, $\theta = \frac{\pi}{2}$ and then seting $h = x$ yields the well-known Maclaurin series for $\arctan(x)$:

$$\arctan(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^{2n-1}}{2n-1}, \quad -1 < x < 1. \quad (3.1)$$

If we put $h = -x$ in (2.1), then $\arctan(x+h) = \arctan(0) = 0$ and hence $0 = \frac{\pi}{2} - \theta - x\sin\theta\sin\theta - \frac{x^2}{2}\sin^2\theta\sin 2\theta - \frac{x^3}{3}\sin^3\theta\sin 3\theta - \dots,$ that is,

$$\begin{aligned} \frac{\pi}{2} - \theta &= \cot\theta\sin\theta\sin\theta + \frac{\cot^2\theta\sin^2\theta\sin 2\theta}{2} + \frac{\cot^3\theta\sin^3\theta\sin 3\theta}{3} + \dots \\ &= \cos\theta\sin\theta + \frac{\cos^2\theta\sin 2\theta}{2} + \frac{\cos^3\theta\sin 3\theta}{3} + \dots, \quad 0 < \theta < \frac{\pi}{2} \quad (3.2) \end{aligned}$$

On using the relation $\frac{\pi}{2} - \theta = \phi$, the equation(3.2) transforms into an elegant alternating series with two positive terms and two negative terms

$$\phi = \sin\phi\cos\phi + \frac{\sin^2\phi\sin 2\phi}{2} - \frac{\sin^3\phi\cos 3\phi}{3} - \frac{\sin^4\phi\sin 4\phi}{4} + + - - \dots, \; 0 < \phi < \frac{\pi}{2} \quad (3.3)$$

Putting $h = -x + \frac{1}{x} = -\cot\theta + \frac{1}{\cot\theta} = -2\cot 2\theta$ in (2.1), we get

$$\arctan\left(\frac{1}{x}\right) = \arctan(x) + (-2\cot 2\theta)\sin\theta\sin\theta - \frac{(-2\cot 2\theta)^2}{2}\sin^2\theta\sin 2\theta + \frac{(-2\cot 2\theta)^3}{3}\sin^3\theta\sin 3\theta - \dots$$

Since $(2\cot 2\theta\sin\theta)^n = \left(\frac{\cos 2\theta}{\cos\theta}\right)^n$, we obtain

$$\arctan(x) - \arctan\left(\frac{1}{x}\right) = \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{\cos 2\theta}{\cos\theta}\right)^n \sin n\theta.$$

As $x = \cot\theta = \tan(\frac{\pi}{2} - \theta)$ and $\frac{1}{x} = \frac{1}{\cot\theta} = \tan\theta,$ it follows that

$$\frac{\pi}{2} - 2\theta = \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{\cos 2\theta}{\cos\theta}\right)^n \sin n\theta, \quad 0 < \theta < \frac{\pi}{2}. \quad (3.4)$$

Observe that

$$\begin{aligned} x-\sqrt{1+x^2} &= \frac{\cos\theta}{\sin\theta}-\frac{1}{\sin\theta}=\frac{\cos\theta-1}{\sin\theta} \\ &= \frac{-2\sin^2(\frac{\theta}{2})}{2\sin(\frac{\theta}{2})\cos(\frac{\theta}{2})}=-\tan\frac{\theta}{2}=\tan\frac{-\theta}{2}, \end{aligned}$$

and similarly, $x+\sqrt{1+x^2}=\cot\frac{\theta}{2}=\tan(\frac{\pi}{2}-\frac{\theta}{2})$.
Hence, putting $h=-\sqrt{1+x^2}$ in (2.1), and cancelling $(\sin\theta)^n$, we get

$$\frac{\pi}{2}-\frac{\theta}{2}=\sum_{n=1}^{\infty}\frac{\sin n\theta}{n}, \quad 0<\theta<2\pi. \tag{3.5}$$

We find this formula in §166, Chapter 6, Part II of Euler's *Foundations of Differential Calculus*[11].
Similarly putting $h=\sqrt{1+x^2}$ in (2.1), we obtain

$$\frac{\theta}{2}=\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\sin n\theta}{n}, \quad -\pi<\theta<\pi. \tag{3.6}$$

Since $\sin m\pi=0$, $\cos(2m+1)\pi=-1$ and $\cos(2m)\pi=1$ for integer m, we can derive (3.6) from (3.5) by simply replacing θ by $\pi-\theta$.

Adding the equation (3.5) and (3.6) yields

$$\frac{\pi}{4}=\sum_{n=1}^{\infty}\frac{\sin(2n-1)\theta}{2n-1}, \quad 0<\theta\le\frac{\pi}{2}. \tag{3.7}$$

It may be pointed out that popular calculus texts such as [8] use Fourier series to get the expansions (3.6) and (3.7). For example, for (2π-periodic) odd function $f(x)$, the French mathematician Jean J. B. Fourier (1768–1830) considered

$$f(x)=b_1\sin x+b_2\sin 2x+b_3\sin 3x+\cdots=\sum_{k=1}^{\infty}b_k\sin kx,$$

and to find the coefficients b_n, he multiplied both sides of the last expression by $\sin nx$ and then integrated over $[0,\,\pi]$ to obtain

$$\int_0^{\pi}f(x)\sin nx\,dx=\sum_{k=1}^{\infty}b_k\int_0^{\pi}\sin kx\,\sin nx\,dx.$$

Since

$$\int_0^{\pi}\sin kx\,\sin nx\,dx=0 \text{ for } k\neq n \text{ and } =\frac{\pi}{2} \text{ for } k=n,$$

$$b_k=\frac{2}{\pi}\int_0^{\pi}f(x)\,\sin nx\,dx.$$

By using a similar analysis, he showed that one can also expand any (periodic) even function, whatever, into a series of cosines of multiple arcs. In fact, he asserted in his work on the propagation of heat that one can represent "arbitrary" functions (not only odd or even functions) of a variable into series of sines and cosines of multiple arcs.

If we set $h = -2x + \sqrt{1+x^2}$, and $h = -2x - \sqrt{1+x^2}$, in (2.1), we get

$$\frac{\pi}{2} - \frac{3\theta}{2} = \sum_{n=1}^{\infty} \frac{1}{n} (2\cos\theta - 1)^n \sin n\theta, \qquad 0 \le \theta \le \frac{\pi}{2}. \tag{3.8}$$

$$\pi - \frac{3\theta}{2} = \sum_{n=1}^{\infty} \frac{1}{n} (2\cos\theta + 1)^n \sin n\theta, \qquad \frac{\pi}{2} \le \theta \le \pi. \tag{3.9}$$

4. BBP-type Formulas for π

We shall now use the expansions to derive formulas for π.

On setting $\theta = \frac{\pi}{4}$ in (3.2) and on putting the values $\cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$, $\sin\frac{\pi}{4}(4n) = 0$, $\sin\frac{\pi}{4}(8n-6) = 1$, $\sin\frac{\pi}{4}(8n-2) = -1$, $\sin\frac{\pi}{4}(8n-7) = \sin\frac{\pi}{4}(8n-5) = \frac{1}{\sqrt{2}}$, $\sin\frac{\pi}{4}(8n-5) = \sin\frac{\pi}{4}(8n-3) = -\frac{1}{\sqrt{2}}$, we get following alternating series with three positive terms and three negative terms as every fourth term vanishes

$$\begin{aligned} \frac{\pi}{4} &= \left(\frac{1}{2}\right)^1 + \frac{1}{2}\left(\frac{1}{2}\right)^1 + \frac{1}{3}\left(\frac{1}{2}\right)^2 - \frac{1}{5}\left(\frac{1}{2}\right)^3 - \frac{1}{6}\left(\frac{1}{2}\right)^3 - \frac{1}{7}\left(\frac{1}{2}\right)^4 + \dots \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{4^n} \left(\frac{2}{4n-3} + \frac{2}{4n-2} + \frac{1}{4n-1} \right), \end{aligned}$$

which, on shifting the summation index, becomes

$$\pi = \sum_{n=0}^{\infty} \left(\frac{-1}{4}\right)^n \left(\frac{2}{4n+1} + \frac{2}{4n+2} + \frac{1}{4n+3} \right). \tag{4.1}$$

Using the standard notation, it becomes $\pi = P(1, -4, 4, (2, 2, 1, 0))$. This can be transformed, by combining two consecutive terms into one, to $\pi = \frac{1}{4} P(1, 16, 8, (8, 8, 4, 0, -2, -2, -1, 0)$, or equivalently to,

$$\pi = 4 \sum_{n=0}^{\infty} \frac{30720n^5 + 90368n^4 + 100064n^3 + 51292n^2 + 11905n + 981}{16^n (8n+1)(8n+2)(8n+3)(8n+5)(8n+6)(8n+7)}.$$

BBP-type formulas can be obtained by assigning Gaussian rational values to the complex number z in the following power series expansion

$$\ln \frac{1}{1-z} = \sum_{n=1}^{\infty} \frac{z^n}{n}, \qquad |z| < 1.$$

For example, formula (4.1) can be derived by setting $z = \frac{1+i}{2}$ and taking imaginary parts.

Further, a BBP-type formula can also be obtained by computing integrals obtained in the following manner, where $|\frac{1}{b}| < 1$, and if permissible swapping of sigma and integral signs is resorted to.

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{b^n(kn+c)} &= b^{c/k}\sum_{n=0}^{\infty}\frac{1}{kn+c}\left(\frac{1}{b^{1/k}}\right)^{kn+c} = b^{c/k}\sum_{n=0}^{\infty}\left[\frac{x^{kn+c}}{kn+c}\right]_0^{\frac{1}{b^{1/k}}}\\ &= b^{c/k}\sum_{n=0}^{\infty}\int_0^{\frac{1}{b^{1/k}}} x^{kn+c-1}\,dx = b^{c/k}\int_0^{\frac{1}{b^{1/k}}}\sum_{n=0}^{\infty}(x^k)^n\,x^{c-1}\,dx\\ &= b^{c/k}\int_0^{\frac{1}{b^{1/k}}}\frac{x^{c-1}}{1-x^k}\,dx.\end{aligned}$$

Thus on putting $b^{1/k}x = y$, we obtain

$$\sum_{n=0}^{\infty}\frac{1}{b^n(kn+c)} = b\int_0^1 \frac{y^{c-1}}{b-y^k}\,dy.$$

This procedure, adapted from [1, 2, 7], has been used by S. K. Lucas to derive such formulas. Adamchik and Wagon derive, on page 854 in [2], an integral $g(i)$ for the case $k = 4$, taking variable z and constant i in place of c. On setting $z = \frac{x}{\sqrt{2}}$ and replacing i by c, their integral becomes

$$\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{4n+c} = 4\int_0^1\frac{x^{c-1}}{4+x^4}\,dx.$$

We find in a standard table of integrals [10, 27]

$$\begin{aligned}\int\frac{dx}{a^4+x^4} &= \frac{1}{4a^3\sqrt{2}}\ln\frac{x^2+ax\sqrt{2}+a^2}{x^2-ax\sqrt{2}+a^2}+\frac{1}{2a^3\sqrt{2}}\arctan\frac{ax\sqrt{2}}{a^2-x^2},\\ \int\frac{x\,dx}{a^4+x^4} &= \frac{1}{2a^2}\arctan\frac{x^2}{a^2},\\ \int\frac{x^2\,dx}{a^4+x^4} &= -\frac{1}{4a\sqrt{2}}\ln\frac{x^2+ax\sqrt{2}+a^2}{x^2-ax\sqrt{2}+a^2}+\frac{1}{2a\sqrt{2}}\arctan\frac{ax\sqrt{2}}{a^2-x^2}.\end{aligned}$$

Computing these integrals from 0 to 1 with $a = \sqrt{2}$, we obtain

$$\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\left(\frac{2}{4n+1}+\frac{2}{4n+2}+\frac{1}{4n+3}\right) = 2\left(\arctan 2+\arctan\frac{1}{2}\right) = \pi.$$

5. Formulae involving convergents of π

Further, we observe that

$$\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{4n+6} = -4\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{4n+2},$$

$$\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{4n+7} = -4\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{4n+3}.$$

Therefore it becomes clear that

$$\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\left(\frac{2}{4n+1}-\frac{6}{4n+2}+\frac{5}{4n+3}-\frac{2}{4n+6}+\frac{1}{4n+7}\right)$$
$$=\frac{2}{3}+\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\left(\frac{2}{4n+1}+\frac{2}{4n+2}+\frac{1}{4n+3}\right)=\pi-\frac{8}{3}.$$

Hence,

$$\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\left(\frac{2}{4n+1}-\frac{6}{4n+2}+\frac{5}{4n+3}-\frac{2}{4n+6}+\frac{1}{4n+7}\right)$$
$$=\pi-\frac{8}{3}-\left(2-3+\frac{5}{3}-\frac{1}{3}+\frac{1}{7}\right)=\pi-\frac{22}{7}.$$

Luckily, collecting all the terms into one leaves only a constant in the numerator yielding the following neat formula for the second convergent of π

$$\pi=\frac{22}{7}+120\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{(4n+1)(4n+2)(4n+3)(4n+6)(4n+7)},$$

which can be transformed into following series with only positive terms.

$$\pi=\frac{22}{7}-15\sum_{n=1}^{\infty}\frac{768n^3+1984n^2+836n+87}{\left(\begin{array}{c}16^{n-1}(8n-3)(8n-2)(8n-1)(8n+1)\\(8n+2)(8n+3)(8n+6)(8n+7)\end{array}\right)}. \tag{5.1}$$

Let me touch upon the formulae involving the convergents of π. Dalzell [9] found in 1944 the following integral and series

$$\pi=\frac{22}{7}-\int_0^1\frac{t^4(1-t)^4}{1+t^2}. \tag{5.2}$$

$$\pi=\frac{22}{7}+\sum_{n=1}^{\infty}\left(\frac{-1}{4}\right)^n\left[\frac{3(4n)!^2}{(8n+1)!}+\frac{(4n+1)!^2}{(8n+3)!}-\frac{1}{2}\frac{(4n+2)!^2}{(8n+5)!}-\frac{(4n+3)!^2}{(8n+7)!}\right]. \tag{5.3}$$

By partial fractions decomposition and simplification, I was able to transform it into a 'cousin' of BBP-type formulas with a rapidly rising additional factor in the denominator

$$\pi=\frac{1}{4^5}\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\frac{1}{\binom{8n}{4n}}\left(\frac{3183}{8n+1}+\frac{117}{8n+3}-\frac{15}{8n+5}-\frac{5}{8n+7}\right) \tag{5.4}$$

which may be written in the form

$$\frac{\pi}{2}=\sum_{n=0}^{\infty}\left(\frac{-1}{4}\right)^n\left[\frac{820n^3+1533n^2+902n+165}{\binom{8n}{4n}(8n+1)(8n+3)(8n+5)(8n+7)}\right].$$

or

$$\pi=\frac{22}{7}-$$
$$\sum_{n=1}^{\infty}\frac{P(n)}{n(4n-1)(8n-1)(8n-3)(16n+1)(16n+3)(16n+5)(16n+7)}\frac{1}{2^{4n-1}\binom{16n}{8n}}, \tag{5.5}$$

$$\begin{aligned}P(n) &= 1717985280n^7 + 747324416n^6 - 238713984n^5 - 103697680n^4 \\ &\quad + 6801420n^3 + 2995544n^2 - 21201n - 10080.\end{aligned}$$

We do not know whether such 'natural' formulas exist for the further convergents of π, but I could derive this 'artificial' formula:

$$\pi = \frac{355}{113} - \frac{1}{8 \cdot 19 \cdot 113} \sum_{n=2}^{\infty} \frac{(9546n^3 - 147727n^2 - 55625n + 409063)}{(2n+1)(2n+2)(2n+3)(2n+4)(2n+5)} \frac{\binom{2n}{n}}{2^{4n}} \quad (5.6)$$

and some more formulas like this for $\frac{22}{7}$

$$\pi = \frac{22}{7} - \frac{1}{7} \sum_{n=1}^{\infty} \frac{n(102n+231)\binom{2n}{n}}{(2n+1)(2n+2)(2n+3)(2n+4)(2n+5)(2n+6)2^{4n-2}}. \quad (5.7)$$

6. BBP-type Formulas for $\pi\sqrt{3}$

If we set $\theta = \frac{\pi}{3}$ in the expansion in (3.2), we get following series with two positive terms followed by two negative terms as every third term vanishes.

$$\pi\sqrt{3} = 18 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{8^n} \left(\frac{2}{3n-2} + \frac{1}{3n-1} \right),$$

which, on shifting the summation index, becomes

$$\pi\sqrt{3} = \frac{9}{4} \sum_{n=0}^{\infty} \left(\frac{-1}{8} \right)^n \left(\frac{2}{3n+1} + \frac{1}{3n+2} \right). \quad (6.1)$$

Using the standard notation, it becomes $\pi\sqrt{3} = \frac{9}{4} P(1, -8, 3, (2, 1, 0))$, which can be transformed into $\pi\sqrt{3} = \frac{9}{32} P(1, 64, 6, (16, 8, 0, -2, -1, 0))$, or equivalently,

$$\pi\sqrt{3} = \frac{9}{8} \sum_{n=0}^{\infty} \frac{1134n^3 + 2097n^2 + 1188n + 193}{64^n(6n+1)(6n+2)(6n+4)(6n+5)}.$$

Formula (4.1), discovered by Adamchik and Wagon [1, 2], finds place as entry (16) with associated entry (15) in the compendium of such formulas [5] while (6.1) in the associated form is entry (18).

If we put $\theta = \frac{\pi}{6}$ in (3.2), we obtain a sluggish alternating series having five positive terms followed by five negative terms as every sixth term vanishes and we get following formula with non-integral base.

$$\pi = \frac{3\sqrt{3}}{64} \sum_{n=0}^{\infty} \left(\frac{-27}{64} \right)^n \left[\frac{16}{6n+1} + \frac{24}{6n+2} + \frac{24}{6n+3} + \frac{18}{6n+4} + \frac{9}{6n+5} \right]. \quad (6.2)$$

We now deduce some formulas which have unit-base though Bailey's compendium includes only those formulas with $b > 1$ for exponential rather than linear rate of convergence.

Putting $\theta = \frac{\pi}{3}$ in (3.4) in one obtains

$$\pi = 3\sqrt{3} \left[\sum_{n=1}^{\infty} \left(\frac{1}{3n-2} - \frac{1}{3n-1} \right) \right]. \quad (6.3)$$

Putting $\theta = \frac{\pi}{2}$ and $\theta = \frac{\pi}{3}$ in (3.5), we get the celebrated Madhava-Leibnitz-Gregory series

$$\frac{\pi}{4} = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{2n-1}. \tag{6.4}$$

$$\pi = \frac{3\sqrt{3}}{2}\left[\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{3n+1}+\frac{1}{3n+2}\right)\right]. \tag{6.5}$$

Putting $\theta = \frac{\pi}{3}$, $\theta = \frac{\pi}{4}$ and $\theta = \frac{\pi}{6}$ in (3.7), we get

$$\pi = 2\sqrt{3}\left[\sum_{n=0}^{\infty}\left(\frac{1}{6n+1}-\frac{1}{6n+5}\right)\right]. \tag{6.6}$$

$$\pi = 2\sqrt{2}\left[\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{4n+1}+\frac{1}{4n+3}\right)\right]. \tag{6.7}$$

$$\pi = 2\left[\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{6n+1}+\frac{2}{6n+3}+\frac{1}{6n+5}\right)\right]. \tag{6.8}$$

If we add the equations (6.4) and (6.7), we obtain

$$\pi = 8(\sqrt{2}-1)\left[\sum_{n=0}^{\infty}\left(\frac{1}{8n+1}-\frac{1}{8n+7}\right)\right]. \tag{6.9}$$

Formulae (6.3) and (6.6) appear in §176 and §177 respectively of Chapter X, Vol. I of Euler's *Analysis*[12]. Formulas (6.7) and (6.9) occur in §179 there.

Putting $\theta = \frac{\pi}{4}$ in (3.8) or $\theta = \frac{3\pi}{4}$ in (3.9) yields following interesting formula.

$$\frac{\pi}{4} = \sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\sqrt{2})^{4n+2}}\left(\frac{2+\sqrt{2}}{4n+1}+\frac{2}{4n+2}+\frac{2-\sqrt{2}}{4n+3}\right). \tag{6.10}$$

Finally, we give following efficient alternating *base-3* or ternary formula deduced by putting $\theta = \frac{\pi}{6}$ in (3.4).

$$\pi\sqrt{3} = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{-1}{27}\right)^n\left(\frac{9}{6n+1}+\frac{9}{6n+2}+\frac{6}{6n+3}+\frac{3}{6n+4}+\frac{1}{6n+5}\right). \tag{6.11}$$

In the standard notation, this is $\pi\sqrt{3} = \frac{1}{3}P(1,-27,6,\,(9,9,6,3,1,0))$. The alternating formula (6.11) transforms into
$\pi\sqrt{3} = \frac{1}{81}P(1,3^6,12,\,(243,243,162,81,27,0,-9,-9,-6,-3,-1,0))$.

Neither of the two forms is in Bailey's compendium which has entry (66) as $\pi\sqrt{3} = \frac{1}{9}P(1,3^6,12,\,(81,-54,0,-9,0,-12,-3,-2,0,-1,0,0))$.

Formula (6.11) can be written as

$$\pi\sqrt{3} = \frac{16}{9}\sum_{m=0}^{\infty} \frac{\begin{pmatrix} 26085556224m^9 + 126116020224m^8 \\ +262617292800m^7 + 307980610560m^6 \\ +223165269504m^5 + 103065777216m^4 \\ +30145041120m^3 + 5342703940m^2 \\ +515537612m + 20359495 \end{pmatrix}}{\begin{pmatrix} 729^n(12m+1)(12m+2)(12m+3) \\ (12m+4)(12m+5)(12m+7)(12m+8) \\ (12m+9)(12m+10)(12m+11) \end{pmatrix}}$$

To compute the five relevant integrals, we can use the following general formula for $m < 2n$ given in [13]

$$\int \frac{x^{m-1}dx}{1+x^{2n}} = -\frac{1}{2n}\sum_{k=1}^{n} \cos\frac{m\pi(2k-1)}{2n} \ln(1 - 2x\cos\frac{\pi(2k-1)}{2n} + x^2)$$
$$+ \frac{1}{n}\sum_{k=1}^{n} \sin\frac{m\pi(2k-1)}{2n} \arctan\frac{x - \cos\frac{\pi(2k-1)}{2n}}{\sin\frac{\pi(2k-1)}{2n}}.$$

Motivated by the formula (6.11), S. K. Lucas discovered following formula with a general parameter e.

$$\pi = \sum_{n=0}^{\infty}\left(\frac{-1}{27}\right)^n \left(\frac{9e}{6n+1} + \frac{2\sqrt{3}-9e}{6n+2} + \frac{2\sqrt{3}-12e}{6n+3} + \frac{2\sqrt{3}-9e}{3(6n+4)} + \frac{e}{6n+5}\right).$$

The choices of $e = 2\sqrt{3}/9$ and $e = \sqrt{3}/6$ eliminate fractions, giving

$$\pi = \frac{2\sqrt{3}}{9}\sum_{n=0}^{\infty}\left(\frac{-1}{27}\right)^n \left(\frac{9}{6n+1} - \frac{3}{6n+3} + \frac{1}{6n+5}\right) = 2\sqrt{3}\sum_{n=0}^{\infty}\frac{(-1)^n}{3^n(2n+1)},$$

which is Abraham Sharp's formula deduced by putting $x = \frac{1}{\sqrt{3}}$ in the Maclaurin's series for $\arctan x$. In fact on combining two consecutive terms, Sharp's series leads to following another BBP-type of formula.

$$\pi\sqrt{3} = 2\sum_{n=0}^{\infty}\left(\frac{1}{9}\right)^n \left(\frac{3}{4n+1} - \frac{1}{4n+3}\right).$$

Two infinite classes of formulas like that of Sharp have been derived by the author in his recent paper[21].

More formulas can be deduced by combining the various results. I am giving only one of these – a 'balanced formula', sum of whose coefficients is zero.
$\pi\sqrt{3} = P(1, -3^3, 6, (9, -3, -6, -1, 1, 0))$,
$\pi\sqrt{3} = \frac{1}{27}P(1, 3^6, 12, (243, -81, -162, -27, 27, 0, -9, 3, 6, 1, -1, 0))$.
This may be combined with the one given in the compendium to deduce
$\pi\sqrt{3} = \frac{1}{27}P(1, 3^6, 12, (243, 0, -324, -27, 54, 36, -9, 12, 12, 5, -2, 0))$.

Bailey records this formula for $\log 7$ as the entry (70) in his compendium $\ln 7 = \frac{2}{3^5} P(1, 3^6, 6, (405, 81, 72, 9, 5, 0))$.

We discovered this pair of formulas using the integrals approach

$$\ln 7 = \frac{2}{3^2} P(1, -3^3, 6(9, 0, 0, 0, -1, 0)), \tag{6.12}$$

$$\ln 7 = \frac{2}{3^5} P(1, 3^6, 12(243, 0, 0, 0, -27, 0, -9, 0, 0, 0, 1, 0)). \tag{6.13}$$

In sigma notation, our 4-term formula is

$$\ln 7 = \frac{2}{243} \sum_{n=0}^{\infty} \frac{1}{729^n} \left(\frac{243}{12n+1} - \frac{27}{12n+5} - \frac{9}{12n+7} + \frac{1}{12n+11} \right).$$

We may mention that the Japanese team led by Y. Kanada of Tokyo University used Machin like arctangent identities for computing π. But while giving similar exponential convergence, these formulae mix bases and thus are not as good as Ramanujan type series or BBP-type formulae.

Our discovery of these BBP-type formulas shows that results *can* be discovered *mathematically* which a sophisticated software may fail to find. However, given the advantage that a computer has in terms of speed, memory and precision in computation, there is no denying the fact that advanced computing technology, in the shape of software such as *Mathematica* and *Maple*, can be used as an adjunct in mathematical research for discovering new mathematical results as well as for validation of empirically obtained conjectures. The author himself used a computer program to discover Machin-like arctangent identities for π[20].

Acknowledgement: The author wishes to thank Prof. S. K. Lucas of James Madison University, Harrisonburg, Virginia (US) for sharing his results with the author. Prof. Paul Levrie, University of Antwerp, Belgium deserves thank for providing computational support as well as for giving helpful suggestions. The author is also grateful to the anonymous referee for some valuable suggestions to improve the write-up.

References

[1] Adamchik, Victor and Wagon, Stan, *π: A 2000-Year Search Changes Direction,* Mathematica in Education and Research, **5**(1) (1996) 11-19. Available at `https://www.cs.cmu.edu/ adamchik/articles/pi/pi.pdf`

[2] Adamchik, Victor and Wagon, Stan, *A Simple Formula for* π, American Mathematical Monthly, **104** (November, 1997), Notes, 852–855.

[3] Bailey, David H., Borwein, Peter B. and Plouffe, Simon, *On The Rapid Computation of Various Polylogarithmic Constants,* Mathematics of Computation, **66** (218) 1997, 903–913.

[4] Bailey, David H., The BBP Algorithm for Pi David H. Bailey 8 Sept 2006. Available at `http://www.experimentalmath.info/bbp-codes/bbp-alg.pdf`

[5] Bailey, David H., *A Compendium of BBP-Type Formulas for Mathematical Constants,* 2013, at `http://crd-legacy.lbl.gov/~dhbailey/ dhbpapers/bbp-formulas.pdf`

[6] Ball, W. W. Rouse, *A Short Account of the History of Mathematics,* Dover, New York, 1960 (originally published, 1908), 258–259.

[7] Borwein, Jonathan M., Bailey David H. & Girgensohn, Roland, *Mathematics by Experiment: Plausible Reasoning in the 21st Century and Experiments in Mathematics: Computational Paths to Discovery,* 2003. Excerpts. pp.24-25 . Available at textttthttp://oldweb.cecm.sfu.ca/Preprints03/2003-212.pdf

[8] Bradley, Gerald L. and Smith, Karl J., *Single Variable Calculus,* Prentice Hall, 2nd ed. 1995, pp.695–696, Technology Window.

[9] Dalzell, D. P. *On* $\frac{22}{7}$*,* J. London Math. Soc. **19** (1944), 133–134.

[10] Dwight, Herbert Bristol, *Tables Of Integrals And Other Mathematical Data,* 3rd ed, 1957, New York: The Macmillan Company, p.38.

[11] Euler, L., Institutiones calculi differentialis cum eius usu in analysi finitorum ac doctrina serierum, Part II, Chapter 6. Originally published in 1755. Availabe as E212 at `http://eulerarchive.maa.org/`

[12] Euler, L., *Introductio in analysin infinitorum,* Lausannae: Apud Marcum-Michaelem Bousquet & socios., 1748. E101 and E102 at `http://eulerarchive.maa.org/`. An english translation of both volumes by John D. Blanton published by Springer-Verlag. Each chapter of vol.I translated and annotated by Ian Bruce is available at `http://www.17centurymaths.com/contents/introductiontoanalysisvol1.htm`

[13] Gradshteyn, I. S. and Ryzhik, I. M., *Table of Integrals, Series, and Products,*, 7th ed., 2007, Elsevier, p.76, 2.146.

[14] Gregory, Duncan Farquharson, *Examples of the Processes of the Differential and Integral Calculus,* J. & J. J. Deighton, Cambridge and John W. Parker, London, 1841, p.52.

[15] Gupta, Radha Charan, *The Mādhava-Gregory Series,* The Mathematics Education, Vol. **VII**, No. 3, Sept., 1973, 67–70.

[16] Haddon, James,*Examples and Solutions in the Differential Calculus,* John Weale, London, 1851, 18–19.

[17] Joseph, George Gheverghese, The Crest of the Peacock: Non-European Roots of Mathematics, 3rd ed., Princeton University Press, 2011, 428–435.

[18] MacLaurin, Colin, *A Treatise of Fluxions in two books,* T. W. and T. Ruddimans, Edinburgh, 1742. Second edition 1801.

[19] De Morgan, Augustus, *Differential and Integral Calculus,* London: Baldwin and Cradock, 1836, p.74, footnote.

[20] Nimbran, A. S., *On the derivation of Machin-like Arctangent identities for computing Pi,* The Mathematics Student **79** Nos. 1–4, (2010), 171–186.

[21] Nimbran, A. S., *Enhancing the Convergence of Ramanujan's Series and Deriving Similar Series for Catalan's Constant and* $\pi\sqrt{3}$, (to appear in) Mathematics Student **84** Nos. 3-4, July-December (2015).

[22] Roy, Ranjan, *The Discovery of the Series Formula for* π *by Leibniz, Gregory and Nilakantha,* Mathematics Magazine **63** No. 5 (Dec., 1990), 291–306.

[23] Stirling, James, *Methodus Differentialis: Sive Tractatus de Summatione Et Interpolatione Serierum Infinitarum,* Londini : G. Bowyer, Impensis G. Strahan, 1730.

[24] Taylor, Brook, *Methodus incrementorum directa & inversa,* Impensis Gulielmi Innys, 2nd ed., 1717, p.23.

[25] Todhunter, Isaac, *A Treatise on the Differential Calculus With Numerous Examples,* Macmillan, London, 1890, p.97.

[26] Tweedie, Charles, *James Stirling: A sketch of his life and works along with his scientific correspondence,* Oxford, London, 1922, pp. 40, 45–46, 59.

[27] Zwillinger, Daniel, *CRC Standard Mathematical Tables and Formulae,* 31st ed., 2003, CRC Press, section 5.4.7.

Amrik Singh Nimbran
Former Director General of Police, Bihar, Patna, India
Email: *amrikn622@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 53-68

WHAT ALL CAN YOU HEAR FROM A DRUM?

BINOY

(Received : 14 - 03 - 2015 ; Revised : 15 - 04 - 2015)

Abstract. This article surveys some results about the spectrum of Laplacian on bounded planar domains.

Sound, which is one of the primary modes of communication, which is varied in an abundant way in nature has always intrigued humans. Even in the ancient Bronze age, people performed experiments and built different shaped drums to understand, experience and to indulge in the sensory delight induced by this simple physical phenomenon. As our knowledge and perception of natural phenomena is evolving in parallel with dynamically growing modern science, our understanding of principles behind sound is also advancing. An elementary observation suggests that sound is produced when materials vibrate, like the vibrations of a string or a membrane whose boundaries are fixed. German researcher Ernst Chladni was one of the first who started a systematic inquiry of sound by studying the vibrating plates in 18th century. His experimental model, which is now referred as Chladni's Plate consists of a metal plate fixed in its middle. When the plate is struck, it vibrates and produces the sound. Performing the experiment by pouring a small amount of sand on the plate, the wave nature of the sound can be physically observed [see figure 1] as the sand accumulates only at those parts of the plate which do not oscillate. When the plate is struck differently not only that different sound is heard but the pattern formed by sand also changes; which suggests its close relation with the sound produced. Its no wonder that in a curious mind this

Figure 1. A pattern on a vibrating Chladni's plate.

2010 Mathematics Subject Classification: Primary 35-02 ; Secondary 57-02
Keywords and Phrases: Eigenvalues of Laplacian, Shape.

simple experiment raises several questions like how exactly this pattern is related to the sound? How are the shape and dimensions of the plate related to the sound we hear and the pattern formed by sand? Is there a unique way in which the pattern of the sand, the shape of the plate and the sound produced are related to each other?

Physical aspect of the study of sound mainly concerns its wave nature. This wave phenomenon of a vibrating membrane or a string is very general in the sense that the underlying mathematical principle forms the basis for several other natural phenomena. General laws of physics related to wave motions reduce this study into analyzing the following mathematical equation called wave equation

$$\frac{\partial^2 F}{\partial t^2} = c\,\Delta F \tag{0.1}$$

where F is a function of space and time variables, c is a real number and Δ is the Laplacian on the space variables. It will be redundant here to explain the process by which physicists reduce a physical phenomenon into abstract equations as any relevant physics book would explain them [5]. Still it is worth mentioning that in its general form wave equation stands at the core of several fields of physics such as hydrodynamics, heat propagation, quantum mechanics etc. To see how helpful this equation is in the study of sound, the best way will be to begin with the simple case of a vibrating string.

1. Strings

Consider a guitar, arguably the most popular musical instrument in the world. The 6 strings on which guitarist brushes his or her fingernail to produce chords are attached firmly at both ends. Guitarist produces different notes on a single string by changing the length of that part of the string which vibrates. It is these notes which are combined in artistic manner by a musician that gives the experience of contentment to a music lover. If the element of sensory pleasure is set aside for a moment from the notes produced by a vibrating string, the relationship between the notes and the length of the string is immediate. It is a stodgy but unavoidable fact that an entry into the the world of abstractness by getting out of the realm of music is necessary to understand this relation better.

So, consider a string of length l with its ends fixed. To simplify the scenario assume that it has a fixed thickness and is made up of a material of uniform density. This string can be modeled as the interval $[0, l]$ where the fixed ends correspond to the points 0 and l. For $x \in [0, l]$ let $f(x, t)$ be the vertical displacement of the point x at time t. Since at the end points vibration seizes to exist, f must satisfy $f(0, t) = f(l, t) = 0$ for all t, which are called as the boundary conditions. The wave equation (0.1)[1] takes the following form in this case:

[1]Without loss of generality, the constant c appearing in the wave equation will be set to 1 throughout this article.

$$\frac{\partial^2 f}{\partial t^2}(x,t) = \frac{\partial^2 f}{\partial x^2}(x,t). \tag{1.1}$$

A careful observation of the waves formed when the string is struck shows that there is a particular type of wave which has a fixed underlying shape but its amplitude changes with time. These waves, which are called standing waves, have the form $f(x,t) = g(x)h(t)$ where g represents the basic shape and h, the amplitude. For this form of waves, the equation (1.1) takes the following form[2]

$$g(x)h''(t) = g''(x)h(t)$$

with the corresponding boundary conditions $g(0) = g(l) = 0$. The following equivalent formulation of above equation

$$g''(x) + \lambda g(x) = 0 \tag{1.2}$$

$$h''(t) + \lambda h(t) = 0 \tag{1.3}$$

immediately gives the solutions as

$$g(x) = A\sin(\sqrt{\lambda}\,x) + B\cos(\sqrt{\lambda}\,x),$$

and

$$h(t) = C\sin(\sqrt{\lambda}\,t) + D\cos(\sqrt{\lambda}\,t),$$

where A, B, C, D are real constants and $\lambda > 0$. The boundary conditions on g, which are used to obtain $\lambda > 0$, further imply that $g(x) = A\sin(\sqrt{\lambda}\,x)$ and $\lambda = \frac{n^2\pi^2}{l^2}$. Notice that h represents the amplitude of the wave and hence the constants C and D are determined by the initial state of the string and hence by the force with which it is struck. Since the frequency of $\sin(\alpha\,t)$ or $\cos(\alpha\,t)$ is $\frac{\alpha}{2\pi}$, the standing waves oscillate with a frequency of $\frac{1}{2l}, \frac{2}{2l}, \frac{3}{2l}, \cdots$. Or in other words, when struck to create a standing wave, the string vibrates with one of the frequencies among $\frac{1}{2l}, \frac{2}{2l}, \frac{3}{2l}, \cdots$. For a musician these are "overtone series", in particular $\frac{1}{2l}$ is the fundamental tone or the pitch of the string.

The above analysis, which employs only elementary mathematical tools, brings out a remarkable property; by hearing the pitch of a vibrating string which is made up of a material of uniform density and has a given thickness, it is possible to know the length of the string. Said in poetical terms; "one can hear the length of a string!" As length determines the frequency of vibration of the string completely, this demonstrates the fact that there is a unique way in which vibrating frequencies and strings are related. Also this prompts to ask the question: What about drums? What all can be heard from a drum?

2. Drums

For a music enthusiast, it is rather easy to recognize the notes produced by different drums in a musical concert. As in the case of guitar, when the drumhead - the tightly stretched membrane over the opening of the drum, which are usually in circular shape - is struck, it causes vibrations of various frequencies. Clearly, the size/volume of the shell - the main cylindrical body - is a major factor which

[2]This is nothing but the familiar separation of variable technique!

determines the sound produced by a drum(ignoring the fact that the material also matters!). A not so obvious fact would be to distinguish the notes produced by two drums which have the same shell volume but with different shaped drumhead, say one circular and other triangular. In fact this illustrates the following profound question which is being investigated by mathematicians for a couple of centuries.

What geometrical properties of a vibrating membrane, like perimeter, area, shape etc. are determined by the tunes it produce?

To model this problem mathematically, the stretched membrane is identified with a domain in plane which is denoted by Ω, and assume that its boundary is fixed along a simple closed curve Γ. Ignoring the material properties like density, elasticity[3] etc., the vibrating motion of the membrane can be seen to satisfy the wave equation (0.1). Similar to the one dimensional case of strings, in this case also the standing waves produced by the membrane has the form $F(x, y, t) = g(x, y)h(t)$; which upon substitution in (0.1) gives the following set of equations:

$$\begin{aligned} \Delta g(x,y) + \lambda g(x,y) = 0 \quad \text{for } (x,y) \in \Omega \\ g(x,y) = 0 \quad \text{for } (x,y) \in \Gamma \end{aligned} \tag{2.1}$$

and

$$h''(t) + \lambda h(t) = 0.$$

In contrast to equation (1.2), the above set of equations (2.1) is much harder to solve. In fact, to prove that there exist a discrete sequence of positive real numbers and corresponding functions which solve problems of type (2.1) was a major challenge in mathematical physics until early 20th century. Such a sequence of numbers are usually referred as the spectrum or eigenvalues and the associated functions as eigenfunctions corresponding to the problem. The non triviality and difficulties encountered by physicists and mathematicians of earlier period in dealing with this seemingly simple problem is well captured in the following remark made by British spectroscopist Sir Arthur Schuster in 1882; *"To find out the different tunes sent out by a vibrating system is a problem which may or may not be solvable in certain special cases, but it would baffle the most skillful mathematicians to solve the inverse problem and to find out the shape of a bell by means of the sounds which it is capable of sending out. And this is the problem which ultimately spectroscopy hopes to solve in the case of light. In the meantime we must welcome with delight even the smallest step in the desired direction."*

Thanks to the theory of integral equations developed in the early 20th century, the existence of spectrum, which is a discrete increasing sequence of positive real numbers, and associated eigenfunctions are guaranteed for problems of type (2.1) even in much more generality [15]. For such an existence, the domain and its boundary need to satisfy certain conditions of technical nature. To avoid an

[3]These properties determine the constant c in equation (0.1) which has been set to 1.

effortful journey through the technical jungles of abstract differential equation theories, the domain and boundary in problem (2.1) will be assumed to satisfy those conditions. This whole theory is not a superfluous addition to the mere pointless intellectual exercise of an affluent arm chair mathematician, as there are real examples to demonstrate it. Among the domains for which a complete description of spectrum is known, easiest is the case when Ω is a rectangle. Recalling that on planar domains $\Delta = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}$, the problem can be reduced into solving a pair of one dimensional equations like (1.2), if the function g is sought to behave separately in each variable, that is, if g has the form $g(x, y) = f(x)h(y)$. In that case the spectrum is obtained as the collection of numbers of the form $\pi^2(\frac{n^2}{a^2} + \frac{m^2}{b^2})$ with the corresponding combinations of trigonometric functions as eigenfunctions where a, b are the side lengths of the rectangle Ω and n, m are positive integers.

At this point, an inquisitive reader must be wondering; whether the spectrum obtained above for a string and rectangle consists of all possible eigenvalues? Because those were gotten by imposing extra conditions on eigenfunctions and there is no reason why a generic eigenfunction satisfies them. This is one of the places where the theory of infinite dimensional Hilbert spaces helps spectral analysis. In early 20th century, German mathematician David Hilbert formulated this theory which gives a general framework for lot of different areas of mathematics. In particular, modern theory of differential equations heavily relies on it. But how does the spectrum and eigenfunctions of problems (1.2) and (2.1) fit into the context of this general theory? A closer look at these problems is needed to see this. Since Laplacian is a linear operator, that is $\Delta(f + g) = \Delta f + \Delta g$ and $\Delta(cf) = c\Delta f$, where c is a constant, the finite sum and constant multiples of solutions are again a solution. In other words a new wave is formed when two standing waves are superimposed. This property qualifies the solution set of (1.2) and (2.1) to be called as a linear space, which together with a properly defined inner product (a notion of product of two function which gives a real number) constitutes an inner product space. The general Hilbert space theory applied to the solution set then guarantees the existence of special subsets of the whole solution space which can be used to represent a generic element. There is no better way to call these subsets as bases as everything else in the space can be built up with them. It turns out that in the case of string, collection of standing waves and in the case of rectangle, collection of eigenfunctions described above are bases and hence the spectrum obtained for them contains all possible eigenvalues!

Returning to the example of a rectangle, it is now clear that the spectrum determines the side lengths and hence the rectangle. Thus among the collection of rectangular drums, each drum produces distinct tones. Another illuminating feature of this example is that it sheds light on the link between the choice of coordinate systems and the geometry of the domain in hand. The fact that the perpendicular sides of a rectangle can be extended to give a coordinate system for the plane, with the corner as origin, is what lies behind the separation of variables technique used to find the quick solution. If the domain were a disc, this technique would not have worked! As the shape suggests, best choice would be a coordinate system which takes into account the rotational symmetry of the disc. For a disc it then becomes natural to choose the polar coordinate system centered at the center of the disc. Indeed when Ω is a disc of radius r such a choice reduces the problem (2.1) to solving a second order ordinary differential equation called Bessel's equation [2]. The theory of Bessel's equation is a well developed branch of mathematics and gives that the first non zero eigenvalue is a constant multiple of $\frac{1}{r^2}$. Hence among the collection of all discs, each disc has its own tones.

Evidently, the domains considered so far, rectangle and disc, are dealt with considerable easiness, but it would be a presumptuous attempt at this point to make any conjecture about spectrum, and its relation to a general domain and work towards it! In fact the enigmatic nature of mathematics reveals itself at one of its best ways when the domain is a triangle as all the elementary tools fail in this case. But, desperation turns into excitement when the versatile mathematician looks into the tool box and finds the proper tool to attack the problem, which in this case is the heat trace.

2.1. **Heated Drums!** Presume that a heat source is kept on the drumhead Ω at time $t = 0$ in such a way that the temperature on the boundary Γ is zero. Then the propagation of heat satisfies the heat equation

$$\begin{aligned} \frac{\partial F}{\partial t}(x,y,t) &= \Delta F(x,y,t) \quad \text{for } (x,y) \in \Omega \\ F(x,y,t) &= 0 \ \text{ for } (x,y) \in \Gamma \text{ and for all } t > 0. \end{aligned} \tag{2.2}$$

Here F represents the temperature at the point (x, y) in Ω at time t. Separating the variables, a solution of the form $F(x, y, t) = g(x, y)h(t)$ can be seen to satisfy

$$\begin{aligned} \Delta g(x,y) + \lambda g(x,y) &= 0 \quad \text{for } (x,y) \in \Omega \\ g(x,y) &= 0 \quad \text{for } (x,y) \in \Gamma \end{aligned} \tag{2.3}$$

and

$$h'(t) + \lambda h(t) = 0.$$

Even though heat dissipation and sound propagation are two different phenomena, both mathematically and physically, occurrence of same set of equations (2.1) and (2.3) in their mathematical models leads to a new way of dealing the problem. Notice that the equation which is dependent on time variable is of first order in this case, solution of which is $h(t) = e^{-\lambda t}$ where λ is one of the possible eigenvalues. As previously mentioned, the theory of integral equations assures the existence of the spectrum which can be arranged as $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots$ and the corresponding eigenfunctions $\phi_1, \phi_2, \phi_3, \cdots$ which forms a basis for the solutions space. Thus the function $e^{-\lambda_k t}\phi_k(x, y)$ is a solution of the heat equation for any k. Moreover, because the collection $\{\phi_k\}_{k=1}^{\infty}$ is a basis, a generic solution can be expressed as $F(x, y, t) = \sum_{k=1}^{\infty} a_k \, e^{-\lambda_k t} \phi_k(x, y)$ where a_k's are constants. The advantage of expressing the solution in this form is that it gives insights about qualitative properties of the heat diffusion. For example, the dominant term in the infinite sum involves only λ_1 and ϕ_1 as other terms wither away quickly with time. This suggests that some information about the distribution of heat, and hence about the shape, nearby the source for initial time t near 0, that is immediately after the source was placed on the drumhead, is encoded in λ_1 and ϕ_1. Considerations similar to this prompts the study of heat trace associated with the domain, which is given by

$$H(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t}, \quad t > 0. \tag{2.4}$$

An elaborate analysis [16], details of which involve interesting and sophisticated mathematical concepts, of heat trace for domains Ω with polygonal boundary Γ shows that

$$H(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} = \frac{b_1}{t} + \frac{b_2}{\sqrt{t}} + b_3 + O(e^{-\frac{c}{t}}) \text{ as } t \to 0$$

for some constant $c > 0$ where

$$b_1 = \frac{A}{4\pi}, \; b_2 = -\frac{L}{8\sqrt{\pi}}, \; b_3 = \frac{1}{24} \sum_i \left(\frac{\pi}{\theta_i} - \frac{\theta_i}{\pi} \right)$$

where $A = \text{Area}(\Omega), L = \text{Length}(\Gamma)$ and θ_i's are the interior angles of the polygon Γ. When Ω is a triangle, $\sum_i \theta_i = \pi$, hence $b_3 = \frac{1}{24} \sum_{i=1}^{3} \frac{1}{\theta_i} - \frac{1}{24}$. So far the analysis shows that area, perimeter, and the sum of reciprocals of angles of a triangle are spectral invariants, i.e., for triangles with identical spectrum, these quantities would be identical. But then, is a triangle determined uniquely by its area, perimeter and the sum of reciprocals of angles? An affirmative answer to this question would prove that like rectangle and disc, triangle is also determined by its spectrum among similar objects.

When it comes to deception it is doubtful that there is another subject which poses a threat to the pole position mathematics holds. It is full of such naive questions which stand unshattered in front of utmost cunning and intricate strategies by brilliant mathematicians for centuries. Weighed against the famous, now solved, problem like Fermat's last theorem which exemplifies the scenario of proportionality in simplicity and difficulty of a statement and its proof, the above mentioned question about triangle would feel very light. Still, its first proof appeared in [6] uses the wave trace, an equation similar to (2.4) for the wave equation, which is much harder to deal with than heat equation. At the same time its second proof [9] legitimizes the metaphysical principle that the beauty of mathematics lies in its simplicity as it uses only accessible elementary tools. This proof underlies on the simple fact that for a triangle

$$\frac{L^2}{4A} = \cot\frac{\alpha}{2} + \cot\frac{\beta}{2} + \cot\frac{\gamma}{2}.$$

It remains to prove that a triple (α, β, γ) of positive real numbers satisfying $\alpha + \beta + \gamma = \pi$ is uniquely determined by the values of $g(\alpha, \beta, \gamma) = \cot\frac{\alpha}{2} + \cot\frac{\beta}{2} + \cot\frac{\gamma}{2}$ and $h(\alpha, \beta, \gamma) = \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma}$, which is done by studying the convexity properties of these functions.

The heat trace gives another wonderful insight into the following related problem: *Suppose that two different drum heads have same area but different shapes. Then how are their tunes related?.* If two such drum heads are heated in a same way for the same amount of time, it is reasonable to expect that the heat should diffuse faster on the one which has bigger boundary (because on the boundary temperature should be zero). Instinctively a related question arise: *for a given area, which planar domain has the minimal perimeter?* A mere glimpse into nature is sufficient to guess the correct answer, in fact the fascinating view of colorful bubbles formed when a soap film is blown is just one among the ubiquitous ways in which nature exhibits the solution of this question. Minimization is one of the prevalent principles of nature and the spherical shape of soap bubbles is due to the fact that among all the domains of given volume, sphere has the minimum surface area. For the planar domain it would be the disc which has this property. By passing the rigorous proof of this statement, and recalling that λ_1 and ϕ_1 dominates the heat trace for large time, if a domain Ω has the same area as that of a disc B, then it is logical to conjecture that

$$\lambda_1(\Omega) \geq \lambda_1(B).$$

This is the famous Faber-Krahn inequality which was proved in 1923 [2]. It is one of the very few sharp inequalities related to spectrum in the sense that if $\text{Area}(\Omega) = \text{Area}(B)$ and $\lambda_1(\Omega) = \lambda_1(B)$, then Ω must be a disc.

Though area of disc and of a domain with polygonal boundary is a spectral invariant, as seen above using heat trace, so far it is not clear whether such a result holds for general domains. If such a result is true, then from the Faber-Krahn inequality it is clear that if a domain has the same spectrum as that of a disc, then it must be the disc. Which is to say; one can hear everything from a drum with circular shaped drumhead! A result to that effect can be proved by studying the heat trace, but this time its the end, not the beginning of the infinite series which comes under the scrutiny. Instead of excursing into the asymptotic world of heat trace to prove such a result, a more elementary approach will be worth exploring as it brings into light some interesting aspects of spectrum.

2.2. **Area.** In the second decade of 20th century, German mathematician Herman Weyl proved one of the earliest asymptotic formula for the spectrum. Inspired by a lecture delivered by the mathematical physicist H. A. Lorentz on black body radiation, Weyl proved the formula, which is referred as Weyl's law, using the techniques developed by Hilbert. Also his proof added a fresh chapter into the folklore of blissful ironies of mathematicians as Hilbert is said to have predicted that such a formula would not be proved in his life time, but within few months his student Weyl proved it!

For the problem (2.1) Weyl's formula takes the following form:

$$N(\lambda) \sim \frac{Area(\Omega)\lambda}{4\pi} \tag{2.5}$$

where for a real number λ, $N(\lambda)$ is the number of eigenvalues less than or equal to λ and the symbol $\sim$ means that the two quantities become equal when λ approaches infinity. An elementary proof of this rests on the answer to the following question: if Ω' is a smaller domain which is contained in Ω, then how are their eigenvalues related? More precisely, if the spectrum of Ω and Ω' are arranged as $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ and $0 < \lambda'_1 \leq \lambda'_2 \leq \cdots$ respectively, then for an integer k how are λ_k and λ'_k related? It is apparent that merely from its implicit appearance in (2.1) or from the formulas like heat trace which involves an infinite sum, the distinctive relations between eigenvalues would be hard to conclude. For such an understanding a more individualistic study of eigenvalues is needed. Earliest such attempts were made by Rayleigh in 1870's when his famous study of vibrating systems began. Rayleigh observed that the conservation of energy principle, when applied suitably for a vibrating system, the maximum of potential energy and kinetic energy of the system should be equal and it is possible to get the natural frequencies by taking the ratio of them [13]. It is easy to check this for the case of string as the solutions of wave equation are explicitly known. For the membrane, suppose that a function of the form $F(x,y,t) = g(x,y)h(t)$ solves the wave equation where g and h satisfies (2.1) for some λ. The velocity of this wave is obtained by taking the time derivative, $\frac{\partial F}{\partial t} = g(x,y)h'(t)$ and since h satisfies $h''(t) + \lambda h(t) = 0$, the maximum velocity is equal to $\sqrt{\lambda} g(x,y)$. So, the maximum kinetic energy of the wave is given by $\lambda \int_\Omega g^2$. Similarly, the maximum potential energy equals $\int_\Omega |\nabla g|^2$(constants are ignored in both cases without loss of generality) and hence

$$\lambda = \frac{\int_\Omega |\nabla g|^2}{\int_\Omega g^2}$$

where $|\nabla g| = \sqrt{g_x^2 + g_y^2}$ and g_x and g_y are partial derivatives of g with respect to x and y respectively. The ratio on RHS in above equation is known as Rayleigh quotient. Note that above formula can be used to find an eigenvalue if and only if the related eigenfunction is known. One of the interesting things Rayleigh proved was if the Rayleigh quotient is computed for a suitable function, then it can not be lower than fundamental frequency, ie; $\lambda_1 \leq \frac{\int_\Omega |\nabla f|^2}{\int_\Omega f^2}$, and in fact

$$\lambda_1 = \inf \frac{\int_\Omega |\nabla f|^2}{\int_\Omega f^2} \tag{2.6}$$

where the infimum is taken over all admissible functions, those f which vanish on boundary (to be precise, among all functions which are compactly supported inside the domain) and the equality is achieved if and only if f is an eigenfunction. This formula at once shows $\lambda_1 \leq \lambda_1'$, as any admissible function on Ω' can be extended to an admissible function on Ω by defining it to be zero in $\Omega \setminus \Omega'$.

Though it seems that this method works only for λ_1, Rayleigh quotients do exist for higher eigenvalues. A short detour into the familiar world of 2×2 matrices would clarify the essential ideas behind them.
Consider a 2×2 diagonal matrix,

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

with $0 < \lambda_1 \leq \lambda_2$. Clearly the eigenvalues of this matrix, which are the roots of its characteristic polynomial, are λ_1 and λ_2. This matrix viewed as a linear map on the plane takes a vector (x, y) into

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda_1 x \\ \lambda_2 y \end{bmatrix},$$

in particular $A((1, 0)) = \lambda_1(1, 0)$ and $A((0, 1)) = \lambda_2(0, 1)$, showing that $(1, 0)$ and $(0, 1)$ are the corresponding eigenvectors. The important observation to make is that the map A merely stretches the vectors $(1, 0)$ and $(0, 1)$ to $\lambda_1(1, 0)$ and $\lambda_2(0, 1)$ respectively with out changing their directions. This implies that

$$\langle A((1, 0)), (1, 0)\rangle = \| A((1, 0)) \| \, \| (1, 0) \| = \lambda_1 \, \| (1, 0) \|^2$$

and

$$\langle A((0, 1)), (0, 1)\rangle = \| A((0, 1)) \| \, \| (0, 1) \| = \lambda_2 \, \| (0, 1) \|^2$$

where the inner product is the standard dot product; $\langle (x_1, y_1), (x_2, y_2)\rangle = x_1x_2 + y_1y_2$ and $\| (x_1, y_1) \| = \sqrt{\langle (x_1, y_1), (x_1, y_1)\rangle}$, the length of the vector (x_1, y_1). This shows that

$$\lambda_1 = \frac{\langle A((1, 0)), (1, 0)\rangle}{\| (1, 0) \|^2} \text{ and } \lambda_2 = \frac{\langle A((0, 1)), (0, 1)\rangle}{\| (0, 1) \|^2}.$$

It can be easily shown that being the smallest eigenvalue, λ_1 is given by

$$\lambda_1 = \inf_{(x,y)} \frac{\langle A((x, y)), (x, y)\rangle}{\| (x, y) \|^2} \tag{2.7}$$

The fact that eigenvector corresponding to λ_2 is perpendicular to $(1, 0)$ would then imply that

$$\lambda_2 = \inf_{(x,y)\perp(1,0)} \frac{\langle A((x, y)), (x, y)\rangle}{\| (x, y) \|^2}.$$

Undoubtedly, order of the matrix is irrelevant as all calculations can be done in a similar way for a $n \times n$ diagonal matrix and also the similarity between equations (2.6) and (2.7) should not be left unnoticed. A careful look would yield the following properties which are crucial to the above calculations

(1) the matrix is diagonal,
(2) eigenvectors corresponding to different eigenvalues form a basis and are perpendicular to each other with respect to the dot product[4],
(3) the linear operator A is acting on a space which is of finite dimension.

As noticed earlier, Laplacian is a linear operator on the solutions space of problem (2.1) which helps to detect a similarity with the case of 2×2 matrices; both deal with eigenvalues of linear operators, one defined on an infinite dimensional space and other on a finite dimensional space. Apart from being infinite dimensional, as of now, what is missing on the space of solutions which prevents similar calculations is an appropriate notion of inner product which guarantees that eigenfunctions corresponding to different eigenvalues of Laplacian form a basis as well as they are perpendicular to each other. Also when the method is applied to the first eigenvalue the result should be consistent with the Rayleigh quotient (2.6). Inner products which incorporate this consistency criteria indeed exist and the solution space with those specially defined inner products is called as the Sobolev space.[5] Even then, these inner products are not an affirmation of the existence of a basis consisting of mutually perpendicular eigenfunctions of the Laplacian. So, at this moment it would take a giant leap of faith to infer anything about the eigenvalues of problem (2.1) from the above glance into linear algebra. But, when such a leap is backed by the famous spectral theory of compact self adjoint operators, the inferences based on faiths can be turned into provable facts [15]. This theory applied in the present context essentially proves all the three properties listed for operators like Laplacian defined on Sobolev spaces of functions defined on suitable domains. It is therefore possible to obtain analogous formula for eigenvalues of problem (2.1) as in the case of matrices; if W denotes the space of admissible functions then,

$$\lambda_1 = \inf_{f \in W} \frac{\int_\Omega |\nabla f|^2}{\int_\Omega f^2}$$

and if W_1 denotes the space of all functions in W which are perpendicular to the eigenfunctions of λ_1, then

$$\lambda_2 = \inf_{f \in W_1} \frac{\int_\Omega |\nabla f|^2}{\int_\Omega f^2}$$

and so on.

[4]Properties 1 and 2 are actually equivalent.

[5]Usually Sobolev space is defined as the subspace of square integrable functions which possess weak derivatives and on it various inner products are defined. This subspace is referred as solution space in this article.

Proceeding as in the case of first eigenvalue, with the help of corresponding Rayleigh quotients comparison for higher eigenvalues can now be obtained [2], i.e., for every integer $k > 0$, $\lambda_k \leq \lambda'_k$. More generally, assume that the domain Ω is filled with pairwise disjoint smaller domains $\Omega_1, \Omega_2, \cdots, \Omega_r$, such that their boundaries when intersecting with $\partial\Omega$ do so transversally. Then if $0 < \nu_1 \leq \nu_2, \cdots$, is the collection of *all* eigenvalues of problem (2.1) on domains $\Omega_1, \Omega_2, \cdots, \Omega_r$ with each eigenvalue repeated according to its multiplicity [6], then for all integers $k > 0$

$$\lambda_k \leq \nu_k. \tag{2.8}$$

With this new information in hand, a possible way to prove the Weyl's law would be to approximate Ω from inside and outside by domains whose eigenvalues and their distributions are explicitly known. As these things are known for squares, the best choice would be to consider a grid as shown in the figure below [1].

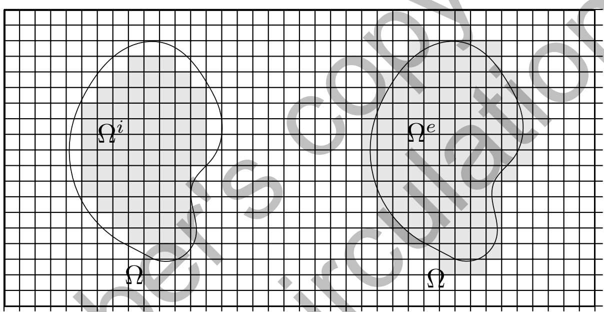


FIGURE 2

If $0 < \lambda_1^i, \leq \lambda_2^i, \cdots$ and $0 < \lambda_1^e \leq \lambda_2^e \cdots$ are the collections of all eigenvalues (repeated according to its multiplicity) of all squares in the domains Ω_i and Ω_e respectively, then $\lambda_k^e \leq \lambda_k \leq \lambda_k^i$ by the above mentioned domain monotonicity relation. This sandwiched state of λ_k when used in conjunction with the easily computable $N(\lambda^i)$ for the domain Ω_i and $N(\lambda^e)$ for Ω_e gives the Weyl's law (2.5).

With the examples of squares, discs, polygons and Weyl's law, which clearly shows that area of a domain is a spectral invariant, now it is time for the question;

3. CAN YOU HEAR THE SHAPE OF A DRUM?

It is very rare that a complex mathematical question can be asked in an aesthetic way and this famous question, which was posed by Mark Kac in 1966 [10], is an example of that. Milnor [11] had already given a negative answer to a similar question for higher dimensional spaces in 1964. Milnor's proof present another intrinsic nature of the problem which is the relationship between geodesics, which are the shortest paths connecting two nearby points in a domain, and the eigenfunctions and eigenvalues. This is plausible as the wave always chooses the

[6]As in the case of matrices, the eigenvalues of Laplacian also have finite multiplicity which can be greater than one.

shortest path to travel. In the plane, obviously straight lines are shortest paths connecting two points. But a bit more ingenuity is needed to understand the notion of geodesics when the space is curved like a torus. The way a torus is usually constructed is by identifying the opposite edges of a rectangle.

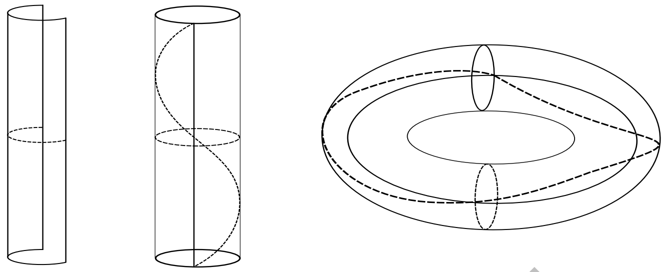


FIGURE 3

As this can be done with out shearing or stretching the rectangle, locally on such a torus the distances and angles between the lines will be same as those in the rectangle, that is to say torus is locally flat like a rectangle. But an attentive effort would show that the geodesics in the rectangle and on the torus behave differently. For example, there is only one geodesic connecting any two points in the rectangle which is not true on torus! Probably the most important difference is that on torus there are closed geodesics[7] which do not exist at all on a rectangle and this implies that unlike on rectangle, waves can travel on closed paths on a torus. This at once suggests that though torus is locally flat like rectangle, solutions of problem (2.1) may be different when considered on them[8] and more importantly the lengths of such closed geodesics could influence the frequency of waves and hence eigenvalues.

Now torus can be constructed by identifying the opposite edges in different ways, like after giving a twist. In these ways from the same rectangle different toruses can be obtained! Such constructions can also be done in an abstract way in higher dimensions by identifying opposite edges of cubes. Milnor carefully chose two such different toruses of dimension 16 whose closed geodesics have the same lengths. Toruses like that were already known to be isospectral, i.e., they have same set of eigenvalues, thus giving the first example of two different spaces with same spectrum. It is also to be mentioned that Milnor's is just one among several fascinating results about the spectrum and its relations with geodesics [3, 4, 12].

Apparently Milnor's ideas could not be used to address the case of planar domains, which motivated Kac to ask his famous question 2 years after Milnor's result. The question remained elusive for another 26 years until in 1992 Gordon, Webb and Wolpert [8] exhibited the two domains[9] in figure 4 which have the same

[7]Geodesics which are closed curves.

[8]Notice that the boundary condition disappears on torus as it has no boundary!

[9]This picture and the brief description given are taken from the expository article [7].

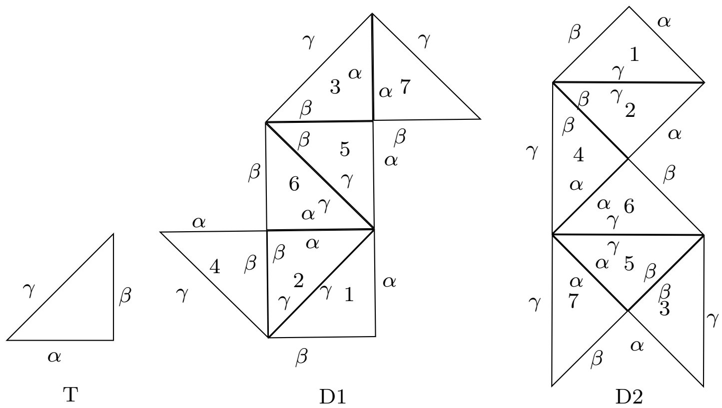


FIGURE 4

spectrum but are not geometrically congruent to each other. They constructed these domains by gluing 7 copies of a model triangle T along its edges in two different ways. If the edges of T are denoted as α, β and γ, then the only condition for gluing is that the copies of it should be glued along matching edges.

If every eigenfunction on domain $D1$ can be transplanted to domain $D2$ and vice versa, then both domains should have same spectrum. But how is it possible to transplant functions from one domain to other? A look back into the case of string gives the needed intuition to perform such an action. Suppose that F is an eigenfunction on the interval $[0, l]$. Then, joining the negative of mirror reflection of F at the end point 0, F can be extended to an eigenfunction on $[-l, l]$. It is the boundary condition imposed on F which enables such a smooth extension. If a planar domain is constructed by joining several pieces together, then an analogous reflection principle may be used to extend eigenfunctions of small pieces beyond their boundaries and hence to the whole domain in a smooth way. This suggests that it may be possible to extend the restriction of an eigenfunction of $D1$ on the model triangle to whole of domain $D2$ in a smooth way. Gordan et al. showed that such a procedure can be performed on these two domains as illustrated in figure 5. The letters on each individual triangle on $D1$ is just a renaming of restrictions of an eigenfunction F to respective pieces and on triangles on $D2$, the formulas represent the pieces of transplanted function which join together to give a smooth function. To see how the joining works, consider the top two triangles 1 and 2 in $D2$, in figure 4. These triangles constitutes the bottom square in $D1$. In figure 5, on the outer boundary of the top triangle in $D2$, the function $G + F - C$ is zero. This is true because an eigenfunction of $D1$ will be zero on its boundary and the edges α and β in the outer boundary are part of the boundary of $D1$. On the common edge, the function takes the values $G + F - C$ and $E - D - G$. But they are equal as on this edge γ, F equals E, because its the shared edge of triangles 1 and 2 in $D1$ and the values of D, G, and C are equal to that of the original

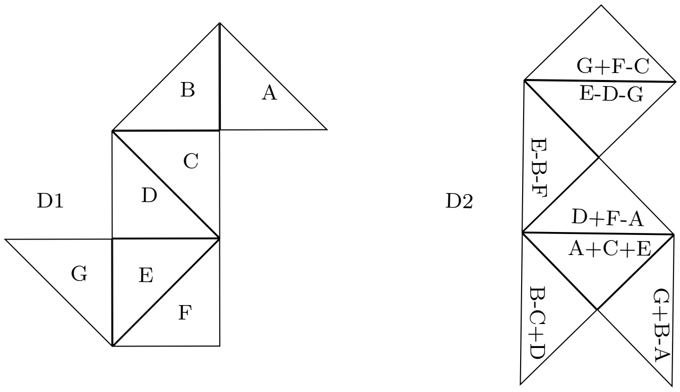


FIGURE 5

eigenfunction F. Proceeding in a similar fashion it is possible to show that the transplanted function on $D2$ is in fact an eigenfunction. This process when done with changing the role of $D1$ and $D2$, shows that both domains have the same eigenfunctions and same eigenvalues.

One may wonder why it took 28 long years for mathematicians to discover this simple method. That's because camouflaged behind this seemingly trivial manipulation is a deep result which justifies all these actions in a rigorous way. An exposition into this result, which was proved by Sunada [14] in 1984 would require to open a new door which connects spectral theory with another area of mathematics, the group theory.

It is the right time to stop this short expedition, as by this time the reader must be convinced that spectral geometry, which began by asking questions about strings and drums has been expanded vastly in last century with many facets which links it to several other areas of mathematics; and that all aspects of it can not be contained in a single article. Even though, several remarkable results were proved in these years, it's far from completely understood. There are still numerous fundamental questions which are yet to be addressed and the best way to finish this article is by mentioning one of them: it is not known whether there exist two geometrically incongruent planar domains with smooth boundaries which are isospectral!

Acknowledgment: The author highly appreciates the comments and suggestions given by two anonymous referees which helped to improve the article.

REFERENCES

[1] Berard, Pierre H., Spectral Geometry: Direct and inverse problems, *Lecture Notes in Mathematics, 1207,* Springer, 1986.

[2] Chavel, I., Eigenvalues in Riemannian Geometry, Academic Press, Inc. 1984

[3] Colin de Verdière, Y., Spectre du Laplacien et longueurs des géodésiques périodiques I, *Comp. Math.* 27, 80-106, 1973.

[4] Colin de Verdière, Y., Spectre du Laplacien et longueurs des géodésiques périodiques II", *Comp.Math.* 27, 159-184, 1973.

[5] Courant, R, and Hilbert, D., Methods of Mathematical Physics, Vol II, Interscience, Wiley, New York, 1962.

[6] Durso, C., Solution of the inverse spectral problem for triangles, *Ph.D. thesis, Massachusetts Institute of Technology*, 1990.

[7] Gordon, C., and Webb, David L., You can't hear the shape of a drum, *American Scientist* 84 Jan.-Feb., 46-55, 1996.

[8] C. Gordon, D. Webb and S. Wolpert, Isospectral plane domains and surfaces via Riemannian orbifolds, *Invent. math.* 110, 1-22, 1992.

[9] Grieser D. and Maronna, S., Hearing the shape of a triangle, *Notices of the AMS.* Volume 60: 1440-1447, 2013.

[10] Kac, M., Can One Hear the Shape of a Drum?, *The American Mathematical Monthly*, Vol. 73, No. 4, Part 2: 1-23, 1966.

[11] Milnor, J., Eigenvalues of the Laplace operator on certain manifolds, *Proc. Nat. Acad. Sci. USA*, 51, 542, 1964.

[12] Otal, J. P., Le spectre marqué des longueurs des surfaces à courbure négative, *Annals of Mathematics*, 131, 151-162, 1990.

[13] Parlett, B. N., The Rayleigh quotient iteration and some generalization for nonnormal matrices, *Mathematics of computation*, Volume 28, No: 127, 679-693, July 1974.

[14] Sunada, T., Riemannian Coverings and Isospectral Manifolds, *Annals of Mathematics*, Second Series, Vol. 121, 169-186, No. 1 Jan., 1985.

[15] Taylor, M. E., Partial Differential Equations 1, Springer, 1996.

[16] Van Den Berb, M. and Srisatkunarajah, S., Heat equation for a region in $\mathbb{R}^{\not\models}$ with polygonal boundary, *J. London Math. Soc.* (2) 31(1):119-127, 1988.

Binoy

E-mail: *binoy1729@gmail.comST*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 69-73

NOTE ON VANISHING POWER SUMS OF ROOTS OF UNITY

NEERAJ KUMAR AND K. SENTHIL KUMAR

(Received : 30 - 03 - 2015 ; Revised : 04 - 05 - 2015)

Abstract. For fixed positive integers m and ℓ, we give a complete list of integers n for which their exist mth complex roots of unity $x_1, \dots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$. This extends the earlier result of Lam and Leung on vanishing sums of roots of unity. Furthermore, we characterize all positive integers n with $2 \le n \le m$, for which there are distinct mth complex roots of unity $x_1, \dots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$.

1. Introduction

Let m be a positive integer. By an mth root of unity, we mean a complex number ζ such that $\zeta^m = 1$. That is, a root of the polynomial $X^m - 1$. One can easily see that the roots of $X^m - 1$ are distinct, in fact there are exactly m, mth roots of unity. Using the relationship between the roots and the coefficients of a polynomial, we see that the sum of all mth roots of unity, which is the coefficient of X^{m-1} in $X^m - 1$, is zero. A natural question is: What are all the positive integers n for which there exist mth roots of unity $x_1, \dots, x_n$ (repetition is allowed) such that $x_1 + \cdots + x_n = 0$. A beautiful result of T. Y. Lam and K. H. Leung [1] gives a complete classification of all such integers. Suppose m has prime factorization $p_1^{a_1} \dots p_r^{a_r}$, where $a_i > 0$, then we have the following theorem due to Lam and Leung.

Theorem 1.1. *Let n be a positive integer. Then there are mth roots of unity $x_1, \dots, x_n$ such that $x_1 + \cdots + x_n = 0$ if and only if n is of the form $n_1 p_1 + \cdots + n_r p_r$ where each n_i is a non-negative integer for $1 \le i \le r$.*

Theorem 1.1 motivate us to ask the following question.

Question 1. Let m and ℓ be positive integers. What are all the positive integers n for which there exist mth roots of unity $x_1, \dots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$?

Note that when $\ell = 1$, the complete answer to Question 1 is given by Theorem 1.1. However, for $\ell \ge 2$, we do not find any results in this direction in the literature. Our objective here is to study the case when $\ell \ge 2$. First, we fix some notations. Let m be a positive integer, and let Ω_m denotes the set of all mth roots of unity.

2010 Mathematics Subject Classification: Primary 11L03 , Secondary 11R18.
Keywords and Phrases: Roots of unity, Power sum.

For a positive integer ℓ, $W_\ell(m)$ denotes the set of all positive integers n for which there exist n-elements $x_1, \ldots, x_n \in \Omega_m$ such that $x_1^\ell + \cdots + x_n^\ell = 0$. When $\ell = 1$, we simply denote $W_\ell(m)$ by $W(m)$. With this notation, Question 1 can be reformulated as follows: *Let m and ℓ be positive integers. What are all the positive integers in the set $W_\ell(m)$?*

It is clear that if m divides ℓ then $W_\ell(m)$ is an empty set. Suppose that there are mth complex roots of unity, say, $x_1, \ldots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$. Since the ℓth power of an mth root of unity is still an mth root of unity, the equation $x_1^\ell + \cdots + x_n^\ell = 0$ with $x_i \in \Omega_m$ can be written in the form $y_1 + \cdots + y_n = 0$ with $y_i \in \Omega_m$. This shows that for any positive integer m and ℓ, $W_\ell(m)$ is a subset of $W(m)$. It follows from Theorem 1.1 that any positive integers in the set $W_\ell(m)$ must be of the form $n_1p_1 + \cdots + n_rp_r$ where each n_i is a non-negative integer for $1 \leq i \leq r$. In Section 2, we give a complete list of integers in the set $W_\ell(m)$ (see Theorem 2.1). Moreover, in Section 3 we find all positive integers $n \in W_\ell(m)$ for which there are distinct mth complex roots of unity $x_1, \ldots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$ (see Theorem 3.1).

There are algebraic aspects why Question 1 is important. For instance, for a positive integer a, denote by p_a the power sum polynomial $X_1^a + \cdots + X_n^a$ of degree a. Let ℓ and k be positive integers such that $\ell < k$. In commutative algebra, one encounters the following situation. To show that the ideal $\langle p_\ell, p_k \rangle$ generated by the polynomials p_ℓ and p_k is a prime ideal in $\mathbb{C}[X_1, \ldots, X_n]$, one needs to show that the power sum polynomial $X_1^\ell + \cdots + X_n^\ell$ does not vanish when one allows the X_i's to take values among the $(\ell - k)$th roots of unity [2, see proof of Theorem 3.8].

Acknowledgment: We are grateful to Professor Ram Murty for his valuable suggestions regarding the paper. We also thank the referee for many useful comments for improving the manuscript. This project was funded by the Department of Atomic Energy (DAE), Government of India.

2. Vanishing of power sums of roots of unity

Let m and ℓ be positive integers. In this section, we completely characterize all the positive integers in the set $W_\ell(m)$. More precisely, we prove the following theorem:

Theorem 2.1. *Let m and ℓ be positive integers. Let $d = (m, \ell)$ be the greatest common divisor of m and ℓ. Then $W_\ell(m) = W(m/d)$.*

In other words, Theorem 2.1 says that: For any positive integer n, $x_1^\ell + \cdots + x_n^\ell = 0$ with $x_i \in \Omega_m$ if and only if $y_1 + \cdots + y_n = 0$ with $y_i \in \Omega_{m/d}$.

Proof. It is well known that Ω_m, that is, the set of all mth roots of unity, form a group with respect to the multiplication of complex numbers. In fact, it is a cyclic group of order m, generated by the complex number $\zeta_m = \cos 2\pi/m + \mathfrak{i} \sin 2\pi/m$.

There is a remarkable property about finite cyclic groups. Namely, if G is a finite cyclic group and ℓ is a positive integer relatively prime to the order of G, then the map

$$x \longmapsto x^\ell \quad (x \in G) \tag{2.1}$$

is an automorphism of G (In fact, all the automorphisms of G are of the form (2.1) for some integer ℓ which is relatively prime to the order of G). It follows that, if ℓ is a positive integer which is relatively prime to m then every element of Ω_m is a ℓth power of some element of Ω_m. Thus, for an integer ℓ which is relatively prime to m, the equation $x_1^\ell + \cdots + x_n^\ell = 0$ with $x_i \in \Omega_m$ can be replaced by $y_1 + \cdots + y_n = 0$ with $y_i \in \Omega_m$, and vice versa. This discussion proves Theorem 2.1 for the case when ℓ is relatively prime to m.

Now assume that $d > 1$. Consider the map

$$\psi_d : \Omega_m \longrightarrow \Omega_{m/d} \tag{2.2}$$

defined by $x \mapsto x^d$ for $x \in \Omega_m$. This map is clearly onto, and the kernel is exactly Ω_d. Thus, $\Omega_m/\Omega_d \cong \Omega_{m/d}$. Now suppose that there are elements $x_1, \ldots, x_n \in \Omega_m$ such that $x_1^\ell + \cdots + x_n^\ell = 0$. Then this sum can be rewritten as $\left(x_1^{\ell/d}\right)^d + \cdots + \left(x_n^{\ell/d}\right)^d = 0$. Since ℓ/d and m are relatively prime, by the above discussion, the latter equation can be rewritten in the form $y_1^d + \cdots + y_n^d = 0$ with $y_i \in \Omega_m$. Finally, using the map ψ_d, the latter sum can be realized as $z_1 + \cdots + z_n = 0$ where $z_i \in \Omega_{m/d}$ for $1 \le i \le n$. In fact, all these steps can be reversed. This completes the proof of Theorem 2.1. $\square$

Combining Theorems 1.1 and 2.1, we have the following corollary.

Corollary 2.1. *Let m, n and ℓ be positive integers. Let $d = (m, \ell)$ be the greatest common divisor of m and ℓ. Then there are mth roots of unity $x_1, \ldots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$ if and only if n is of the form $n_1q_1 + \cdots + n_sq_s$ where each n_i is a non-negative integer for $1 \le i \le s$ and $q_1, \ldots, q_s$ are distinct prime divisors of m/d.*

Example 2.1. Let $m = 60$, and let ℓ be an integer with $1 \le \ell < 60$. By Theorem 2.1, $W_\ell(m) = W(m/d)$ where d is the greatest common divisor of m and ℓ. When d varies over the divisors of m, m/d also varies over the divisors of m. Thus $W_\ell(m)$ coincides with $W(d)$ for some divisor d of m. On the other hand, by Theorem 1.1, $W(d) = \sum_{i=1}^s q_i\mathbb{N}$ where $d = q_1^{b_1} \ldots q_s^{b_s}$ is the prime factorization of d. Here $\mathbb{N}$ denotes the set of non-negative integers. We thus have the following table which describe $W(d)$ for all positive divisors d of $m = 60$.

3. VANISHING OF POWER SUMS OF DISTINCT ROOTS OF UNITY

Let m and ℓ be two positive integers. For an integer $n \in W_\ell(m)$, the *height* $H(n; \ell, m)$ of n is defined to be the smallest positive integer h for which there are mth roots of unity $x_1, \ldots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$ and the maximum among

d	$W(d)$
1	$\emptyset$
2	$2\mathbb{N}$
3	$3\mathbb{N}$
4	$2\mathbb{N}$
5	$5\mathbb{N}$
6	$2\mathbb{N} + 3\mathbb{N} = \mathbb{N} \setminus \{1\}$
10	$2\mathbb{N} + 5\mathbb{N} = \mathbb{N} \setminus \{1, 3\}$
12	$2\mathbb{N} + 3\mathbb{N} = \mathbb{N} \setminus \{1\}$
15	$3\mathbb{N} + 5\mathbb{N} = \mathbb{N} \setminus \{1, 2, 4, 7\}$
20	$2\mathbb{N} + 5\mathbb{N} = \mathbb{N} \setminus \{1, 3\}$
30	$2\mathbb{N} + 3\mathbb{N} + 5\mathbb{N} = \mathbb{N} \setminus \{1\}$
60	$2\mathbb{N} + 3\mathbb{N} + 5\mathbb{N} = \mathbb{N} \setminus \{1\}$

the repetition of x_i's is h, that is, h is the maximum among the h_i, where h_i is the number of times x_i appears in the list $x_1, \ldots, x_n$. When $\ell = 1$, we denote $H(n; \ell, m)$ by $H(n; m)$. Note that $H(n; m) = 1$ provided $2 \leq n \leq m$. Gary Sivek [3] refined the work of Lam and Leung by proving that for any integers $m \geq 2$ and $2 \leq n \leq m$, $H(n; m) = 1$ if and only if both n and $m - n$ are expressible as a linear combination of the prime factors of m with non-negative integer coefficients. Here we extend Sivek's result to vanishing of power sums of distinct roots of unity.

Theorem 3.1. *Let m and ℓ be positive integers, and let n be an integer such that $2 \leq n \leq m$. Let d be the greatest common divisor of m and ℓ. Then $H(n; \ell, m) = 1$ if and only if $H(n; m/d) \leq d$.*

Proof. Let $\Omega_{m/d} = \{z_1, \ldots, z_{m/d}\}$. Suppose that there are distinct mth roots of unity $x_1, \ldots, x_n$ such that $x_1^\ell + \cdots + x_n^\ell = 0$. Since d is the greatest common divisor of ℓ and m, this equation can be rewritten in the form $y_1^d + \cdots + y_n^d = 0$ with $y_1, \ldots, y_n$ are mth roots of unity. Using the map ψ_d, the latter equation can be written as $\sum_{i=1}^{m/d} a_i z_i = 0$ where a_i is the cardinality of the set $\{y_1, \ldots, y_n\} \cap \psi_d^{-1}(z_i)$ for $1 \leq i \leq m/d$. On the other hand, $\psi_d^{-1}(z)$ has exactly d elements for each $z \in \Omega_{m/d}$. It follows that $H(n; m/d) \leq \max\{a_1, \ldots, a_{m/d}\} \leq d$. This proves that if $H(n; \ell, m) = 1$ then $H(n; m/d) \leq d$.

Conversely, suppose that $H(n; m/d) \leq d$. Then there is a partition $(a_1, \ldots, a_{m/d})$ of n into non-negative integers a_i with $a_i \leq d$ for $1 \leq i \leq m/d$ and $\sum_{i=1}^{m/d} a_i z_i = 0$. Let y_i be any element of $\psi_d^{-1}(z_i)$ for $1 \leq i \leq m/d$. Then $\psi_d^{-1}(z_i) = y_i \Omega_d = \{ y_i x \mid x \in \Omega_d\}$. Since $a_i \leq d$, one can replace $a_i z_i$ by $y_i^d(x_1^d + \cdots + x_{a_i}^d)$ where $x_1, \ldots, x_{a_i}$ are distinct elements of Ω_d. Hence $H(n; \ell, m) = H(n; d, m) = 1$ since $\sum_{i=1}^{m/d} a_i = n$. This completes the proof of Theorem 3.1. $\square$

References

[1] Lam, T. Y. and Leung, K. H., *On Vanishing sums of roots of unity.* J. Algebra **224** (2000), no. 1, 91–109.

[2] Kumar, Neeraj, *Prime ideals and regular sequences of symmetric polynomials.* arXiv:1309.1098[math.AC].

[3] Sivek, G., *On vanishing sums of distinct roots of unity* Integers, **10** (2010), pp. A31, 365–368.

Neeraj Kumar
The Institute of Mathematical Sciences
4th cross road, CIT Campus, Taramani
Chennai-600 113, India
E-mail: *neerajkr@imsc.res.in*

K. Senthil Kumar
The Institute of Mathematical Sciences
4th cross road, CIT Campus, Taramani
Chennai-600 113, India
E-mail: *senthilkk@imsc.res.in*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 75-94

SNELL'S LAW AT THE BOUNDARIES OF REAL ELASTIC MEDIA*

M. D. SHARMA

(Received : 14 - 04 - 2015 ; Revised : 25 - 04 - 2015)

Abstract. An ideal elastic medium represents an isotropic, perfectly (lossless) elastic, single phase medium but a real one includes the presence of anisotropy, porosity (multi-phase), dissipation or attenuation. The present study considers the propagation of harmonic plane waves across the plane boundaries of real elastic media. Snell's law for the phenomenon of reflection comes from the boundary conditions at the free surface of the elastic medium. The boundary conditions at the common interface between two elastic media yield the same law for the reflection/refraction phenomenon. Snell's law ensures that the propagation of reflected / refracted waves confine to the plane formed by the incident wave and the normal to the boundary (at the point of incidence). It is used further to derive the propagation characteristics (velocities, directions of propagation/attenuation) of reflected/refracted waves from those of the incident wave. It may be noted that an ideal description is a particular case of the real one but the real description can not be a simple and straightforward extension of the ideal description.

1. Introduction

The concepts of anisotropy, dissipation, anelasticity and porosity in elastic media have gained much attention in recent years. Applications of the relevant studies cover a variety of fields including civil engineering, non-destructive testing, soil mechanics, earthquake preparation, underwater acoustics, exploration seismology and geophysics. The structural properties of in-situ rocks are obtained mainly from the seismic properties such as travel times (or velocities), amplitude information (reflection/refraction coefficients) and wave polarizations (motion of constituent particles). These measurable quantities are affected by the presence of anisotropy, anelasticity and porosity. Ocean covers a nearly three-fourth of earth's surface and sediments at ocean-bottom provide the most appropriate examples of

* This article is dedicated to the memory of my Ph.D. supervisor late Professor M. L. Gogna (1931-2010).

2010 Mathematics Subject Classification: 74B05; 74E10; 74J10; 74J20; 74L05.
Keywords and Phrases: Snell's law, reflection/refraction, anisotropy, dissipation, attenuation, inhomogeneous waves.

saturated porous materials in the crust. In such water-saturated particulate materials, energy loss occurs due to friction in fluid flow and/or anelasticity of skeletal frame. The texture of an ocean sediment is complicated because it is a coupled multi-phase composite consisting of a solid phase of oriented grains and fluid phases of pore water and gases. Dissipation of energy happens due to the viscous losses at the points of inter-granular contact when there is any relative motion between pore-fluids and skeletal frame. Then, with the presence of frame anelasticity in addition to the anisotropy, the mechanics of ocean sediments becomes further complicated. This demands more attention to the theoretical studies of wave propagation in more realistic models of porous media.

The assumption of isotropy in the propagation of seismic waves is made for mathematical convenience. On the contrary, seismic anisotropy is widespread in many types of rocks at many depths in many geological and tectonic environments and it has widespread applications in the various areas of economic and scientific interest ([12]). Lithological and crystal alignments, pre-stresses, aligned cracks, periodic thin layering and fluid-filled pores are identified as the major causes of seismic anisotropy. The absence of point symmetry in pore distribution results in the anisotropy of general type with arbitrary symmetry. Gurevich ([14]) investigated that the fluid-saturated porous rocks with aligned fractures, in an isotropic porous background, can be represented as anisotropic poroelastic aggregate.

The crustal rocks are always subjected to stresses. The slow process of creep inside the earth creates a differential stress environment in the crust, which is responsible for the preferential alignments in the Earth, ranging from mineral orientations, grains, or microcracks to sedimentary folds or regional fractures. The difference between confined tectonic stress and pore-fluid pressure conducts the flow of fluid to a reservoir through the connected cracks. This signifies that initial stress and crack-induced anisotropy not only coexist in the crust but will, also, be affecting the quality/quantity of each other. For the presence of pre-stress or aligned cracks, an elastic medium behaves anisotropic to wave propagation ([12]).

Piezoelectric materials show a unique electromechanical coupling characteristic which produce electrical fields under the application of mechanical loads and elastic deformation on applying electrical loads. Piezoelectric composites are developed to improve upon their applications in acoustic transducers/sensors, medical imaging and non-destructive evaluation. These composite materials behave anisotropic to wave propagation ([15]). The reinforcements and laminations are the major sources of anisotropy in these materials. Physics of granular media ([13]) represents an active area of current research activity due to the presence of stress-induced anisotropy existing there. Hence, any study of anisotropic elasticity finds its relevance in understanding the mechanical behaviour of composite or granular materials ([18], [5]). A latest book by Carcione ([6]) explains the importance of anisotropy for wave propagation studies in real materials.

The accurate analysis of observed seismic attenuation is important for the advancement in knowing the structure of the earth. In sedimentary rocks, the anelastic nature alone may not be able to explain the observed attenuation. Confining stresses and wave-induced flow of viscous-fluid in pores and cracks are other important factors. The presence of viscosity in pore-fluid implies relaxation phenomenon, which causes wave-field dissipation. This is particularly true in exploration industry and in the investigation of tectonic stress and failure where small scale fracturing and flow of fluid into the fractures are important. The confined stresses and shale anisotropy helps in predicting flow path for improved oil recovery and designing hydraulic fracturing schemes. The flow mechanism to equilibrate fluid pressure produces a great deal of seismic attenuation at high frequency.

Whatever be the sources of attenuation of seismic waves in sedimentary rocks, mathematical models are often required to explain the decay of amplitude of waves propagating away from the source ([2]). This demands a much deeper insight into the process of wave propagation in dissipative models of sedimentary rocks in the crust. These rocks can, more closely, be modeled as fluid-saturated porous solids pervaded by aligned cracks. The fluid-saturated micro-cracks are highly compliant and crustal rocks respond immediately to the small changes in differential stress. Hence, a pre-stressed anisotropic porous solid makes a much realistic geophysical model to be used for seismic characterization of sedimentary or reservoir rocks ([14]). In particular, such a composite physical model facilitates the parametric studies of the influence of various measurable physical properties (i.e., porosity, permeability, pore-fluid viscosity, frame anelasticity, initial-stress, elastic/hydraulic anisotropy, crack density, etc.) on the wave characteristics.

Importance of inhomogeneous waves in single-phase viscoelastic media are found in Borcherdt ([2], [3]) and Cerveny and Psencik ([10], [11]). Experimental results ([3]) confirm the generation and existence of inhomogeneous body waves and the differences in their physical characteristics from elastic body waves. Carcione ([7]) suggested that the differences in amplitude variation with offset (AVO) of waves transmitted at ocean bottom depend not only on the properties of the medium but also on the inhomogeneity of the wave. An earlier study of the author ([20]) shows that, compared to velocity, attenuation is more sensitive to the inhomogeneity of waves propagating in dissipative anisotropic poroelastic medium. This implies that the propagation of inhomogeneous waves may be able to explain, mathematically, the larger attenuation of seismic waves in sedimentary regions.

Temperature variations play a significant role in the modification of cracks and the flow of fluid ([19]). These modifications in microcracks are responsible for the dynamism around geothermal reservoirs and the sedimentary basins. Theory of thermoelasticity is used to understand such dynamical systems that involve interactions between mechanical work and thermal changes. With the introduction of relaxations in temperature field, few generalised theories explain the attenuation of waves in thermoelastic medium. However, a more realistic scheme is defined with memory effects allowed to all the constitutive properties of thermoelastic coupling. Analogous to the correspondence principle in classical elasticity, the complex values of appropriate constitutive quantities may define the time-harmonic material dissipation in a thermoviscoelastic medium ([9]). Through a wonderful coincidence, Biot presented the theories of, both, thermoelasticity and poroelasticity, in the same year of 1956. Indeed, he was very much aware of the connection between them ([1]). Using the correspondence between these two theories, the results from thermoelasticity are often translated to solve the problems of wave propagation in poroelasticity ([17]). This analogy between thermoelasticity and poroelasticity theories can be extended to three-dimensional wave propagation in a general anisotropic medium ([21]).

The present article illustrates the use of Snell's law in deriving the propagation characteristics (direction, velocity, attenuation) of reflected/refracted waves from those of the incident wave. The whole discussion is presented through 8 sections. The section 2 explains Snell's law for reflection/ refraction as applicable in optics. Snell's law for scattering of elastic waves is explained in section 3. The section 4 illustrates the use of this law in deciding the propagation of reflected/refracted waves at the boundaries of an isotropic elastic medium. For the presence of dissipation in isotropic elastic medium, a modified form of this law is illustrated in section 5. The anisotropic interpretation of the law in the absence of dissipation is discussed in section 6. In section 7, Snell's law in anisotropic viscoelastic medium is interpreted to explain the inhomogeneous propagation of attenuated waves. The last section in this article illustrates the extension of the various relations/derivations to poroelastic and thermoelastic media.

2. Snell's Law

Origin of Snell's law lies with optics and it is named after Dutch astronomer Willebrord Snellius (1580-1626). It follows from Fermat's principle of least travel time, which in turn follows from the propagation of light as waves. It is used to determine the direction of light rays through refractive media with varying indices of refraction. Refractive index ($r = c/V$) of a medium is defined as the factor by which the speed of light-ray traveling through a refractive medium (V) decreases from its velocity in a vacuum (c), i.e., $V = c/r$. Hence, the denser the medium, the slower the speed of light, the higher the refractive index.

a) For reflection at a plane boundary (i.e., opaque surface), Snell's law is interpreted as follows.

i) Angle of reflection is same as the angle of incidence.

ii) Reflected ray, incident ray and normal to boundary lie in same plane.

b) For reflection and refraction at a plane (transparent) boundary, Snell's law yields the following results.

i) Reflected ray, refracted ray, incident ray and normal to boundary lie in same plane.

ii) Velocities (V_R) and propagation directions (θ_R) of reflected / refracted rays relate to the velocity (V_I) and direction (θ_I) of incident ray as follows.

$\frac{\sin\theta_R}{V_R} = \frac{\sin\theta_I}{V_I}$ or $\frac{\sin\theta_R}{\sin\theta_I} = \frac{V_R}{V_I} = \frac{r_I}{r_R}$.

Note that for reflected ray, $V_R = V_I$ yields $\theta_R = \theta_I$, i.e., angle of reflection is same as the angle of incidence.

3. Elastic Waves

Elastic waves are the agents to extract subsurface information about the internal structure of any material body, be it a construction or a natural rock. Velocity and attenuation are two important characteristics of wave propagation in crustal rocks. Let an elastic medium be specified through a Cartesian coordinate system (x_1, x_2, x_3). For the harmonic propagation of plane waves in this medium, displacement (u_1, u_2, u_3) of material particles is written as follows.

$$u_j = S_j \exp\{\imath\omega(p_k x_k - t)\}, \ (j = 1, 2, 3), \tag{3.1}$$

where ω is angular frequency of the harmonic oscillations of the particles. Repeated index ($'k'$) implies summation over $k = 1, 2, 3$. Slowness vector $\vec{p} = (p_1, p_2, p_3) = \hat{n}/V$ provides the propagation velocity V and the propagation direction $\hat{n}$. Magnitude (direction) of vector (S_1, S_2, S_3) defines the amplitude (polarisation) of the displaced particles.

3.1. **Snell's law.** The displacements (3.1) make a general solution of the wave equation, given by

$$\nabla^2\vec{u} = \frac{1}{V^2}\frac{\partial^2\vec{u}}{\partial t^2}, \ \vec{u} = (u_1, u_2, u_3). \tag{3.2}$$

This solution needs to satisfy some conditions at the boundary of the medium for propagation of disturbance in a bounded continuum. For an elastic medium, boundary conditions, in general, are the continuity of disturbance (particle vibrations and stresses) at every point on its boundary. The matching of magnitudes in the boundary conditions relate the amplitudes (and energies) of reflected/refracted waves to that of the incident wave. The phase-matching at boundary requires that the projections (i.e., tangential components) of the slowness vectors on the boundary must coincide. This yields Snell's law, for the propagation of elastic waves in the considered elastic medium. It is interpreted to check that

a) the propagation directions of reflected / refracted waves lie in the plane formed by the incident wave and the normal to the boundary (at the point of incidence).
b) the directions and velocities of reflected / refracted waves could be derived from the direction and velocity of the incident wave.
To be more specific, we consider an incident wave, which is defined through its

i) propagation direction,

ii) attenuation direction (if medium is dissipative), and

iii) phase velocity.

Then, the relevant boundary conditions provide Snell's law, which is used further to find

i) the propagation direction,

ii) the attenuation direction (if medium is dissipative), and

iii) the phase velocity of each reflected / refracted wave at the boundary.

4. Isotropic Non-dissipative Medium

To start with an ideal medium, the propagation of disturbance is considered in a homogeneous isotropic elastic solid. Such a medium is represented through two elastic constants (i.e., Lame's moduli, say λ, μ) and density (say, ρ). Elastic waves in this medium are either dilatational waves (propagating with velocity $\alpha = \sqrt{(\lambda+2\mu)/\rho}$) or rotational waves (propagating with velocity $\beta = \sqrt{\mu/\rho}$). The dilatational waves represent particle motion parallel to the propagation direction ($\hat{n}$) and hence are also identified as longitudinal (or P) waves. Similarly, the rotational waves represent particle motion normal to the propagation direction ($\hat{n}$) and are also identified as transverse (or S) waves.

4.1. **Reflection.** In Cartesian coordinate system (x_1, x_2, x_3), the elastic medium occupies a half-space, which is bounded by the stress-free plane boundary $x_3 = 0$. In this medium, displacement are specified as follows.
Incident wave: $\vec{u}^{(I)} = \vec{S}^{(I)} \exp\{\imath\omega(p_k^{(I)}x_k - t)\}$; $p_k^{(I)} = n_k^{(I)}/V_I$, and $V_I = \alpha$ or β.
Reflected waves: $\vec{u}^{(R)} = \vec{S}^{(R)} \exp\{\imath\omega(p_k^{(R)}x_k - t)\}$; $p_k^{(R)} = n_k^{(R)}/V_R$, and $V_R = \alpha,\ \beta$.

Boundary conditions (stresses vanish at $x_3 = 0$) yields Snell's law as $p_1^{(I)} = p_1^{(R)}$, $p_2^{(I)} = p_2^{(R)}$. In terms of propagation directions and velocities, we have $n_1^{(I)}/V_I = n_1^{(R)}/V_R$, and $n_2^{(I)}/V_I = n_2^{(R)}/V_R$. These relations imply that $n_2^{(R)}/n_1^{(R)} = n_2^{(I)}/n_1^{(I)}$, or, in spherical polar coordinates (r, θ, ϕ), we get $\tan\phi_R = \tan\phi_I$, i.e., same azimuth or same (vertical) plane for incident wave and reflected waves. Then $n_1^{(I)}/V_I = n_1^{(R)}/V_R$ is reduced to $\sin\theta_R/V_R = \sin\theta_I/V_I$, which provides the directions of reflected waves as $\theta_R = \sin^{-1}(\sin\theta_I V_R/V_I)$. The direction-independent velocities for reflected waves in this isotropic elastic medium (i.e., $V_R = \alpha,\ \beta$) are calculated from elastic/dynamical constants ($\lambda,\ \mu,\ \rho$) for the medium.

4.2. **Reflection & refraction.** In Cartesian coordinate system (x_1, x_2, x_3), the plane $x_3 = 0$ represents a welded boundary between two different isotropic elastic

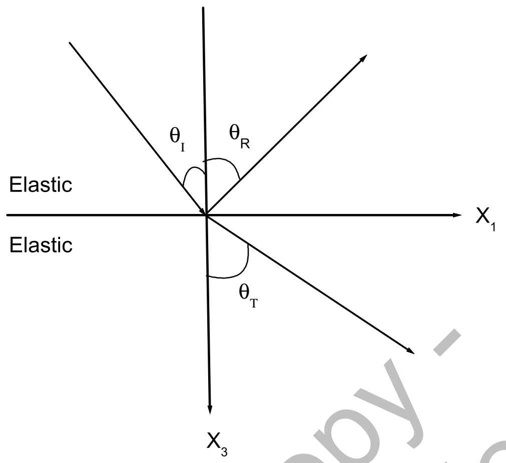


FIGURE 1

solids (FIGURE 1). For various waves in two connecting media, the displacements are specified as follows.

Incident wave: $\vec{u}^{(I)} = \vec{S}^{(I)} \exp\{\imath\omega(p_k^{(I)} x_k - t)\}$; $V_I = \alpha_1$ or β_1.

Reflected waves: $\vec{u}^{(R)} = \vec{S}^{(R)} \exp\{\imath\omega(p_k^{(R)} x_k - t)\}$; $V_R = \alpha_1,\ \beta_1$.

Refracted (transmitted) waves: $\vec{u}^{(T)} = \vec{S}^{(T)} \exp\{\imath\omega(p_k^{(T)} x_k - t)\}$; $V_T = \alpha_2,\ \beta_2$.

Boundary conditions (continuity of stresses and displacement) at $x_3 = 0$ yield Snell's law as $p_1^{(I)} = p_1^{(R)} = p_1^{(T)}$, $p_2^{(I)} = p_2^{(R)} = p_2^{(T)}$. This implies that $n_2^{(I)}/n_1^{(I)} = n_2^{(R)}/n_1^{(R)} = n_2^{(T)}/n_1^{(T)}$, i.e., same (vertical) plane for incident, reflected and refracted waves.

Propagation directions for reflected (P and S) waves and refracted (P and S) waves are related to the incident direction as follows.
For reflected waves: $\theta_R = \sin^{-1}(\sin\theta_I V_R/V_I)$, $V_R = \alpha_1,\ \beta_1$.
For refracted waves: $\theta_T = \sin^{-1}(\sin\theta_I V_T/V_I)$, $V_T = \alpha_2,\ \beta_2$.

Note that the velocities $(\alpha_j, \beta_j;\ j = 1, 2)$ are calculated from the Lame's moduli (λ_j, μ_j) and the densities (ρ_j) of the corresponding elastic media.

5. ISOTROPIC VISCOELASTIC MEDIUM

Linear viscoelastic effect in an isotropic medium is represented through the complex values for the Lame's moduli $(\lambda,\ \mu)$. With real density (ρ), we get complex-valued velocities $(\alpha,\ \beta)$ for P and S waves. Harmonic propagation of plane waves with real frequency ω, in (x_1, x_2, x_3) space, is given by

$$\vec{u} = \vec{S} \exp\{\imath\omega(p_k x_k - t)\}. \tag{5.1}$$

Any wave propagating in a dissipative medium undergoes attenuation, which is specified through the complex slowness vector $\vec{p} = (p_1, p_2, p_3)$. For an attenuated wave, we write $\vec{p} = \vec{N}/V$ to determine its complex velocity $V = 1/\sqrt{(\vec{p}.\vec{p})}$ and a complex (dual) phase vector $\vec{N}$ such that $\vec{N}.\vec{N} = 1$. For such a wave, the resolved complex $\vec{p} = \vec{P} + \imath\vec{A}$ defines its propagation vector $\vec{P}$ and attenuation vector $\vec{A}$. The phase vector $\vec{N} = (N_1, N_2, N_3)$ in complex slowness vector $\vec{P} + \imath\vec{A} = \vec{N}/V$ decides the (in)homogeneous propagation of the attenuated wave.

When all N_j are real: both $\vec{P}$ and $\vec{A}$ are directed along unit vector $\vec{N}$. Hence, $\vec{p}$ defines the homogeneous propagation of attenuated waves.

When any of the N_j is non-real, i.e., $\vec{N}$ is a complex vector: $\vec{P}$ and $\vec{A}$ have different directions. Hence, complex $\vec{N}$ represents the inhomogeneous propagation of attenuated waves. To be more specific, inhomogeneity is measured as the deviation from homogeneity, which accounts for the difference in directions of $\vec{P}$ and $\vec{A}$. In other words, the angle between $\vec{P}$ and $\vec{A}$, measures the inhomogeneity of an attenuated wave.

5.1. **Reflection of attenuated wave.** In Cartesian coordinate system (x_1, x_2, x_3), isotropic viscoelastic solid occupies the half-space $x_3 < 0$. In this medium, we consider the propagation of a harmonic attenuated plane wave, which become incident at the plane boundary $x_3 = 0$ (FIGURE 2). Displacement for this incident wave is given by

$$\vec{u}^{(I)} = \vec{S}^{(I)} \exp\{\imath\omega(p_k^{(I)} x_k - t)\}, \tag{5.2}$$

where $\vec{p}^{(I)} = \vec{P}^{(I)} + \imath\vec{A}^{(I)}$, for propagation vector $\vec{P}^{(I)}$ and attenuation vector $\vec{A}^{(I)}$.

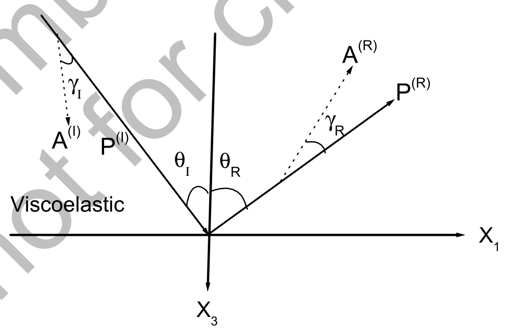


FIGURE 2

Isotropy allows to study propagation in sagittal (say x_1-x_3) plane, without any loss. Further, the orientations of various angles are not taken into account. Then, from FIGURE 2, the horizontal slowness can be written as

$$p_1^{(I)} = |\vec{P}^{(I)}| \sin\theta_I + \imath|\vec{A}^{(I)}| \sin(\theta_I - \gamma_I). \tag{5.3}$$

where θ_I is the angle, which the propagation vector of the incident (P or SV) wave makes with the normal to boundary. The angle $\gamma_I \in [0, \pi/2]$ measures the deviation of propagation direction from the direction of maximum attenuation. Note that the incident attenuated wave turns homogeneous with $\gamma_I = 0$. In other words, non-zero γ_I measures the inhomogeneity in the propagation of the (incident) attenuated wave in this dissipative medium. The general plane waves propagating in a dissipative medium are inhomogeneous waves ([3]).

For an incident wave (represented by the chosen values of θ_I and γ_I) of complex velocity V_I, we have $\vec{p}^{(I)} = \vec{P}^{(I)} + \imath\vec{A}^{(I)}$, where

$$1/V_I^2 = \vec{p}^{(I)}.\vec{p}^{(I)} = |\vec{P}^{(I)}|^2 + |\vec{A}^{(I)}|^2 + 2\imath|\vec{P}^{(I)}||\vec{A}^{(I)}|\cos\gamma_I. \tag{5.4}$$

These relations are solved into

$$\begin{aligned} 2|\vec{P}^{(I)}|^2 &= \Re(V_I^{-2}) + \sqrt{\Re(V_I^{-2})^2 + \Im(V_I^{-2})^2 \sec^2\gamma_I}, \\ 2|\vec{A}^{(I)}|^2 &= -\Re(V_I^{-2}) + \sqrt{\Re(V_I^{-2})^2 + \Im(V_I^{-2})^2 \sec^2\gamma_I}, \end{aligned} \tag{5.5}$$

which enables the relation (5.3) to calculate the horizontal slowness for attenuated incident wave.

For propagation in x_1-x_3 plane, Snell's law ensures that slowness of reflected waves along x_1-direction ($p_1^{(R)}$) is same as that of the incident wave ($p_1^{(I)}$), i.e.,

$$p_1^{(R)} = |\vec{P}^{(R)}|\sin\theta_R + \imath|\vec{A}^{(R)}|\sin(\theta_R - \gamma_R) = |\vec{P}^{(I)}|\sin\theta_I + \imath|\vec{A}^{(I)}|\sin(\theta_I - \gamma_I) = p_1^{(I)}$$

or

$$|\vec{P}^{(R)}|\sin\theta_R = |\vec{P}^{(I)}|\sin\theta_I, \tag{5.6}$$

$$|\vec{A}^{(R)}|\sin(\theta_R - \gamma_R) = |\vec{A}^{(I)}|\sin(\theta_I - \gamma_I). \tag{5.7}$$

This common horizontal slowness (i.e., $p_1^{(I)} = p_1^{(R)} = s$, say) is used to calculate the vertical slowness $p_3^{(R)} = p.v.\sqrt{V_R^{-2} - s^2}$ for each reflected wave ($p.v.$ stands for principal value of complex radical). Then, the slowness vectors $\vec{p}^{(R)} = (s, 0, p_3^{(R)})$ for reflected waves yield

$$\vec{P}^{(R)} = \Re(\vec{p}^{(R)}) = \Re(s, 0, p_3^{(R)}),\ \vec{A}^{(R)} = \Im(\vec{p}^{(R)}) = \Im(s, 0, p_3^{(R)}). \tag{5.8}$$

Now, for any reflected wave, its direction of propagation θ_R, (angle with normal to boundary) and attenuation direction γ_R, (angle between propagation direction & attenuation direction) are calculated as follows.

$$s = |\vec{P}^{(R)}|\sin\theta_R + \imath|\vec{A}^{(R)}|\sin(\theta_R - \gamma_R);$$

$$\sin\theta_R = |\vec{P}^{(R)}|^{-1}\Re s, \tag{5.9}$$

$$\sin(\theta_R - \gamma_R) = |\vec{A}^{(R)}|^{-1}\Im s. \tag{5.10}$$

5.2. **Reflection & refraction at elastic-viscoelastic interface.** One of the two elastic media in section 4.2 is replaced with a viscoelastic medium. Then, at elastic-viscoelastic boundary, the incident wave keeps the option of traveling either in elastic medium or in viscoelastic medium. Obviously, in each case, the phenomenon of reflection and refraction will be different as illustrated below.

a) Incidence through elastic medium (FIGURE 3):

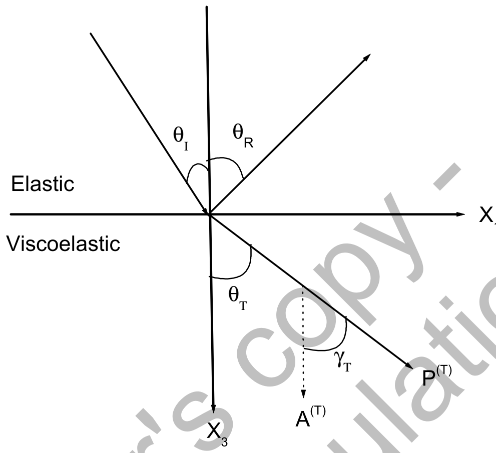


FIGURE 3

The incident wave is a non-attenuated elastic wave with real velocity (V_I). The propagation of this wave along a chosen direction (i.e., angle θ_I) is represented through a real slowness vector $\vec{p}^{(I)} = \sin\theta_I/V_I$. Consequently, for pre-critical incidence, real $\vec{p}^{(R)} = \sin\theta_R/V_R$ specifies the non-attenuated reflected waves. Even for critical incidence, the horizontal slowness of any reflected wave will be real. However, the continuing medium across the boundary is a dissipative medium, which can afford the attenuation of refracted waves, i.e., complex slowness vector $\vec{p}^{(T)}$. Snell's law ensures that the common horizontal slowness (s) is real, i.e., the relation

$$\begin{aligned} p_1^{(I)} &= \sin\theta_I/V_I = \sin\theta_R/V_R \\ &= p_1^{(R)} = \ s \ = p_1^{(T)} = |\vec{P}^{(T)}|\sin\theta_T + \imath|\vec{A}^{(T)}|\sin(\theta_T - \gamma_T) \end{aligned} \tag{5.11}$$

must follow that $\gamma_T = \theta_T$. That means a fixed direction (normal to the boundary) for attenuation of waves refracted to dissipative medium. Propagation directions of reflected / refracted waves, from (5.11), are given by

$$\theta_R = \sin^{-1}(sV_R),\ \theta_T = \sin^{-1}(s|\vec{P}^{(T)}|). \tag{5.12}$$

b) Incidence through viscoelastic medium (FIGURE 4):
The incident wave originates in viscoelastic medium. This wave must be an attenuated wave represented through a complex slowness vector $\vec{p}^{(I)}$. Such a wave

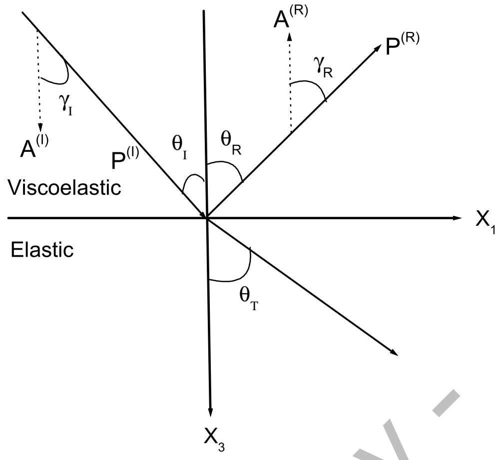


FIGURE 4

is specified through a propagation direction (i.e., angle θ_I), an attenuation direction (γ_I) and complex velocity (V_I). Consequently, reflected waves will also be attenuated, represented by complex slowness vectors $\vec{p}^{(R)}$. However, the continuing medium at the boundary is a non-dissipative elastic medium, which cannot afford the attenuation of refracted waves, i.e., real slowness vector $\vec{p}^{(T)}$. The horizontal slowness of such a refracted wave can be determined from its propagation direction (θ_T) and velocity (V_T). The demand of Snell's law for common horizontal slowness (real s) at the boundary is met through the relation

$$\begin{aligned} p_1^{(I)} &= |\vec{P}^{(I)}| \sin\theta_I + \imath |\vec{A}^{(I)}| \sin(\theta_I - \gamma_I) = s \\ &= p_1^{(R)} = |\vec{P}^{(R)}| \sin\theta_R + \imath |\vec{A}^{(T)}| \sin(\theta_R - \gamma_R) \;=\; \sin\theta_T / V_T \;=\; p_1^{(T)}. \end{aligned} \tag{5.13}$$

The above relation yields $\gamma_I = \theta_I$ and $\gamma_R = \theta_R$. This implies that incident wave must have a fixed direction of attenuation (normal to the boundary). Consequently, the same fixed attenuation direction applies to all the reflected waves as well. Then, we use

$$2|\vec{P}^{(I)}|^2 = \Re(V_I^{-2}) + \sqrt{\Re(V_I^{-2})^2 + \Im(V_I^{-2})^2 \sec^2\theta_I} \tag{5.14}$$

to calculate the slowness $s = |\vec{P}^{(I)}| \sin\theta_I$. Finally, the directions of reflected and refracted waves are given by

$$\theta_R = \sin^{-1}(s|\vec{P}^{(R)}|^{-1}),\ \theta_T = \sin^{-1}(sV_T), \tag{5.15}$$

where $\vec{P}^{(R)} = \Re(\vec{p}^{(R)}) = \Re(s, 0, p.v.\sqrt{V_R^{-2} - s^2})$.

5.3. **Reflection & refraction at viscoelastic-viscoelastic interface.** Incident wave will be an attenuated wave propagating in x_1-x_3 plane, as shown in FIGURE

5. Snell's law is defined through the common value of slowness along x_1-direction (say, s), for all the waves at the boundary.

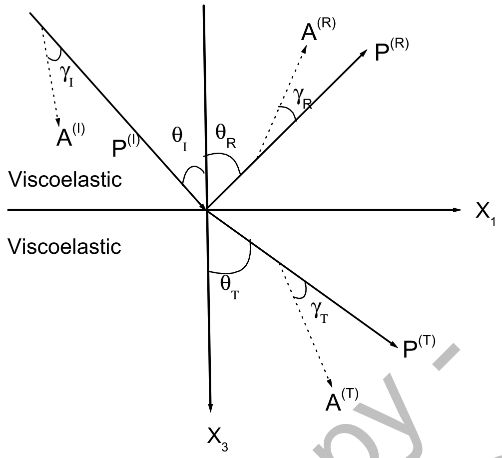


FIGURE 5

For incident wave with chosen propagation direction (θ_I) and attenuation angle (γ_I), the slowness s is defined through the relations (5.3) and (5.5). Vertical slowness for any reflected or refracted wave is calculated as $p_3^{(k)} = p.v.\sqrt{(V_k^{-2} - s^2)}$, $k = R, T$.

Snell's law in this case is given by $s = p_1^{(I)} = p_1^{(R)} = p_1^{(T)}$. Then $\vec{P}^{(k)} = \Re(\vec{p}^{(k)}) = \Re(s, 0, p_3^{(k)})$ and $\vec{A}^{(k)} = \Im(\vec{p}^{(k)}) = \Im(s, 0, p_3^{(k)})$ are used to calculate the propagation and attenuation directions of reflected and refracted waves as follows.

$$\sin\theta_k = |\vec{P}^{(k)}|^{-1}\Re s,\ \sin(\theta_k - \gamma_k) = |\vec{A}^{(k)}|^{-1}\Im s. \tag{5.16}$$

6. Anisotropic Elastic Medium (AEM)

Every medium is anisotropic upto some extent. Isotropy is an ideal assumption made only for mathematical convenience. In AEM with anisotropy up to azimuthal isotropy, the propagation will confine to the symmetry plane only. Else, the energy propagation in a general anisotropic medium is a three-dimensional phenomenon. Wave motion in such a medium is not confined to a particular plane. For propagation in a symmetry plane, the quasi-shear waves may be called as qSP (or qSR) waves, depending upon the particle motion parallel (or normal) to the symmetry plane. The prefix $'q'$ ($\equiv$ quasi) refers to the particle motion when polarization is not exactly longitudinal or transverse. The notations qSV and qSH are also used for near-vertical and near-horizontal polarised shear waves, respectively ([16]). These notations may be adequate only for propagation in a symmetry plane but are not suitable in off-symmetry directions. In case of general anisotropy without any symmetry, three coupled body waves propagate along a general direction in three-dimensional space. These waves are identified as qP, $qS1$ and $qS2$, where $qS1$ is faster among two split-shear waves ([22]).

For the propagation of harmonic (angular frequency ω) plane waves, we write

$$u_j = S_j \exp\{\imath\omega(p_k x_k - t)\}, \ (j = 1, 2, 3). \tag{6.1}$$

In slowness vector $(p_1, p_2, p_3) = \vec{p} = \hat{n}/V$, the phase velocity $V(\hat{n})$ will be a function of the propagation direction $\hat{n}$. The vector (S_1, S_2, S_3) defines the polarisation of material particles. For these displacements, the equations of motion in AEM yield an eigen-system, say

$$(A_{ik} - V^2\delta_{ik})u_k = 0; \ A_{ik} = c_{ijkl}n_j n_l/\rho, \tag{6.2}$$

where c_{ijkl} is elastic moduli tensor for the medium of density ρ. Note that the same system, called Christoffel system, can also be obtained from the Fourier transform of equations of motion. For non-trivial solution, this system is solved into a cubic equation in V^2, say

$$C_3V^6 + C_2V^4 + C_1V^2 + C_0 = 0. \tag{6.3}$$

Three roots of this equation define the phase velocities (say, $V_1 > V_2 > V_3$) of three waves (qP, $qS1$, $qS2$) propagating along $\hat{n}$ in AEM. Hence, three real eigen-values (orthogonal eigen-vectors) of the Christoffel system define the velocities (polarisations) of qP, $qS1$ and $qS2$ waves. Note that A_{ij} involve n_j. This implies that the velocities/polarisations of the waves depend on the direction of propagation ($\hat{n}$). We see that the quasi-longitudinal (qP) wave propagate with particle motion not exactly (but nearly) parallel to $\hat{n}$. Similarly, the two quasi-transverse ($qS1$, $qS2$) waves propagate with particle motion not exactly (but nearly) normal to $\hat{n}$.

6.1. **Reflection.** In Cartesian coordinate system (x_1, x_2, x_3), AEM occupies a half-space, which is bounded by the stress-free plane boundary $x_3 = 0$. In this medium, we consider the propagation of a quasi-longitudinal (or quasi-transverse) wave along $\hat{n}$, which become incident at the plane boundary $x_3 = 0$. This incidence results in three reflected (qP, $qS1$, $qS2$) waves, propagating along different directions ($\hat{n}^{(R)}$) in AEM.

For incident wave: $\vec{u}^{(I)} = \vec{S}^{(I)} \exp\{\imath\omega(p_k x_k - t)\}$, $p_k = n_k/V_I(n_k)$, $I = qP$ or $qS1$ or $qS2$. For reflected waves: $\vec{u}^{(R)} = \vec{S}^{(R)} \exp\{\imath\omega(p_k^{(R)} x_k - t)\}$, $p_k^{(R)} = n_k^{(R)}/V_R(n_k^{(R)})$, $R = qP$, $qS1$, $qS2$.

In (x_1, x_2, x_3) space, let θ is the polar angle with x_3-axis and ϕ is the azimuth in the x_1-x_2 plane measured from the x_1-axis to the x_2-axis. So, for an incident wave propagating along (θ_I, ϕ_I), $\hat{n} = (\sin\theta_I \cos\phi_I, \sin\theta_I \sin\phi_I, \cos\theta_I)$ represents the direction of incidence. This $\hat{n}$ is used to calculate $V_I(\hat{n})$ as a root of the cubic equation (6.3). Boundary conditions are the vanishing of stresses (at $x_3 = 0$), which yield Snell's law as $p_1 = p_1^{(R)}$, $p_2 = p_2^{(R)}$. In terms of velocities and phase directions, we have

$$\frac{n_i}{V_I(\vec{n})} = \frac{n_i^{(R)}}{V_R(\vec{n}^{(R)})}, \ (i = 1, 2), \tag{6.4}$$

where, the superscript R represents each of the three waves $(qP, qS1, qS2)$ in AEM. This form of Snell's law helps to deduce the following points.
i) $n_2/n_1 = n_2^{(R)}/n_1^{(R)}$, imply that $\phi_R = \phi_I$. This implies that the phase directions of all the reflected waves lie in the same vertical plane (fixed ϕ), which contains the phase direction of incident wave. Then, for fixed ϕ_I, Snell's law reduces to $\frac{\sin\theta_R}{V_R(\theta_R,\phi_I)} = \frac{\sin\theta_I}{V_I(\theta_I,\phi_I)}$.
ii) Snell's law and $n_k^{(R)} n_k^{(R)} = 1$, yield

$$V_R^2 \sin^2\theta_I - V_I^2 \sin^2\theta_R = 0. \tag{6.5}$$

For chosen $(\theta_I,\ \phi_I)$ and thus calculated $V_I(\theta_I, \phi_I)$, the equation (6.5) relates θ_R and V_R^2.
iii) Recall that, the cubic equation (6.3) from Christoffel system (6.2) also relates V_R^2 to θ_R (for known $\phi_R = \phi_I$). On using this relation in (6.5), we get a transcendental equation in θ_R, which is solved for (real θ_R) using a numerical method (bisection method or Newton's method). Then, this real value of θ_R alongwith ϕ_R $(= \phi_I)$ is used to calculate the (real) phase velocity V_R of reflected wave.
iv) It may be noted that polar angle θ_R for each reflected wave is derived from θ_I. As incidence reaches the critical angle for any of the reflected waves, this reflected wave propagates along the boundary with velocity $v_c = V_R(\pi/2,\ \phi_I)$. Such a critical angle is determined from the non-linear equation

$$v_c \sin\theta_I - V_I(\theta_I,\ \phi_I) = 0. \tag{6.6}$$

This equation is solved for θ_I, with a given value of ϕ_I, numerically, for each of the reflected waves. A valid root of this equation, say θ_c, divides the incidence plane into two parts. For incidence beyond this angle (i.e., $\theta_I \geq \theta_c$), no real θ_R is obtained through step iii). Then, we will not be able to calculate the direction and hence velocity of reflected waves for all incident directions. That means, the suggested procedure calculates the directions and velocities of reflected waves provided the incidence guarantees the homogeneous reflected waves (i.e., pre-critical incidence).

6.2. **Inhomogeneous (generalised) reflection.** Incident wave is chosen with a propagation direction $\hat{n}$. It is used to calculate velocity $V_I = V(\hat{n})$ from (6.3) and then slowness $p_k = n_k/V_I,\ (k = 1, 2, 3)$, of the incident wave. The vector $\vec{u} = \vec{S}\exp\{\imath\omega(p_1x_1 + p_2x_2 + p_3x_3 - t)\}$ defines the displacement for the incident wave. According to Snell's law, the tangential slowness (p_1, p_2) of the incident wave should remain same for all the reflected waves as well. Then, displacement for the any wave in AEM is given by

$$\vec{u} = \vec{S}\exp\{\imath\omega(p_1x_1 + p_2x_2 + qx_3 - t)\}, \tag{6.7}$$

where p_1 and p_2 are known (from incident wave). The unknown q represents the vertical slowness, which should identify the different waves at the boundary.

Using the displacement (6.7) in equations of motion yields the Christoffel system of equations in vertical slowness $'q'$. Non-trivial solution of this system requires q to satisfy an algebraic equation of degree six. Roots of this equation (i.e., six values of q) define the existence of six waves at the boundary of AEM. For three (out of six) values of q (say q_1, q_2, q_3), the directions of propagation are found to be away from the boundary and hence identify the three reflected waves. One of the remaining three q's is $p_3 = n_3/V_I$, the vertical slowness of the incident wave. For each of the three reflected waves, the relevant value of q is combined with p_1 and p_2 to define its slowness vector. So, the directions of propagation of the reflected waves are obtained from

$$\hat{n}^{(R)} = (p_1, p_2,\ q_j)/\sqrt{p_1^2 + p_2^2 + q_j^2},\ (j = 1, 2, 3). \tag{6.8}$$

The velocities of reflected waves along their propagation directions are given by

$$V_j = 1/\sqrt{(p_1^2 + p_2^2 + q_j^2)}\ (j = 1, 2, 3). \tag{6.9}$$

6.3. **Reflection & refraction.** The previous section illustrates the procedure to define the reflected waves when incidence takes place in AEM. The tangential slowness of incident wave (p_1, p_2) is known. Snell's law ensures the same (p_1, p_2) for refracted waves transmitting to the continuing medium, which may be isotropic or anisotropic. In case of isotropic refraction, this (p_1, p_2) are used directly to calculate the directions of propagation of refracted waves propagating with known (constant) velocities. For refraction in AEM, this (p_1, p_2) are used to calculate the six q-values and then identify the three slowness vectors for refracted waves (propagating away from boundary). On the other hand, for incidence from an isotropic elastic medium, the incidence direction and constant velocity defines the horizontal slowness (p_1, p_2) for the incident wave as well as the reflected and refracted waves. Again, this (p_1, p_2) are used to calculate the slowness vectors for refracted waves in AEM.

7. Anisotropic Viscoelastic Medium (AVM)

For propagation of plane waves in a viscoelastic anisotropic medium, various relations in the section 6 remain the same but for different definitions of variables and constants. The viscoelastic properties of the medium are represented through the complex values for the elastic constants c_{ijkl}. The density of the medium and the circular frequency retain their real values. In a viscoelastic medium, the components of the slowness vector (p_1, p_2, p_3) are, in general, complex. In order to keep the expressions (6.1) for the displacement components, we write the slowness vector $\vec{p} = \vec{N}/V$, where $\vec{N} = (N_1, N_2, N_3)$ is the complex (dual) vector such that $N_j N_j = 1$. The variable $V = 1/\sqrt{p_j p_j}$ gets a complex value with the dimension of velocity. With $\vec{p} = \vec{N}/V$, the coefficients in the cubic equation (6.3) become the complex functions of $\vec{N}$. Three roots of this equation define the complex velocities $V = V(\vec{N})$ for the three attenuated waves in AVM. The corresponding complex eigensystem (6.2) define the complex (elliptic) polarizations of these waves.

Let two unit vectors $\hat{n}$ and $\hat{a}$, in real space, denote the propagation direction and direction of maximum attenuation, respectively. In terms of these real unit vectors, the complex slowness vector for an attenuated wave is expressed as

$$\vec{p} = \frac{1}{v}\hat{n} + \imath\Lambda\hat{a}, \tag{7.1}$$

where, v denotes the phase velocity and Λ is the attenuation coefficient. The angle (γ) between the propagation direction and the attenuation direction is defined as

$$\hat{n}.\hat{a} = \cos\gamma. \tag{7.2}$$

The attenuated plane wave is called homogeneous for $\gamma = 0$ and inhomogeneous for $\gamma \neq 0$. However, for a given $\hat{n}$, an arbitrary $\hat{a}$ (or γ) may yield a negative value for v. This is considered a forbidden direction for anisotropic propagation in dissipative media ([8]).

7.1. **Inhomogeneous propagation.** To avoid forbidden directions (due to arbitrary attenuation angle γ), wave inhomogeneity is defined through a non-dimensional parameter $\delta \in [0, 1)$ ([23]). $\delta = 0$ represents the homogeneous wave. For a chosen propagation direction $\hat{n}$, an orthogonal direction $\hat{m}$ and the inhomogeneity parameter δ, we write

$$\vec{p} = \frac{1}{v}[\hat{n} + \imath\eta\hat{n} + \imath\delta\hat{m}] = \vec{P} + \imath\vec{A}, \tag{7.3}$$

where the attenuation coefficient η and the phase velocity v are to be determined. In relation to (7.1), the attenuation part, given by

$$\Lambda\hat{a} = (\frac{\eta}{v}\hat{n} + \frac{\delta}{v}\hat{m}), \tag{7.4}$$

implies total attenuation (Λ) as the vector sum of two orthogonal vectors. The part $(\eta/v)\hat{n}$ defines the attenuation along $\hat{n}$, i.e. homogeneous wave. Other part $(\delta/v)\hat{m}$ represents inhomogeneous propagation.

7.2. **Reflection.** Let the AVM half-space is bounded above by the plane $x_3 = 0$ in right-hand Cartesian coordinate system (x_1, x_2, x_3). Let x_3-axis be the outward normal to its boundary. A unit vector $\hat{n} = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)$ defines the propagation of a wave in the medium along $(\theta,\ \phi)$.

7.2.1. *$\vec{p}$ for inhomogeneous incident wave.* Displacement for incident wave is given by $\vec{u} = \vec{S}\exp\{\imath\omega(p_1x_1 + p_2x_2 + p_3x_3 - t)\}$. For $\vec{p} = \vec{N}/V$, the resulting Christoffel system is solved into a complex cubic equation, which provides the complex velocities $(V_j(\vec{N}),\ j = 1, 2, 3)$ for three attenuated waves in AVM. So, we have a system of ten real unknowns in $\vec{N},\ V,\ v,\ \eta$, which are related through the ten real scalar relations, given by

$$\vec{p} = \frac{\vec{N}}{V} = \frac{1}{v}[\hat{n} + \imath\eta\hat{n} + \imath\delta\hat{m}],\ \vec{p}.\vec{p} = 1/V^2,\ V = V(\vec{N}). \tag{7.5}$$

These relations are solved to obtain

$$\begin{aligned} \vec{N} &= [\hat{n}(1+\imath\eta) + \imath\delta\hat{m}]/\sqrt{[(1+\imath\eta)^2 - \delta^2]}, \\ \frac{v^2}{V^2} &= (1+\imath\eta)^2 - \delta^2. \end{aligned} \tag{7.6}$$

The phase vector $\vec{N}$ is a vector function of η. This implies that $V^2 = F(\eta)$, for some complex function F. Then, we get

$$v^2 = F(\eta)[(1+\imath\eta)^2 - \delta^2]. \tag{7.7}$$

Since v is real, the imaginary part of the right hand side in equation (7.7) vanishes, i.e.,

$$Im\{F(\eta)[(1+\imath\eta)^2 - \delta^2]\} = 0. \tag{7.8}$$

The real equation (7.8) is solved numerically for the value of real unknown η. Thus obtained η is used in the real part of the equation (7.7) to calculate the value of v^2 and hence v. From (7.7) and (7.8), the relation

$$v^2 = [(1-\eta^2-\delta^2)^2 + 4\eta^2]Re(V^2)/(1-\eta^2-\delta^2), \tag{7.9}$$

yields a positive v^2, provided $\eta^2 < 1 - \delta^2$. Sign of δ does not affect the values of η and v. Thus, $0 \le \delta < 1$ ensures that both η and v are real.

Finally, an incident wave in AVM is specified with a propagation direction $\hat{n}$, orthogonal direction $\hat{m}$ and inhomogeneity parameter δ. The values of η and v, as calculated above, are used in (7.3) to derive the complex slowness vector $\vec{p}$ for the (in)homogeneous incident wave.

7.2.2. *$\vec{p}$ for reflected waves.* According to Snell's law, the reflected waves keep the tangential slowness of the incident wave. That means, the vector $(p_1, p_2,\ q)$ defines the slowness for a reflected wave. The values of q come from the roots of a six-degree algebraic equation, which is obtained from the Christoffel system for AVM in vertical slowness (q). Hence, the six values of q are obtained for the values of p_1 and p_2 extracted from the $\vec{p}$, which represents the incident wave.

i) For propagation in a half-space, only three among these six q-values correspond to the three reflected waves (decaying with distance from boundary).

ii) For each of these three values of q, the slowness vector $\vec{p}^{(R)} = (p_1, p_2,\ q)$ is resolved as $\vec{N}/V_R$ so that $\vec{N}.\vec{N} = 1$.

iii) For this $\vec{N}$, complex velocities for three waves in the medium come from the cubic equation in V^2, which is obtained from Christoffel system for AVM in V.

iv) Then, one of these three velocities must be equal to V_R. This associates the slowness vector $\vec{p}^{(R)}$ to a particular wave-type.
v) For each reflected wave, the corresponding $\vec{p}^{(R)} = \frac{1}{v}[\hat{n} + \imath\eta\hat{n} + \imath\delta\hat{m}]$ is resolved to find its phase velocity $v = 1/|\Re(\vec{p}^{(R)})|$, propagation direction $\hat{n} = v\Re(\vec{p}^{(R)})$ and attenuation $(\eta\hat{n} + \delta\hat{m})/v$.

7.3. **Reflection & refraction.** The phenomenon of reflection and refraction is considered at the interface of AVM with an isotropic viscoelastic medium. For incidence in AVM, the section 7.2 illustrates the procedure to calculate the slowness vectors of three reflected waves at the boundary of AVM. Snell's law ensures that the tangential slowness (p_1, p_2) of incident wave remains valid for refracted waves as well. With known (p_1, p_2), the derivations in section 5 are used to derive the velocities and propagation/attenuation directions of the refracted waves. On the other hand, any incidence from the continuing isotropic medium provides (p_1, p_2), which are used to calculate the six q-values in AVM. The resulting slowness vectors can be analysed to identify the three refracted waves (decaying away from the boundary). Combining these two situations, we can study the reflection and refraction at the boundary between two dissimilar anisotropic viscoelastic media.

8. Concluding Remarks

In this study, Snell's law is illustrated in relating the propagation characteristics of reflected/refracted waves to those of the incident wave. Medium for propagation starts with an ideal one, i.e., isotropic elastic, and proceeds upto the real one, i.e., anisotropic viscoelastic composite. Anisotropy represents the presence of pre-stresses, aligned cracks, asymmetric pores, fibre-reinforcement, piezoelectricity, lamination or periodic thin layering. Viscoelasticity accounts for dissipation from the presence of skeletal-frame anelasticity, pore-fluid viscosity or thermal relaxation. Various derivations in this study can be extended / modified to study the wave propagation at the boundaries of poroelastic as well as thermoelastic media.

Any presence of saturated pores in an isotropic elastic solid is represented through the propagation of an additional dilatational (longitudinal) wave. That means, an isotropic poroelastic solid supports the propagation of two dilatational (fast P or P_f and slow P or P_s) waves alongwith a shear (or S) wave. Velocities of these waves are derived from elastic constants and densities ([27]). The relations emerging from Snell's law for isotropic elastic solids extend to the additional wave as well. Hence, for incidence in poroelastic solids, we have three options for incident wave, i.e., P_f, P_s or S wave. Consequently, there will be three reflected / refracted waves in poroelastic media.

In an analogous manner, anisotropic poroelastic media support the propagation of four waves ([24], [25]). Two of these waves are quasi-longitudinal (qP_1, qP_2)

and two are quasi-transverse (qS_1, qS_2). Propagation of these waves is governed by a modified Christoffel system same as (6.2) but is solved into a quartic equation in V^2. Similarly, the Christoffel system for vertical slowness (q) translates into an algebraic equation of degree eight ([26]). However, the procedures to calculate the velocities, propagation directions and attenuations remain unaltered but to accommodate the propagation of additional wave. The replacement of slow-P wave in porous solids with a thermal diffusive phase in thermoelastic media yields the thermoelastic-poroelastic analogy. Hence, the procedure applicable to poroelastic propagation can be extended to thermoelastic propagation as well.

References

[1] Biot, M. A., Thermoelasticity and irreversible thermodynamics, J. Appl. Phys. 27 (1956), 240-253.

[2] Borcherdt, R. D., Reflection and refraction of type-II S waves in elastic and anelastic media, Bull. Seism. Soc. Am. 67 (1977), 43-67.

[3] Borcherdt, R. D., Reflection-refraction of general P and type-I S waves in elastic and anelastic solids, Geophys. J. R. astr. Soc. 70 (1982), 621-638.

[4] Borcherdt, R. D., Glassmoyer, G. and Wennerberg, L. Influence of welded boundaries in anelastic media on energy flow and characteristics P, S-I and S-II waves: Observational evidence for inhomogeneous body waves in low-loss solids, J. geophys. Res. 91 (1986), 11503-11518.

[5] Braga, A. M. B., Wave Propagation in Anisotropic Layered Composites, Ph.D. dissertation, Stanford University Stanford CA, 1990.

[6] Carcione, J. M., Wave fields in real media: Wave propagation in anisotropic anelastic, porous and electromagnetic media, Elsevier, Amsterdam, 2007.

[7] Carcione, J. M., The effects of vector attenuation on AVO of off-shore reflections, Geophysics 64 (1999), 815-819.

[8] Carcione, J. M. and Cavallini, F. Forbidden directions for inhomogeneous pure shear waves in dissipative anisotropic media, Geophysics 60 (1995), 522-530.

[9] Caviglia, G. and Morro, A., Harmonic waves in thermoviscoelastic solids, Int. J. Engng. Sci. 43 (2005), 1323-1336.

[10] Cerveny, V. and Psencik, I., Plane waves in viscoelastic anisotropic media -I. Theory, Geophys. J. Int. 161 (2005), 197-212.

[11] Cerveny, V. and Psencik, I., Plane waves in viscoelastic anisotropic media. Part 2: Numerical examples, Geophys. J. Int. 161 (2005), 213-229.

[12] Crampin, S., Geological and industrial implications of extensive-dilatancy anisotropy, Nature 328 (1987), 491-496.

[13] Fan, C. W. and Hwu, C., Rigid stamp indentation on a curvilinear hole boundary of an anisotropic elastic body, J. Appl. Mech. 65 (1998), 389-397.

[14] Gurevich, B., Elastic properties of saturated porous rocks with aligned fractures, J. Appl. Geophys. 54 (2003), 203-218.

[15] Gurevich, B., Bisegna, P. and Luciano, R., On methods for bounding the overall properties of periodic piezoelectric fibrous composites, J. Mech. Phys. Solids 45 (1997), 1329-1356.

[16] Keith, C. M. and Crampin, S., Seismic body waves in anisotropic media : reflection and refraction at a plane interface, Geophys. J. R. astr. Soc. 49 (1977), 181-208.

[17] Norris, A. N., On the correspondence between poroelasticity and thermoelasticity, J. Appl. Phys. 71 (1992), 1138-1141.

[18] Norris, A. N. and Johnson, D. L., Nonlinear elasticity of granular media, J. Appl. Mech. 64 (1997), 39-49.

[19] Paulsson., B. N. P., Meredith, J. A., Wang, Z. and Fairborn, W., The Steepbank crosswell seismic project : Reservoir definition and evaluation of steamflood technology and Alberta tar sands, Lead. Edge 13 (1994), 737-747.

[20] Sharma, M. D. Propagation of inhomogeneous plane waves in dissipative anisotropic poroelastic solids, Geophys. J. Int. 163 (2005), 981-990.

[21] Sharma, M. D., Wave propagation in anisotropic generalized thermoelastic medium, J. Therm. Stresses 29(2006), 329-342.

[22] Sharma, M. D., Group velocity along a general direction in a general anisotropic medium, Int. J. Solids Struct. 39 (2002), 3277-3288.

[23] Sharma, M. D., Propagation of inhomogeneous plane waves in viscoelastic anisotropic media, Acta Mech., 200 (2008), 145-154.

[24] Sharma, M. D., Three-dimensional wave propagation in a general anisotropic poroelastic medium: Phase velocity, group velocity and polarisation, Geophys. J. Int. 156 (2004), 329-344.

[25] Sharma, M. D., Wave propagation in a general anisotropic poroelastic solid with anisotropic permeability, Int. J. Solids Struct. 41 (2004), 4587-4597.

[26] Sharma, M. D., Propagation of harmonic plane waves in a general anisotropic porous solid, Geophys. J. Int. 172 (2008), 982-994.

[27] Sharma M. D. and Gogna, M. L., Propagation of elastic waves in a cylindrical bore in a liquid saturated porous solids, Geophys. J. Int., 103 (1990), 47-54.

M. D. Sharma
Department of Mathematics
Kurukshetra University, Haryana, India-136 119
Email: *mdsharma@kuk.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 95-110

THE SADDLE POINT METHOD AND ITS APPLICATIONS

KANNAPPAN SAMPATH

(Received : 12 - 05 - 2015 ; Revised : 03 - 09 - 2015)

Abstract. We study the method of steepest descent useful in deriving asymptotic expansions of Laplace-type integrals. We hope that this exposition will enable presenting this technique in an undergraduate course thereby making the diverse applications of this technique more accessible.

1. Introduction

In many physical and mathematical problems, it is desirable to understand the asymptotic behaviour of functions as the independent parameter tends to a limit. These functions may arise as solutions to differential equations, as integral transforms of functions; or perhaps arise from combinatorial considerations as generating functions of different kinds.

In his memoir "Théorie Analytique des Probabilités" in 1812, Laplace made a fundamental advance in this subject with what appears to be the most natural strategy in retrospect: he postulates that the largest contribution to an integral must come from neighbourhoods of points where the integrand attains its maximum. He also suggested the discrete analogue of this philosophy: the asymptotic behaviour of a series of positive terms in which the terms steadily increase upto a certain point and then steadily decrease can be obtained by studying the order of magnitude of the largest term in the series.

An extension of this philosophy to the complex plane was found by Siegel in some unpublished manuscripts of Riemann to study the error term in the approximate functional equation for the Riemann zeta function. Riemann [25, p. 428] had also used this method in his study of hypergeometric functions. Debye [11, p. 583] used this method to derive asymptotic expansions of Bessel functions of higher order.

Voluminous literature has since evolved that surveying the historical developments in these few pages is out of the question. One of the points to be emphasized is that these methods are not cookbook recipes but rather an indication of the

2010 Mathematics Subject Classification: 30E15, 41A60.
Keywords and Phrases: Laplace method, saddle point method, method of steepest descent, Stirling's formula.

ideas that should be brought to bear on the problem of determining asymptotic expansions.

In order to keep the exposition reasonably short, our examples are chosen so that they are easy to grasp and devoid of many subtleties. For a more extensive study, we suggest the excellent accounts by Olver [20], Wong [32], Bender and Orszag [2]. Our exposition is self-contained and is inspired by [15] and [32] beginning with the basic definitions in Section 2 and a discussion of Laplace's Method in Section 3. The method of steepest descent is discussed at length in Section 4 closing with an illustration of this method by proving Stirling's formula for the complex Gamma function in Section 5.

Acknowledgement: I would like to thank Prof. Ram Murty for initiating me into this study, guiding me with references and providing me with very insightful remarks that improved the clarity of this exposition. Various parts of this exposition is also a part of my masters' project report[1] written under the supervision of Prof. Ram Murty. In this connection, it is a pleasure to thank Prof. Jamie Mingo and Prof. Ivan Dimitrov for their comments on the report (and consequently, this paper itself).

2. Preliminaries

2.1. **Asymptotic Expansions.** A function $f(z)$ which is analytic at $z = \infty$ can be expanded in its Laurent series, more plainly, it is given by a power series in z^{-1} converging in some annulus of the form $|z| > R$. Recall the elementary yet remarkable property of the partial sums $p_n(z)$ of the Laurent series of $f(z)$:

$$f(z) = p_n(z) + o(1) \tag{2.1}$$

not only as $n \to \infty$ but also as $|z| \to \infty$ for a fixed n. Such expansions do not exist when the function f is not analytic at $z = \infty$. However, there are *asymptotic expansions* which do a similar job: these are series in z^{-1} which are not convergent, but, for a fixed n and for z coming from a certain sector in the complex plane, we do have that

$$f(z) = p_n(z) + O(z^{-(n+1)}), \quad |z| \to \infty.$$

Curious and misleading terms such as "divergent" series, "semi-convergent" series[2] and "convergently beginning" series[3] have often been applied to these series and consequently, there has been skepticism from leading mathematicians in the past in discussing these series. We shall however not pursue these historical remarks further.

The definition of an asymptotic expansion that is now widely used is commonly attributed to Poincaré (1886).

[1] The report is available on my homepage: `http://www.mast.queensu.ca/~knsam/`

[2] Due to Stieltjes, 1856–1894.

[3] Due to Fritz Emde, 1873–1951.

Definition 1. Let $f : \Omega \to \mathbf{C}$ be functions on a domain $\Omega \subseteq \mathbf{C}$. We say that

$$\sum_{n=0}^{\infty} a_n z^{-n} \tag{2.2}$$

is an asymptotic expansion for the function f as $|z| \to \infty$ and write

$$f(z) \sim \sum_{n=0}^{\infty} a_n z^{-n} \tag{2.3}$$

if there is a sector S in the complex plane such that for every $n \geqslant 0$, we have

$$f(z) = \sum_{k=0}^{n} a_k z^{-k} + o(z^{-n}), \quad |z| \to \infty \text{ in } \Omega \cap S. \tag{2.4}$$

This is to be interpreted in the obvious degenerate manner if the function f is real valued.

The usefulness of an asymptotic expansion is *not* its (lack of!) convergence properties but the fact that, for every $n \geqslant 0$, the error incurred in truncating an asymptotic expansion after first $n-1$ terms is $O(z^{-n})$. Indeed, we have from (2.4) that

$$f(z) = \sum_{k=0}^{n-1} a_k z^{-k} + O(z^{-n}) \text{ as } |z| \to \infty \text{ in } \Omega \cap S. \tag{2.5}$$

To amplify the difference between convergent series and asymptotic expansions, we note that, for a fixed z, it is usually the case that the implied constant in (2.5) grows indefinitely as $n \to \infty$ so that it is typically not possible to determine $f(z)$ with arbitrary precision by adding more terms of the expansion.

2.2. **Uniqueness of asymptotic expansions.** *If a function f admits an asymptotic expansion, the coefficients $\{a_n\}$ of that expansion are determined by f.* Indeed, we have the following recursive formula

$$a_0 = \lim_{z\to\infty} f(z) \text{ and} \tag{2.6}$$

$$a_m = \lim_{z\to\infty} z^m \left(f(z) - \sum_{n=0}^{m-1} a_n z^{-n} \right) \text{ for } m > 0. \tag{2.7}$$

However, two different functions may have the same asymptotic expansion: for example, the function $e^{-|z|}$ has the identically zero asymptotic expansion: $e^{-|z|} \sim 0$ as $|z| \to \infty$ in a sector of the form $\{z \in \mathbf{C} : |\arg(z)| < \pi/2 - \delta < \pi/2\}$.

Notation. Typically, it is the case that a function f by itself may not admit an asymptotic expansion as defined above but there is a function g such that f/g admits an asymptotic expansion. In such cases, we will write

$$f(z) \sim g(z) \times (\text{asymptotic expansion of } f/g)$$

by an abuse of notation.

For a discussion of the calculus of asymptotic expansions, see the texts cited in the introduction.

2.2.1. *An Example.* Here is a simple example to illustrate the definition.

Example 2.1 (Exponential Integral)**.** Consider the integral[4]

$$E_1(x) = \int_x^\infty \frac{e^{-z}}{z}\,\mathrm{d}z, \quad |\arg(x)| < \pi$$

where the integral can be taken over any path from x to ∞ (since the integrand is holomorphic on the given domain). A repeated integration by parts proves that

$$E_1(x) = \frac{e^{-x}}{x}\sum_{k=0}^{n}\frac{(-1)^k k!}{x^k} + (-1)^{n+1}(n+1)!\int_x^\infty \frac{e^{-z}}{z^{n+2}}\,\mathrm{d}z. \tag{2.8}$$

We claim that, in the sector $|\arg(x)| \leqslant \pi - \delta < \pi$ (for any $\delta > 0$)

$$E_1(x) \sim \frac{e^{-x}}{x}\sum_{k=0}^{\infty}\frac{(-1)^k k!}{x^k}. \tag{2.9}$$

To prove this, we need an estimate on the integral in (2.8). Fix an $n \geqslant 0$ and let $R > 0$ be given. Let $x = \sigma + it$ such that $|x| > R$ and $|\arg(x)| < \pi$ be arbitrary. We use the ray $\Im(z) = t$, $\Re(z) \geqslant \sigma$ extending from σ to ∞ parallel to the real line as the path of integration for the integral in (2.8). On this path $|z| > t(= |x|\sin\delta)$ where $0 < \delta < \pi$ is such that $|\arg(x)| < \pi - \delta$. Thus, we have

$$\left|(-1)^{n+1}(n+1)!\int_x^\infty \frac{e^{-z}}{z^{n+2}}\,\mathrm{d}z\right| \leqslant \frac{(n+1)!}{(\sin\delta)^{n+2}}\frac{|e^{-x}x^{-1}|}{|x|^{n+1}}.$$

This completes the proof.

2.3. **Watson's Lemma.** We recall the "extended" probability integral: for $x \in \mathbf{C}$ with $\Re(x) > 0$, we have

$$I(x) := \int_{-\infty}^{\infty}\exp\left(-\frac{xt^2}{2}\right)\,\mathrm{d}t = \sqrt{\frac{2\pi}{x}} \tag{2.10}$$

where $x^{1/2}$ is its principal value. To see this "extended" version, one may argue as usual to obtain that $I(x)^2 = \frac{2\pi}{x}$; it suffices to note that $I(x)$ is analytic in the open right half-plane $\{z \in \mathbf{C} : \Re(z) > 0\}$ by Morera's theorem and the principle of analytic continuation finishes the proof.

Lemma 2.1 (Watson's Lemma, Elementary Version)**.** *Let $g(z)$ be analytic and bounded on a domain containing the real line and let $x \in \mathbf{C}$ be fixed with $\Re(x) > 0$. Writing a_n for the Taylor coefficients of g around 0 so that*

$$g(z) = \sum_{n=0}^{\infty} a_n z^n, \tag{2.11}$$

we have the following expansion as $|x| \to \infty$

[4]**Important Notation:** Here and in what follows, all instances of 'arg' mean the principal branch of the argument.

$$\int_{-\infty+0i}^{\infty+0i} \exp\left(-\frac{x^2z^2}{2}\right) g(z)\,\mathrm{d}z \sim \sqrt{2\pi}\left(\sum_{n=0}^{\infty}\frac{a_{2n}}{x^{2n+1}}\prod_{k=1}^{n}(2k+1)\right) \tag{2.12}$$

in the open right half-plane $\{x \in \mathbf{C} : |\arg(x)| < \frac{\pi}{2}\}$.

Proof. We begin by noting that

$$\frac{g(z) - \sum_{k=0}^{2n-1} a_k z^k}{z^{2n}} \tag{2.13}$$

is bounded and analytic in a domain containing the real axis.

For $n \geqslant 2$, integration by parts gives us

$$\begin{aligned}
&\int_{-\infty}^{\infty} \exp\left(-\frac{x^2z^2}{2}\right) z^n\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} z^{n-1}\exp\left(-\frac{x^2z^2}{2}\right) z\,\mathrm{d}z \\
&= z^{n-1}\frac{-1}{x^2}\exp\left(-\frac{x^2z^2}{2}\right)\Big|_{z=-\infty}^{z=\infty} + \frac{n-1}{x^2}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right) z^{n-2}\,\mathrm{d}z.
\end{aligned}$$

Now, the first term vanishes thanks to the negative exponential and our hypothesis that $\Re(x) > 0$. Rearranging what remains, we get, for $n \geqslant 2$

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2z^2}{2}\right) z^n\,\mathrm{d}z = \frac{n-1}{x^2}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right) z^{n-2}\,\mathrm{d}z. \tag{2.14}$$

Now, plugging $2n$ for n, we have the following for $n \geqslant 1$

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2z^2}{2}\right) z^{2n}\,\mathrm{d}z = \frac{2n-1}{x^2}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right) z^{2n-2}\,\mathrm{d}z. \tag{2.15}$$

By induction on n, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} \exp\left(-\frac{x^2z^2}{2}\right) z^{2n}\,\mathrm{d}z &= \frac{\prod_{k=1}^{n}(2k-1)}{x^{2n}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right)\mathrm{d}z \\
&= \frac{\sqrt{2\pi}\prod_{k=1}^{n}(2k-1)}{x^{2n+1}}.
\end{aligned}$$

Now, on the one hand, we have

$$\left|\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right)\left(g(z) - \sum_{k=0}^{2n-1} a_k z^k\right)\mathrm{d}z\right| \tag{2.16}$$

$$= O(1)\left|\int_{-\infty}^{\infty}\exp\left(-\frac{x^2z^2}{2}\right) z^{2n}\,\mathrm{d}z\right|$$

$$= O(1)\frac{\sqrt{2\pi}\prod_{k=1}^{n}(2k-1)}{|x|^{2n+1}}, \tag{2.17}$$

while on the other hand, the integrals in (2.16) vanish for odd powers of z leaving us with the following expression for the left hand side of (2.16)

$$\left| \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 z^2}{2}\right) g(z)\,\mathrm{d}z - \sum_{k=0}^{n-1} \frac{a_{2k}\sqrt{2\pi}}{x^{2k+1}} \prod_{m=1}^{k} (2m-1) \right|. \tag{2.18}$$

Multiplying both sides of (2.17) by x^{2n} and letting $|x| \to \infty$, the desired result follows. □

Corollary 2.1. *Let $a, b \in \mathbf{R} \cup \{\pm\infty\}$ such that $-\infty \leqslant a < 0 < b \leqslant \infty$. Let $g(z)$ be bounded and analytic in a domain containing the real line and let x be as in Lemma 2.1. Then, we have the following asymptotic expansion as $|x| \to \infty$*

$$\int_a^b \exp\left(-\frac{x^2 z^2}{2}\right) g(z)\,\mathrm{d}z \sim \sqrt{2\pi} \left(\sum_{n=0}^{\infty} \frac{a_{2n}}{x^{2n+1}} \prod_{k=1}^{n} (2k+1) \right) \tag{2.19}$$

in any sector $\{x \in \mathbf{C} : |\arg(x)| \leqslant \frac{\pi}{2} - \delta\}$ $(\delta > 0)$.

Proof. This follows from the following classical estimate for the tail of the normal distribution

$$\int_b^{\infty} e^{-\frac{t^2}{2}}\,\mathrm{d}t \leqslant \int_b^{\infty} \frac{t}{b} e^{-\frac{t^2}{2}} = \frac{1}{b}\left[-e^{-\frac{t^2}{2}}\right]_{t=b}^{t=\infty} = \frac{1}{b} e^{-\frac{b^2}{2}} \tag{2.20}$$

The details are routine. □

2.3.1. *Simple examples.*

Example 2.2 (Some Examples due to Ramanujan)**.** Here are two examples taken from [3] due to Ramanujan:

$$\int_1^{\infty} e^{-au} \log u\,\mathrm{d}u \sim \sum_{k=0}^{\infty} \frac{(-1)^k k^k}{a^{k+1}}$$

as $a \to \infty$. This follows easily from a generalised Watson's Lemma ([3]). For the second example, we define

$$I(\alpha) = \alpha^{-1/4} \left(1 + 4\alpha \int_0^{\infty} \frac{x e^{-\alpha x^2}}{e^{2\pi x} - 1}\,\mathrm{d}x \right).$$

Ramanujan proves that if $\alpha, \beta > 0$ are such that $\alpha\beta = \pi^2$, then, $I(\alpha) = I(\beta)$ and goes on to suggest the following asympotic formula

$$I(\alpha) \sim \left(\frac{1}{\alpha} + \frac{\alpha}{\pi^2} + \frac{2}{3} \right)^{1/4}, \quad \alpha \to \infty.$$

However, we note that the expansion obtained from Watson's lemma is as follows

$$I(\alpha) \sim \alpha^{-1/4} + \sum_{n=0}^{\infty} I_n \text{ as } \alpha \to \infty,$$

where I_n is given by the formula

$$I_n = \frac{4\alpha^{3/4} B_n (2\pi)^{n-1}}{n!} \int_0^{\infty} e^{-\alpha x^2} x^n\,\mathrm{d}x$$

in which B_n is the nth Bernoulli number. These can be shown to be very close to each other by Taylor expanding Ramanujan's expansion.

Example 2.3 (Asymptotic Expansion of the Beta function)**.** For x, y with $\Re(x)$, $\Re(y) > 0$, consider the integral

$$B(x,y) = \int_0^1 z^{x-1}(1-z)^{y-1}\,\mathrm{d}z = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{2.21}$$

We view the above integral as an integral in x for a fixed y. Now, substitute $z = e^{-u}$ so that we are in the situation of (a generalised) Watson's Lemma. Indeed, we get

$$\begin{aligned} B(x,y) &= \int_0^\infty e^{-ux}(1-e^{-u})^{y-1}\,\mathrm{d}u \\ &= \int_0^\infty e^{-ux}\left(\sum_{k=1}^\infty (-1)^{k+1}\frac{u^k}{k!}\right)^{y-1}\,\mathrm{d}u \\ &\sim \frac{\Gamma(y)}{x^y} \text{ for a fixed } y \text{ as } |x| \to \infty \end{aligned}$$

in a sector $\{x \in \mathbf{C} : |\arg(x)| \leqslant \frac{\pi}{2} - \delta\}$ $(\delta > 0)$.

3. Laplace's Method

This method is used in determining an asymptotic expansion of functions given by certain real integrals of the form

$$\int_a^b \varphi(x,t)\,\mathrm{d}t \tag{3.1}$$

which holds for large positive values of x. It is a powerful tool that finds applications in contemporary research in pure and applied mathematics alike. As an example of the potential of this method, we recommend the papers of Dewar and Murty [12, 13] where this method was used to study the Hardy-Ramanujan asymptotic formula for the number of partitions $p(n)$ of n and the Fourier coefficients of a weakly holomorphic modular form; see also [18] where the particular case of the elliptic modular j-function is studied using this method. This method consists in approximating the integral based on the heuristic that the maximum contribution to the integral comes from neighbourhoods of points where $\varphi(x,t)$ attains its largest values; the success of this method is more pronounced if this largest contribution becomes more and more dominant as $x \to \infty$.

While this philosophy works for a large class of functions $\varphi(x,t)$, it is often useful to rewrite the integrand so that the integral under consideration becomes

$$\int_a^b e^{xh(t)}g(t)\,\mathrm{d}t. \tag{3.2}$$

The point is that, this way, we can exploit the exponential decay of the integrand as we move away from the points where $h(t)$ attains its maximum.

Let us illustrate this method by determining an asymptotic expansion of the real Gamma function as $x \to \infty$. Consider the real Gamma function

$$\Gamma(x+1) = \int_0^\infty e^{-t} t^x \,\mathrm{d}t, \qquad x > -1. \tag{3.3}$$

Recasting the integral as in (3.2), we have:

$$\Gamma(x+1) = \int_0^\infty e^{-t + x \ln t} \,\mathrm{d}t. \tag{3.4}$$

The function $f(t) = -t + x \ln t$ attains its maximum at $t = x$. Thus, the point at which the function attains its maximum keeps varying as x varies; furthermore, the mass of the function $f(t)$ spreads more and more around the point $t = x$ as x increases (see Figure 1). This motivates the substitution $t = ux$ so that, for

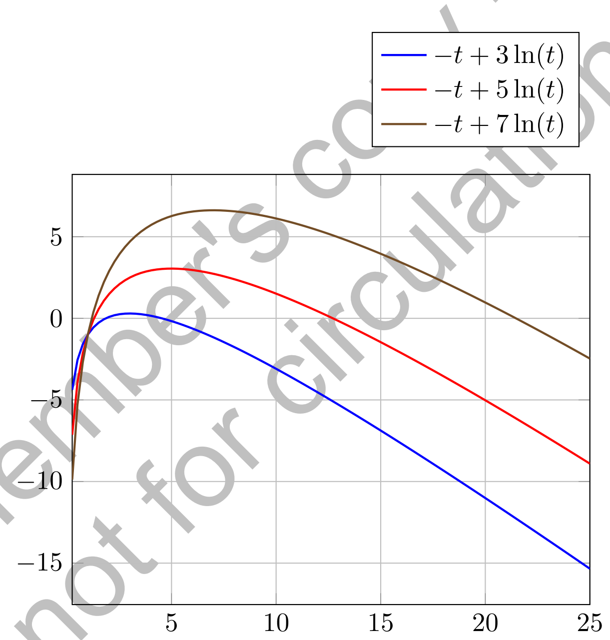


Figure 1. Graph of the function $f(t) = -t + x\ln(t)$ for $x = 3, 5, 7$.

$x > 0$, the integral becomes

$$x^{x+1} \int_0^\infty e^{x(\ln(u) - u)} \,\mathrm{d}u. \tag{3.5}$$

The function $h(u) = \ln(u) - u$ attains its maximum at $u = 1$. Thus, we may replace the function h by its Taylor expansion around 1 and extend the lower limit to $-\infty$ since significant contribution comes only from the range of integration around 1 anyway, we must have that

$$\Gamma(x+1) \sim x^{x+1} \int_{-\infty}^\infty e^{x\left(-1 - \frac{(u-1)^2}{2}\right)} \,\mathrm{d}u. \tag{3.6}$$

Simplifying using the probability integral, we have

$$\Gamma(x+1) \sim x^{x+\frac{1}{2}} e^{-x} \sqrt{2\pi}. \tag{3.7}$$

That this heuristic argument always leads to an asymptotic formula for (3.2) for sufficiently nice functions g and h can be proven [23, §1.1.5].

Remark 3.1. For a proof of Stirling's formula (3.7) from much more elementary considerations, see [19].

4. The method of steepest descent

The method of steepest descent (sometimes also known as the saddle point method) is an extension of Laplace's method to the complex plane.

Let Ω be a simply connected domain in $\mathbf{C}$. Let f and g be two holomorphic functions on Ω and let $\gamma : [0,1] \to \Omega$ be a piecewise differentiable curve in Ω. The method of steepest descent aids in deriving an asymptotic formula for integrals the form

$$F(t) = \int_\gamma e^{tf(\xi)} g(\xi)\,\mathrm{d}\xi \tag{4.1}$$

for large positive values of t. For $\xi = x + iy$, write $f(\xi) = u(x,y) + iv(x,y)$. If we are to appeal to Laplace's philosophy, since $|e^{tf(\xi)}| = e^{tu(x,y)}$, we should seek to deform the contour γ to another contour γ' in Ω which consists of points (x_0, y_0) where $u(x,y)$ is large. Note first that this deformation does not change the value of the integral as the integrand is analytic in Ω (Cauchy's theorem). However, the key difficulty that we are presented with is as follows: "heuristically", we see that for large values of t, the function $e^{itv(x,y)}$ might oscillate rapidly even for small displacements along the curve γ' that the positive and negative swings in the values taken by the function will tend to cancel out.

However, this suggests that there is a reasonable chance of success if we can find a contour γ' such that e^{tu} attains large values on those parts of the contour where v is constant and whenever v varies, the magnitude e^{tu} is small. One can argue that on a path where v is constant and u attains its maximum at an interior point, say z_0, we must have that $f'(z_0) = 0$. Thus, we seek a contour γ' such that:

(1) γ' passes through one or more zeroes of $f'(z)$.
(2) $v(x,y)$ is constant on γ'.

A few remarks are in order about the choice of γ'. Consider the surface $S(x,y,u(x,y))$ in $\mathbf{R}^3$. If $z_0 = x_0 + iy_0$ is a zero of $f'(z)$, then, Cauchy-Riemann equations

$$f'(z) = u_x - iu_y \tag{4.2}$$

imply that $u_x(x_0,y_0) = u_y(x_0,y_0) = 0$. In other words, the point (x_0,y_0) is a critical point of the harmonic function u. Therefore, the point (x_0,y_0) is either on the boundary of the domain Ω or is a saddle point. When this point is a saddle point of u, the surface is saddle-shaped near z_0 (see Figure 2) and this point z_0 is in turn called a saddle point of $f(z)$. Let us now set up some language before we

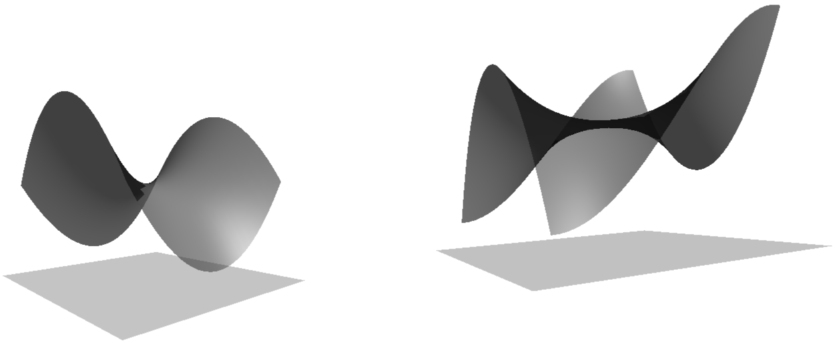


FIGURE 2. Saddle surfaces near saddle points of order 1 and 2.

proceed. Let $z_0 = x_0 + iy_0$ be a saddle point of f. The regions in the plane where $u(x, y) > u(x_0, y_0)$ are called *hills* and $u(x, y) < u(x_0, y_0)$ are called *valleys*. We say that the saddle point z_0 is of order m, $m \geqslant 1$, if

$$f'(z_0) = \cdots = f^{(m)}(z_0) = 0 \text{ and } f^{(m+1)}(z_0) \neq 0. \tag{4.3}$$

Let ψ be a curve through the point z_0 with the parameter s giving its unit speed parametrisation. That is,

$$\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2 = 1, \tag{4.4}$$

and the direction θ of the curve at the point s is given by

$$\cos\theta(s) = \frac{\mathrm{d}x}{\mathrm{d}s}(s) \text{ and } \sin\theta(s) = \frac{\mathrm{d}y}{\mathrm{d}s}(s). \tag{4.5}$$

Call a contour ψ a *steepest path* if the points $(x(s), y(s))$ on ψ are points at which the direction $\frac{\mathrm{d}u}{\mathrm{d}s}$ of the path $u \circ \psi$ on the surface S attains its extremum as $\theta(s)$ varies. The contour ψ is called a *path of steepest descent* (resp. *path of steepest ascent*) if the points are points of maxima (resp. minima) for $\frac{\mathrm{d}u}{\mathrm{d}s}$.

The key fact is that the paths of constant phase (cf. (2) of our conditions on γ') are also steepest paths in the regions where $f' \neq 0$: indeed,

$$\frac{\mathrm{d}u}{\mathrm{d}s} = \cos\theta\frac{\partial u}{\partial x} + \sin\theta\frac{\partial u}{\partial y} \tag{4.6}$$

attains its extrema at points where the derivative of $\frac{\mathrm{d}u}{\mathrm{d}s}$ with respect to θ

$$-\sin\theta\frac{\partial u}{\partial x} + \cos\theta\frac{\partial u}{\partial y} = -\sin\theta\frac{\partial v}{\partial y} - \cos\theta\frac{\partial v}{\partial x} = -\frac{\mathrm{d}v}{\mathrm{d}s} \tag{4.7}$$

vanishes and the second derivative of $\frac{\mathrm{d}u}{\mathrm{d}s}$ with respect to θ

$$-\cos\theta\frac{\partial u}{\partial x} - \sin\theta\frac{\partial u}{\partial y} = \frac{\mathrm{d}u}{\mathrm{d}s} \tag{4.8}$$

is non-zero. Note further that γ' is a path of steepest ascent (resp. descent) if $\frac{\mathrm{d}u}{\mathrm{d}s} > 0$ (resp. $\frac{\mathrm{d}u}{\mathrm{d}s} < 0$).

Remark 4.1. The discussion above can also be paraphrased in terms of gradient of multivariate function: a path ψ would then be called a steepest path if the tangent vector $(\frac{\mathrm{d}x}{\mathrm{d}s}, \frac{\mathrm{d}y}{\mathrm{d}s})$ at s of the path ψ is parallel to ∇u at $\psi(s)$; equivalently, we require that the normal vector $(-\frac{\mathrm{d}y}{\mathrm{d}s}, \frac{\mathrm{d}x}{\mathrm{d}s})$ and ∇u are perpendicular (cf. (4.7)).

Suppose now that z_0 is a saddle point of order m, $m \geqslant 1$. Then, the Taylor expansion of f at $z = z_0 + re^{i\theta}$ around z_0 is given by

$$f(z) = f(z_0) + \frac{r^{m+1}}{(m+1)!} a^{i(m+1)\theta+\varphi} + \cdots \tag{4.9}$$

where we have let $f^{(m+1)}(z_0) = ae^{i\varphi}$ $(a > 0)$. This gives us:

- the direction of the curves $v(x,y) = v(x_0,y_0)$ of constant phase at z_0, viz, the solutions to the equation $\sin((m+1)\theta+\varphi) = 0$:
$$\theta = \frac{1}{m+1}(-\varphi + k\pi), \quad k = 0, \ldots, 2m+1 \tag{4.10}$$
- the direction of the level curves of u, viz, the solutions to the equation $\cos((m-1)\theta+\varphi) = 0$:
$$\theta = \frac{1}{m+1}\left(-\varphi + (2k+1)\frac{\pi}{2}\right), \quad k = 0, \ldots, 2m+1. \tag{4.11}$$

This analysis tells us that the region of the plane near z_0 is divided into $2(m+1)$ regions by the $2(m+1)$ level curves; the regions bounded between these curves alternate between hills and valleys of the function u. In each valley (and similarly, in each hill), there is a path of steepest descent (resp. ascent) meeting a certain level curve at z_0 at an angle of $\frac{\pi}{2(m+1)}$. (See Figure 3.) Suppose now

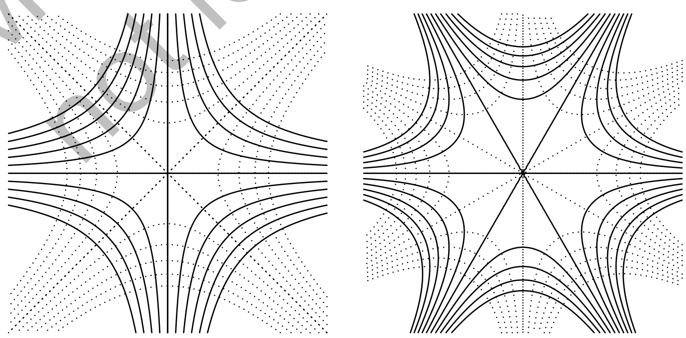


FIGURE 3. Lines of constant phase (solid) and level curves of u (dotted) near saddle points of order 1 and 2.

that γ' is a steepest path through z_0. Then, the real part is either monotonically increasing to or decreasing from $u(x_0,y_0)$ on this path γ'. The paths γ' lying in valleys (viz, those along which u decreases) are paths of steepest decent.

The thesis of the method of steepest descents is the following: by Taylor expanding f near a saddle point z_0, the expansion obtained by term-wise integration is asymptotic if we can shift our contour γ to a contour γ' lying in valleys and passing through a saddle point between valleys.

4.1. **Contribution from a saddle point.** Throughout what follows, we shall assume that z_0 is a saddle point of order 1 for f. The key feature of this assumption is that, in this case, there are two paths of steepest descent at an angle of $\pm\pi$ from each other (so oppositely directed rays emanating from z_0).

Expanding f in Taylor series around z_0, we have (since $f'(z_0) = 0$):

$$f(z) = f(z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots . \tag{4.12}$$

We cut down to a small neighbourhood of z_0 in which

$$f_1(z) := \frac{f(z) - f(z_0)}{(z - z_0)^2}$$

is analytic. If $\widetilde{\gamma}$ is the part of the curve γ in this neighbourhood, then, the saddle point z_0 contributes

$$\int_{\widetilde{\gamma}} \exp\left(t f_1(z_0)(z - z_0)^2\right) g(z)\, \mathrm{d}z \tag{4.13}$$

to the integral (4.1). Since $(z - z_0)^2 f''(z_0)$ is real and negative on a path of the steepest descent, we put $-\frac{1}{2}\zeta^2 = f_1(z)(z - z_0)^2$ so that ζ is determined as the principal value of the square root of the appropriate quantity. This change of variable leaves us with:

$$e^{tf(z_0)} \int_{\widetilde{\gamma}} \exp\left(-\frac{1}{2}\zeta^2\right) g(z) \frac{\mathrm{d}z}{\mathrm{d}\zeta}\, \mathrm{d}\zeta. \tag{4.14}$$

This puts us in the situation of Watson's lemma showing the existence of an asymptotic expansion. Now, we are faced with the problem of inverting the given power series $\zeta(z)$ to get $z(\zeta)$. This is usually cumbersome.

However, getting just the leading term in the series $z(\zeta)$ is not hard and gives us the first term of the asymptotic expansion (Watson's lemma). Write $z - z_0 = r(z)e^{-i\varphi}$ (recall that $\varphi = \arg(f''(z_0))$) so that $\zeta = \pm r\sqrt{|f''(z_0)|}$ (the correct sign being given by the sign of the rate of change of the real part on $\widetilde{\gamma}$: thus, $+$ would be the right sign if the real part is increasing and $-$ otherwise). Via Watson's Lemma, we conclude that the amount $I(z_0)$ contributed by the saddle point z_0 is

$$I(z_0) = \pm \frac{g(z_0)e^{tf(z_0)}\sqrt{2\pi}e^{-i\varphi}}{\sqrt{|tf''(z_0)|}} \tag{4.15}$$

This completes our discussion of the saddle point method.

5. Stirling's Formula

We apply the method of steepest descent to the complex Gamma function defined as follows

$$\Gamma(z+1) := \int_0^\infty e^{-s} s^z \, \mathrm{d}s, \quad \text{for } \Re(x) > -1. \tag{5.1}$$

We rewrite the integrand so that we have

$$\Gamma(z+1) = \int_0^\infty e^{-s+z\ln s} \, \mathrm{d}s. \tag{5.2}$$

For $z \in \mathbf{C}$ with $\Re(z) > 0$, put $s = z(1+t)$ so that $\mathrm{d}s = z\mathrm{d}t$ and

$$\Gamma(z+1) = z^{z+1} e^{-z} \int_{-1}^\infty \exp(-tz + z\ln(1+t)) \, \mathrm{d}t. \tag{5.3}$$

Now, analysing the function $f(t) = \ln(1+t) - t$, we see that the only saddle point $t = 0$ of order 1 has the property that $f''(0) = -1$. In this case, the real line is already a path of steepest descent and the sign in $I(0)$ given by the contour is $+$. Putting these facts together, we have

$$I(0) = \sqrt{2\pi/z}. \tag{5.4}$$

This gives us the Stirling's formula for the complex Gamma function

$$\Gamma(z+1) \sim \sqrt{2\pi z}\left(\frac{z}{e}\right)^z \tag{5.5}$$

when $|z| \to \infty$ in any sector $\{z \in \mathbf{C} : |\arg(z)| \leqslant \frac{\pi}{2} - \delta < \frac{\pi}{2}\}$.

Appendix A. Historical Remarks on Stirling's Formula

In this appendix, we indicate some references to curious historical events and some (and by no means all!) instances of proofs of "Stirling's" formula in the literature.

Our story begins with Abraham de Moivre. In the second edition of his "The Doctrine of Chances" [10, pp. 243 – 254], he considers the problem of finding an asymptotic expression for the binomial coefficients; he introduces the normal distribution as an error distribution and applies it to derive the normal approximation to the binomial distribution. This special case of the central limit theorem is widely known today as de Moivre-Laplace Theorem. In particular, he announces that

$$\lim_{n\to\infty} \frac{\sqrt{2n}}{2^{2n}} \binom{2n}{n} \tag{A.1}$$

is a certain number $2B^{-1}$ where B satisfies

$$\log(B) = 1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} + \frac{1}{1680} + \cdots. \tag{A.2}$$

He proceeds to indicate that his initial investigations were limited to a study of the magnitude of this number B until his "worthy and learned Friend Mr. James Stirling [...] found that the quantity B did denote the square-root of the circumference of a circle whose radius is unity". We must note that de Moivre proves this using Wallis's product formula for π in his book Miscellanea Analytica [8].

James Stirling describes his approximation to $\log(n!)$ in his book "Methodus Differentialis" [28, Proposition 28, Example 2, pp. 149 – 152] (see also [31, 269–275] for an annotated English translation of this work). Stirling, from more general considerations, proves that

$$\log(n!) = \left(n+\frac{1}{2}\right)\log\left(n+\frac{1}{2}\right)+\frac{1}{2}\log(2\pi)-\left(n+\frac{1}{2}\right) - \frac{1}{24\left(n+\frac{1}{2}\right)}+\frac{7}{288\left(n+\frac{1}{2}\right)^3},+\cdots. \tag{A.3}$$

Curiously, the leading term in this expansion is attributed to W. Burnside [4]. In the literature (as also in this paper), the following modified asymptotic expansion obtained by de Moivre [9]

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n}-\frac{1}{360n^3}+\frac{1}{1260n^5}-\frac{1}{1680n^7}+\cdots\right) \tag{A.4}$$

has come to be known as Stirling's series [14] and the leading term Stirling's formula. We recommend the books [30, 31] and the paper [29] by Tweddle for more historical remarks and insightful comparisons between these asymptotic expansions.

Numerous variations on Stirling's formula has since been worked out. We close with one due to Ramanujan

$$n! \sim \sqrt{\pi}\left(\frac{n}{e}\right)^n \sqrt[6]{8n^3+4n^2+n+\frac{1}{30}}, \quad n\to\infty, \tag{A.5}$$

who also gives the following inequality [24] which holds for all $x \geqslant 1$

$$\sqrt{\pi}\left(\frac{x}{e}\right)^n \sqrt[6]{8x^3+4x^2+x+\frac{1}{100}} < \Gamma(x+1) < \sqrt{\pi}\left(\frac{x}{e}\right)^n \sqrt[6]{8x^3+4x^2+x+\frac{1}{30}} \tag{A.6}$$

(this is to be compared with Question 754 submitted by Ramanujan to the J. of Indian Math. Soc., Vol. 8, p.80).

A.1. **Proofs in the literature.** There are numerous proofs of this ubiquitous formula in the literature involving results of varying levels of sophistication. While there are some elementary proofs [6, 26, 14], there are also some proofs in the literature that appeal to Lebesgue's dominated convergence theorem, see [27, Chapter 8, §6] and [21]. Some proofs use various properties of the real or complex Gamma functions [1, 17]. Some proofs [22, 16] motivated by probability theory have also been given. In fact, in a remarkable reversal of historic timeline, the central limit theorem is proven today without using Stirling's formula and applying this theorem to specific distributions to derive Stirling's formula is now so well-known [5, 16]. A recent addition to this list is a paper with R. Murty [19] in the last issue of "The Mathematics Student" where a very short proof of Stirling's formula is presented using only an elementary knowledge of Calculus.

A.2. **The constant $\sqrt{2\pi}$.** As we point out in our paper [19], it is non-trivial to obtain the constant $\sqrt{2\pi}$ in Stirling's formula. It might therefore be of interest to see how this constant is obtained in these different proofs. Table 1 summarises how the constant π enters the proofs of Stirling's formula cited here. It is conceivable [31, p.271] that Stirling's proof that $C = \sqrt{2\pi}$ involves Wallis's product formula too.

Idea	References
Wallis' Product formula for $\frac{\pi}{2}$	[6], [14]
The Gaussian integral	[17], [21], [22], [27], [16]
$\Gamma(\frac{1}{2}) = \sqrt{\pi}$	[1]
de Moivre's formula (A.1)	[19]

TABLE 1. The constant $\sqrt{2\pi}$

In this connection, we remark that A. J. Coleman [7] computes the Gaussian integral using Wallis's formula.

REFERENCES

[1] Artin, E., *The gamma function.* Holt, Rinehart and Winston, New York-Toronto-London, 1964.

[2] Bender, C. M. and Orszag, S. A., *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory.* Springer-Verlag, 1978.

[3] Berndt, B. C. and Evans, R. J., Some elegant approximations and asymptotic formulas of Ramanujan. *J. Comp. and App. Math.*, 37:35–41, 1991.

[4] Burnside, W., A rapidly converging series for $\log(N!)$. *Messenger Math.*, 46:157 – 159, 1917.

[5] Chatterji, S. D., Asymptotic formulae derived from the central limit theorem. *Confer. Sem. Mat. Univ. Bari*, (234):37pp, 1990.

[6] Coleman, A. J., A simple proof of Stirling's formula. *American Mathematical Monthly*, 58(5):334 – 336, May 1951.

[7] Coleman, A. J., The Probability Integral. *American Mathematical Monthly*, 61(10):710 – 711, Dec 1954.

[8] De Moivre, A., *Miscellanea Analytica de Seriebus et Quadraturis.* Touson and Watts, London, England, 1730.

[9] De Moivre, A., *Miscellaneis Analyticis Supplementum.* London, England, 1730.

[10] De Moivre, A., *The Doctrine of Chances: Or, A Method of Calculating Probabilities of Events in Play.* A. Millar, London, England, Third edition, 1756.

[11] Debye, P., *Collected Works.* Interscience Publishers, Inc., 1953.

[12] Dewar, M. and Murty, M. R., A derivation of the Hardy-Ramanujan formula from an arithmetic formula. *Proceedings of the American Mathematical Society*, 141:1903 – 1911, 2012.

[13] Dewar, M. and Murty, M. R., An asymptotic formula for the coefficients of $j(z)$. *International Journal of Number Theory*, 9:1 – 12, 2013.

[14] Impens, C., Stirling's Series Made Easy. *American Mathematical Monthly*, 110(8):730 – 735, October 2003.

[15] Jeffreys, H. and Jeffreys, B. S., *Methods of Mathematical Physics*. Cambridge University Press, 1946.

[16] Khan, R. A., A Probabilistic proof of Stirling's formula. *American Mathematical Monthly*, 81(3):366–369, 1974.

[17] Michel, R., The $(n + 1)$th proof of Stirling's formula. *American Mathematical Monthly*, 115(9):844 – 845, Nov. 2008.

[18] Murty, M. R. and Kannappan, S., On the asymptotic formula for the Fourier coefficients of j-function. Accepted at Kyushu J. Math., June 2015.

[19] Murty, M. R. and Kannappan, S., A very simple proof of Stirling's formula. *The Mathematics Student*, 82(1-2), 2015.

[20] Olver, F., *Asymptotics and Special Functions*. AKP Classics. Taylor & Francis, 1997.

[21] Patin, J. M., A very short proof of Stirling's formula. *American Mathematical Monthly*, 96(1):41 – 42, Jan 1989.

[22] Pinsky, M. A., Stirling's Formula via the Poisson Distribution. *American Mathematical Monthly*, 114(3):256 – 258, March 2007.

[23] Pinsky, M. A., *Introduction to Fourier Analysis and Wavelets*, volume 102 of *Graduate Studies in Mathematics*. American Mathematical Society, 2009.

[24] Ramanujan, S., *The lost notebook and other unpublished papers*. Narosa Publ. H.-Springer, New Delhi - Berlin, 1988.

[25] Riemann, B., *Gesammelte Mathematische Werke*. Dover Publications, Inc., 1953.

[26] Robbins, H., A remark on Stirling's formula. *American Mathematical Monthly*, 62(1):26 – 29, Jan 1955.

[27] Schwartz, L., *Mathematics for the physical sciences*. Hermann, Paris; Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1966.

[28] Stirling, J., *Methodus Differentialis: sive Tractus de Summatione et Interpolatione Serierum Infinitarum*. J. Whiston and B. White, London, England, 1730.

[29] Tweddle, I., Approximating $n!$. Historical origins and error analysis. *American Journal of Physics*, 52:487 – 488, 1984.

[30] Tweddle, I., *James Stirling: 'This about series and such things'*. Scottish Academic Press, Edinburgh, 1988.

[31] Tweddle, I., *James Stirling's Methodus Differentialis: An Annotated Translation of Stirling's Text*. Springer-Verlag, London, England, 2003.

[32] Wong, R., *Asymptotic Approximation of Integrals*. Academic Press, 1989.

Kannappan Sampath

Dept. of Mathematics, Jeffery Hall, 99 University Avenue

Queen's University, Kingston, Ontario, K7L 3N6, Canada.

E-mail: *knsam@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 111-112

AN ELEMENTARY PROOF OF THE CONNECTEDNESS OF THE GENERAL LINEAR GROUP $GL_{\mathbf{n}}(\mathbb{C})$

SURJIT KUMAR

(Received : 06 - 05 - 2015)

Abstract. The purpose of this note is to popularize an elementary proof of the connectedness of $GL_n(\mathbb{C})$. The only non-trivial fact required in this proof is a simple consequence of the Polynomial Factor Theorem.

There are many simple deductions of the connectedness of the general linear group $GL_n(\mathbb{C})$ of invertible, complex $n \times n$ matrices. However, most of these are not really elementary. For instance, the proofs given in [2] and [4] rely on upper triangularization and polar decomposition of matrices respectively. The purpose of this note is to popularize a simple and elementary proof of the connectedness of $GL_n(\mathbb{C})$. The idea of the proof is certainly known (see [3], Proposition 1). The only non-trivial fact required in the proof is the following consequence of the Polynomial Factor Theorem [1]:

Lemma 1. *Let p be a polynomial in x over a field K of degree n. Then p has at most finitely many zeros in K.*

We will derive a general fact about the zero set of complex polynomials from the preceding lemma. First a notation:

Let p be an analytic polynomial in n complex variables $z_1, \cdots, z_n$. The *zero set* $Z(p)$ of p is given by

$$Z(p) := \{(z_1, \cdots, z_n) \in \mathbb{C}^n : p(z_1, \cdots, z_n) = 0\}.$$

Theorem 1. *Let p denote an analytic polynomial in the complex variables $z_1, \cdots, z_n$. Then $\mathbb{C}^n \setminus Z(p)$ is path-connected.*

Proof. Let $z, w \in \mathbb{C}^n \setminus Z(p)$. Consider the straight-line path

$$\gamma(t) = (1-t)z + tw \ (t \in \mathbb{C}).$$

2010 Mathematics Subject Classification: Primary 54D05, 54H11; Secondary 12D05, 30C 15

Keywords and Phrases: Connectedness, Polynomial Factor Theorem, General linear group.

Note that $\{t \in \mathbb{C} : \gamma(t) \in Z(p)\}$ is precisely the zero set $Z(p \circ \gamma) := Z$. By Lemma 1, Z is a finite subset of $\mathbb{C}$. Thus γ maps the path-connected set $\mathbb{C} \setminus Z$ continuously into $\mathbb{C}^n \setminus Z(p)$. In particular, z and w belong to the path-connected subset $\gamma(\mathbb{C} \setminus Z)$ of $\mathbb{C}^n \setminus Z(p)$. $\square$

Remark 1. One may imitate the proof of Theorem 1, and appeal to the Identity Theorem from $\mathbb{C}$-analysis (instead of Lemma 1) to obtain the following: *Let $f : \mathbb{C}^n \to \mathbb{C}$ denote a holomorphic function in the complex variables $z_1, \cdots, z_n$. Then $\mathbb{C}^n \setminus Z(f)$ is path-connected.*

Corollary 1. *The general linear group $GL_n(\mathbb{C})$ is path-connected.*

Proof. If one identifies in a natural way the space of complex $n \times n$ matrices with $\mathbb{C}^{n^2}$ then $GL_n(\mathbb{C})$ can be seen as $\mathbb{C}^{n^2} \setminus Z(\det)$, where det is the analytic polynomial which sends a matrix to its determinant. Now the desired conclusion is immediate from Theorem 1. $\square$

Acknowledgment: I realized the proof presented in this article during discussions with Sameer Chavan. This research work is supported by CSIR grant no. 09/092(0748)/2010-EMR-I.

References

[1] Gallian, Joseph, *Contemporary Abstract Algebra*, Narosa Publishing House, New Delhi, 1999.

[2] Hall, Brian C, *Lie groups, Lie algebras, and representations. An elementary introduction*, Graduate Texts in Mathematics, 222. Springer-Verlag, New York, 2003.

[3] Narasimhan, Raghavan, *Several Complex Variables*, Univ. of Chicago Press, Chicago, 1995.

[4] Wong, Yung-chow and Au-yeung, Yik-hoi, *An elementary and simple proof of the connectedness of the classical groups*, Amer. Math. Monthly **74** (1967) 964-966.

Surjit Kumar
Department of Mathematics
Indian Institute of Science, Bangalore 560012, India
E-mail: *surjitk15@math.iisc.ernet.in*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 113-122

THE JORDAN CURVE THEOREM

AMIYA MUKHERJEE

(Received : 15 - 05 - 2015 ; Revised : 13 - 07 - 2015)

Abstract. In this article we outline a proof of the Jordan curve theorem which is an unpublished work of a British mathematician Graham Robert Allan (1936-2007), Reader in Functional Analysis and Vice-Master of Churchill College at Cambridge University, England. To the best of my knowledge, this is the simplest proof which uses only elementary point-set topology and complex analysis (see Andrew Browder [2]).

1. Introduction

A Jordan curve is a non-intersecting continuous loop in the complex plane $\mathbb{C}$. Another name is simple closed curve, which is a homeomorphic image of the unit circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. The celebrated theorem of Jordan states that a simple closed curve Γ in the plane $\mathbb{C}$ divides the complement $\mathbb{C} - \Gamma$ into two connected nonempty open sets, exactly one of which is bounded. The bounded region is called the interior (or the inner region) of the curve Γ. The unbounded component is the exterior (or the outer region) of Γ. The situation can be visualised easily for the case of a circle in the plane. But the simple closed curve may be very complicated, like a jigsaw puzzle, and it may be difficult to ascertain the inside and outside of the curve.

The theorem has an interesting history. It is named after the mathematician Camille Jordan (1838-1922). The proof of the theorem was given by Jordan in his lectures on real analysis, and was published in his book "Cours d'analyse de l'École Polytechnique (Gauthier-Villars, Paris, 1887, Vol. 3)", from which a whole generation of mathematicians learned the modern concept of rigour in analysis. For decades, it was generally thought that the proof was imperfect. According to Courant and Robbins [4], "the proof given by Jordan was neither short nor simple, and the surprise was even greater when it turned out that Jordan's proof was invalid and that considerable effort was necessary to fill the gaps in his reasoning. The first rigorous proofs of the theorem were quite complicated and hard to understand, even for many well-trained mathematicians. Only recently have comparatively simple proofs been found". Morris Kline, famous writer on history,

2010 Mathematics Subject Classification : Primary 30-01, Secondary 55-01, 55M05.
Key words and phrases : Complex logarithm, Jordan curve.

philosophy, and a populariser of mathematics, wrote that "Jordan himself and many distinguished mathematicians gave incorrect proofs of the theorem". The first rigorous and correct proof was given by Oswald Veblen [8]. Veblen complained that Jordan's proof "is unsatisfactory to many mathematicians. It assumes the theorem without proof in the important special case of a simple polygon and of the argument from that point on, one must admit at least that all details are not given".

The theorem is a standard result in algebraic topology. A complete proof may be found in Allen Hatcher, Algebraic Topology (2002), and in other texts such as Edwin Spanier, Algebraic Topology (1966). Recently, a proof checker was used by a Japanese-Polish team to create a "computer-checked" proof of the theorem.

Acknowledgement. We thank the referee for a careful reading of the article and for pointing out a couple of typos.

2. Preliminaries

Unlike the real case, the complex exponential function is not one-to-one. We have $e^z = e^{z+2\pi ni}$ for a complex number z, and any integer n. However, restricting the domain of e^z to certain region, we may get a one-to-one function $e^z = w$, and define the inverse function $\log w = z$.

Note that since $e^z \neq 0$ for any $z \in \mathbb{C}$, we can not define $\log 0$. Therefore, suppose that $e^z = w$ and $w \neq 0$. If $z = x + iy$, then $w = e^x \cdot e^{iy}$. Therefore $|w| = e^x$, and $y = \arg w + 2\pi n$, $n \in \mathbb{Z}$ (the set of integers). Thus the set $\{\log |w| + i(\arg w + 2\pi n) : n \in \mathbb{Z}\}$ is the set of solutions for $e^z = w$.

Definition 2.1. *If $U \subset \mathbb{C}$ is a connected open set, and $f : U \to \mathbb{C}$ is a continuous function such that $z = e^{f(z)}$ for all $z \in U$, then f is called a branch of the logarithm* $\log z$ *on U. Note that $0 \notin U$.*

Proposition 2.2. *If $U \subset \mathbb{C}$ is a connected open set and f is a branch of* $\log z$ *on U, then the set of branches of* $\log z$ *consists of the functions $g(z) = f(z) + 2\pi ni$, $n \in \mathbb{Z}$.*

Proof. If f is a branch of the logarithm, and n is any integer, then $g(z) = f(z) + 2\pi ni$ is also a branch of the logarithm, because $e^{g(z)} = e^{f(z)}$. Conversely, if f and g are both branches of $\log z$, then for each $z \in U$, $f(z) = g(z) + 2\pi ni$ for some $n \in \mathbb{Z}$. In fact, n is the same for each z. To see this, consider the function $h(z) = \frac{1}{2\pi i}[f(z) - g(z)]$. This is a continuous function $U \to \mathbb{Z}$. Since U is connected, $h(U)$ must also be connected. Therefore there is an $n \in \mathbb{Z}$ such that $f(z) + 2\pi ni = g(z)$ for all $z \in U$. This completes the proof. □

If X is a compact Hausdorff space, we denote by $C(X)$ the space of continuous functions $f : X \to \mathbb{C}$, and by $C^*(X)$ the space of nowhere zero continuous functions $f : X \to \mathbb{C}^*$, where $\mathbb{C}^* = \mathbb{C} - \{0\}$. Thus $C^*(X)$ is the space of invertible functions in $C(X)$. Let

$$\exp C(X) = \{e^f : f \in C(X)\}.$$

This is the space of continuous functions $X \to \mathbb{C}^*$ which admit continuous logarithm.

Then $C^*(X)$ is a commutative group with pointwise multiplication, and $\exp C(X)$ is a subgroup of $C^*(X)$.

Definition 2.3. *Define an equivalence relation $\sim$ in $C^*(X)$ by $f \sim g$ iff (if and only if) $f/g \in \exp C(X)$, that is, $f/g = e^h$ for some $h \in C(X)$. Thus $f \sim g$ iff f and g belong to the same coset of* $\exp C(X)$, *in other words, f and g represent the same element of the quotient group $H_X = C^*(X)/\exp C(X)$.*

Lemma 2.4. *If $f, g \in C(X)$, and $|g| < |f|$. then $f + g \sim f$.*

Proof. The condition $|g| < |f|$ implies that f and $f + g$ are never zero. If $h = g/f$, then $|h| < 1$, and so $1 + h$ maps X into the right half-plane

$$\mathcal{R} = \{z = re^{i\theta} \ : \ r > 0, \ -\pi/2 < \theta < \pi/2\}.$$

This means that there is a continuous function $k : X \to \mathbb{C}$ such that $1 + h = e^k$. To see this suppose that $b : \mathcal{R} \to \mathbb{C}$ is a continuous branch of $\log w$ on the right half-plane $\mathcal{R}$ so that $w = e^{b(w)}$ for $w \in \mathcal{R}$. Let $k = b \circ (1 + h)$. Then, for $z \in X$, $(1+h)(z) = e^{b(1+h)(z)} = e^{k(z)}$ Thus $f + g = f(1+h) = fe^k$, that is, $f + g \sim f$. $\square$

The space $C(X)$ is a metric space with the standard metric ρ given by

$$\rho(f, g) = \sup_{z \in X} |f(z) - g(z)|.$$

Lemma 2.5. *$C^*(X)$ is an open set of $C(X)$.*

Proof. It is sufficient to show that if $f \in C^*(X)$, there exists a $\epsilon > 0$ such that

$$\{g : \rho(f, g) < \epsilon\} \subset C^*(X).$$

The function $|f(z)|$, $z \in X$, is a real valued continuous function on the compact set X. Therefore it has an absolute minimum $a \in X$ so that $|f(a)| < |f(z)|$ for all $z \in X$. Since $|f(a)| \neq 0$, there is an ϵ such that $0 < \epsilon < |f(a)|$. Then, if $|f(z) - g(z)| < \epsilon$, g is never zero. Therefore $g \in C^*(X)$. $\square$

Corollary 2.6. *For any $f \in C^*(X)$, there is an $\epsilon > 0$ such that if $g \in C^*(X)$ is such that $\rho(f, g) < \epsilon$, then $f \sim g$.*

Therefore the set $[f] = \{g \in C^*(X) : g \sim f\}$ is both open and closed in $C^*(X)$.

Proof. Choose $\epsilon > 0$ such that $0 < \epsilon < |f(z)|$. Then $|g(z) - f(z)| < \epsilon < |f(z)|$ implies $f \sim g$, by Lemma 2.4.

Clearly, the set $[f]$ is open. It is also closed, since its complement in $C^*(X)$ is the union of the other cosets which are open. $\square$

It may be noted that $\exp C(X)$ is the connected component of the identity element 1 in the group $C^*(X)$.

Theorem 2.7. *Let $f, g \in C^*(X)$. Then $f \sim g$ iff f and g are homotopic, that is, there is a continuous map $h : X \times [0,1] \to C^*(X)$ such that $h(x,0) = f(x)$ and $h(x,1) = g(x)$. (The map h is called a homotopy between f and g.)*

Proof. If $f \sim g$, then $f = ge^h$ for some $h \in C(X)$. The continuous map $h(x,t) = f(x)e^{-th(x)}$ is then a homotopy between f and g. Conversely, if $h : X \times [0,1] \to C^*(X)$ is a homotopy between f and g, consider the continuous map $\sigma : [0,1] \to C^*(X)$ given by $\sigma(t) = \sigma_t \in C^*(X)$, where $\sigma_t(x) = h(x,t)$, $t \in [0,1]$, $x \in X$. Then, by Corollary 2.6, the set $\sigma^{-1}[f] = \{t \in [0,1] : \sigma_t \sim \sigma_0\}$ is both open and closed in the interval $[0,1]$, and hence it must be the whole of $[0,1]$, and so $\sigma_1 \sim \sigma_0$, or $g \sim f$. □

Note that if c is a nonzero constant function, then $c \sim 1$, so $c \in \exp C(X)$

Let $D(a,r) = \{z \in \mathbb{C} : |z - a| < r\}$ be the open disk with centre $a \in \mathbb{C}$ and radius r. We denote the closed unit disk centred at $0 \in \mathbb{C}$,

$$\overline{D(0,1)} = \{z \in \mathbb{C} : |z| \leq 1\},$$

by Δ, and its boundary $\{z \in \mathbb{C} : |z| = 1\}$ (the unit circle) by T.

Corollary 2.8. *If $X = \overline{D(a,r)}$, then $C^*(X) = \exp C(X)$, and the group H_X is trivial.*

The result also holds for $X = \Delta$, and also for $X = [0,1]$.

Proof. If $f \in C^*(X)$, define $h : X \times [0,1] \to C^*(X)$ by $h(x,t) = f(tx + (1-t)a)$. Then $h(x,0)$ is the constant map to $f(a)$, and $h(x,1) = f(x)$, so $f \sim 1$, by Theorem 2.7.

For the second part, take $h(x,t) = f(tx)$. □

Lemma 2.9. *If n is an integer, then the function $z^n : T \to \mathbb{C}^*$ does not admit a continuous logarithm unless $n = 0$.*

The result also holds when Δ and T are replaced by the disk $\overline{D(a,r)}$ and its boundary circle $\Gamma = \{z \in \mathbb{C} : |z - a| = r\}$, and the function z^n by $(z-a)^n$.

Proof. It is required to show that if $\log z^n$ exists, then $n = 0$. So suppose that $z^n = e^{2\pi i f(z)}$ for some $f \in C(T)$. Define

$$h : [0,1] \to \mathbb{C}$$

by $h(t) = f(e^{2\pi i t})$, $t \in [0,1]$. Then $h(0) = f(1) = h(1)$. Define $g : [0,1] \to \mathbb{C}$ by $g(t) = h(t) - nt$. Then for all $t \in [0,1]$,

$$e^{2\pi i g(t)} = e^{2\pi i (h(t) - nt)} = e^{2\pi i f(e^{2\pi i t})} \cdot e^{-2\pi i n t} = 1,$$

since $e^{2\pi i f(e^{2\pi i t})} = (e^{2\pi i t})^n$. Therefore $g(t)$ takes only integer values. Since g is continuous, and the interval $[0,1]$ is connected, g must be constant. Then $g(0) - g(1) = h(0) - h(1) + n = n$, or $n = 0$.

For the second part note that there is a homeomorphism $\alpha : \Gamma \to T$ given by $\alpha(z) = (z-a)/r = \zeta$, say. Then $(z-a)^n = e^{f(z)}$, $f \in C(\Gamma)$, implies $\zeta^n = e^{g(\zeta)}$, where $g \in C(T)$ is given by $g(\zeta) = -\log(r^n) + f(a + \zeta r)$. □

Corollary 2.10. *If n is a nonzero integer, then the function $z^n : T \to \mathbb{C}^*$ is not the restriction to T of any continuous map $f : \Delta \to \mathbb{C}^*$.*

Similarly, if $n \neq 0$, then the map $(z-a)^n : \Gamma \to \mathbb{C}^*$ cannot be extended to a continuous map $f : \overline{D(a,r)} \to \mathbb{C}^*$.

Proof. Suppose that $f : \Delta \to \mathbb{C}^*$ is a continuous map such that $f = z^n$ on T. Then $f \in C^*(\Delta) = \exp C(\Delta)$, by Corollary 2.8. Therefore $f = e^g$ for some $g \in C(\Delta)$, or $f|T = e^{g|T} = e^{z^n}$, or the restriction of g to T is the continuous logarithm of z^n. This contradicts Lemma 2.9.

The second part is obvious. □

Theorem 2.11. *Let U be a bounded open set in $\mathbb{C}$, and $B = \overline{U} - U$ be the boundary of U. Then for any $a \in U$ and any nonzero integer n, the function $(z-a)^n$ on B does not admit an extension to a continuous map $f : \overline{U} \to \mathbb{C}^*$.*

Proof. Choose $r > 0$ such that $U \subset D(a,r)$. Suppose that we have a map $f : \overline{U} \to \mathbb{C}^*$ such that $f = (z-a)^n$ on B. Define a map $F : \overline{D(a,r)} \to \mathbb{C}^*$ by $F = f$ on $\overline{U}$, and $F = (z-a)^n$ on $\overline{D(a,r)} - U$. Then F is continuous, since the definitions agree on B. Also $F = (z-a)^n$ on Γ. This contradicts Corollary 2.10. □

Recall that Tietze's extension theorem states that if X is a metric space and A is a closed subset of X, then any bounded continuous map $f : A \to \mathbb{C}$ extends to a continuous map $F : X \to \mathbb{C}$ (see Dieudonné [4], p.89).

Lemma 2.12. *Let K be a compact subset of $\mathbb{C}$, and $f \in \exp C(K)$. Then $f : K \to \mathbb{C}^*$ can be extended to a nonvanishing continuous map F on $\mathbb{C}$.*

Proof. We have $f = e^g$ for some continuous $g : K \to \mathbb{C}$. By Tietze's extension theorem, g has a continuous extension $G : \mathbb{C} \to \mathbb{C}$. Then $F = e^G$ is the required extension of f, and it is nonvanishing. □

Theorem 2.13. *Let X be a compact subset of $\mathbb{C}$. Then a sufficient condition for the connectedness of the set $\mathbb{C} - X$ is that the group H_X is trivial*

Proof. Suppose that H_X is trivial, but $\mathbb{C} - X$ is not connected. Then there exists a bounded component U of $\mathbb{C} - X$. Let $a \in U$, and K be a compact neighbourhood of $B = \overline{U} - U$ in $\mathbb{C} - \{a\}$. Then $B \subset K \subset \mathbb{C} - \{a\}$, and the function $z - a$ does not vanish on K. Therefore $z - a \in \exp C(K)$, and, by Lemma 2.12, the function $z - a$ extends to a continuous function $F : \mathbb{C} \to \mathbb{C}^*$. That is, the function $z - a$ on B extends to $F : \overline{U} \to \mathbb{C}^*$. This contradicts Lemma 2.11. □

We shall see in Corollary 3.8 that the condition of Theorem 2.13 is also necessary.

3. Jordan curve theorem

Let $X \subset \mathbb{C}$ be compact. Then, since $\mathbb{C} - X$ is second countable, it is a union of a countable number of connected components, each of which is both open and closed. Recall that the one-point compactification $\mathbb{C}^+$ of $\mathbb{C}$ is the union of $\mathbb{C}$ and a point ∞ not in $\mathbb{C}$, whose topology consists of all open sets of $\mathbb{C}$, and all sets of the form $\mathbb{C}^+ - K$ where K is a compact subset of $\mathbb{C}$. This is the unique topology on $\mathbb{C}^+$ which makes it a compact Hausdorff space. A component of $\mathbb{C} - X$ is unbounded iff it is a neighbourhood of the point ∞ in $\mathbb{C}^+$. Therefore there can be only one unbounded component of $\mathbb{C} - X$. Let $U_0, U_1, \dots, U_k, \dots$ be the components of $\mathbb{C} - X$, with U_0 the unbounded component. Fix points $a_k \in U_k$ for each k.

Lemma 3.1. *If points a and b belong to the same component of $\mathbb{C} - X$, and $z \in X$, then $z - a \sim z - b$. Moreover, if $a \in U_0$, then $z - a \sim 1$.*

Proof. Since a and b belong to the same path component of $\mathbb{C} - X$, there is a path $\sigma : [0,1] \to \mathbb{C} - X$ such that $\sigma(0) = a$, and $\sigma(1) = b$. Define a continuous map $F : X \times [0,1] \to \mathbb{C}^*$ by $F(z,t) = z - \sigma(t)$. Then $F(z,0) = z - a$ and $F(Z,1) = z - b$. Then, by Theorem 2.7, $z - a \sim z - b$.

Next, if $a \in U_0$, choose $R > \sup\{|z| : z \in X\}$ such that R is in the same component U_0 of a. Then $\operatorname{Re}(z - R) \neq 0$ on X. So $z - R$ has a continuous logarithm, that is, $z - R \sim 1$. Since a and R lie in the same component of $\mathbb{C} - X$, $z - a \sim z - R \sim 1$, by the first part of the lemma. □

Lemma 3.2. *A continuous function $f : X \to \mathbb{C}^*$ given by*

$$f = \prod_{k=1}^{N} (z - a_k)^{n_k}, \quad z \in X,$$

where $n_1, n_2, \dots, n_N$ are integers, does not admit a continuous logarithm unless $n_k = 0$ for every k.

Proof. If $f = e^h$ for some $h \in C(X)$, then, by Lemma 2.12, f has an extension $F : \mathbb{C} \to \mathbb{C}^*$. Let V_1 be an open set such that $a_1 \in V_1 \subset \overline{V_1} \subset U_1$. Then the function

$$g = \prod_{k=2}^{N} (z - a_k)^{n_k}$$

on X extends to a continuous function $X \cup \overline{V_1} \to \mathbb{C}$, which is the restriction of F. In other words, the function $(z - a_1)^{n_1} = f/g : X \to \mathbb{C}^*$ extends to a continuous map $X \cup \overline{V_1} \to \mathbb{C}^*$. Therefore $n_1 = 0$, by Theorem 2.11. Similarly, $n_k = 0$ for every k. □

Theorem 3.3. *If $f : X \to \mathbb{C}$ is a rational function with no zero or pole in X, and N is the number of bounded components of $\mathbb{C} - X$, which is either an integer ≥ 0 or ∞, then*

$$f \sim \prod_{k=1}^{N} (z - a_k)^{n_k},$$

where n_k are uniquely determined integers which are zero for all but finitely many k.

Proof. The proof follows from Lemmas 3.1 and 3.2. □

Let $U \subset \mathbb{C}$ be open. Let $C^1(U)$ denote the set of all $f : U \to \mathbb{C}$ that have continuous first order partial derivatives. Let $C_c^1(U)$ be the set of those $f \in C^1(U)$ which have compact support. Note that the support of f, denoted by $\operatorname{supp} f$, is the closure of the set of points of U where f is nonzero.

If $f \in C^1(U)$, define

$$\overline{\partial} f = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right).$$

Note that if $f = u + iv$, then $\overline{\partial} f = 0$ implies the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \ \ \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

Therefore $\overline{\partial} f = 0$ implies that f is holomorphic (or analytic).

Lemma 3.4. *For $g \in C_c^1(U)$, define a function Ag on $\mathbb{C}$ by*

$$(Ag)(\zeta) = \frac{1}{\pi}\int\int_{\mathbb{C}} \frac{g(z)}{\zeta - z} dxdy, \quad z = x + iy.$$

Then $Ag \in C^1(\mathbb{C})$, $A(\overline{\partial} g) = g$, and $\overline{\partial}(Ag) = g$.

Proof. The fact that $Ag \in C^1(\mathbb{C})$ follows by direct computations of partial derivatives. Alternatively, we may rewrite the definition as

$$(Ag)(\zeta) = \frac{1}{\pi}\int\int_{\mathbb{C}} \frac{g(\zeta - z)}{z} dxdy,$$

then $Ag \in C^1(\mathbb{C})$, because by the dominated convergence theorem, we may differentiate under the sign of integration.

By a “Cauchy formula”

$$\frac{1}{\pi}\int\int_{\mathbb{C}} \frac{\overline{\partial} g(z)}{\zeta - z} dxdy = g(\zeta)$$

(see Rudin [6], Lemma 20.3, p. 384). The proof is simple. To see this, put $z = \zeta + re^{i\theta}$ where $r > 0$ and θ is real, and put $\phi(r, \theta) = g(\zeta + re^{i\theta})$. Then, by chain rule

$$(\overline{\partial} g)(z) = \frac{1}{2}\, e^{i\theta}\left(\frac{\partial}{\partial r} + \frac{i}{r}\frac{\partial}{\partial \theta}\right)\phi(r, \theta).$$

Then

$$\frac{1}{\pi}\int\int_{\mathbb{C}} \frac{\overline{\partial} g(z)}{\zeta - z} dxdy = \lim_{\epsilon\to 0} \frac{1}{2\pi}\int_{\epsilon}^{\infty}\int_{0}^{2\pi}\left(\frac{\partial}{\partial r} + \frac{i}{r}\frac{\partial}{\partial\theta}\right)\phi(r,\theta)d\theta d\phi.$$

For each $r > 0$, ϕ is periodic of period 2π, and so the integral of $\partial\phi/\partial\theta$ is zero.

Therefore

$$\lim_{\epsilon\to 0}\frac{1}{2\pi}\int_0^{2\pi} d\theta \int_{\epsilon}^{\infty}\frac{\partial\phi}{\partial r}dr = \lim_{\epsilon\to 0}\frac{1}{2\pi}\int_0^{2\pi}\phi(\epsilon,\theta)d\theta.$$

As $\epsilon \to 0$, $\phi(\epsilon,\theta) \to g(\zeta)$ uniformly. This gives $A(\overline{\partial} g) = g$.

Finally, if $h = Ag$, then $\overline{\partial} h = \overline{\partial}(Ag) = A(\overline{\partial} g) = g$, differentiating under the integral sign. $\square$

Theorem 3.5 (**Runge's theorem**. Approximation by rational function)**.** *If $g \in C^1(\mathbb{C})$ and $\overline{\partial} g = 0$ on a neighbourhood of a compact set K, then for any $\epsilon > 0$, there exists a rational function $R(z)$ with poles outside K such that*

$$|R(z) - g(z)| < \epsilon$$

for all $z \in K$.

See Rudin [6], Theorem 13.6, p. 256 for proof.

For the next theorem we also need smooth Urysohn's lemma. (See also Rudin [6], Lemma 2.12, p. 39).

Lemma 3.6. *If $K \subset U \subset \mathbb{C}$, where K is closed and U open, then there is a smooth function $f : \mathbb{C} \to \mathbb{R}$ such that $0 \le f \le 1$, $f|K = 1$, and* $\operatorname{supp} f \subset U$.

Proof. The open sets $U_1 = U$ and $U_2 = \mathbb{C} - K$ form an open covering of $\mathbb{C}$. Since $\mathbb{C}$ is paracompact, $\mathbb{C}$ admits a smooth partition of unity subordinate to this covering. This means that there exist smooth functions $\lambda_1 : \mathbb{C} \to \mathbb{R}$ and $\lambda_2 : \mathbb{C} \to \mathbb{R}$ such that for each $j = 1, 2$, (1) $0 \le \lambda_j(z) \le 1$ for all $z \in \mathbb{C}$, (2) $\operatorname{supp} \lambda_j \subset U_j$, and (3) $\lambda_1(z) + \lambda_2(z) = 1$ for all $z \in \mathbb{C}$. Then λ_1 is a solution f of the problem. $\square$

Theorem 3.7. *Let $f \in C^*(X)$. Then there exist uniquely determined integers $n_1, n_2, \ldots$ such that $n_k = 0$ for all but finitely many k such that*

$$f \sim \prod_{k=1}^{N}(z - a_k)^{n_k}.$$

Proof. By Weierstrass approximation theorem ([7], Theorem 7.24), there exists a polynomial p in the real coordinate functions x and y such that $|p - f| < \epsilon$, where $\epsilon = \min_X |f|$. Since $\epsilon < |f|$, p does not vanish on X. Let U be a neighbourhood of X on which p is never zero .

By the Urysohn's lemma, there is a $\phi \in C_c^1(U)$ such that $\phi = 1$ in a neighbourhood V of X. Let $g = \phi \cdot \frac{\overline{\partial} p}{p}$. Then $g \in C^1(\mathbb{C})$. By Lemma 3.4, there exists $h \in C^1(\mathbb{C})$ such that $\overline{\partial} h = g$. Let $F = e^{-h}p$. Then

$$\overline{\partial} F = e^{-h}\overline{\partial} p - pe^{-h}\overline{\partial} h = e^{-h}(\overline{\partial} p - pg) = 0$$

on V, as $\phi = 1$ on V. By Runge's theorem, there exists a rational function $R(z)$ with poles outside X such that $|F(\zeta) - R(\zeta)| < \min_X |F(\zeta)|$. Denote the restrictions of F, p, and R to X by the same letters. Then $f \sim p$, $p \sim F$, and $F \sim R$ (first and third following from Lemma 2.4). Thus we have found a rational function R with neither zero nor pole in X such that $f \sim R$. By Theorem 3.3

$$R \sim \prod_{k=1}^{N} (z - a_k)^{n_k}.$$

This completes the proof. $\square$

Corollary 3.8. *The group $H_X = C^*(X)/\exp C(X)$ is a free abelian group with N generators, where N is the number of bounded components of $\mathbb{C} - X$.*

The proof follows from Lemmas 3.1, 3.2, and Theorems 3.3, and 3.7. The elements $[f_k] \in H_X$, where $f_k : X \to \mathbb{C}$ is the function $f_k(z) = z - a_k$, $a_k \in U_k$, freely generate the group H_X.

Corollary 3.9. *If X and Y are homeomorphic compact sets in $\mathbb{C}$, then X and Y have the same number of complementary components.*

Note that if $\phi : X \to Y$ is a homeomorphism, then the map $\phi_* : H_Y \to H_X$ defined by $\phi_*([f]) = [f \circ \phi]$, where $f \in C^*(Y)$ and $[f] \in H_Y$ is the coset of f, is an isomorphism.

Corollary 3.10. *If J is a Jordan curve in $\mathbb{C}$, that is, $J = \phi(T)$, where ϕ is a homeomorphism of the unit circle T into $\mathbb{C}$, then $\mathbb{C} - J$ has exactly one bounded component. Moreover, $\mathbb{C} - J$ is connected when J is the homeomorphic image of $[0, 1]$.*

The first part is the Jordan curve theorem. This now holds trivially for T, and therefore for any homeomorphic image of T. The second part follows from Corollary 2.8 for $X = [0, 1]$.

Remark 3.11. *Although algebraic topology is beyond the scope of the presents lectures, we may mention another application of the theory, which is Alexander duality theorem for the plane. A theorem of Bruschlinsky [3] says that if X is compact, the group H_X is isomorphic to $\check{H}^1(X, \mathbb{Z})$, which is the first Čech cohomology group of X with integer coefficients $\mathbb{Z}$. The proof uses a little sheaf theory. With this theorem in hand, Corollary 3.8 can obviously be interpreted in modern language as Alexander duality theorem in the plane in the following form.*
Alexander Duality Theorem. *If X is a compact subset of $\mathbb{R}^2$, then the Čech cohomology group $\check{H}^1(X, \mathbb{Z})$ is a free abelian group with N generators, where N is the number of bounded components of the complement $\mathbb{R}^2 - X$. In other words, $\check{H}^1(X, \mathbb{Z})$ is isomorphic to the homology group $H_0(\mathbb{R}^2 - X, \mathbb{Z})$.*

The proof of the main idea of this theorem is given in Eilenberg's doctoral thesis, and published in [5]. The arguments described above are much simpler.

References

[1] Ahlfors, L,, *Complex Analysis*, 1st ed., McGraw-Hill, New York, 1953.

[2] Browder, A., *Topology in the complex plane*, Amer. Math. Monthly, **107** (2000), 393-401.

[3] Bruschlinsky, N., *Stetige Abbildungen und Bettische Gruppen der Dimensionzahl 1 und 3*, Math. Ann. **109** (1934), 525-537.

[4] Dieudonné, J., *Foundations of Modern Analysis*, Academic Press, New York, London, 1969.

[5] Eilenberg, S., *Transformations continues en circonférence et la topologie du plan*, Fund. Math. **26** (1936), 61-112.

[6] Rudin, W., *Real and Complex Analysis*, 2nd ed., Tata McGraw-Hill, New Delhi, New York, 1974.

[7] Rudin, W., *Principles of Mathematical Analysis*, 2nd ed., McGraw-Hill Book Company, New York, 1964.

[8] Veblen, O., *Theory of plane curves in non-metrical analysis situs*, Trans. Amer. Math. Soc., **6**(1) (1905), 83-98.

Amiya Mukherjee
Math-Stat Division, Indian Statistical Institute Calcutta
203 Barrackpore Trunk Road, Kolkata-700108, West Bengal, India
E-mail: *mukherjee.amiya@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 123-125

MMXV-NUMBERS

MIROSLAV KUREŠ

(Received : 26 - 05 - 2015 ; Revised : 06 - 06 - 2015)

Abstract. Special integers corresponding with the current year are considered and the question of their primality is raised. To demonstrate a way of a possible generalization, the squiggly function of a given function of two variables is introduced.

1. The problem

The current year is 2015 and since the number 2015 can be expressed by an interesting formula, we try to look (only recreationally) at the number 2015 through a very little of number theory. In particular, we ask: what about other numbers satisfying this formula? Let us start.

Let p, q, $p < q$ be two prime numbers and consider the number $m_{p,q}$ given by the formula

$$m_{p,q} = p^{q^p+p} - q \cdot (p^q + q).$$

Observe that for $p = 2$, $q = 3$ we have $m_{2,3} = 2015$ which is represented by MMXV in Roman numerals; and hence, we will call the numbers $m_{p,q}$ the $MMXV-numbers$. One can easily calculate MMXV-numbers using the following *Mathematica* code:

```
Clear[p, q, mmxv];
mmxv=p^(q^p+p)-q*(p^q+q);
maxp=3;
maxq=5;
For[p=2,p<=maxp,p++,For[q=p+1,q<=maxq,q++,If[PrimeQ[p]
&& PrimeQ[q],
Print[Subscript[m,p,q]," = ",mmxv]]]]
```

The following table contains the list of all the MMXV-numbers up to 10^{100} (in the fourth column, the last two rows constitute the single number $m_{p,q}$); and, it may be observed that none of these MMXV-numbers is a prime number. For example, $m_{2,7} = 2251799813684303 = 424777 \cdot 5301134039$.

2010 Mathematics Subject Classification : 11A51, 11A99.
Key words and phrases : Elementary number theory, Primality, Recreational mathematics.

i	p	q	$m_{p,q}$
1	2	3	2015
2	2	5	134217543
3	2	7	2251799813684303
4	2	11	10633823966279326983230456482242733959
5	2	13	2993155353253689176481146537402947624255349847908183
6	3	5	11790184577738583171520872861412518665678211592275841109095721
7	2	17	3978585891278293137243057985174566720803649206378781739523711815145275976100267002035935

In view of the fact that MMXV-numbers are very big numbers, a computer testing of their primality has limitations. We present (approximate) values of MMXV-numbers $m_{p,q}$ for the lowest p and the first lowest q to them in the next table. It is verified by us that none of the MMXV-

p	q	$m_{p,q}$
2	3	2015
3	5	1.179×10^{61}
5	7	1.213×10^{11751}
7	11	$8.575 \times 10^{16468575}$
11	13	?

numbers with $p = 2$, $q \leq 1499$; $p = 3$, $q \leq 103$; $p = 5$, $q \leq 19$ and $p = 7$, $q \leq 13$ is a prime number. In particular, there is no prime MMXV-number less than 10^{500000}.

Is it true in general that none of the MMXV-numbers is a prime number?

(Is it well known that for a large integer n, the probability that a random integer not greater than n is prime is very close to $\frac{1}{\ln n}$. As the MMXV-numbers grow so fast, the conjecture about primality seems to be almost certainly true, but one never knows.)

2. A wider mathematical background and ideas for a generalization

Observe that when one uses eight instances of two parameters, with four operations and parentheses, it is not surprising that one can hardly get a target number. Therefore, let us investigate the structure of our "nice" formula more closely. Our formula for $m_{p,q}$ can be expressed as

$$m_{p,q} = p^{f(q,p)} - q \cdot f(p,q), \qquad \text{where } f(p,q) = p^q + q.$$

Clearly, for certain choices of the function $f(p,q)$, say $f(p,q) = pq$, it is easy to see that numbers $m_{p,q}$ are always composite. On the other hand, one can not expect that there is a simple choice for the function $f(p,q)$ for which $m_{p,q}$ are always prime numbers (Let us recall that Legendre had already proved that there is no rational algebraic function which always gives primes; some other formulas are known, but they are extremely inefficient, depending on some unknown real constant, etc.)

We have seen that finding of examples of f such that $m_{p,q}$ are always composite numbers is not a difficult task. Similarly, it is easy to find f providing sometimes composite numbers and sometimes primes. (We leave it to the reader.) However, it may be difficult to decide, for a given f, whether $m_{p,q}$ are always composite numbers or not. Therefore, we introduce some definitions.

If p, q, $p < q$ are two prime numbers and $f(p,q)$ is an integer valued function, then the function $m_{p,q} = p^{f(q,p)} - q \cdot f(p,q)$ will be called the *squiggly function* of the function f. Further, a squiggly function m will be called *purely composite* if its values are composite numbers for all p, q.

We can now reformulate our main problem in the following way

Is the squiggly function of $f(p,q) = p^q + q$ purely composite?

To partially solve the problem, we can also search for special primes for which the answer is positive, though it is uncertain whether it will lead us to success. For instance, let $p \equiv -1 (\mathrm{mod}\ 3)$ and $q \equiv 1 (\mathrm{mod}\ 3)$ (or analogously $p \equiv 1 (\mathrm{mod}\ 3)$ and $q \equiv -1 (\mathrm{mod}\ 3)$). Then

$$f(p,q) = (3M-1)^{3N+1} + 3N + 1 = 3M^{3N} - \cdots + 3M - 1 + 3N + 1$$

which is evidently divisible by 3 and hence $f(p,q)$ too is divisible by 3. Nevertheless, $m_{p,q}$ may have such a decomposition that it may not have any dependence on this fact; see $m_{2,7}$ above or

$$m_{2,13} = 19 \cdot 881 \cdot 14957 \cdot 527741 \cdot 2265345353295053005003345143557685258l.$$

Acknowledgment: The research is supported by Brno University of Technology, the specific research plan No. FSI-S-14-2290. The author is also grateful to the reviewer for valuable comments which have contributed in improving of this paper.

Miroslav Kureš
Institute of Mathematics
Brno University of Technology
Technická 2, 61669 Brno, Czech Republic
E-mail: *kures@fme.vutbr.cz*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 127-134

A SIMPLE PROOF OF WIENER-IKEHARA TAUBERIAN THEOREM*

AKSHAA VATWANI

(Received : 28 - 05 - 2015 ; Revised : 12 - 08 - 2015)

ABSTRACT. We present a simple and elegant proof of the Wiener-Ikehara Tauberian theorem, relying only upon the technique of contour integration. We also discuss some of its applications in number theory.

1. INTRODUCTION

To motivate the notion of Tauberian theorems, let us begin with a brief discussion of Abel's theorem. Let $\sum_{n=0}^{\infty} a_n x^n$, $x \in \mathbb{R}$ be a power series centered at 0 having radius of convergence 1. At the boundary of the region of convergence, i.e. at $|x| = 1$, the series may converge or diverge. Abel's theorem states that if the series converges at a boundary point, then it is reasonably well behaved in the sense that it is continuous at that point. More precisely, if

$$\sum_{n=0}^{\infty} a_n = A, \tag{1}$$

then

$$\lim_{x \to 1^-} \sum_{n=0}^{\infty} a_n x^n = A. \tag{2}$$

Broadly speaking, Tauberian theorems are conditional converses of Abel's theorem. They derive their name from a theorem of A. Tauber [8] published in 1897, which states that if (2) is satisfied and we have the growth condition $a_n = o(1/n)$ on the coefficients of the power series, then (1) holds. These growth conditions were subsequently relaxed, most notably by Hardy and Littlewood.

Some of the most interesting applications of Tauberian theorems pertain to analytic number theory. In this context, Tauberian results can be thought of as estimates for the partial sums of coefficients of certain Dirichlet series. An important result of this type is the Wiener-Ikehara theorem. Introduced by Ikehara [1] in 1931, it generalizes a theorem of Landau [3], by applying a Tauberian result

---

*Research is partially supported by an Ontario Graduate Scholarship.
2010 Mathematics Subject Classification: 11M45, 40E05.
Keywords and Phrases: The Wiener-Ikehara Theorem, Tauberian theorems, the prime number theorem.

obtained by Wiener. Proofs of this and other Tauberian theorems in the literature are usually found to be quite involved.

A well known application of the Wiener-Ikehara theorem is to the derivation of the prime number theorem. In 1980, Newman [4] gave an ingenious short proof of the prime number theorem. We modify Newman's proof to derive the Wiener-Ikehara Tauberian theorem. That this can be done was also recognized by Korevaar [2]. However, our presentation is simplified and our theorem more general. We derive as a consequence an assortment of prime number theorems following the arrangement of Serre [7].

2. The analytic theorem

The following analytic theorem of Newman [4], is the key result that will be used to prove the Tauberian theorem. The proof is an application of Cauchy's residue theorem. Newman's novel idea was the insertion of a new kernel into the relevant integral, playing a role similar to that of the Fejér kernel in standard proofs of the Tauberian theorem.

Theorem 1. *For $t \geq 0$, let $f(t)$ be a bounded and locally integrable function and let $g(s) := \int_0^\infty f(t)e^{-st}dt$ for $\mathrm{Re}(s) > 0$. If $g(s)$ has an analytic continuation to $\mathrm{Re}(s) \geq 0$, then $\int_0^\infty f(t)dt$ exists and equals $g(0)$.*

Proof. For $T > 0$, let $g_T(s) = \int_0^T f(t)e^{-st}dt$. This integral converges for all values of s and it is easy to see that $g_T(s)$ is an entire function. We need to show that

$$\lim_{T\to\infty} g_T(0) = g(0).$$

We will denote $\mathrm{Re}(s)$ by σ. Fix $R > 0$ and consider the positively oriented contour $\mathscr{C}$ shown in Figure 1 below. Here $\delta > 0$ (depending on R) is chosen small enough so that $g(s)$ is analytic on $\mathscr{C}$. Indeed, as $g(s)$ is analytic on the line $\sigma = 0$, one can cover the vertical strip from $(0, R)$ to $(0, -R)$ with open balls, on each of which $g(s)$ is analytic. Compactness of this strip allows one to obtain a finite subcover, which then gives the desired δ.

We use the following notations

$$\mathscr{C}_+ = \mathscr{C} \cap \{s : \sigma > 0\}, \quad \mathscr{C}_- = \mathscr{C} \cap \{s : \sigma < 0\}.$$

We also denote the semicircle of radius R to the left of the line $\sigma = 0$ by C_-. We will use the big O notation, treating everything other than the variables T, R and σ as constants.

Cauchy's theorem gives us

$$I_{\mathscr{C}} := \frac{1}{2\pi i}\int_{\mathscr{C}} (g(s) - g_T(s))e^{sT}\left(1 + \frac{s^2}{R^2}\right)\frac{ds}{s} = g(0) - g_T(0), \tag{3}$$

as the integrand is analytic inside $\mathscr{C}$ except for a simple pole at $s = 0$. We denote the corresponding integrals over $\mathscr{C}_+$ and $\mathscr{C}_-$ as $I_{\mathscr{C}_+}$ and $I_{\mathscr{C}_-}$ respectively. Let

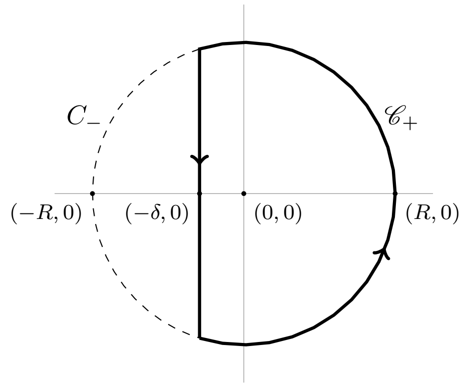


FIGURE 1. The contour $\mathscr{C}$

$M = \sup_{t\geq 0}|f(t)|$. On $\mathscr{C}_+$, as $\sigma > 0$, we have

$$|g(s) - g_T(s)| = \left|\int_T^\infty f(t)e^{-st}dt\right| \leq M\int_T^\infty e^{-\sigma t}dt \ll \frac{e^{-\sigma T}}{\sigma}.$$

Using $s = Re^{i\theta}$ and $R\cos\theta = \sigma$ on $\mathscr{C}_+$, we obtain the following estimate for the kernel

$$\left|e^{sT}\frac{1}{s}\left(1+\frac{s^2}{R^2}\right)\right| = e^{\sigma T}\left|\frac{1}{Re^{i\theta}} + \frac{e^{i\theta}}{R}\right| = e^{\sigma T}\left|\frac{2\cos\theta}{R}\right| \ll e^{\sigma T}\frac{|\sigma|}{R^2}. \tag{4}$$

Thus, the contribution to (3) from the path $\mathscr{C}_+$ of length πR is

$$|I_{\mathscr{C}_+}| \ll \frac{1}{R^2}\left|\int_{\mathscr{C}_+} ds\right| \ll \frac{1}{R}.$$

On $\mathscr{C}_-$, we examine $g_T(s)$ and $g(s)$ separately. Consider first the integral

$$I_1 := \frac{1}{2\pi i}\int_{\mathscr{C}_-} g_T(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{ds}{s}$$

As $g_T(s)$ is entire and the rest of the integrand is analytic to the left of $\sigma = 0$, we have by Cauchy's theorem,

$$I_1 = \frac{1}{2\pi i}\int_{C_-} g_T(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{ds}{s}.$$

That is, we can integrate over the semicircle C_- instead of $\mathscr{C}_-$, with C_- oriented in the same manner as $\mathscr{C}_-$. Then, noting that $\sigma < 0$ in this case, we have

$$|g_T(s)| = \left|\int_0^T f(t)e^{-st}dt\right| \leq M\int_0^T e^{-\sigma t}dt \ll \frac{e^{-\sigma T}}{|\sigma|},$$

and the estimate (4) holds on C_- exactly as it did on $\mathscr{C}_-$. We obtain $|I_1| \ll 1/R$ in the same way as done for $|I_{\mathscr{C}_+}|$ above. This leaves us with the integral

$$I_2 := \frac{1}{2\pi i}\int_{\mathscr{C}_-} g(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{ds}{s}.$$

As $\mathscr{C}_-$ is contained in a compact set on which $g(s)$ is analytic, $|g(s)|$ can be bounded in terms of R on $\mathscr{C}_-$. As the estimate (4) holds on the arcs of $\mathscr{C}_-$, the integrand in this region is of the order of

$$|\sigma|e^{\sigma T} \text{ as } T \to \infty,$$

with the implicit constant depending on R. Recalling that $\sigma < 0$ in this region, the above quantity can be compared to the real valued function xe^{-x}, which attains a global maximum of e^{-1} (as can be checked by standard derivative tests). Thus,

$$|\sigma|e^{\sigma T} \leq e^{-1}/T,$$

giving a bound of $O_R(1/T)$ for the integrand over the arcs of $\mathscr{C}_-$. As the length of the arcs is again a function of R which gets absorbed into the implied constant, we see that the contribution to $|I_2|$ from the arcs of $\mathscr{C}_-$ is $O_R(1/T)$ as $T \to \infty$. On the vertical strip of $\mathscr{C}_-$, as $\sigma = -\delta$, we have

$$|e^{sT}| = e^{-\delta T}.$$

The rest of the integrand of I_2 is analytic in this region and hence absolutely bounded in terms of R. The contribution to $|I_2|$ from this strip is thus $O_R(e^{-\delta T})$. Putting everything together, we have obtained, as $T \to \infty$,

$$\begin{aligned} |g(0) - g_T(0)| = |I_{\mathscr{C}}| &\leq |I_{\mathscr{C}_+}| + |I_1| + |I_2| \\ &\ll O\left(\frac{1}{R}\right) + O_R\left(\frac{1}{T}\right) + O_R(e^{-\delta T}). \end{aligned}$$

As R is arbitrary, the right hand side can be made as small as needed. This completes the proof. □

3. The proof of the Tauberian theorem

We establish the following version of the Tauberian theorem, applicable in many settings.

Theorem 2. *Let*

$$G(s) = \sum_{n=1}^{\infty} b_n/n^s$$

be a Dirichlet series with non-negative coefficients, satisfying

(a) $G(s)$ *is absolutely convergent for* $\mathrm{Re}(s) > 1$.
(b) *The function* $G(s)$ *extends meromorphically to the region* $\mathrm{Re}(s) \geq 1$, *having no poles except possibly a simple pole at* $s = 1$ *with residue* R.
(c) $B(x) := \sum_{n \leq x} b_n = O(x)$.

Then, as $x \to \infty$,

$$B(x) = Rx + o(x).$$

We begin by making some elementary observations. For any $\epsilon > 0$,

$$\sum_{n \le x} b_n \le \sum_{n=1}^{\infty} b_n \left(\frac{x}{n}\right)^{1+\epsilon}$$

The right hand side is $x^{1+\epsilon}G(1+\epsilon)$ which is of the order of $x^{1+\epsilon}/\epsilon$ since $G(1+\epsilon) \ll 1/\epsilon$. Choosing $\epsilon = (\log x)^{-1}$ gives $B(x) \ll x \log x$. Note that this estimate does not use any information about the behaviour of $G(s)$ on $\mathrm{Re}(s) = 1$, except at $s = 1$. Normally (c) is not needed in the general Wiener-Ikehara Tauberian theorem. One can deduce it from the other assumptions, as indicated in the concluding remarks. However, in practically all applications, this condition is found to be readily available and we retain it for the sake of a shorter proof.

A natural starting point for this and indeed most proofs of the Tauberian theorem is what is known as Abel's trick: for $\mathrm{Re}(s) > 1$, we have

$$G(s) = s \int_1^{\infty} \frac{B(x)}{x^{s+1}} dx. \tag{5}$$

This can be derived using partial summation, as is done in Exercise 2.1.5 of [5]. We proceed to prove the above theorem.

Proof of Theorem 2. Without loss of generality, we may suppose $R > 0$. Indeed, if $R \le 0$, it is enough to prove the result for $G(s) + m\zeta(s)$, where ζ is the Riemann-zeta function and m is an integer greater than $|R|$. For $R > 0$, replacing b_n by b_n/R if needed, we may assume $R = 1$. From our discussion above, we have for $\mathrm{Re}(s) > 1$,

$$\frac{G(s)}{s} - \frac{1}{s-1} = \int_1^{\infty} \frac{B(x) - x}{x^{s+1}} dx \tag{6}$$

After the change of variable x to e^u and then s to $s+1$, we have for $\mathrm{Re}(s) > 0$,

$$\frac{G(s+1)}{s+1} - \frac{1}{s} = \int_0^{\infty} \frac{B(e^u) - e^u}{e^u} e^{-su} du,$$

which is suitable for application of Theorem 1 because the function

$$f(u) := (B(e^u) - e^u)/e^u$$

is bounded on account of (c) and the left hand side has an analytic continuation to $\mathrm{Re}(s) \ge 0$ by (b). Hence, by Theorem 1, the integral

$$\int_0^{\infty} \frac{B(e^u) - e^u}{e^u} du = \int_1^{\infty} \frac{B(t) - t}{t^2} dt \tag{7}$$

converges. We will show that $B(x) \sim x$ as $x \to \infty$. Suppose not. Then either $\lim_{x\to\infty} B(x)/x$ does not exist or does not equal 1 if it exists. In either case, we see that $\limsup_{x\to\infty} B(x)/x > 1$ or $\liminf_{x\to\infty} B(x)/x < 1$. Suppose the former inequality holds (the latter case can be treated similarly). Then there exists some

$\lambda > 1$ such that $B(x) \geq \lambda x$ for infinitely many x. As there exists x arbitrarily large with $B(x) \geq \lambda x$ and $B(x)$ is an increasing function, we have

$$\int_x^{\lambda x} \frac{B(t)-t}{t^2}dt \geq \int_x^{\lambda x} \frac{\lambda x - t}{t^2}dt$$
$$= \int_1^{\lambda} \frac{\lambda x - vx}{(vx)^2}xdv = \int_1^{\lambda} \frac{\lambda - v}{v^2}dv,$$

which is a positive quantity $c(\lambda)$ (say) depending only on λ. This gives

$$\left| \int_x^{\infty} \frac{B(t)-t}{t^2}dt - \int_{\lambda x}^{\infty} \frac{B(t)-t}{t^2}dt \right| = c(\lambda)$$

For fixed λ, as $x \to \infty$, the above integrals are tails of the convergent integral (7) and can be made arbitrarily small, thereby giving a contradiction. This completes the proof. □

The result can be extended to Dirichlet series with complex coefficients as follows.

Corollary 3. *Let*

$$F(s) = \sum_{n=1}^{\infty} a_n/n^s$$

be a Dirichlet series with complex coefficients. Let $A(x)$ denote the partial sum of the coefficients:

$$A(x) = \sum_{n \leq x} a_n.$$

Suppose there exists a Dirichlet series $G(s) = \sum_{n=1}^{\infty} b_n/n^s$ with non-negative coefficients, such that

(a) $|a_n| \leq b_n$ *for all* n.
(b) $G(s)$ *is absolutely convergent for* $\mathrm{Re}(s) > 1$.
(c) *The function $G(s)$ (resp. $F(s)$) extends meromorphically to the region* $\mathrm{Re}(s) \geq 1$, *having no poles except for a simple pole at $s = 1$ with residue R (resp. r).*
(d) $B(x) := \sum_{n \leq x} b_n = O(x)$.

Then, as $x \to \infty$,

$$A(x) = rx + o(x).$$

Proof. If a_n's are real, we consider the series $G(s) - F(s)$, which has non-negative coefficients and satisfies the conditions of Theorem 2, giving

$$\sum_{n \leq x} (b_n - a_n) = (R - r)x + o(x),$$

as $x \to \infty$. As $B(x) = Rx + o(x)$, this proves the result in the case of real coefficients. If the coefficients a_n are not real, we define

$$F^*(s) = \sum_{n=1}^{\infty} \bar{a}_n/n^s$$

so that

$$F = \frac{F + F^*}{2} + i\left(\frac{F - F^*}{2i}\right).$$

and apply the result for real coefficients separately to the real and imaginary part above after checking that the necessary conditions are satisfied. □

4. Applications

In this section we demonstrate some applications of the Tauberian theorem, following the treatment of Serre [7] who gives a general set-up for the same in the context of equidistribution.

We make this more precise in an abstract setting as follows. Let G be a compact group and X be the space of conjugacy classes of G. Let x_v be a family of elements of X, indexed by a countably infinite set $\mathcal{P}$. Let $N : \mathcal{P} \to \mathbb{Z}$ be a function taking values ≥ 2, ρ an irreducible complex representation of G with character χ. We define

$$\zeta_{\mathcal{P}}(s) = \prod_{v\in\mathcal{P}} \left(1 - \frac{1}{(Nv)^s}\right)^{-1}, \qquad L(s,\rho) = \prod_{v\in\mathcal{P}} \det\left(1 - \frac{\rho(x_v)}{(Nv)^s}\right)^{-1}.$$

Thus, for the trivial representation $\rho = 1$, $L(s,1) = \zeta_{\mathcal{P}}(s)$.

Theorem 4. *Suppose $L(s,\rho)$ is absolutely convergent for $\mathrm{Re}(s) > 1$ and extends to a meromorphic function on $\mathrm{Re}(s) \geq 1$ with no zeros or poles except for a pole of order c_χ at $s = 1$. Then,*

$$\sum_{Nv\leq n} \chi(x_v) = (1 + o(1))c_\chi \frac{n}{\log n}.$$

The proof of the above theorem follows by applying the Tauberian theorem to L'/L. We refer the reader to the appendix of Chapter 1 of [7] for the same. If Theorem 4 holds for all irreducible representations $\rho \neq 1$ with $c_\chi = 0$, then the Peter-Weyl theorem allows us to deduce that the x_v's are equidistributed with respect to the normalized Haar measure of G. Special cases of this theorem lead to important results, among them being the prime number theorem, Chebotarev density theorem and the Sato-Tate theorem. An excellent reference for the interested reader wishing to delve deeper into these topics is [6].

5. Concluding remarks

As remarked earlier, the added condition (c) in Theorem 2 is not restrictive for most practical purposes. However, it is possible to eliminate this condition altogether. We give a brief sketch of the argument. The key idea is to notice that the known bound $B(x) \ll x\log x$ implies that for any $\epsilon > 0$, the function

$$f_\epsilon(t) := \frac{f(t)}{e^{\epsilon t}} = \frac{B(e^t)}{e^{t(1+\epsilon)}} - \frac{1}{e^{\epsilon t}}$$

is bounded and satisfies the conditions of Theorem 1. Applying this theorem to $f_\epsilon(t)$ and following an elementary argument that exploits the increasing behaviour of the function $B(e^t)/e^{t(1+\epsilon)}$, one obtains a uniform bound on $\sup_{t\geq 0}|f_\epsilon(t)|$. Letting $\epsilon \to 0$, we see that $f(t)$ must be bounded. A more detailed proof can be found in [2].

Acknowledgment: I would like to thank Professor M. Ram Murty for several useful discussions and helpful comments.

References

[1] Ikehara, S., *An extension of Landau's theorem in the analytic theory of numbers*, Journal of Math. and Phys. of the Mass. Inst. of Technology, **10** (1931), 1-12.

[2] Korevaar, J., *The Wiener-Ikehara theorem by complex analysis*, Proc. Amer. Math. Soc, **134**, 4 (2005), 1107-1116.

[3] Landau, E., *Über die Betdeutung einiger neuerer Grenzwertsätze der Herren Hardy und Axer*, Prace mat.-Fiz., **21** (1910), 97-177.

[4] Newman, D. J., *Simple analytic proof of the prime number theorem*, Amer. Math. Monthly, **87** (1980), 693-696.

[5] Ram Murty, M., *Problems in Analytic number theory*, 2nd Edition, Graduate Texts in Mathematics, Springer, New York, 2008.

[6] Ram Murty, M. and Kumar Murty, V., *Non-vanishing of L-functions and applications*, Modern Birkhäuser Classics, Birkhäuser/Springer Basel AG, Basel, 1997.

[7] Serre, J. P., *Abelian l-Adic Representations and Elliptic curves*, Lectures at McGill University, New York-Amsterdam, W. A. Benjamin Inc., 1968.

[8] Tauber, A., *Ein Satz aus der Theorie der unendlichen Reihen*, Monatshefte f. Math., **8** (1897), 273-277.

Akshaa Vatwani
Department of Mathematics and Statistics
Queen's University, Kingston, Ontario K7L 3N6, Canada
E-mail: *akshaa@mast.queensu.ca*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 135-139

AN ANCIENT DIOPHANTINE PROBLEM FROM THE BIJAGANITA OF BHASKARACHARYA

AJAI CHOUDHRY

(Received : 17 - 06 - 2015 ; Revised : 01 - 08 - 2015)

Abstract. This paper is concerned with the ancient diophantine problem of finding two perfect squares such that both, their sum and difference, when increased by unity, become perfect squares. Bhaskaracharya, in his 12$^{\text{th}}$ century treatise on algebra, had given an elegant method of finding integer solutions of this problem. Subsequently, only one parametric solution has been published. We obtain an infinite sequence of parametric solutions of this problem. All of these parametric solutions readily yield integers with the desired property. We also obtain the complete solution in rational numbers of this diophantine problem.

1. Introduction

Bhaskaracharya, in his 12$^{\text{th}}$ century treatise on algebra entitled *Bijaganita*, gives an elegant method of finding integer solutions of the following two diophantine problems which he attributes to an ancient author [1, pp. 257–259]:

"Calculate and tell, if you know, two numbers, the sum and difference of whose squares, with one added to each, are squares: or which are so, with the same subtracted."

If the two numbers are taken as x and y, the first problem requires that we solve the simultaneous diophantine equations,

$$x^2 + y^2 + 1 = u^2, \tag{1.1}$$

$$x^2 - y^2 + 1 = v^2, \tag{1.2}$$

while for the second problem, we must solve the simultaneous equations,

$$x^2 + y^2 - 1 = u^2, \tag{1.3}$$

$$x^2 - y^2 - 1 = v^2. \tag{1.4}$$

It is interesting to note that, seven centuries later, the first problem was posed as a problem for solution in the reputed American Mathematical Monthly. A parametric solution was published in the same journal by Drummond, namely,

2010 Mathematics Subject Classification: 11D09.
Keywords and Phrases: two quadratic functions made perfect squares.

$x = 2n^2,\ y = 2n$ [3]. Bhaskaracharya was aware of this solution and in addition, his method generates infinitely many numerical solutions in integers that cannot be obtained by Drummond's parametric solution (eg. $x = 38,\ y = 34$). No other solutions of the first problem seem to have been published. In fact, Piezas [4] has posed the question whether there are any other parametric solutions of the problem.

With respect to the second problem, a fairly comprehensive solution is known. Several authors have given parametric solutions that yield integer solutions of the problem while Genocchi and Pepin have independently given the complete solution of the problem in rational numbers (as quoted by Dickson [2, pp. 479–480]).

This paper accordingly focuses on the first problem mentioned by Bhaskaracharya. We obtain a sequence of infinitely many parametric solutions that yield integer solutions of the problem, the first solution in the sequence being the parametric solution given by Drummond. We also obtain the complete solution of the problem in rational numbers.

2. Integer Solutions of the simultaneous equations $x^2 + y^2 + 1 = u^2,\ x^2 - y^2 + 1 = v^2$

The only integer solutions of equations (1.1) and (1.2) with $y = 0$ are readily seen to be the trivial solutions given by $(x, y, u, v) = (0, 0, \pm 1, \pm 1)$. To obtain solutions of (1.1) and (1.2) in which $y \neq 0$, we take their difference when we get,

$$2y^2 = u^2 - v^2. \tag{2.1}$$

The complete solution of Eq. (2.1), using the obvious solution $(u, v, y) = (1, 1, 0)$, as the initial known solution, is readily obtained and is given by,

$$u = (2p^2 + q^2)t, \quad v = (2p^2 - q^2)t, \quad y = 2pqt, \tag{2.2}$$

where p, q, t are arbitrary integer parameters. Since $y \neq 0$, the three parameters p, q, r must all necessarily be taken as nonzero.

Substituting the values of u, v and y given by (2.2) in (1.1), we get, after suitable transpositions, the following equation:

$$x^2 - (4p^4 + q^4)t^2 = -1. \tag{2.3}$$

To obtain integer solutions of (2.3), we take $q = 1$ when (2.3) may be considered as the Pell's equation $x^2 - dy^2 = -1$ where $d = 4p^4 + 1$ with an obvious solution being given by $x = 2p^2,\ t = 1$. We will use the obvious known solution of (2.3) to obtain a sequence of infinitely many parametric solutions $(x_n, t_n),\ n = 1, 2, 3, \ldots....$ of Eq. (2.3). We take $(x_1, t_1) = (2p^2, 1)$ as the first solution of the sequence so that

$$x_1^2 - dt_1^2 = -1, \tag{2.4}$$

where $d = 4p^4 + 1$. Let (x_n, t_n) be the n^{th} solution of the sequence so that

$$x_n^2 - dt_n^2 = -1. \tag{2.5}$$

We recall the well-known identity related to composition of forms, namely,

$$(X_1^2 - dX_2^2)(Y_1^2 - dY_2^2) = (X_1Y_1 + dX_2Y_2)^2 - d(X_1Y_2 + X_2Y_1)^2, \tag{2.6}$$

and apply it twice so that,

$$\begin{aligned}(x_n^2 - dt_n^2)(x_1^2 - dt^2)^2 = \{&(x_1^2 + dt_1^2)x_n + 2dt_1t_nx_1\}^2 \\ &- d\{(x_1^2 + dt_1^2)t_n + 2t_1x_1x_n\}^2.\end{aligned} \tag{2.7}$$

Now, on writing,

$$x_{n+1} = (x_1^2 + dt_1^2)x_n + 2dt_1t_nx_1, \quad t_{n+1} = (x_1^2 + dt_1^2)t_n + 2t_1x_1x_n, \tag{2.8}$$

it follows from (2.7), (2.4) and (2.5) that,

$$x_{n+1}^2 - dt_{n+1}^2 = (-1)^3 = -1. \tag{2.9}$$

Since $(x_1, t_1) = (2p^2, 1)$ is a known parametric solution of Eq. (2.3), we can now generate a sequence of parametric solutions of Eq. (2.3) using the recurrence relations (2.8). These recurrence relations may be written explicitly by substituting $x_1 = 2p^2$, $t_1 = 1$, and $d = 4p^4 + 1$ in (2.8) when we get the following relations:

$$\begin{aligned} x_{n+1} &= (8p^4 + 1)x_n + 4p^2(4p^4 + 1)t_n, \\ t_{n+1} &= 4p^2x_n + (8p^4 + 1)t_n. \end{aligned} \tag{2.10}$$

Finally, using (2.2) and noting that we have taken $q = 1$, we can obtain a sequence (x_n, y_n), $n = 1, 2, 3, \ldots\ldots$, of univariate polynomials in terms of the parameter p such that (x_n, y_n) give a solution to our diophantine problem. The first solution of this sequence is readily seen to be $(x_1, y_1) = (2p^2, 2p)$ which is precisely the parametric solution given by Drummond. It follows from (2.10) and (2.2) that the sequence of solutions (x_n, y_n), $n = 1, 2, 3, \ldots\ldots$, is given by the recurrence relations,

$$\begin{aligned} x_{n+1} &= (8p^4 + 1)x_n + 2p(4p^4 + 1)y_n, \\ y_{n+1} &= 8p^3x_n + (8p^4 + 1)y_n, \end{aligned} \tag{2.11}$$

where p is an arbitrary parameter.

As already noted, the first parametric solution of this sequence of solutions is given by $(x_1, y_1) = (2p^2, 2p)$. The next three solutions of the sequence are as follows:

$x_2 = 2p^2(16p^4 + 3)$, $\quad$ $y_2 = 2p(16p^4 + 1)$,

$x_3 = 2p^2(256p^8 + 80p^4 + 5)$, $\quad$ $y_3 = 2p(256p^8 + 48p^4 + 1)$,

$x_4 = 2p^2(4096p^{12} + 1792p^8 + 224p^4 + 7)$, $\quad$ $y_4 = 2p(4096p^{12} + 1280p^8 + 96p^4 + 1)$,

where, in each case, p is an arbitrary parameter.

It readily follows by induction that x_n and y_n are given by polynomials of degrees $4n - 2$ and $4n - 3$ respectively, and the leading coefficient of both these polynomials is 2^{4n-3}. As the degrees of the polynomials x_n form a strictly monotonically increasing sequence of positive integers, we are assured of an infinite sequence of distinct parametric solutions of Bhaskaracharya's problem.

As a numerical example, taking $p = 1$, we get the four solutions, $(x_1, y_1) = (2, 2)$, $(x_2, y_2) = (38, 34)$, $(x_3, y_3) = (682, 610)$ and $(x_4, y_4) = (12238, 10946)$.

3. The complete solution in rational numbers of the simultaneous equations $x^2 + y^2 + 1 = u^2$, $x^2 - y^2 + 1 = v^2$

We will first obtain all rational solutions of equations (1.1) and (1.2) in which $y = 0$. On substituting $y = 0$ in these equations, we get $x^2+1 = u^2$ and $x^2+1 = v^2$. These equations are readily solved and their complete solution in rational numbers is given by

$$x = (m^2 - 1)/(2m), \quad u = \pm(m^2 + 1)/(2m), \quad v = \pm(m^2 + 1)/(2m), \qquad (3.1)$$

where m is an arbitrary nonzero rational parameter.

We will now obtain all rational solutions of equations (1.1) and (1.2) in which $y \neq 0$. As in Section 2, we obtain the relations (2.2) and (2.3) where, as before, the three parameters p, q and t are necessarily nonzero. Now (2.3) may be considered as a quadratic equation in the variables x and t, and an obvious rational solution of this equation is given by $(x, t) = (2p^2/q^2, \ 1/q^2)$. All solutions, in rational numbers, of (2.3) may now be obtained by writing,

$$x = rz + 2p^2/q^2, \quad t = sz + 1/q^2, \qquad (3.2)$$

where r, s and z are arbitrary parameters such that $(r, s) \neq (0, 0)$ and $z \neq 0$, these conditions being necessary as otherwise we just get the obvious solution mentioned above. With these values of x and t, Eq. (2.3) reduces to a linear equation in z, which is readily solved, and we thus obtain the following solution of (2.3):

$$\begin{aligned} x &= -\frac{2(p^2r^2 - (4p^4 + q^4)rs + (4p^4 + q^4)p^2s^2}{q^2(r^2 - (4p^4 + q^4)s^2)}, \\ t &= \frac{r^2 - 4p^2rs + (4p^4 + q^4)s^2}{q^2(r^2 - (4p^4 + q^4)s^2)}. \end{aligned} \qquad (3.3)$$

We note here that on substituting $r = 4p^4 + q^4$ and $s = 2p^2$ in (3.3), we get the initial known solution $(x, t) = (2p^2/q^2, \ 1/q^2)$ which is thus included in the solution (3.3).

Now, on using (2.2), we obtain the following solution in rational numbers of equations (1.1) and (1.2):

$$\begin{aligned} x &= -\frac{2(p^2r^2 - (4p^4 + q^4)rs + (4p^4 + q^4)p^2s^2}{q^2(r^2 - (4p^4 + q^4)s^2)}, \\ y &= \frac{2p(r^2 - 4p^2rs + (4p^4 + q^4)s^2)}{q(r^2 - (4p^4 + q^4)s^2)}, \\ u &= \frac{(2p^2 + q^2)(r^2 - 4p^2rs + (4p^4 + q^4)s^2)}{q^2(r^2 - (4p^4 + q^4)s^2)}, \end{aligned} \qquad (3.4)$$

$$v = \frac{(2p^2 - q^2)(r^2 - 4p^2rs + (4p^4 + q^4)s^2)}{q^2(r^2 - (4p^4 + q^4)s^2)},$$

where p, q, r, and s are arbitrary rational parameters such that $pq \neq 0$ and $(r, s) \neq (0, 0)$.

It is to be noted that the elliptic curves $Y^2 = 4X^4 + 1$ and $Y^2 = X^4 + 4$ are both of rank 0, and the only rational points on these curves are given by $X = 0$. It follows that when the parameters p, q, r, s satisfy the conditions $pq \neq 0$ and $(r, s) \neq (0, 0)$, we will necessarily have $r^2 - (4p^4 + q^4)s^2 \neq 0$, and thus, under these conditions, the denominators of the values of x, y, u, v given by (3.4) are all nonzero, and hence (3.4) always generates well-defined rational solutions. Further, it is readily seen that, under the given conditions, the numerator of the value of y given by (3.4) is always nonzero since $p \neq 0$ and the discriminant of the factor $r^2 - 4p^2rs + (4p^4 + q^4)s^2$, considered as a quadratic function of r and s, is $-4q^4 < 0$ so that this factor is always positive. Thus, the value of y given by (3.4) is always nonzero.

We now have the complete solution in rational numbers of equations (1.1) and (1.2). All rational solutions of equations (1.1) and (1.2) in which $y = 0$ are given by (3.1) and all rational solutions in which $y \neq 0$ are given by (3.4).

As a numerical example, taking $(p, q, r, s) = (1, 1, -1, 1)$ in (3.4), we get the rational numbers 11/2 and 5 as a solution to our problem.

Acknowledgement: I am grateful to the anonymous referee for his comments which have led to improvements in the paper. I also wish to thank the Harish-Chandra Research Institute, Allahabad, for providing me all necessary facilities that have helped me to pursue my research work in mathematics.

References

[1] Colebrooke, H. T., *Algebra with arithmetic and mensuration, from the Sanscrit of Brahmegupta and Bhascara*, John Murray, London, 1817.

[2] Dickson, L. E., *History of the Theory of Numbers Vol. 2*, Chelsea Publishing Company, 1992, reprint.

[3] Drummond, J. H., *Problem 98, Solution by Hon. Josiah H. Drummond*, Amer. Math. Monthly **9** (1902), 232.

[4] Piezas, T., *A collection of algebraic identities*, website at https://sites.google.com/ site/ tpiezas/ 007 (accessed on 15 June 2015).

Ajai Choudhry
13/4 A Clay Square, Lucknow - 226001, UP, India
E-mail: *ajaic203@yahoo.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 141-145

COMPLETENESS AND INVERTIBILITY

S. H. KULKARNI

(Received : 03 - 07 - 2015)

Abstract. We show that two very important concepts in Functional Analysis, namely the completeness of a normed linear space and invertibility of a bounded linear map are related to each other. This gives a possibly new characterization of completeness.

1. Introduction

Answers to many important questions in Functional Analysis depend upon knowing whether a certain normed linear space is complete or/and whether a certain bounded linear map has a bounded linear inverse. Usually students do not think that these two important ideas in Functional Analysis, namely completeness and invertibility, have anything to do with each other. In this note, we try to draw the attention of students to connections between these ideas.

The following well known theorem is given in many textbooks of Functional Analysis. (See for example, [1].)

Theorem 1.1. *Let T be a bounded(continuous) linear map from a Banach space X to a normed linear space Y. Then the following are equivalent:*
1. T has a bounded inverse.
2. T is bounded below and the range of T is dense in Y.

It is natural to ask what happens if the hypothesis of completeness of X is dropped. Somehow, this question is not discussed in the textbooks. It is obvious that (1) would still imply (2) even without completeness. But the converse is false and it is easy to construct a counterexample. We give such an example. Further, it is interesting to note that (2) is equivalent to the following even without the completeness of X.
3. The transpose T' of T has a bounded inverse.
Even more interesting is the fact that the completeness of X is equivalent to the invertibility of every bounded linear map satisfying (2).

2. preliminaries

We recall a few standard notations, definitions and results that are used in the next section. For normed linear spaces X, Y, we denote by $BL(X, Y)$ the set of

2010 AMS Subject Classification: 46B99, 47A05.
Key words and phrases: Completeness, invertibility, transpose, bounded below, spectrum

all bounded linear operators from X to Y. For an operator $T \in BL(X,Y)$, $N(T)$ denotes the null space of T and $R(T)$ denotes the range of T. Thus

$$N(T) = \{x \in X : T(x) = 0\} \qquad \text{and} \qquad R(T) = \{T(x) : x \in X\}.$$

Further, T is said to be *bounded below* if there exists $\alpha > 0$ such that $\|T(x)\| \geq \alpha\|x\|$ for all $x \in X$ and T is said to be *invertible* if there exists $S \in BL(Y,X)$ such that $ST = I_X$, the identity map on X, and $TS = I_Y$, the identity map on Y. The *dual space* X' of X, is the set of all bounded linear functionals on X, that is, $X' = BL(X, \mathbb{K})$, where $\mathbb{K}$ is the underlying field of real or complex numbers. For a subset $A \subseteq X$, the *annihilator* A^0 is the set of all continuous linear functionals that vanish on A, that is, $A^0 := \{\phi \in X', \phi(a) = 0 \text{ for all } a \in A\}$. If A is a subspace of X, then it follows by the Hahn-Banach Theorem, that A is dense in X, if and only if $A^0 = \{0\}$. The *transpose* T' of $T \in BL(X,Y)$ is the operator in $BL(Y', X')$ defined by
$(T'\psi)(x) := \psi(T(x))$ for all $x \in X$ and $\psi \in Y'$. All the other notations (including the notations for sequence spaces c_{00}, ℓ^1 etc.) are as in [1] and [2]. We shall make use of the following well known results:
1. $(R(T))^0 = N(T')$.
2. Every normed linear space X can be viewed as a dense subspace of a Banach space which we shall denote by X_c. (More precisely, there is a linear isometry of X onto a dense subspace of X_c.) The Banach space X_c is called the *completion* of X. These results can be found in any book on Functional Analysis, for example [1] and [2].

3. Notes

We begin with an example.

Example 3.1. Let $X := (c_{00}, \|.\|_1)$, $Y := \ell^1$ and $T : X \to Y$ be given by $T(x) = x$ for $x \in X$. Clearly, T is bounded below, range of T is dense in Y, but T is not onto and hence not invertible. More generally, we can consider the inclusion map from a proper dense subspace of a normed linear space.

Remark 3.2. Note that in the above example, though T is not invertible, its transpose T' is invertible. In fact, both the dual spaces X' of X and Y' of Y can be identified with ℓ^∞ in the usual way (See [2] for details.) and with respect to this identification T' becomes the identity operator on ℓ^∞.

This leads to some natural observations. First we consider some elementary results. The following elementary result is given as an Exercise in some books.

Lemma 3.3. *Let T be a bounded(continuous) linear map from a normed linear space X to a normed linear space Y. Then $R(T)$ is dense in Y if and only if T' is injective.*

Proof. Recall that $R(T)$ is dense in Y if and only if $\{0\} = (R(T))^0 = N(T')$ if and only if T' is injective. $\square$

Lemma 3.4. *Let T be a bounded(continuous) linear map from a normed linear space X to a normed linear space Y. Then the following statements are equvivalent:*
1. T has a bounded inverse from $R(T)$ to X.
2. T is bounded below.
3. T' is onto.

Proof. (1) implies (2): This is easy. Suppose $S : R(T) \to X$ is a bounded inverse of T. Then for each $x \in X$,
$\|x\| = \|ST(x)\| \leq \|S\|\|T(x)\|$, that is, $\|T(x)\| \geq \frac{1}{\|S\|}\|x\|$.

(2) implies (1) and (3): Since T is bounded below, there exists $\alpha > 0$ such that $\|T(x)\| \geq \alpha\|x\|$ for all $x \in X$. In particular, T is injective. Hence we can define a map $S : R(T) \to X$ by $S(y) = x$ for $y = T(x) \in R(T)$. This is well defined since T is injective. It is easy to see that S is linear. Also
$\|S(y)\| = \|x\| \leq \frac{1}{\alpha}\|T(x)\| = \frac{1}{\alpha}\|y\|$.
Hence S is bounded. This proves (1).

Next let $\phi \in X'$ and $y \in R(T)$. There exists unique $x \in X$ such that $y = T(x)$. Define ψ by $\psi(y) := \psi(T(x)) = \phi(x)$. This defines ψ as a linear functional on $R(T)$. Further,
$|\psi(y)| = |\psi(T(x))| = |\phi(x)| \leq \|\phi\|\|x\| \leq \|\phi\|\frac{1}{\alpha}\|T(x)\| = \|\phi\|\frac{1}{\alpha}\|y\|$.
This shows that ψ is bounded on $R(T)$ and hence has a bounded (norm preserving) extension to Y by the Hahn-Banach Theorem. We denote this extension also by the same symbol ψ. Thus $\psi \in Y'$ and $\phi = T'(\psi)$. This shows that T' is onto.

(3) implies (2): Let $x \in X$. By the Hahn-Banach Theorem, there exists $\phi \in X'$ such that $\phi(x) = \|x\|$ and $\|\phi\| = 1$. Further, since T' is onto, there exists $\psi \in Y'$ such that $\phi = T'(\psi)$. Now
$\|x\| = \phi(x) = T'(\psi)(x) = \psi(T(x)) \leq \|\psi\|\|T(x)\|$, that is, $\|T(x)\| \geq \frac{1}{\|\psi\|}\|x\|$
This shows that T is bounded below. □

We now give the main theorem.

Theorem 3.5. *Let T be a bounded(continuous) linear map from a normed linear space X to a normed linear space Y. Consider the following statements:*
1. T has a bounded inverse.
2. T is bounded below and the range of T is dense in Y.
3. T' is invertible.

Then (2) and (3) are equivalent and each is implied by (1). If, in addition, X is a Banach space, then all the three statements are equivalent.

Proof. (1) implies (2): Obvious. Since T has a bounded inverse, $R(T) = Y$. Also T is bounded below by Lemma 3.4.

(2) if and only if (3): By Lemma 3.3, $R(T)$ is dense in Y, if and only if T' is injective. Further, by Lemma 3.4, T is bounded below, if and only if, T' is onto. Thus (2) is equivalent to the following:

$T' : Y' \to X'$ is a bijection.
Hence T' is invertible by the Closed Graph Theorem as X', Y' are Banach spaces.

Finally, if X is a Banach space, then (2) implies (1) by Theorem 1.1 and hence all the three statements are equivalent. $\square$

Remark 3.6. We may further note that completeness of X is , in fact, equivalent to the invertiblity of every bounded linear map satisfying (2). In other words, a normed linear space X is a Banach space if and only if every bounded linear map T from X to any normed linear space Y such that T is bounded below and the range of T is dense in Y, is invertible. The only if part is already proved above. To prove the if part, consider $Y = X_c$, the completion of X. Then there is a linear isometry T of X onto a dense subspace Y_0 of Y. (See [2] for details.) Obviously, this T is bounded below and $R(T) = Y_0$ is dense in Y. Hence by the hypothesis, T is invertible and, in particular, onto. Thus X is linearly isometric to Y and hence complete.

Remark 3.7. It is known that the invertibility of an operator is closely related to its spectrum. Let X be a complex normed linear space and $T \in BL(X, X)$. Recall that the *spectrum* $\sigma(T)$ of T is the set of all complex numbers λ such that $\lambda I - T$ is not invertible. Applying Theorem 3.5 to $\lambda I - T$, we obtain the known result that $\sigma(T') \subseteq \sigma(T)$ and the equality holds if X is a Banach space. (See [2]) (The inclusion can be strict if X is not a Banach space. See the next example.) A natural question is whether the converse holds. In other words, can the completeness be also characterized in terms of spectra as follows: A complex normed linear space X is complete if and only if $\sigma(T') = \sigma(T)$ for all $T \in BL(X, X)$? Another formulation of the same question is as follows: Given an incomplete normed linear space X, does there exist $T \in BL(X, X)$ such that T is bounded below, its range is dense in X and T is not invertible (that is, not onto)? Note that the above examples and remarks do not answer this question as the spaces X and Y considered there are different.

Example 3.8. This example shows that the inclusion $\sigma(T') \subseteq \sigma(T)$ can be strict if X is not complete.

Let $X := (c_{00}, \|.\|_2)$, and $T : X \to X$ be the right shift operator given by $T(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$ for $x := (x_1, x_2, \ldots) \in X$. Then the dual space X' can be identified with ℓ^2 and the transpose T' of T can be identified with the left shift operator. (See [2] for details.) Then it can be shown that
$\sigma(T') = \{z \in \mathbb{C} : |z| \leq 1\}$, the closed unit disc. On the other hand, it is easy to see that the equation $(\lambda I - T)x = e_1 = (1, 0, 0, \ldots)$ has no solution $x \in X$ for any complex number λ. In other words, $\lambda I - T$ is not onto. Thus $\sigma(T) = \mathbb{C}$.

References

[1] Bollobás, B., *Linear analysis*, Cambridge Univ. Press, Cambridge, 1990. MR1087297 (92a:46001)

[2] Limaye, B. V. *Functional analysis*, Second edition, New Age, New Delhi, 1996. MR1427262 (97k:46001)

S. H. Kulkarni
Department of Mathematics
Indian Institute of Technology Madras
Chennai 600036, Tamilnadu, India
E-mail: *shk@iitm.ac.in*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 147

A CONJECTURE ON PRIME NUMBERS

NEELABH DEKA

(Received : 11 - 07 - 2015)

Abstract. A new conjecture on prime numbers is proposed in this short note.

Conjecture 1. *Let p_n denote the n-th prime number. If we choose n consecutive natural numbers from the interval $[2, p_n^2 - n]$ such that they are divisible by $p_1, p_2, \cdots, p_n$ not necessarily in order, then the next set of n consecutive natural numbers would contain at least one prime number.*

For example, if $n = 3$, then $p_n = 5$; and we choose three consecutive natural numbers between 2 and 22, say $8, 9$ and 10. Here 2 divides 8, 3 divides 9 and 5 divides 10. Then in the next set of three consecutive natural numbers we get 11 which is a prime.

A related conjecture would be to guess that the previous set of n consecutive natural numbers before our choice would also contain at least one prime number.

If this conjecture is true, then we may hope to apply it to other conjectures like Brocard's Conjecture or Grimms's Conjecture.

References

[1] Andrica, D., Note on a conjecture in prime Number Theory. Elsevier, Amsterdam, 2007.
[2] Erdös, P., Beweis eines Satzes von Tschebyschef. 2007.
[3] Paz, G. A., *On Legendre's, Brocard's, Andrica's and Oppermann's Conjectures*, arxiv:1310.1323v2 [math.NT]. Apr 2, 2014.

Neelabh Deka
Assam, India
E-mail : *neelabhdeka1997@gmail.com*

2010 Mathematics Subject Classification: Primary 11A41. Secondary 11B25, 11N05, 11N13.
Key words and phrases : Prime number, consecutive natural numbers.

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 149-158

COLORED TOPOLOGICAL TVERBERG THEOREM OF BLAGOJEVIĆ, MATSCHKE AND ZIEGLER

SATYA DEO*

(Received : 31 - 08 - 2015 ; Revised : 10 - 10 - 2015)

Abstract. This is an expository article dealing with the colored topological Tverberg thereom formulated and proved by Pavle Blagojević, Benjamin Matschke and Günter Ziegler. We present the simplest of their proofs of this theorem using the degree concept of the equivariant maps and the well known configuration space/ test map method of the topological combinatorics.

1. Introduction

Let $d \geq 1$ and $r \geq 2$ and $N = (d+1)(r-1)$. The classical Tverberg Theorem says that if we have $N+1$ points in the Euclidean space $\mathbb{R}^d$ in general position, then we can decompose these points into r disjoint subsets $F_1, F_2, \cdots, F_r$ such that the intersection of their convex hulls viz $\cap_{i=1}^{r}\text{conv }(F_i)$, is nonempty [15]. This theorem is true for any d and any r as stated above. For the case $d = 2$, it was proved by B. Birch [5] in 1959. For higher dimensional space $\mathbb{R}^d$, the above theorem was conjectured by Birch and proved by H. Tverberg [15] in 1966. It is easily seen that the above theorem can be stated as follows.

Theorem 1.1. *Let $\triangle_N$ denote the $N-$ simplex and $f : \triangle_N \to \mathbb{R}^d$ be any linear map. Then one can find r disjoint faces $\sigma_1, \sigma_2, \cdots, \sigma_r$ of $\triangle_N$ such that*

$$\cap_1^r f(\sigma_i) \neq \phi.$$

Looking at the above version of the Tverberg theorem one is naturally tempted to ask the following topological question:

Question: If $f : \triangle_N \to \mathbb{R}^d$ is any continuous map, not necessarily a linear map, can we still find a family of r disjoint faces $\sigma_1, \sigma_2, \cdots, \sigma_r$ of $\triangle_N$ such that

$$\cap_1^r f(\sigma_i) \neq \phi\,?$$

The above question was first answered by BSS [4] in 1981, who proved that the answer to above question is in affirmation provided the number r is prime. In

* The author was a Senior Scientist of the National Academy of Sciences, India, at Harish-Chandra Research Institute, Allahabad when this work was done.

2010 AMS Subject Classification: Primary 55N05, Secondary 15A66.

Key words and phrases: Tverberg theorem, chessboard complexes, degree of an equivariant map, configuration space/test map method, deleted join

order to prove this result, BSS made use of the Borsuk-Ulam Theorem for $\mathbb{Z}_p$-actions from algebraic topology. Later, M. Özaydin [14] proved that the answer to the above question remains affirmative even if r is a prime power, say $r = p^k$ for any $k \geq 1$. These results became known as topological Tverberg theorems. The general question for arbitrary r remained as one of the most challenging problems of topological combinatorics until very recently when in February 2015, Florian Frick [11] gave an example to show that the answer to the above question, when r is not a power of prime is in negative. In fact, Frick produced an example of a continuous map $f : \triangle_N \to \mathbb{R}^d,\ r,\ d$ suitably chosen, such that whenever we take any family $\sigma_1, \sigma_2, \cdots, \sigma_r$ of r disjoint faces of $\triangle_N$, $\cap_1^r f(\sigma_i) = \phi$. This solved one of the most important problems of topological combinatorics and the position of topological Tverberg theorem became remarkably clear for all possible value of r.

What would be the colored version of the topological Tverberg Theorem ?

This question was first studied by Bárány, Füredi and Lovász [2] in 1990, who proved, among other things, that if we take 21 points in the plane in general position such that 7 are red, 7 are blue and 7 are green, then we can decompose them into 7 triangles, each triangle having vertices of different colors, such that their intersection is non empty. Somewhat later in 1992, Bárány and Larman [3] studied the above question further and proved the following general result:
If we take $3r$ points in the plane in general position where r are of red color, r are of blue color and r are of green color, then we can find r disjoint triangles, each triangle having vertices of different colors, such that their intersection will be non empty.
For higher dimensional Euclidean spaces $\mathbb{R}^d$, Bárány and Larman posed the following problem:
Given r and $N = (d+1)t \geq (d+1)r$ points in $\mathbb{R}^d$, determine the smallest t such that if we take $d+1$ color classes of size $\leq t$, we can find a family of r disjoint d-simplices having vertices of different colors, such that their intersection is nonempty.

Živaljević and Vrećica [18] (ZV in short) introduced the concept of chessboard complexes and showed that when r is a prime, the above colored Tverberg Theorem holds for $t \geq 2r-1$. They also indicated that the theorem is true for $t \geq 4r-3$ for any $r \geq 2$ due to Bertrand's postulate that there is a prime between r and $2r$. This colored Tverberg Theorem of ZV attracted a lot of attention, but could not be considered very satisfactory for two reasons. The first reason was that the classical Tverberg Theorem does not appear as a special case of this Theorem, and the second reason was that several colored points were left out unaccounted. In 2009, Blagojević, Matschke and Ziegler (see [6]) formulated and proved a remarkable new colored topological Tverberg Theorem which is stated as follows.

Theorem 1.2. *(BMZ) Let $d \geq 1,\ r \geq 2$ prime and $N = (d+1)(r-1)$. Consider the N-simplex $\triangle_N$ whose vertices are colored into disjoint color classes*

$$C_0, C_1, \cdots, C_m \quad m \geq d+1$$

such that for each i, the size $|C_i|$ of these colored classes is at most $r-1$. Then given any continuous map $f : \triangle_N \to \mathbb{R}^d$ there exists a family of r disjoint rainbow faces $\sigma_1, \sigma_2, \cdots, \sigma_r$ of $\triangle_N$ such that

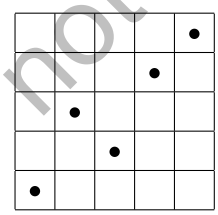
$$f(\sigma_1) \cap \cdots \cap f(\sigma_r) \neq \phi.$$

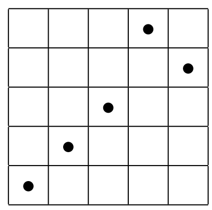
Here a face σ_i is said to be a **rainbow** face if each vertex of σ_i is of different color, i.e., $|\sigma_i \cap C_j| \leq 1 \ \ \forall\, i,\ 1 \leq i \leq r$ and $\forall\, j,\ \ 1 \leq j \leq m$. Note that a rainbow face need not use all colors.

The above Theorem, to be called the BMZ Theorem in the sequel, turns out to be very interesting topological colored Tverberg Theorem since the classical Tverberg Theorem becomes a special case of this Theorem when the size of each color class is one, and also since all points of $\triangle_N$ are accounted for. There are at least three different proofs of the BMZ theorem. In this paper we will discuss and explain the simplest of the three proofs which uses the concept of chessboard complexes and the degree of an equivariant map.

2. Chessboard Complexes and the Degree of an Equivariant Map

For a simple motivation, let us consider a 5×5 chessboard. Suppose we have a number of rooks, not only two, that we want to put on this board so that no rook threatens any other rook. This means in any row there can be only one rook and similarly in any column there can be only one rook. Let us now consider all possible placements of rooks on the above chessboard. Several placements are clearly possible such as





Note that we can place a maximum of 5 rooks on the above chessboard. Since a subset of a valid placement of rooks is also a valid placement, the set of all possible valid rook placements forms a simplicial complex. The vertex set is clearly the set of all squares and consists of 25 points. More generally, we can consider a rectangular chessboard having $m \times k$ squares so that we get a simplicial complex which has $m.k$ vertices and whose simplices are all possible valid rook placements of the above chessboard. Such a simplicial complex is known as *chessboard complex* and is denoted by $\triangle_{m,k}$. These simplicial complexes are very interesting and useful mathematical objects and they have several different interpretations (see [1] [18] [19] for details). One of the most interesting results about these simplicial complexes is their k-connectedness property for suitable k. The following examples and observations are obvious.

(i) $\triangle_{m,1}$ has only m vertices and no other simplex. It is clearly a discrete set.

(ii) $\triangle_{3,2}$ is a connected simplicial complex giving a triangulation of circle. To see this suppose $\Delta_{3,2}$ is represented by the entries of the following matrix

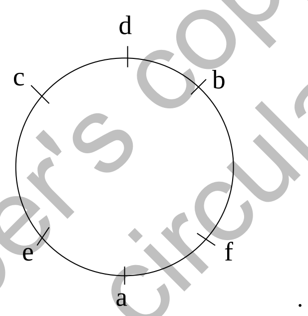
d	e	f
a	b	c

.

Then all the maximal $1-$simplexes are

$$< a, e >, < a, f >, < b, d >, < b, f >, < c, d >, < c, e >$$

which can be represented as



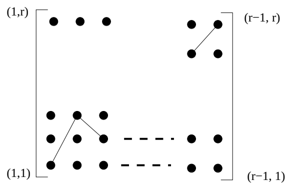
.

(iii) $\triangle_{3,3} \supset \triangle_{3,2}$ as a subcomplex.

(iv) $\triangle_{m,k}$ is simplicially isomorphic to $\triangle_{k,m}$.

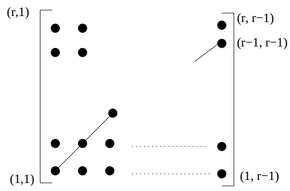
(v) $\triangle_{m,k}$ is a $(k-1)$-dimensional simplicial complex if $m \geq k$.

In this paper we will need only the special chessboard complex $\triangle_{r,r-1}$ where r is a prime number. We can visualize these simplicial complexes with vertices as the entries of a $r \times (r-1)$ matrix with r rows and $r-1$ columns as follows:

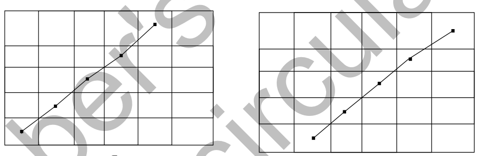


A maximal simplex of $\triangle_{r,r-1}$ is shown as entries joined by lines, and such a simplex has $r-1$ vertices. It follows that $\triangle_{r,r-1}$ is a $(r-2)$-dimensional simplicial complex. We have

Proposition 2.1. *The chessboard complex $\triangle_{r,r-1}$ is a connected $(r-2)$-dimensional orientable pseudomanifold.*

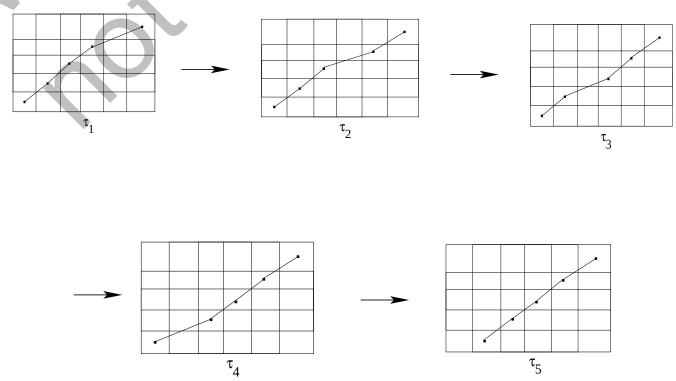


Proof. Consider the $r \times (r-1)$ matrix:
A maximal simplex has $r-1$ vertices and so will occupy $r-1$ rows and $r-1$ columns. Hence $\triangle_{r,r-1}$ is $(r-2)$-dimensional. Moreover any $(r-3)$-simplex will occupy $r-2$ rows and $r-2$ columns leaving one column and two rows vacant. Therefore it can be extended to a $(r-2)$-simplex only in two ways by occupying either of the two vacant entries in the two rows and the vacant column. This means any $(r-3)$-simplex is a face of exactly two $(r-2)$-simplices. Furthermore, given any two $(r-2)-$simplexes σ_1, σ_2 of $\Delta_{r,r-1}$, it is easy to see that we can connect them by a sequence of $(r-2)-$simplexes $\sigma_1 = \tau_0, \tau_1, ..., \tau_n = \sigma_2$ such that the intersection of any two consecutive simplexes is a $(r-3)-$simplex. The following example is enough to formulate a general argument:



$\sigma_1 = \tau_0$ $\sigma_2 = \tau_5$

Now we consider a $(r-1) \times r$ matrix with $r = 6$. The two $4-$simplices shown above can be connected by the following sequence of $4-$simplices so that their intersection is a $3-$simplex.



Thus $\triangle_{r,r-1}$ is a pseudomanifold.

To see the orientability of the complex $\Delta_{r,r-1}$, we observe that we have r rows and $r-1$ columns. Let us order the r rows as $(1, 2, ..., r)$. Now a $(r-2)-$simplex is determined by entries in any $r-1$ rows with one entry in each column fixed.

Let us indicate that $(r-2)-$simplex as $(1,2,...,\hat{i},..r)$ where the index $\hat{i}$ is omitted. Now we give positive orientation to $(1,2,...,\hat{i},..r)$ when i is even, and negative orientation when i is odd. Then, as in the case of $n-$sphere $\partial\sigma^{n+1}$, we can see easily that any $(r-3)-$simplex of $\Delta_{r,r-1}$ will receive opposite orientations from the two $(r-2)-$ simplexes of which it is a common face. Hence the above orientation is coherent and so the pseudomanifold $\Delta_{r,r-1}$ is orientable. $\square$

Let us also recall the idea of $r-$fold deleted join of a simplicial complex.

Definition 2.1. Let K be a simplicial complex. The $r-$fold deleted join, denoted by K^{*r}_{Δ}, is a simplicial complex whose vertices are the $r-$fold disjoint union of the vertices of K and the simplices are given by

$$K^{*r}_{\Delta} = \{\sigma_1 \uplus \sigma_2 \uplus \cdots \uplus \sigma_r : \sigma_1, \sigma_2, ..., \sigma_r \in K, \sigma_1 \cap \sigma_2 \cap \cdots \cap \sigma_r = \phi\}.$$

The polyhedron of K^{*r}_{Δ} can be written as

$$||K^{*r}_{\Delta}|| = \{t_1x_1 \oplus t_2x_2 \oplus ... \oplus t_rx_r : supp(x_1) \cap supp(x_2)... = \phi, t_1 + t_2 + ..t_r = 1\}.$$

One can easily see that

(1) $(\sigma^0)^{*2}_{\Delta}$ consists of two points.
(2) $(D_2)^{*2}_{\Delta}$ is a disjoint union of two edges.
(3) $(\sigma^1)^{*2}_{\Delta}$ is the perimeter of a square.

3. Degree of an Equivariant map

Let σ^{r-1} be the $(r-1)-$simplex with vertex set $[r] = \{1, 2...., r\}$. Then the boundary complex $\partial(\sigma^{r-1})$ is a triangulation of the $(r-2)-$sphere $\mathbb{S}^{r-2}$ for $r \geq 2$. Its simplices can be written as $< 1, 2, \cdots, \hat{i}, \cdots, r >$ where $\hat{i}$ means i is omitted. It is well known that if we orient $\mathbb{S}^{r-2}$ by above ordering then the fundamental homology class (orientation cycle) of $\mathbb{S}^{r-2}$ is given by the simplicial chain ([9] p.130)

$$[\partial\sigma^{r-1}] = [\textstyle\sum(-1)^i < 1, 2, \cdots, \hat{i}, \cdots, r >]. \tag{2.1}$$

On the other hand, since $\triangle_{r,r-1}$ is also an orientable pseudomanifold of dimension $(r-2)$ by Proposition 2.1, we can write the orientation cycle of $\triangle_{r,r-1}$ as

$$[\triangle_{r,r-1}] = \left[\sum_{\pi \in S_r} (-1)^{\text{sign}\pi} < (\pi(1), 1), (\pi(2), 2), \cdots, (\pi(r-1), r-1) > \right]. \tag{2.2}$$

Here S_r denotes the permutation group on r symbols.

Now let us define a map $\xi : \triangle_{r,r-1} \to \partial\sigma^{r-1}$ simply by projecting the simplex $< (i_1, j_1), (i_2, j_2), \cdots, (i_p, j_p) >$ of $\triangle_{r,r-1}$ onto the simplex $< i_1, i_2, \cdots, i_p >$ belonging to $\partial\sigma^{r-1}$. Also note that the group S_r acts on $\triangle_{r,r-1}$ simply by k permuting the rows of $\triangle_{r,r-1}$ whereas it acts on $[r]$ by permuting the vertices of σ^{r-1}. With these actions of S_r, the map ξ is a S_r-map. Under this map, the maximal simplices of $\triangle_{r,r-1}$ go to the maximal simplices of $\partial\sigma^{r-1}$. Let us see what will be inverse of the simplex $< 1, 2, \cdots, \hat{i}, \cdots, r >$. The preimages of this simplex under ξ will be all permutations $< \pi(1), \pi(2), \cdots, \pi(\hat{i}), \cdots, \pi(r) >$ where $\hat{i}$ has been omitted and their number is clearly $(r-1)!$ Since the orientation cycle of the pseudomanifold $\triangle_{r,r-1}$ is given by the simplicial chain

$$\sum_{\pi \in S_r} (-1)^{\mathrm{sign}\pi} < (\pi(1), 1), (\pi(2), 2), \cdots, (\pi(r-1), r-1) >,$$

we find that in homology

$$\xi_*[\triangle_{r,r-1}] = (r-1)![\partial\sigma^{r-1}].$$

This proves the following

Proposition 3.1. *If ξ denotes the projection map $\xi : \triangle_{r,r-1} \to \partial\sigma^{r-1}$, then*

$$\deg \xi = (r-1)!.$$

Let us also recall the following result on the degree of equivariant maps (see [19], page 9).

Proposition 3.2. *Let M be a triangulated compact orientable n-dimensional pseudomanifold. Suppose G is a finite group acting freely and simplicially on M. Suppose $S(W)$ is a $G-$invariant sphere in a real $(n+1)-$dimensional $G-$representation W. Suppose S and M both have the same orientation character, i.e., each element of G either preserves the orientation of both M and S or reverses the orientation. Then for any two G-maps $f, g : M \to S$*

$$\deg f = \deg g \mod |G|.$$

The following proposition is a consequence of the preceding two propositions.

Proposition 3.3. *For any $\mathbb{Z}_r$-map $h : \triangle_{r,r-1} \to \partial\sigma^{r-1}$ the degree of h,*

$$\begin{aligned} \deg h &\equiv \deg \xi \mod r \\ &\equiv \pm 1 \mod r \\ i.e., \deg h &\neq 0. \end{aligned}$$

4. Colored Topological BMZ Theorem

In this section we will present the simplest proof of the BMZ Theorem using the degree theoretic method. In fact, BMZ gave three different proofs of their Theorem (see [7]). One of the proofs uses the obstruction theory of algebraic topology whereas the other uses the concepts of ideal theoretic G-index theory. Both of these proofs are somewhat more involved and technical. The degree theoretic proof (see [7]) which we are going to present turns out to be the simplest of all. However, the basic result needed for this proof is the following Reduction Lemma (see [10] for a motivational proof of this lemma).

Lemma 1. *If the BMZ theorem is true for the special coloring $C_0, C_1 \cdots, C_{d+1}$ where $|C_i| = r-1$ for all $i \neq d+1$ and $|C_{d+1}| = 1$, then the BMZ theorem is true for all coloring $C_0, C_1 \cdots, C_m,\ m \geq d+1$, where $|C_i| \leq r-1$.*

Proof of the BMZ Theorem

Let $C_0, C_1 \cdots, C_{d+1}$ be the $d+2$ color classes of the vertices of the simplex $\triangle_N$ where $|C_i| = r-1$ for all $i \neq d+1$ and $|C_{d+1}| = 1$. We are interested only in rainbow faces of $\triangle_N$ such that the vertices of these faces must come from different color classes. If we regard each color class C_i as discrete space, then their join $K = C_0 * C_1 * \cdots * C_{d+1}$ is a simplicial complex and the vertices of its simplices will have different colors. Therefore these simplices will be rainbow faces of $\triangle_N$. In fact, a face of $\triangle_N$ will be rainbow face iff it is a face of the simplicial complex K which is a proper subcomplex of the simplicial complex $\triangle_N$.

Now we consider the group $\mathbb{Z}_r = [g]$ to be generated by the cyclic permutation g and consider the r-fold join K^{*r} of K. Clearly $\mathbb{Z}_r$ acts on K^{*r} by cyclically permuting the coordinates,.i.e., by the action

$$g(t_1x_1 \oplus t_2x_2 \oplus \cdots \oplus t_rx_r) = t_2x_2 \oplus t_3x_3 \oplus \cdots \oplus t_1x_1.$$

Hence K^{*r} is $\mathbb{Z}_r$-space which need not be free. However, the pairwise deleted join $K^{*r}_{(2)}$ is a free $\mathbb{Z}_r$-space.

Now suppose $f : \triangle_N \to \mathbb{R}^d$ is any continuous map. This map restricts to a continuous map say $f : K \to \mathbb{R}^d$ and hence induces a continuous map in their r-fold join

$$f^{*r} : K^{*r} \to (\mathbb{R}^d)^{*r}. \tag{4.1}$$

Note that $\mathbb{Z}_r$ acts on $\mathbb{R}^r = \{(xe_1 + \cdots + xe_r)|\ x_i \in \mathbb{R}\}$ by cyclicly permuting the coordinates. This gives the regular representation of $\mathbb{Z}_r$ on $\mathbb{R}^r$. The diagonal $\triangle$ of $\mathbb{R}^r$ is fixed under this action and so its orthogonal compliment viz., the plane

$$\{(x_1, \cdots, x_r) \in \mathbb{R}^r|\ \sum x_i = 0\}$$

is invariant under the above action of $\mathbb{Z}_r$. The orthogonal plane say, W_r is therefore a $(r-1)-$dimensional representation of $\mathbb{Z}_r$. Now if $p : \mathbb{R}^r \to W^r$ is the projection map, then this defines a $\mathbb{Z}_r$-map $(\mathbb{R}^d)^{*r} \to (W^r)^{d+1}$ by (4.2)

$$\begin{aligned} p(t_1x_1 \oplus t_2x_2 \oplus \cdots \oplus t_rx_r) \quad = \quad & (p(t_1, \cdots, t_r), \\ & p(t_1x_{11}, \cdots, t_rx_{r1}), \\ & \vdots \\ & p(t_1x_{1d}, \cdots, t_rx_{rd})). \end{aligned}$$

The map f^{*r} in (4.1) is evidently a $\mathbb{Z}_r$-map. Note that a simplex of LHS is a formal convex combination of r-tuple $(\sigma_1, \cdots, \sigma_r)$ of rainbow faces of $\triangle_N$ and its image under f^{*r} is a formal convex combination of r-tuple of subsets $(f(\sigma_1), \cdots, f(\sigma_r))$ of $\mathbb{R}^d$. Clearly $f(\sigma_1) \cap \cdots \cap f(\sigma_r) \neq \phi$ iff $(f(\sigma_1), \cdots, f(\sigma_r))$contains an element of the diagonal of $(\mathbb{R}^d)^{*r}$. Composing the map of (4.1) with (4.2), we get a $\mathbb{Z}_r$-map

$$f^{*r} : (C_0 * C_1 * \cdots * C_{d+1})^{*r} \to (W^r)^{d+1}.$$

The RHS is a real vector space of dimension $(d+1)(r-1)$.

Now let us assume that the map $f : \triangle_N \to \mathbb{R}^d$ violates the BMZ Theorem. This means whenever $\sigma_1, \cdots, \sigma_r$ is a family of r-pairwise disjoint rainbow faces of $\triangle_N$ the intersection of their images $f(\sigma_1), \cdots, f(\sigma_r)$ will be empty, i.e., their images under f^{*r} will never be zero. Since the $\mathbb{Z}_r$-space $(W^r)^{d+1} - \{0\}$ is $\mathbb{Z}_r$-homotopy equivalent to the $(d+1)(r-1) - 1 = N - 1$ dimensional sphere $S((W^r)^{d+1})$, we get a $\mathbb{Z}_r$-map

$$g : (C_0 * C_1 * \cdots * C_{d+1})^{*r}_{\triangle_{(2)}} \to S^{(d+1)(r-1)-1}.$$

Since the join operation commutes with the delete join ([13]. p.108), we get a $\mathbb{Z}_r$-map

$$g : (C_0)^{*r}_{\triangle_{(2)}} * \cdots * (C_d)^{*r}_{\triangle_{(2)}} * (C_{d+1})^{*r}_{\triangle_{(2)}} \to S^{(d+1)(r-1)-1}.$$

Now observe that $(C_i)^{*r}_{\triangle_{(2)}} \simeq \triangle_{r,|C_i|}$ is the chessboard complex for each $i = 1, 2 \cdots, d+1$. Furthermore, $(C_{d+1})^{*r}_{\triangle_{(2)}} = [r]$ is simply the discrete set with r points. Let us denote

$$L = (C_0 * C_1 * \cdots * C_d)^{*r}_{\triangle_{(2)}}.$$

Then the LHS of the map g is $L * [r]$ which has $L * \{1\}$ as a subcomplex and the latter is a cone over L. Finally consider the restriction h of the $\mathbb{Z}_r$-map g to L, viz.,

$$h : L \to S^{(d+1)(r-1)-1}.$$

Note that $\triangle_{r,r-1} = \triangle_{r,|C_i|}$ is an oriented pseudomanifold of dimension $r-2$ (Proposition 2.1). Hence $L = \triangle_{r,r-1} * \cdots * \triangle_{r,r-1}$ ($(d+1)$ copies) is also an orientable pseudomanifold of dimension $(d+1)(r-2)+d = (d+1)(r-1)-1$. Thus the dimension of L = dimension of $S^{(d+1)(r-1)-1}$ and so we can talk of the degree of the map $h : L \to S^{(d+1)(r-1)-1}$. Since h extends to a $\mathbb{Z}_r$-map onto the cone $L * \{1\}$ over L, we find that $\deg h = 0$. But since $\mathbb{Z}_r$ acts freely on both the spaces, $\deg h \neq 0$ by Proposition 3.3. This contradiction completes the proof of the BMZ Theorem for the special coloring $\{C_0, C_1, \cdots, C_{d+1}\}$. Hence, by the Reduction Lemma, the BMZ Theorem is true for arbitrary colorings $\{C_0, C_1, \cdots, C_m\}$ where $m \geq d+1$ and $|C_i| \leq r-1$. This completes the proof.

Acknowledgement: The author is thankful to the referee for a careful reading of the paper pointing out pedagogical improvements and typos.

References

[1] Björner, A., Lovász, L., Vrećica, S. and Zivaljević, R. T., *Chessboard Complexes and Matching Complexes*, J. London Math. Soc., (2) **49** (1994), 25-39.

[2] Bárány, I., Füredi, Z. and Lovász, L., *On the number of halving planes*, Combinatorica, **10** (1990), 175-183.

[3] Bárány, I. and Larman, D. G., *A colored version of Tverberg Theorem,* J. London Math. Soc., **45** (1992), 314-320.

[4] Bárány, I., Shlossman, S. B. and Szücs, A., *On the topological genaralization of a theorem of Tverberg,* J. London Math. Soc., (2) **23** (1981), 158-164.

[5] Birch, B. J., *On 3N points in a plane*, Math. Proc. Cambridge Phil. Soc., **55** (1959), 158-164.

[6] Blagojević, P. V. M., Matschke, B. and Ziegler, G. M., *Optimal bounds for the colored Tverberg problem,* J. European Math Soc., **17** (2015), 739-754.

[7] Blagojević, P. V. M., Matschke, B. and Ziegler, G. M., *Optimal bounds for a colored Tverberg-Vrećica type problem,* Advances in Maths, **226** (2011), 5198-5215.

[8] Blagojević, P. V. M., Matschke, B. and Ziegler, G. M., *A tight colored Tverberg Theorem for maps to manifolds,* Top. and Appl., **158** (2011) 1445-1452.

[9] Deo, Satya, *Algebraic Topology, a Primer,* Hindustan Book Agency, TRIM series No 29 (2006).

[10] Deo, Satya, *Topological colored Tverberg Theorem and the Reduction Lemma*, arXiv 1502.04805, Feb 2015.

[11] Frick, F., *Counterexamples to the topological Tverberg conjecture*, arXiv:1502.00947 Feb 2015.

[12] Longueville, M. de, *Notes on topological Tverberg theorem,* Discrete Math **241** (2001), 207-233.

[13] Matoušek, J., *Using Borsuk-Ulam theorem,* Universitext, Springer-Verlag, 2003.

[14] Özaydin, M., *Equivariant Maps for the Symmetrtic Groups,* (preprint).

[15] Tverberg, H., *A generalizations of Radon's theorem,* J. London Math. Soc., **41** (1966), 123-128.

[16] Tverberg, H. and Vrećica, S., *On generalization of Radon's theorem and the ham sandwich theorem,* Europ. J. Combinatorics, **14** (1993), 259-264.

[17] Volovikov, A. Yu. *On a topological generalization of Tverberg theorem* Math. Notes, (1) **59** (1996), 324-326.

[18] Vrećica, S. and Zivaljević, R. T., *The colored Tverberg's problem and complex of injective functions,* J. Combinatorial Theory, Ser A, **61** (1992), 309-318.

[19] Vrećica, S. and Zivaljević, R. T., *Chessboard complexes indomitable,* arXiv:0911.3512v3, April 2011.

Satya Deo
Harish-Chandra Research Institute
Chhatnag Road, Jhusi,, Allahabad - 211 019, U.P. India
E-mail : *sdeo@hri.res.in*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 159-171

LAWS OF COMPOSITION AND ARITHMETIC STATISTICS: FROM GAUSS TO BHARGAVA

ARUL SHANKAR AND XIAOHENG WANG

(Received : 21 - 11 - 2015)

1. Gauss's law of composition

In his seminal work *Disquisitiones Arithmeticae* of 1801, Karl Friedrich Gass studied the action of $\mathrm{SL}_2(\mathbb{Z})$, the group of integral 2×2 matrices with determinant 1, on the space of integral binary quadratic forms $f(x,y) = ax^2 + bxy + cy^2$ $(a,b,c \in \mathbb{Z})$*. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on integral binary quadratic forms by linear substitution of variable as follows:

$$\gamma \cdot f(x,y) = f((x,y)\cdot\gamma),$$

where $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and f is an integral binary quadratic form. This action preserves the *discriminant* $\Delta = b^2 - 4ac$ of binary quadratic forms, i.e., $\Delta(f) = \Delta(\gamma\cdot f)$. Gauss defined a composition law on the orbits for this action: given two integral binary quadratic forms f and g with the same discriminants, Gauss described a method to construct a third integral binary quadratic form $h = f\circ g$, also with the same discriminant. Furthermore, if f' and g' are two integral binary quadratic forms equivalent to f and g respectively under the action of $\mathrm{SL}_2(\mathbb{Z})$, then $f'\circ g'$ is also equivalent to $f\circ g$ under the action of $\mathrm{SL}_2(\mathbb{Z})$. In fact, Gauss proved that the set of $\mathrm{SL}_2(\mathbb{Z})$-orbits on integral binary quadratic forms having a fixed discriminant D form a finite abelian group under composition. We will denote this group by $\mathrm{Cl}(D)$ and the size of $\mathrm{Cl}(D)$ by $h(D)$.

Gauss formulated several conjectures regarding $h(D)$ which have played an enormous part in shaping number theory. The most famous of them is the celebrated class number one conjecture.

Conjecture 1.1 (Gauss)**.** *The list of positive D such that $h(-D) = 1$ is: 3, 4, 7, 8, 11, 19, 43, 67, and 163.*

Conjecture 1.1 has a long and illustrious history which is beautifully detailed by Goldfeld in [23]. It is now a theorem due independently to Baker [1] and Stark

* Gauss considered only forms where b is even; however we will follow the modern point of view and allow all three coefficients a, b, and c to be arbitrary integers.

2010 AMS Subject Classification: 11A15, 11D25, 11D41, 11E04, 11E08, 11E41, 11E76, 11H06

Key words and phrases: composition laws, arithmetic statistics.

[40], [41]. It is worth noting that Heegner [26] had previously published an almost complete proof of Conjecture 1.1, but his paper contained errors and was not complete. Shortly after the work of Baker and Stark, Deuring [21] filled in the gaps in Heegner's work. Another conjecture regarding $h(D)$ for negative D is the Gauss class number conjecture.

Conjecture 1.2 (Gauss). *The number of positive integers D such that $h(-D)$ is equal to any fixed integer is finite.*

Conjecture 1.2 is a result of the combined work of Hecke [29] (Landau published the theorem, which he attributed to a lecture by Hecke) and Heilbronn [27]. Hecke proves that $h(-D) \to \infty$ as $D \to \infty$ if the generalized Riemann hypothesis is true; Heilbronn proves that $h(-D) \to \infty$ as $D \to \infty$ if the generalized Riemann hypothesis is false! The celebrated result of Siegel [33] provides a rate of growth for $h(-D)$:

Theorem 1.3 (Siegel). *For every $\epsilon > 0$, there exists a constant $c > 0$ such that*

$$h(-D) > cD^{1/2-\epsilon} \tag{1}$$

for every positive integer D.

Siegel proof is ineffective, which is to say that it does not provide a method of computing c given ϵ—only a proof that such a c exists!

Very little is known about the sizes $h(D)$ for positive D. They are expected to behave very differently from the sizes when D is negative. For example, the following is widely believed (though completely unknown).

Conjecture 1.4 (Gauss). *There exist infinitely many positive integers D such that $h(D) = 1$.*

Gauss also made conjectures on the behaviour of $h(D)$ on *average*.

Conjecture 1.5 (Gauss). *For large enough real number X:*

(a) $$\sum_{-X<D<0} h(D) \sim \frac{\pi}{18} \cdot X^{3/2};$$

(b) $$\sum_{0<D<X} h(D) \log \varepsilon_D \sim \frac{\pi^2}{18} \cdot X^{3/2};$$

here $\varepsilon_D = (t + u\sqrt{D})/2$, where t, u are the smallest positive integral solutions of $t^2 - Du^2 = 4$.

It is worth noting that such t, u do not exist if D is negative. Part (a) of Conjecture 1.5 is a result of Mertens [31] and part (b) is a result of Siegel [34]. It is the analogues of these conjectures that we will focus on in this article.

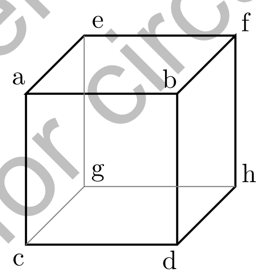
2. Bhargava's laws of composition

In this section, for the sake of convenience, we will restrict ourselves to considering integers D that are *fundamental discriminants*, i.e., integers D such that either $D \equiv 1 \pmod 4$ and is squarefree or $D = 4m$ where $m \equiv 2$ or $3 \pmod 4$ and m is squarefree. For such D, the group $\mathrm{Cl}(D)$ is isomorphic to the *narrow class* *group* of the quadratic number field $\mathbb{Q}(\sqrt{D})$ and $h(D)$ is the *narrow class number* of $\mathbb{Q}(\sqrt{D})$. This narrow class group is the same as the class group for imaginary quadratic fields (the $D < 0$ case) but can be twice as big for real quadratic fields (the $D > 0$ case). The precise definition of class groups and narrow class groups will not be necessary for us. All number fields have class groups, and these class groups measure the failure of the rings of integers of these number fields to be principle ideal domains.

Narrow class groups of number fields have a naturally occuring abelian group structure and Gauss's law of composition on integral binary quadratic forms having discriminant D corresponds to addition in the narrow class group. Thus, the law of composition can be stated as the following theorem.

Theorem 2.1 (Gauss)**.** *Let $D \neq 0$ be a fundamental discriminant. Then there exists a natural bijection between the set of $\mathrm{SL}_2(\mathbb{Z})$-orbits on integral binary quadratic forms with discriminant D and the narrow class group of $\mathbb{Q}(\sqrt{D})$.*

Two centuries after Gauss described his law of composition, Bhargava discovered several new laws of compositions. In what follows, we describe some of them. In our exposition, we closely follow Bhargava's paper [2]. Let $V(\mathbb{Z})$ denote the space $\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$ of cubes whose vertices are integers. We represent elements of $V(\mathbb{Z})$ as 8-tuples (a, b, c, d, e, f, g, h) viewed as vertices of a cube as follows:



We may slice this cube in three different ways obtaining these three pairs of integral matrices:

$$M_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad N_1 = \begin{bmatrix} e & f \\ g & h \end{bmatrix},$$
$$M_2 = \begin{bmatrix} a & c \\ e & g \end{bmatrix}, \quad N_2 = \begin{bmatrix} b & d \\ f & h \end{bmatrix}, \tag{2}$$
$$M_3 = \begin{bmatrix} a & e \\ b & f \end{bmatrix}, \quad N_3 = \begin{bmatrix} c & g \\ d & h \end{bmatrix}.$$

An element $A \in V(\mathbb{Z})$ thus yields three pairs of integral 2×2-matrices (M_i, N_i), $1 \leq i \leq 3$. By considering the discriminant quadratic form

$$Q_i^A := -\det(M_i x - N_i y) \tag{3}$$

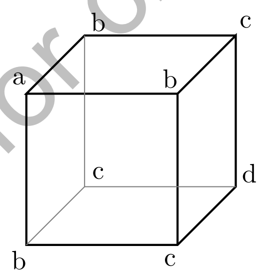
for $1 \leq i \leq 3$, we also obtain three integral binary quadratic forms. Bhargava proves that these three integral binary quadratic forms have the same discriminant! We define the discriminant of A, denoted $\Delta(A)$, to be to this common discriminant of these three binary quadratic forms.

The group $\Gamma = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ acts on $V(\mathbb{Z})$: for $1 \leq i \leq 3$, the i'th component $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ of an element in Γ sends (M_i, N_i) to $(pM_i + qN_i, rM_i + sN_i)$. Furthermore, Γ preserves the discriminant of elements in $V(\mathbb{Z})$. That is, we have $\Delta(A) = \Delta(\gamma \cdot A)$ for $A \in V(\mathbb{Z})$ and $\gamma \in \Gamma$. With Theorem 2.1 as our template to describing laws of composition, we may state Bhargava's first new law of composition.

Theorem 2.2 (Bhargava)**.** *Let $D \neq 0$ be a fundamental discriminant. Then there exists a natural bijection between the set of Γ-orbits on elements in $V(\mathbb{Z})$ with discriminant D and $\mathrm{Cl}^+(\mathbb{Q}(\sqrt{D}))^2$, where $\mathrm{Cl}^+(\mathbb{Q}(\sqrt{D}))$ is the narrow class group of the quadratic field $\mathbb{Q}(\sqrt{D})$.*

The incredible richness of this new composition law is perhaps best described by demonstrating how it gives rise to even more composition laws. First, Bhargava shows that for $A \in V(\mathbb{Z})$ giving rise to the three quadratic forms (3), the sum of Q_1^A, Q_2^A, and Q_3^A, with respect to Gauss composition, is 0. In particular, this law (termed the cube law by Bhargava) is enough to recover all of Gauss composition!

Second, Bhargava shows that the law of composition on integer cubes in $V(\mathbb{Z})$ restricts to *triply symmetric* integer cubes, i.e., those of the form



with $a, b, c, d \in \mathbb{Z}$. These triply symmetric cubes are preserved by the action of $\mathrm{SL}_2(\mathbb{Z})$ embedded diagonally in Γ, and the orbits having fixed fundamental discriminant D form an abelian group under composition. Amazingly, $\mathrm{SL}_2(\mathbb{Z})$-orbits on triply symmetric cubes are in natural bijection with $\mathrm{SL}_2(\mathbb{Z})$-orbits on integral binary cubic forms, whose middle coefficients are multiples of 3: the binary cubic form corresponding to the above triply symmetric integer cube is $ax^2 + 3bx^2y + 3cxy^2 + dy^3$. This leads to another composition law:

Theorem 2.3 (Bhargava)**.** *Let $D \neq 0$ be a fundamental discriminant. Then there exists a natural bijection between the set of $\mathrm{SL}_2(\mathbb{Z})$-orbits on triplicate integral binary cubic forms with discriminant D and $\mathrm{Cl}(\mathbb{Q}(\sqrt{D}))[3]$, where $\mathrm{Cl}(\mathbb{Q}(\sqrt{D}))[3]$ is the 3-torsion subgroup of the class group of the quadratic field $\mathbb{Q}(\sqrt{D})$.*

Details of these and many other composition laws are in the beautiful series of papers [2], [3], [4], [5] by Bhargava.

3. Statistics of number fields

Arithmetic statistics concerns the study of the statistics of arithmetic objects. Two of the foundational questions in the subject are the following.

(1) How are the discriminants of degree-n number fields distributed?
(2) How are the class groups of degree-n number fields distributed?

When $n = 2$, the first question is easy to answer, since an integer D occurs as the discriminant of a quadratic field if and only if D is a fundamental discriminant. Furthermore, each such integer D occurs as the discriminant of exactly one quadratic field. Theorem 2.1 in conjunction with the work of Mertens and Siegel resolving Conjecture 1.5 provides a partial answer to the second question. Siegel in [34] provides a much fuller answer to the second question by computing *all* moments of the sizes of the class groups of number fields. However, even that landmark work does not provide a complete answer. This is because the sizes of class groups do not take into account their group structure. In this regard, the highly influential work of Cohen and Lenstra [17] formulates a detailed series of conjectures that predict the behaviour of class groups of quadratic number fields on average. The most well known of their conjectures is the following:

Conjecture 3.1 (Cohen–Lenstra). *Let p be an odd prime. Then*

(a) *The average size of the p-torsion subgroup in the class group of real quadratic fields is* $1 + 1/p$.
(a) *The average size of the p-torsion subgroup in the class group of imaginary quadratic fields is* 2.

Their conjecture goes much further and in fact predicts the distribution of class groups of quadratic number fields. Very little is proved of their conjecture. The only known case is that of $p = 3$, which is due to Davenport and Heilbronn in work that predates the conjecture of Cohen and Lenstra.

Theorem 3.2 (Davenport–Heilbronn). *We have*

(a) *The average size of the* 3*-torsion subgroup in the class group of real quadratic fields is* $4/3$.
(a) *The average size of the* 3*-torsion subgroup in the class group of imaginary quadratic fields is* 2.

When $n \geq 3$, very little is known of either of the two questions posed in the beginning of this section. Before the work of Bhargava, the only complete result in this regard was the following theorem of Davenport and Heilbronn [20].

Theorem 3.3 (Davenport–Heilbronn). *Let $N_3(\xi,\eta)$ denote the number of cubic fields K, up to isomorphism, that satisfy $\xi < \mathrm{Disc}(K) < \eta$. Then*

$$\begin{aligned} N_3(0,X) &= \frac{1}{12\zeta(3)}X + o(X); \\ N_3(-X,0) &= \frac{1}{4\zeta(3)}X + o(X), \end{aligned} \tag{4}$$

where ζ denotes the Riemann-Zeta function.

As far as the second question is concerned, the Cohen–Lenstra heuristics have been modified by Cohen and Martinet [18] to obtain conjectures for higher degree number fields. Before the work of Bhargava (described in the next section), no case of this conjecture had been proven.

4. Bhargava's advances in Arithmetic Statistics

Davenport, using geometry-of-numbers techniques, proved the following theorem [19] which was a key input in Theorems 3.2 and 3.3.

Theorem 4.1 (Davenport). *Let $N(\xi,\eta)$ denote the number of $\mathrm{GL}_2(\mathbb{Z})$-equivalence classes of irreducible integer-coefficient binary cubic forms f satisfying $\xi < \mathrm{Disc}(f) < \eta$. Then*

$$N(0,X) = \frac{\pi^2}{72}\cdot X + O(X^{15/16});\; N(-X,0) = \frac{\pi^2}{24}\cdot X + O(X^{15/16}).$$

This furthers the works of Siegel and Mertens resolving Conjecture 1.5 by counting $\mathrm{SL}_2(\mathbb{Z})$-orbits on integral binary quadratic forms. It had previously been understood through works of [42] that the statistics of quartic fields can be studied via the representation of $G_4 = \mathrm{GL}_2 \times \mathrm{SL}_3$ on the space $W_4 = (\mathrm{Sym}^2\mathbb{Z}^3 \otimes \mathbb{Z}^2)^*$ of pairs of ternary quadratic forms. However, two major obstacles remained: first, though the rational orbits for the action of $G_4(\mathbb{Q})$ on $W_4(\mathbb{Q})$ were understood to correspond to quartic extensions of $\mathbb{Q}$, there was no corresponding interpretation of the integral orbits of $G_4(\mathbb{Z})$ acting on $W_4(\mathbb{Z})$. Second, the combinatorial difficulties in counting $\mathrm{SL}_2(\mathbb{Z})$-orbits on the space of integral binary quadratic forms (a 2-dimensional space) and in counting $\mathrm{GL}_2(\mathbb{Z})$-orbits on the space of integral binary cubic forms (a 3-dimensional space) grow infinitely when dealing with $G_4(\mathbb{Z})$-orbits on $W_4(\mathbb{Z})$ which is 12-dimensional!

The first difficulty was resolved by Bhargava in [4], where he proves the following remarkable theorem.

Theorem 4.2 (Bhargava). *There is a canonical bijection between the set of $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$-orbits on the space $(\mathrm{Sym}^2\mathbb{Z}^3 \otimes \mathbb{Z}^2)^*$ of pairs of integral ternary quadratic forms and the set of ismorphism classes of pairs (Q,R), where Q is a quartic ring and R is a cubic resolvent ring of Q.*

It is important to note that a very slightly modified representation is shown by Bhargava to yield another important law of composition.

Theorem 4.3 (Bhargava)**.** *There is a bijection between the set of* $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$*-orbits on the space* $\mathrm{Sym}^2\mathbb{Z}^3 \otimes \mathbb{Z}^2$ *and the set of equivalence classes of triples* (R, I, δ)*, where* R *is a nondegenerate cubic ring over* $\mathbb{Z}$*,* I *is an ideal of* R *having rank* 3 *as a* $\mathbb{Z}$*-module, and* δ *is an invertible element of* $R \otimes \mathbb{Q}$ *such that* $I^2 \subset (\delta)$ *and* $N(I)^2 = N(\delta)$*.*

Introducing fundamental new tools, Bhargava transformed the reach of the geometry-of-numbers methods used by Mertens, Siegel, and Davenport. These new tools made it possible to resolve enormous combinatorial difficulties and determine asymptotics for the number of absolutely irreducible $G_4(\mathbb{Z})$-orbtis on $W_4(\mathbb{Z})$, where an orbit is said to be *absolutely irreducible* if it corresponds to a quartic integral domain R such that the Galois closure of the fraction field of R over $\mathbb{Q}$ is S_4. This led to the following results proved in [6].

Theorem 4.4 (Bhargava)**.** *Let* $N_4^{(i)}(\xi, \eta)$ (*resp.* $M_4^{(i)}(X)$) *denote the number of* S_4*-quartic fields* K (*resp. quartic orders* $\mathcal{O}$ *contained in* S_4*-quartic fields*) *having* $4 - 2i$ *real embeddings such that* $|\mathrm{Disc}(\mathcal{O})| < X$*. Then*

(a) $\displaystyle \lim_{X \to \infty} \frac{N_4(X)}{X} = \frac{1}{n_i} \frac{\zeta(2)^2 \zeta(3)}{\zeta(5)},$

(a) $\displaystyle \lim_{X \to \infty} \frac{M_4(X)}{X} = \frac{1}{n_i} \prod_p (1 + p^{-2} - p^{-3} - p^{-4}),$

where $n_0 = 48$*,* $n_1 = 8$*, and* $n_2 = 16$*.*

These methods in conjunction with Theorem 4.3 also lead to the following theorem, which is the first, and thus far only, result proving an instance of the Cohen-Lenstra-Martinet heuristics involving fields having degree greater than 2.

Theorem 4.5 (Bhargava)**.** *We have*

(a) *The average size of the* 3*-torsion subgroup in the class group of cubic fields having positive discriminants is* 5/4*.*

(a) *The average size of the* 3*-torsion subgroup in the class group of cubic fields having negative discriminants is* 3/2*.*

Quintic fields had been known to correspond to $G_5(\mathbb{Q})$-orbits on $W_5(\mathbb{Q})$, where $G_5 = \mathrm{GL}_4 \times \mathrm{SL}_5$ and W_5 is the space of 4-tuples of 5×5-altering forms. The space W_5 is 50-dimensional, which is a large increase over the 12-dimensional space W_4. The combinatorial difficulties, both in understanding what the integral orbits parameterize and in counting the integral orbits, are correspondingly larger. However, Bhargava resolves them both in [5] and [7], respectively, yielding the following theorems.

Theorem 4.6 (Bhargava)**.** *There is a canonical bijection between the* $\mathrm{GL}_4(\mathbb{Z}) \times \mathrm{SL}_5(\mathbb{Z})$*-orbits on the space* $\mathbb{Z}^4 \otimes \wedge^2\mathbb{Z}^5$ *of quadruples of* 5×5 *skew-symmetric matrices and the set of isomorphism classes of pairs* (R, S)*, where* R *si a quintic ring and* S *is a sextic resolvent of* R*.*

Theorem 4.7 (Bhargava)**.** *Let* $N_5^{(i)}(\xi,\eta)$ (*resp.* $M_5^{(i)}(X)$) *denote the number of quintic fields* K (*resp. quintic orders* $\mathcal{O}$ *contained in quintic fields*) *having* $4-2i$ *real embeddings such that* $|\mathrm{Disc}(\mathcal{O})| < X$*. Then*

(a) $\displaystyle\lim_{X\to\infty}\frac{N_5(X)}{X} = \frac{1}{n_i}\prod_p (1+p^{-2}-p^{-4}-p^{-5},$

(a) $\displaystyle\lim_{X\to\infty}\frac{M_4(X)}{X} = \frac{\alpha}{n_i},$

where

$$\alpha = \prod_p \Big(\frac{p-1}{p}\sum_{[R_p:\mathbb{Z}_p]=5}\frac{1}{|\mathrm{Aut}_{\mathbb{Z}_p}(R_p)|}\cdot\frac{1}{\mathrm{Disc}_p(R_p)}\Big),$$

$n_0 = 240$*,* $n_1 = 24$*, and* $n_2 = 16$*.*

5. Elliptic curves

An elliptic curve E over $\mathbb{Q}$ is given by the equation

$$E : y^2 = x^3 + Ax + B, \tag{5}$$

where the discriminant $-(4A^3 + 27B^2)$ is nonzero. The rational points $E(\mathbb{Q})$ of an elliptic curve, along with the point at infinity, form an abelian group. The following famous result is due to Mordell:

Theorem 5.1 (Mordell)**.** *The group* $E(\mathbb{Q})$ *is finitely generated as an abelian group.*

This implies that we have the isomorphism

$$E(\mathbb{Q}) \cong T \oplus \mathbb{Z}^r,$$

where T is a finite abelian group and r is denoted the *rank* of E. The following remarkable result of Mazur [30] gives all the possibilities for the group T.

Theorem 5.2 (Mazur)**.** *Let* E *be an elliptic curve over* $\mathbb{Q}$*. Then the torsion subgroup* T *of* $E(\mathbb{Q})$ *can only be* $\mathbb{Z}/n\mathbb{Z}$ *for* $1 \le n \le 12$ *or* $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/m\mathbb{Z}$ *for* $m = 2, 4, 6$ *or* 8*.*

The rank r, on the other hand, is much less understood. Given an elliptic curve E over $\mathbb{Q}$, it is possible to associate an L-function to it. These L-functions have many of the same features as the Riemann zeta function. They have Euler products, functional equations, and a meromorphic continuation to the entire complex plane. We will normalize these L-functions so that the line of symmetry of their functional equation is $\mathrm{Re}(s) = 1$. The form of these functional equations is the following.

$$\Lambda(E,s) = \omega(E)N^{1-s}\Lambda(E,2-s), \tag{6}$$

where Λ is the completed L-function of E, N is the conductor of E, and $\omega(E)$ is the root number of E. The precise definitions of Λ, N, and $\omega(E)$ will not be important to us. However, it is important to note that $\omega(E)$ is ± 1. The order of the zero of L (equivalently the zero of Λ) at $s = 0$ is called the *analytic rank* of E. Then the Birch–Swinnerton-Dyer conjecture is the following:

Conjecture 5.3 (Birch–Swinnerton-Dyer)**.** *The rank of an elliptic curve E is equal to the analytic rank of E.*

The parity of the analytic rank of E is determined by the root number; the analytic rank is even or odd depending on whether $\omega(E)$ is 1 or -1, respectively. Thus, the Birch–Swinnerton-Dyer conjecture implies that the parity of the rank of E is determined by $r(E)$.

It is widely believed (though yet unproven) that $r(E)$ is 1 half the time and -1 half the time. Together with the Birch–Swinnerton-Dyer conjecture, this would imply that the rank is even half the time and odd half the time. In conjunction with a general belief that elliptic curves should have as few rational points as they can get away with, with have the following "minimalist" conjecture due to Goldfeld and Katz–Sarnak.

Conjecture 5.4 (Goldfeld, Katz–Sarnak)**.** *The average rank of elliptic curves is* $1/2$. *The proportion of elliptic curves having rank* 0 *is* 50%; *the proportion having rank* 1 *is* 50%.

There are two important points to make regarding the above conjecture. The first is that the believed 0% of elliptic curves having rank greater than or equal to 2 still constitutes infinitely many curves! The second is that statements about the average rank of elliptic curves, or statements concerning proportions of elliptic curves cannot be made precisely without first ordering elliptic curves in some way. This can be done in several natural ways: we may order elliptic curves by conductor or discriminant or some sort of height. Note the elliptic curve $E_{A,B}$: $y^2 = x^3+Ax+B$ is isomorphic to the elliptic curve $E_{u^4A,u^6B} : y^2 = x^3+u^4Ax+u^6B$ under the map $(x, y) \mapsto (u^2x, u^3y)$ for any rational number u and we certainly do not want to count both. Hence we may scale A, B so that they are both integers and that there is no prime p such that $p^4 \mid A$ and $p^6 \mid B$. Because of this apparent "weight" of A and B (which will appear again in the definition of the height for hyperelliptic curves in the next section), we define the "naive" height of an elliptic curve $E_{A,B}$ to be

$$H(E_{A,B}) = \max\{4A^3, 27B^2\}. \tag{7}$$

The extra factors of 4 and 27 are there to balance their contribution to the discriminant and have no effect on the average behavior.

Conditional on the generalized Riemann hypothesis, Brumer [16] showed that the average analytic rank of elliptic curves, when ordered by height, is finite and bounded by 2.3. Still assuming the generalized Riemann hypothesis, this constant was improved by 2 and 1.79 by Heath-Brown [25] and Young [43], respectively. However, no unconditional results were proven about the finiteness of the average analytic rank (or the average rank) of elliptic curves.

The geometry-of-numbers methods developed by Bhargava may be applied to the following representations that are intimately connected to ranks of elliptic curves:

$$\begin{array}{ccc} \mathrm{GL}_2(\mathbb{Z}) & \to & \mathrm{End}(\mathrm{Sym}^4(\mathbb{Z}^2)) \\ \mathrm{GL}_3(\mathbb{Z}) & \to & \mathrm{End}(\mathrm{Sym}^3(\mathbb{Z}^3)) \\ \mathrm{GL}_2(\mathbb{Z}) \times \mathrm{GL}_4(\mathbb{Z}) & \to & \mathrm{End}(\mathbb{Z}^2 \otimes \mathrm{Sym}^2(\mathbb{Z}^4)) \\ \mathrm{GL}_5(\mathbb{Z}) \times \mathrm{GL}_5(\mathbb{Z}) & \to & \mathrm{End}(\mathbb{Z}^5 \otimes \wedge^2(\mathbb{Z}^5)). \end{array}$$

In joint work with the first named author, this yielded the following theorem, which was a result of a series of papers [11], [12], [13], [14].

Theorem 5.5. *When elliptic curves over $\mathbb{Q}$ are ordered by height, their average rank is $< .885$; a density of at least 83.75% have rank 0 or 1; a density of at least 20.62% have rank 0.*

Extending these methods still further, and using the famous Gross–Zagier formula [24], Kolyvagin's theory of Euler systems [28], and recent works of Dokchitser–Dockchitser [22] and recent Skinner, Urban, and Zhang [35], [36], [37], [38], [39], Bhargava, Skinner, and Zhang obtain the following stunning result [15].

Theorem 5.6. *When elliptic curves over $\mathbb{Q}$ are ordered by height, at least 66.48% of them satisfy the Birch–Swinnerton-Dyer conjecture; at least 16.50% of elliptic curves over $\mathbb{Q}$ have algebraic and analytic rank zero; at least 20.68% have algebraic and analytic rank one.*

6. Hyperelliptic curves

Finally, we consider hyperelliptic curves, i.e., (projective) curves defined by an equation of the form

$$y^2 = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n, \tag{8}$$

where the polynomial on the right hand side is assumed to have no repeated factors. The genus of the above curve where $n = 2g+1$ or $2g+2$ is g and so when $n \geq 2$ these curves have genus at least 2. Curves having genus 0 have either no rational points or infinitely many rational points all of which can be parameterized algebraically using one parameter just like the parametrization of Pythagorean triples. Curves of genus 1 can have no rational points. When they do have rational points, they are elliptic curves and as we have seen in the previous section, they can have finitely many rational points (when $r = 0$), or infinitely many rational points (when $r \geq 1$). Curves with genus 2 or higher are addressed by a tremendously powerful theorem of Faltings (originally a conjecture of Mordell):

Theorem 6.1 (Faltings)**.** *When $n \geq 5$, equation* (8) *has finitely many rational solutions.*

However, the above theorem does not address the question of *how many* points $C(\mathbb{Q})$ has. In fact, Theorem 6.1 is ineffective, and does not provide any bound on the size of $C(\mathbb{Q})$. Effectivising Theorem 6.1 is open and would be a major breakthrough, but we can apply the philosophy of Arithmetic Statistics and ask instead: what is the average number of rational points in families of curves having genus $g \geq 2$. In this regard, there have been a slew of recent results, many of which crucially use Bhargava's methods. We describe three of these results below.

First, we consider the family of monic odd hyperelliptic curves, i.e., curves cut out by the equation (8) with $a_0 = 1$ and n odd. Each such curve has at least one rational point (the point at infinity). We order these curves C by height which is defined as follows:

$$H(C) = \max\{|a_1|, |a_2|^{1/2}, \ldots, |a_n|^{1/n}\}. \tag{9}$$

In [9], Bhargava and Gross study the Selmer groups of Jacobians of these curves, and using their results Poonen and Stoll prove the following result in [32].

Theorem 6.2 (Poonen–Stoll)**.** *For each odd $n \geq 7$, a positive proportion of curves in the family of monic degree-n hyperelliptic curves, when ordered by height, have exactly one rational point. Furthermore, this proportion tends to 1 as n tends to infinity.*

Next we consider the family of even hyperelliptic curves, i.e., curves C cut out by (8) with even $n \geq 6$. We order curves in this family by the following height:

$$H(C) = max\{|a_i|\}.$$

Bhargava proves the following result in [8]:

Theorem 6.3. *For each even $n \geq 8$, a positive proportion of curves in the family of degree-n hyperelliptic curves, when ordered by height, have no rational points. Furthermore, this proportion tends to 1 as n tends to infinity.*

Finally, in [10], Bhargava, Gross, and Wang prove the following stunning results.

Theorem 6.4. *For any even $n \geq 2$, a positive proportion of curves in the family of degree-n hyperelliptic curves, when ordered by height, have no points over any odd degree extension of $\mathbb{Q}$.*

Theorem 6.5. *Fix any $m > 0$. Then as $n \to \infty$, a proportion approaching 1 of degree-n hyperelliptic curves have no points defined over any extension of $\mathbb{Q}$ having odd degree $\leq m$.*

References

[1] Baker, A., Imaginary quadratic fields with class number 2, *Annals of Math.* (2) **94** (1971), 139–152.

[2] Bhargava, M., Higher composition laws I: A new view on Gauss composition, and quadratic generalizations, *Ann. of Math.* **159** (2004), no. 1, 217–250.

[3] Bhargava, M., Higher composition laws. II. On cubic analogues of Gauss composition. *Ann. of Math. (2)* **159** (2004), no. 2, 865–886.

[4] Bhargava, M., Higher composition laws III: The parametrization of quartic rings, *Ann. of Math.* **159** (2004), 1329–1360.

[5] Bhargava, M., Higher composition laws. IV. The parametrization of quintic rings. *Ann. of Math. (2)* **167** (2008), no. 1, 53–94.

[6] Bhargava, M., The density of discriminants of quartic rings and fields. *Ann. of Math. (2)* **162** (2005), no. 2, 1031–1063.

[7] Bhargava, M., The density of discriminants of quintic rings and fields, *Ann. of Math. (2)* **172** (2010), no. 3, 1559–1591.

[8] Bhargava, M., Most hyperelliptic curves over Q have no rational points, http://arxiv.org/abs/1308.0395.

[9] Bhargava, M. and Gross, B. H., The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point, *Automorphic representations and L-functions, Tata Inst. Fundam. Res. Stud. Math.* , **22**, Tata Inst. Fund. Res., Mumbai, (2013), 23–91.

[10] Bhargava, M., Gross, B. H. and Wang, X., Pencils of quadrics and the arithmetic of hyperelliptic curves, http://arxiv.org/abs/1310.7692.

[11] Bhargava, M. and Shankar, A., Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves, *Ann. of Math. (2)* **181** (2015), no. 1, 191–242.

[12] Bhargava, M. and Shankar, A., Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0, *Ann. of Math. (2)* **181** (2015), no. 2, 587–621.

[13] Bhargava, M. and Shankar, A., The average number of elements in the 4-Selmer groups of elliptic curves is 7, http://arxiv.org/abs/1312.7333

[14] Bhargava, M. and Shankar, A., The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1, http://arxiv.org/abs/1312.7859

[15] Bhargava, M., Skinner, C. and Zhang, W., A majority of elliptic curves over $\mathbb{Q}$ satisfy the Birch and Swinnerton-Dyer conjecture, http://arxiv.org/pdf/1407.1826v2.pdf.

[16] Brumer, A., The average rank of elliptic curves. I. *Invent. Math.* **109** (1992), no. 3, 445–472.

[17] Cohen, H. and Lenstra, H. W., Heuristics on class groups of number fields, Number theory (Noordwijkerhout, 1983), 33–62, *Lecture Notes in Math.* **1068**, Springer, Berlin, 1984.

[18] Cohen, H. and Martinet, J., Étude heuristique des groupes de classes des corps de nombres, *J. Reine Angew. Math.* **404** (1990), 39–76.

[19] Davenport, H., On the class number of binary cubic forms I and II, *J. London Math. Soc.* **26** (1951), 183–198.

[20] Davenport, H. and Heilbronn, H., On the density of discriminants of cubic fields II, *Proc. Roy. Soc. London Ser. A* **322** (1971), no. 1551, 405–420.

[21] Deuring, M., Imaginare quadratische Zahlkorper mit der Klassenzahl Eins, *Invent. Math.* **5** (1968), 169-179.

[22] Dokchitser, T. and Dokchitser, V., On the Birch–Swinnerton-Dyer quotients modulo squares, *Ann. of Math. (2)* **172** (2010), no. 1, 567–596.

[23] Goldfeld, D., The Gauss class number problem for imaginary quadratic fields, *Bull. Amer. Math. Soc. (N.S.)* **13** (1985), no. 1, 23–37.

[24] Gross, B. H. and Zagier, D., Heegner points and derivatives of L-series, *Invent. Math.* **84** (1986), no. 2, 225–320.

[25] Heath-Brown, D. R., The average analytic rank of elliptic curves. *Duke Math. J.* **122** (2004), no. 3, 591–623.

[26] Heegner, K., Diophantische Analysis und Modulfunktionen, *Math. Z.* **56** (1952), 227–253.

[27] Heilbronn, H., On the class number in imaginary quadratic fields, *Quart. J. Math. Oxford Ser. 2* **5** (1934), 150–160

[28] Kolyvagin, V. A., Euler Systems, *The Grothendieck Festschrift*, Progr. in Math. **87**, Birkhauser Boston, Boston, MA (1990).

[29] Landau, E., Uber die Klassenzahl imaginar-quadratischer Zahlkorper, *Gottinger Nachr.* (1918), 285–295.

[30] Mazur, B., Modular curves and the Eisenstein ideal, *Publications Mathmatiques de l'IHS* (*1*) **47**, 33–186.

[31] Mertens, F., Ueber einige asymptotische Gesetze der Zahlentheorie, *J. reine angew Math.* **77** (1874), 289–338.

[32] Poonen, B. and Stoll, M., Most odd degree hyperelliptic curves have only one rational point, *Ann. of Math.* (*2*) **180** (2014), no. 3, 1137–1166.

[33] Siegel, C. L., Uber die Classenzahl quadratischer Zahlkorper, *Acta Arith.* **1** (1935), 83-86.

[34] Siegel, C. L., The average measure of quadratic forms with given determinant and signature, *Ann. of Math. (2)* **45** (1944), 667–685.

[35] Skinner, C., A converse to a theorem of Gross–Zagier–Kolyvagin, http://arxiv.org/abs/1405.7294

[36] Skinner, C., Multiplicative reduction and the cyclotomic main conjecture for modular forms, http://arxiv.org/abs/1407.1093.

[37] Skinner, C. and Urban, E., The Iwasawa main conjectures for GL(2), *Invent. Math.*, **195** (2014), no. 1, 1–277.

[38] Skinner, C. and Urban, E., forthcoming.

[39] Skinner, C. and Zhang, W., Indivisibility of Heegner points in the multiplicative case, http://arxiv.org/abs/1407.1099.

[40] Stark, H., A complete determination of the complex quadratic fields of class number one, *Mich. Math. J.* **14** (1967), 1–27.

[41] Stark, H., A transcendence theorem for class number problems I, II, *Annals of Math.* (2) **94** (1971), 153–173; *ibid.* **96** (1972), 174–209.

[42] Wright, D. and Yukie, A., Prehomogeneous vector spaces and field extensions,*Invent. Math.* **110** (1992), no. 2, 283–314.

[43] Young, M. P., Low-lying zeros of families of elliptic curves. *J. Amer. Math. Soc.* **19** (2006), no. 1, 205–250.

Arul Shankar
Department of Mathematics, FAS, Harvard University
One Oxford Street, Cambridge MA 02138, USA
arul.shnkr@gmail.com
and
Xiaoheng Wang
Princeton Math Department, Fine Hall
Washington Road, Princeton, NJ 08544, USA
xw5@math.princeton.edu

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 173-176

BOOK-REVIEW

TITLE : **HOW TO STUDY FOR A MATHEMATICS DEGREE**
Author : **Lara Alcock**, Mathematics Education Center, Loughborough University
Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK; 2013
Hard/Soft cover; Soft cover: xvi + 272 pages
ISBN No. 978-0-19-966132-9 ;
Library of Congress Control Number : 2012940939.
Reviewer : **Shabd Sharan Khare**

Millions of students, every year, enter different universities all over the world for mathematics degree. Most of them, including even some good ones, struggle with the demands of transition from school-level to university-level mathematics which is hard in terms of complexity of the subject matter, the rigour of definitions, theorems and proofs, and the need to be able to study much more independently. There is major shift from mechanical procedures and calculations using formulas to rigorous definitions and proofs.

This wide ranging book aims at engaging with all these issues and many more very successfully giving a very helpful insight into what is coming for beginning undergraduates in mathematics and mathematics related disciplines like statistics and economics etc. It translates mathematics education research based insights into practical advices for students entering universities for mathematics degree or existing students undergoing mathematics in universities at undergraduate level. It covers every aspect of studying for mathematics degree, starting from the most abstract intellectual challenges to everyday business of interacting with lecturers and making good use of study time.

It is not a popular mathematics book and is less focused on mathematical curiosities, applications and contents. It is more focused on how to engage students with academic content of mathematics courses and on challenges of coping with formal and abstract undergraduate mathematics. This book is about how to keep on top of study while enjoying mathematics.

The book has two parts. Part 1 describes the nature of university mathematics, discusses how it differs from school mathematics and offers advices/tips about things students could do to understand it in a better way and come out successfully in mathematics degree. It contains suggestions on how to interact with mathematical content at university level. At school level, the focus of mathematics is on applying mathematical formulas and procedures to calculate answers of standard questions. Some students enjoy this. However, some students dislike this aspect of mathematics and they get more satisfaction from learning about why the various procedures work and how they can be modified in changed situations. At the university level, the lecturers expect from the students to be fluent in using procedures (like solving algebraic equations, differentiating or integrating given functions and so on), they have learned at the school level. Therefore students are supposed to brush up class 11-12 relevant procedures and formulas prior to arriving undergraduate mathematics classes. In undergraduate mathematics, there is focus on underlying mathematics in different procedures and formulas. The suggestion given by the author to turn ordinary exercises into opportunities for reflection and further thinking by asking following questions within oneself is excellent.

1. Why did that procedure work?
2. What could be changed in the question so that the same procedure still works?
3. What could be changed in the question so that the same procedure does not work?
4. Could one modify the procedure to work for the changed situation?

Another major difference between school-level and university-level mathematics is that in contrast to school-level mathematics, students have to decide which procedure or trick needs to be applied in a particular problem, for example integration and differential equation etc. Learning procedures mechanically has a disadvantage that it is easier to forget them, to misapply them and to mix them up.The author has rightly stressed that developing a proper understanding of mathematics underlying a procedure is supportive to flexible and accurate reasoning.

In Part 1, the author has also explained the notion of axioms, definitions, and theorems with nice examples. Chapter 4 deals with quantifiers like $\forall, \exists, \Rightarrow, \rightarrow$ with simple examples. Very often, most of the students do not understand these quantifiers and their proper use. The author has explained these notions and their use in detail with lot of examples. Chapters 5 and 6 describe common undergraduate proof structures and different useful strategies and tricks for tackling them including proofs by contradiction and proofs by induction. Chapter 7 explains the strategies with examples that are useful in reading for understanding, for synthesis and for memory. The author also tries to discuss how to build on these strategies while preparing for examination. The book also emphasizes the importance of developing the habit of reading mathematical stuff independently at under graduate level. Practice and use of diagrams are crucial for memory, while reading mathematics.

Chapter 8 is about the importance of good mathematical writing and proper presentation of mathematical ideas. Mathematical arguments need to be presented very clearly. The author has given some useful tips on avoiding common mistakes while using mathematical symbols like $\forall, \exists, \Rightarrow, \rightarrow$ and brackets etc.

Part 2 of the book deals about how to get the best out of undergraduate lectures in mathematics and how to organize the time for studies so that students can keep up with mathematics successfully and enjoy as well. This part is not only useful for students entering graduate programs in mathematics or mathematics related disciplines but also for students who are already undergoing undergraduate program in mathematics. Perhaps, the latter ones might like to jump on to part 2 first. In part 2, Chapters 9 and 10 are about what to expect from lectures and tutorial classes at the undergraduate level. It advises how to get the best out of being an independent learner in a lecture-based environment and about how to deal with common problems faced in mathematics lecture. These days, there is much varieties in the mode of lecturing/ teaching by different teachers. Some use chalks/whiteboards pens, some use power point presentation, some provide hand outs before or after lectures and some use university virtual learning systems. One of the key advice given by the author to make the best use of lecture in mathematics is to go through the concerned portion one day before the lecture from the books/ handouts/ notes given by teacher/ university virtual learning system and mark the steps or portions or exercise not clear, be extra careful in the class to understand when the teacher comes to those steps or portions and ask the teacher politely to explain those points, in case the points were not clear even after the teacher had explained. Many other useful tips have been given in the book that might be of great help in understanding and enjoying lectures. The students need to be very particular while giving feedback by filling the feedback form given by the teacher, giving constructive comments. The tutorial classes should be taken as the best opportunity to remove doubts and difficulties. It may be good practice for student to discuss mathematics with classmates, specially good ones. Some of the difficulties can be discussed with the teacher on email.

Chapter 11 is about how to organize study time in mathematics so that students can keep up with mathematics and avoid the stress of falling behind. The chapter has provided practical suggestions for making realistic plans that can be stuck to without having a feeling of getting bogged down in studies. Making a term planner may be of great help to get an overview of what is happening when, to plan for weekend activities, and to schedule in coursework and test preparation. Further, planning for typical week will help in allowing one to think realistically about when to study what. In addition to this, including social activities in plan will help to ensure students to enjoy as well. Making a list of current study jobs for different modules may be of great help in prioritizing and deciding how to allocate time in a particular week. Most of the students feel that they have to work on something until they finish it. However, in reality this is likely to be ineffective. The book has made some good suggestions on planning for shorter bursts of study time. Most of the students have periodic failure of time management. The author has given useful tips on how to get back on track quickly depending on the aim of the student.

Some students lag behind the studies due to slackening or illness or some traumatic events in life. Due to this, they find themselves in a state of panic. Chapter 12 is about such students and gives some tips to sort out the panic situation and restore their confidence. The crux of the tip is to list urgent items in each of the current modules in a sheet along with the associated deadlines, to list four items from the sheet for each day of the week in a separate sheet giving some reasonable time for each item and then to follow meticulously.

Chapter 13 aims to dispel some common misunderstanding about the notion of success in mathematics at undergraduate level and to replace them with its realistic notion. Due to vast change in the nature and approach of mathematics and due to faster pace of teaching at undergraduate level, many students find it difficult to cope up and develop a wrong feeling that mathematics is not their cup of tea. According to the author, keeping up might be easier, if they focus on central ideas in a module or on those ideas which are likely to be used in the upcoming lectures. Being good in mathematics is not necessarily same as being fast at it. As mentioned in this chapter, taking some time to understand new notions properly is key to success in mathematics at undergraduate level. Also, keeping in view the fact that it is normally not possible to understand everything by the time of examination, students should think about how to distribute their effort in order to do well in examination.

On the whole, this book may help students in giving a head-start in mathematics at undergraduate level, which may pay them in years to come. It is an excellent book of great value to any one embarking on mathematics at university level as well as to students already undergoing mathematics or mathematics related courses at undergraduate level. It is of great help in bridging the vast gap between school-level and undergraduate-level mathematics. One of the key feature of the book is that it is written more in narrative and conversational style, which is quite in contrast with other mathematics books. Due to this, the book is more friendly, readable and self-help. It is one of the best books a budding mathematician could read before going to university for mathematics. I wish this book had been available 50 years ago, when I entered university for undergraduate program.

Shabd Sharan Khare
Ex Pro Vice Chancellor, NEHU, Shillong
E-mail : *kharess1947@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3-4, July-December (2015), 177

BOOK-REVIEW

TITLE : **POLYHEDRA PRIMER**
Authors : **Peter Pearce and Susan Pearce**
Digital reproduction: Pearce Publications, 2015; Originally: Van Nostrand Reinhold, 1978; Subsequently by Dale Seymour Publications.
Soft cover/Paperback: viii + 134 pages
ISBN-10: 1507686226 ; ISBN-13: 978-1507686225
Reviewer : **N. K. Thakare**

This book is a wonderful and useful addition to the existing literature on polyhedra. I am struck by the lucidity and simplicity in the approach to the study of the geometry of three dimensional space and its application to building systems: Through seven chapters spread up over 134 pages and 250 captioned drawings, the authors begin with as elementary concepts as points, lines, angles, polygons etc. They demonstrate through constructions how combining of polygons can lead to tessellation (i.e. tiling) without overlaps or gaps. And the tiling can be done with nonuniform and nonperiodic regular polygons as well as nonregular polygons. They discuss polyhedra that are formed by enclosing a portion of three-dimensional space with four or more polygons. Through constructions they explain the formations of dual polyhedra. They further illustrate how space filling occurs by packing polyhedra together. They explain the procedure of uniform space filling with one or more kinds of polyhedra. The nice explanation of how open packing can be derived from the space filling systems is breathtakingly engrossing. In the final chapter titled 'constructions', the authors elaborate basic constructions such as bisecting a line, an angle etc, that lead to the construction of polyhedra models.

The book shall be well understood by high school students, undergraduates and laymen. The book shall be not only meaningful but immensely useful to the practitioners in professions such as architecture, planning, engineering, industrial designs, arts, etc.

N. K. Thakare
C/o: Center for Advanced Study in Mathematics
Savitribai Phule Pune University
Pune-411 007, (MS), India
E-mail : *nkthakare@gmail.com*

The Mathematics Student ISSN: 0025-5742
Vol. 84, Nos. 3 - 4, July-December, (2015), 179 - 186

PROBLEM SECTION

We are delighted to announce that there was reasonably good participation from the floor to our very first Problem Section restarted with The Mathematics Student, Vol. 84, Nos. 1-2, January-June (2015) in which 11 problems were proposed.

We received a total of 7 responses to the problems posted in MS-2015, Nos. 1-2: One submitted solution *each to the problems*-3, 5, 9 and 11 and three submitted solutions to Problem 4; of these solutions, only the solution to problem 3 turned out to be wrong.

Besides these, there was also a problem proposal submitted for possible inclusion in the new problem set.

In this issue, we present 7 new problems, including the one just mentioned above. Also, readers can try their hand on the remaining problems from MS-2015, Nos.1-2 because as a matter of policy proposer's solution to those problems for which no correct solution is received from the floor will be provided in the corresponding issue of the following year (thus, for example, solutions to those problems from MS-2015, Nos.1-2 for which no solution is received will be provided in the MS-2016, Nos. 1-2 issue).

Solutions from the floor received till April 30, 2016, if approved by the Editorial Board, will be published in The Mathematics Student Vol. 85, Nos. 1-2, January-June (2016).

The seven new problems are as under followed by the proposer's solution or the Editorial Board approved submitted solution from the floor to those problems of MS-2015, Nos. 1-2 , for which correct solutions are received.

MS-2015, Nos. 3-4: Problem-12: Proposed by Atul Dixit.

> Let ABC be a triangle with medians AD, BE, and FC. Construct semicircles with diameters BD, DC, and BC outwardly. Let T_A be the circle tangent to the three semicircles with diameters BD, DC, and BC, and let A' be the center of T_A. Define B' and C' cyclically. Then give a synthetic (pure geometric) proof of the fact that the segments AA', BB', and CC' are concurrent.

MS-2015, Nos. 3-4: Problem-13: Proposed by B. Sury.

Let $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Determine the *semigroup* generated by these two matrices (note that we are not allowed to take negative powers of the matrices).

MS-2015, Nos. 3-4: Problem-14: Proposed by B. Sury.

Let $f \in \mathbb{Q}[X]$ be a polynomial of degree at least 2. If f gives a bijection on a subset S of $\mathbb{Q}$, then prove that S must be finite.

MS-2015, Nos. 3-4: Problem-15: Proposed by B. Sury.

If N is a positive integer which is a multiple of a number of the form $99\cdots 9$ where 9 is repeated n times, then show that N has at least n non-zero digits. More generally, prove that if $b > 1$ and $b^n - 1$ divides a, then the base-b expression of a has at least n non-zero digits.

MS-2015, Nos. 3-4: Problem-16: Proposed by Mathew Francis, ISI Bangalore; (may be an old problem); submitted through L. Sunil Chandran.

Let $S_1, S_2, \ldots, S_k$ be subsets of a universe $U = \bigcup S_i$ such that the union of no t of them covers U. Then prove that there exist $t+1$ sets among them that are pairwise incomparable.

MS-2015, Nos. 3-4: Problem-17: Proposed by Mahender Singh, IISER Mohali; submitted through C. S. Aravinda.

Let G be a finite group of order n and let ϕ be the Euler's totient function. It is easy to see that if G is cyclic then $|Aut(G)| = \phi(n)$. Is the converse true?

MS-2015, Nos. 3-4: Problem-18: Proposed from the floor by Amrik Singh Nimbran, Patna, Bihar, India.

Evaluating the product, prove that

$$\frac{\sqrt{2}}{1+\sqrt{2}} \cdot \frac{\sqrt{2+\sqrt{2}}}{1+\sqrt{\sqrt{2}}} \cdot \frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{1+\sqrt{\sqrt{\sqrt{2}}}} \cdots = \frac{\ln 4}{\pi}.$$

Solution by the Proposer: MS-2015, Nos. 1-2: Problem 4:

Show that $n \mid 2^n - 1$ over $\mathbf{Z}$ if and only if $n = 1$.

Solution. Suppose there is an $n > 1$ for which $n \mid 2^n - 1$. Then n is odd. Let n_0 be the least such n. By Euler's theorem, $2^{\varphi(n_0)} \equiv 1 \bmod n_0$ where φ is Euler's function. Put $g = \gcd(n_0, \varphi(n_0))$. Then $1 \leqslant g < n_0$ and by the Euclidean algorithm, there are integers x and y so that $g = n_0 x + \varphi(n_0)y$. Hence $2^g \equiv 1 \bmod n_0$. As $g \mid n_0$, we get $2^g \equiv 1 \bmod g$. By the minimality of n_0, we deduce that $g = 1$. But then $1 \equiv 2^g \equiv 2 \bmod n_0$ so that $n_0 = 1$, a contradiction.

Correct solution received from the floor from:

Manjil Saikia (Diploma Student, Abdus Salam International Center for Theoretical Physics, Trieste 34151, Italy; *msaikia@ictp.it* ; received on 22-06-2015)

Subhash Chand Bhoria (Corporal, Air Force Station, Bareilly, Technical Flight, UP-243002, India; *scbhoria@yahoo.com* ; received on 07-08-2015).

Prahlad Sharma (BSc. 2nd Year Student, Institute of Mathematics and Applications, Andharua, Bhubaneswar, Odisha, *livecrunklife@gmail.com*; received on 27-06-2015)

Solution from the floor to MS-2015, Nos. 1-2: Problem 5: Prove that

$$\int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x}\left(\frac{x^4\cos^3 x - x\sin x + \cos x}{x^2\cos^2 x}\right)\mathrm{d}x$$
$$= \frac{1}{12\pi}\left(e^{\frac{4+\pi}{4\sqrt{2}}}(48\sqrt{2} - 3\pi^2) + e^{\frac{3+\pi\sqrt{3}}{6}}(4\pi^2 - 72)\right).$$

(Submitted on 07-08-2015 by **Subhash Chand Bhoria**; Corporal, Air Force Station, Bareilly, Technical Flight, UP-243002, India; *scbhoria@yahoo.com*).

Solution. Let I be the given integral and $I = I_1 + I_2$, where

$$I_1 = \int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x} x^2 \cos x \,\mathrm{d}x$$

and

$$I_2 = \int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x}\left(\frac{-x\sin x + \cos x}{x^2\cos^2 x}\right)\mathrm{d}x.$$

Observe that

$$\begin{aligned} I_1 &= \int_{\pi/4}^{\pi/3} x \frac{\mathrm{d}}{\mathrm{d}x}\left(e^{x\sin x + \cos x}\right)\mathrm{d}x \\ &= \left[xe^{x\sin x + \cos x}\right]_{\pi/4}^{\pi/3} - \int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x}\,\mathrm{d}x \\ &= \frac{\pi}{3}e^{\left(\frac{3+\pi\sqrt{3}}{6}\right)} - \frac{\pi}{4}e^{\left(\frac{4+\pi}{4\sqrt{2}}\right)} - \int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x}\,\mathrm{d}x \end{aligned} \tag{0.1}$$

and we can express I_2 as

$$I_2 = \int_{\pi/4}^{\pi/3} e^{x\sin x + \cos x}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{-1}{x\cos x}\right)\mathrm{d}x$$

$$
\begin{aligned}
&= \left[\frac{-e^{x\sin x+\cos x}}{x\cos x}\right]_{\pi/4}^{\pi/3} + \int_{\pi/4}^{\pi/3} e^{x\sin x+\cos x}\,\mathrm{d}x \\
&= \frac{-6}{\pi}e^{\left(\frac{3+\pi\sqrt{3}}{6}\right)} + \frac{4\sqrt{2}}{\pi}e^{\left(\frac{4+\pi}{4\sqrt{2}}\right)} + \int_{\pi/4}^{\pi/3} e^{x\sin x+\cos x}\,\mathrm{d}x. \qquad (0.2)
\end{aligned}
$$

Adding I_1 and I_2 we get

$$
\begin{aligned}
I &= I_1 + I_2 \\
&= \frac{\pi}{3}e^{\left(\frac{3+\pi\sqrt{3}}{6}\right)} - \frac{\pi}{4}e^{\left(\frac{4+\pi}{4\sqrt{2}}\right)} + \frac{-6}{\pi}e^{\left(\frac{3+\pi\sqrt{3}}{6}\right)} + \frac{4\sqrt{2}}{\pi}e^{\left(\frac{4+\pi}{4\sqrt{2}}\right)} \\
&= \frac{1}{12\pi}\left(e^{\frac{4+\pi}{4\sqrt{2}}}(48\sqrt{2}-3\pi^2) + e^{\frac{3+\pi\sqrt{3}}{6}}(4\pi^2-72)\right),
\end{aligned}
$$

as required.

Solution from the floor to MS-2015, Nos. 1-2: Problem 9: Consider the polynomial $f = x^3 - 15x^2 + 75x - 120$. Let $f_1 = f$, $f_2 = f \circ f$ and, in general, $f_n = f \circ f_{n-1}$. Determine the zeroes of the polynomial f_n

(Submitted on 27-06-2015 by **Prahlad Sharma**; BSc. 2nd Year Student, Institute of Mathematics and Applications, Andharua, Bhubaneswar, Odisha, India. *livecrunklife@gmail.com*).

Solution. Note that $f(x) = (x-5)^3 + 5 \Rightarrow f(x) - 5 = (x-5)^3$.

Substituting $x = f_n(x)$, we get a recurrence relation $a_{n+1} = a_n^3$, where $a_n = f_n - 5$ The general term for this simple recurrence is $a_n = a_1^{3^{n-1}}$. Hence $f_n - 5 = (f_1 - 5)^{3^{n-1}} \Rightarrow f_n = (x-5)^{3^n} + 5$ from which it follows that the only zero of f_n is $5 - 5^{\frac{1}{3^n}}$.

Solution from the floor to MS-2015, Nos. 1-2: Problem 11*: Let G be a finite p-group of order p^n and $Aut(G)$ be the group of all automorphisms of G. It is easy to see that if $n = 1$ then $|Aut(G)| = (p-1)$, and if $n = 2$ then $|Aut(G)| = (p^2 - p)$ or $(p^2-1)(p^2-p)$.

Now suppose that $n \geq 3$ and G is abelian. Then show that $|G|$ divides $|Aut(G)|$. It is an open conjecture that, in this case, $|G|$ divides $|Aut(G)|$ even if G is non-abelian.

* The Editors regret the inadvertent omission of the extra condition in the statement of problem 11 that it is seeking a solution for '*non cyclic*' groups. This was also rightly pointed out by 'Prahlad Sharma' who happens to have submitted a correct solution that we publish below.

(Submitted on 27-06-2015 by **Prahlad Sharma**; BSc. 2nd Year Student, Institute of Mathematics and Applications, Andharua, Bhubaneswar, Odisha, India. *livecrunklife@gmail.com*).

Solution. We **Claim** : Let G be an abelian *non-cyclic* p-group with $|G| = p^n$, $n \geq 3$. Then $|G| = p^n$ divides $|Aut(G)|$.

Proof of the claim : Throughout the proof, many standard properties of abelian groups are used without mentioning. Since G is abelian and $|G| = p^n$, it is well-known (Herstein, I. N., *Topics in Algebra*, 2nd ed., Wiley, India, 2014; Theorem 2.14.1, p.109) that G can be written as the internal direct product of its cyclic subgroups $N_1, N_2, \cdots, N_k$, say

$$G = N_1 N_2 \cdots N_k, \tag{0.3}$$

where

$$N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \cdots N_k) = (e) \tag{0.4}$$

for each $i = 1, 2, \cdots, k$.

Let $N_i = \langle a_i \rangle$. Then $o(a_i) = p^{n_i}$ since this must divide p^n. Without loss of generality we may assume that $n_1 \leq n_2 \leq \cdots \leq n_k$. Clearly $n_1 + n_2 + \cdots + n_k = n$, from (0.3) and (0.4). We prove the following lemma.

Lemma: For G as described above, a homomorphism $\phi : G \to G$, is an automorphism if and only if the images $\phi(a_i) = b_i$ are such that $o(a_i) = o(b_i)$ and

$$M_i \cap (M_1 M_2 \cdots M_{i-1} M_{i+1} \cdots M_k) = (e), \tag{0.5}$$

where $M_i = \langle b_i \rangle$ for each $i = 1, 2 \cdots k$.

Proof of the lemma: Suppose the map ϕ is an automorphism. Then clearly $o(a_i) = o(\phi(a_i)) = o(b_i)$; and also

$$\begin{aligned} & N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \cdots N_k) = (e) \\ \Longrightarrow\ & \phi(N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \ldots N_k)) = (e) \\ \Longrightarrow\ & \phi(N_i) \cap (\phi(N_1)\phi(N_2) \cdots \phi(N_{i-1})\phi(N_{i+1}) \cdots \phi(N_k)) = (e) \end{aligned}$$

because ϕ is an automorphism. It follows that

$$M_i \cap (M_1 M_2 \cdots M_{i-1} M_{i+1} \cdots M_k) = (e)$$

for each $i = 1, 2 \cdots, k$ as $\phi(N_i) = M_i$ for each $i = 1, 2 \cdots, k$.

In the other direction, suppose that the map ϕ satisfies the properties stated in the claim. For any $g = a_1^{r_1} a_2^{r_2} \cdots a_k^{r_k} \in G$

$$\begin{aligned} & \phi(g) = e \\ & \Rightarrow \phi(a_1^{r_1} a_2^{r_2} \cdots a_k^{r_k}) = e \\ & \Rightarrow b_1^{r_1} b_2^{r_2} \cdots b_k^{r_k} = e \\ & \Rightarrow b_i^{r_i} = e, i = 1, 2 \cdots k \text{ (using (0.5))} \\ & \Rightarrow a_i^{r_i} = e \text{ (since } o(a_i) = o(b_i)) \\ & \Rightarrow g = a_1^{r_1} a_2^{r_2} \cdots a_k^{r_k} = e. \end{aligned}$$

Hence ϕ is an isomorphism. G being finite, this implies ϕ is an automorphism.

We now proceed to our main problem.

Case 1: $n_k = 1$.

This implies $n_1 = n_2 = \cdots = n_k = 1$, and hence in this case every element of G is of order p. Using the lemma, it is then not difficult to see that a mapping $\phi : G \to G$ is an automorphism iff

$$\begin{aligned}
&\phi(a_1) = b_1, && b_1 \in G \setminus (e); \\
&\phi(a_2) = b_2, && b_2 \in G \setminus M_1 \quad \text{in which} \quad M_1 = \langle b_1 \rangle\,; \\
&\phi(a_3) = b_3, && b_3 \in G \setminus M_1 M_2 \quad \text{in which} \quad M_2 = \langle b_2 \rangle\,; \\
&\quad\vdots \\
&\phi(a_k) = b_k, && b_k \in G \setminus M_1 M_2 \cdots M_k \quad \text{in which} \quad M_k = \langle b_{k-1} \rangle\,;
\end{aligned} \tag{0.6}$$

and for any $g = a_1^{q_1} a_2^{q_2} \cdots a_k^{q_k} \in G$

$$\phi(a_1^{q_1} a_2^{q_2} \cdots a_k^{q_k}) = b_1^{q_1} b_2^{q_2} \cdots b_k^{q_k}. \tag{0.7}$$

From this it follows that

$$|Aut(G)| = \prod_{i=0}^{n-1} (p^n - p^i)$$

which is divisible by p^n if $n \geq 3$.

Case 2: $n_k \geq 2$.

Note that since G is non-cyclic, $k \geq 2$. We show that there is a subgroup of $Aut(G)$ whose order is divisible by p^n. Consider that mapping $\phi : G \to G$, where

$$\begin{aligned}
&\phi(a_1) = a_1^{r_1} a_k^{s_1 p^{n_k - 1}} = b_1, \qquad \phi(a_k) = a_1^{r_k p^{n_1 - 1}} a_k^{s_k} = b_k, \\
&\phi(a_i) = a_1^{r_i p^{n_1 - 1}} a_i^{t_i} a_k^{s_i p^{n_k - 1}} = b_i \text{ for } i = 2, \cdots, k-1, \ \text{(omit this step if k=2)}
\end{aligned} \tag{0.8}$$

where r_1, s_k, t_i are non-negative integers such that $(r_1, p) = (s_k, p) = (t_2, p) = (t_3, p) \cdots (t_{k-1}, p) = 1$ and all other r_i and s_i are any non-negative integers. Further for any $g = a_1^{q_1} a_2^{q_2} \cdots a_k^{q_k} \in G$,

$$\phi(a_1^{q_1} a_2^{q_2} \cdots a_k^{q_k}) = b_1^{q_1} b_2^{q_2} \cdots b_k^{q_k}.$$

Note that $b_i^{p^{n_i}} = e$ and since $(t, p) = 1$, for $0 < r < n_i$ we have

$$b_i^{p^r} = a_i^{tp^r} \neq e.$$

Hence $o(b_i) = o(a_i) = p^{n_i}$ for each i and it can be shown using the fact $N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \cdots N_k) = (e)$, that $M_i \cap (M_1 M_2 \cdots M_{i-1} M_{i+1} \ldots M_k) = (e)$ for each $i = 1, \cdots, k$, where $M_i = \langle b_i \rangle$. Hence ϕ is an automorphism in view of the lemma. Let A be the collection of all such automorphisms obtained by varying r_i, s_i and t_i satisfying the mentioned conditions. Let us now compute $|A|$.

For computation we may take $0 \leq r_1 < p^{n_1}$; for $r_1 \geq p^{n_1} \implies a_1^{r_1} = a_1^r$, where $r = r_1 \mod p^{n_1}$ because $o(a_1) = p^{n_1}$. For similar reason, we may take $0 \leq t_i < p^{n_i}$, $i = 2, 3, \cdots, k-1$ and $0 \leq s_k < p^{n_k}$. And for $s_i, i = 1, 2, 3, \cdots, k-1$ we may take $0 \leq s_i < p$, because $s_i > p$ implies $a_k^{s_i p^{n_k - 1}} = a_k^{r p^{n_k - 1}}$, where

$r = s_i \mod p$ because $o(a_k) = p^{n_k}$). Similarly, we take $0 \leq r_i < p$ for $i = 2, 3, 4, \cdots, k$.
Summing up these and the initial conditions on $r_i's,\ s_i's$ and $t_i's$ we have
$0 \leq r_1 < p^{n_1}$ and $(r_1, p) = 1$, $\quad 0 \leq r_i < p\ ,\quad i = 2, 3, \cdots, k$,
$0 \leq s_k < p^{n_k}$ and $(s_k, p) = 1$, $\quad 0 \leq s_i < p,\ i = 1, 2, \cdots, k-1$,
and $0 \leq t_i < p^{n_i}$ and $(t_i, p) = 1,\ i = 2, 3, \cdots, k$.
Let $R_i,\ S_i,\ T_i$ be the collection of all such $r_i,\ s_i,\ t_i$ respectively, then
$|R_1| = p^{n_1-1}(p-1)$, $\quad |R_i| = p,\ i = 2, 3, \cdots, k$,
$|S_k| = p^{n_k-1}(p-1)$, $\quad |S_i| = p,\ i = 1, 2, 3, \cdots, k-1$,
$|T_i| = p^{n_i-1}(p-1)$.
For $k = 2$, note that each of the 4-tuples $(r_1, r_2, s_1, s_2),\ r_i \in R_i,\ s_i \in S_i$ uniquely identifies an automorphism in A. Also each automorphism in A corresponds to such a 4-tuple. Hence

$$|A| = |R_1|.|R_2|.|S_1|.|S_2| = p^{n_1+n_2}(p-1)^2 = p^n(p-1)^2. \tag{0.9}$$

Similarly for $k > 2$, each of $3k-1$-tuple $(r_1, r_2, \cdots, r_k, s_1, s_2, \cdots, s_k, t_2, t_3, \cdots, t_k)$, $r_i \in R_i, s_i \in S_i, t_i \in T_i$ uniquely determines an automorphism in A and vice-versa. Hence

$$\begin{aligned} |A| &= \left(\prod_{i=1}^{k} |R_i|\right)\left(\prod_{i=1}^{k} |S_i|\right)\left(\prod_{i=2}^{k-1} |T_i|\right) \\ &= \left(p^{k-1}.p^{n_1-1}(p-1)\right)\left(p^{k-1}.p^{n_k-1}(p-1)\right)\left(p^{n_2+n_3+\cdots+n_{k-1}-k+2}(p-1)^{k-2}\right) \\ &= p^{n_1+n_2+\cdots+n_k+k-2}(p-1)^k = p^{n+k-2}(p-1)^k. \end{aligned} \tag{0.10}$$

which is divisible by p^n since $k > 2$. Now all that remains is to show A is a subgroup of $Aut(G)$. For this we use a known result: *If a nonempty finite subset H of a group G is closed under multiplication then it is a subgroup of G* (Herstein, I. N., Lemma 2.4.2, p.38). To see if this holds for A, pick any $\phi,\ \psi \in A$ and let $(p_1, p_2, \cdots, p_k, q_1, q_2, \cdots, q_k, t_2, \cdots, t_k)$ and $(m_1, m_2, \cdots, m_k, w_1, w_2, \cdots, w_k, l_2, \cdots, l_k)$ be the tuples associated with ϕ and ψ respectively, then

$$\begin{aligned} \phi \circ \psi(a_1) = \phi(\psi(a_1)) &= \phi\left(a_1^{m_1} a_k^{w_1 p^{n_k-1}}\right) \\ &= (\phi(a_1))^{m_1} (\phi(a_k))^{w_1 p^{n_k-1}} \\ &= \left(a_1^{p_1} a_k^{q_1 p^{n_k-1}}\right)^{m_1} \left(a_1^{p_k p^{n_1-1}} a_k^{q_k}\right)^{w_1 p^{n_k-1}} \\ &= a_1^{p_1 m_1 + p_k w_1 p^{n_1+n_k-2}} a_k^{(q_1 m_1 + q_k w_1) p^{n_k-1}} \\ &= a_1^{p_1 m_1} a_k^{(q_1 m_1 + q_k w_1) p^{n_k-1}} \end{aligned} \tag{0.11}$$

because $n_k \geq 2$ implies $n_1 + n_k - 2 \geq n_1$ and $o(a_1) = p^{n_1}$. Now since $(m_1, p) = (p_1, p) = 1 \Rightarrow (p_1 m_1, p) = 1$, we see that $\phi \circ \psi(a_1)$ has the required property.

Similarly,

$$\begin{aligned}
\phi \circ \psi(a_k) = \phi(\psi(a_k)) &= \phi\left(a_1^{m_k p^{n_1-1}} a_k^{w_k}\right) \\
&= (\phi(a_1))^{m_k p^{n_1-1}} (\phi(a_k))^{w_k} \\
&= \left(a_1^{p_1} a_k^{q_1 p^{n_k-1}}\right)^{m_k p^{n_1-1}} \left(a_1^{p_k p^{n_1-1}} a_k^{q_k}\right)^{w_k} \\
&= a_1^{(p_1 m_k + p_k w_k) p^{n_1-1}} a_k^{q_k w_k + q_1 m_k p^{n_k+n_1-2}},
\end{aligned} \tag{0.12}$$

where $\left(q_k n_k + q_1 m_k p^{n_k+n_1-2}, p\right) = 1$ since $n_k + n_1 - 2 \geq 1$.
Finally for $i = 2, \cdots, k-1$

$$\begin{aligned}
\phi \circ \psi(a_i) &= \phi\left(a_1^{m_i p^{n_1-1}} a_i^{l_i} a_k^{w_i p^{n_k-1}}\right) \\
&= a_1^{(m_i r_1 + l_i r_i) p^{n_1-1}} a_i^{l_i t_i} a_k^{(m_i s_1 p^{n_1-1} + l_i s_i + w_i s_k) p^{n_k-1}} \\
&= a_1^{d_1 p^{n_1-1}} a_i^{d_2} a_k^{d_3 p^{n_k-1}};
\end{aligned} \tag{0.13}$$

and since $(l_i, p) = (t_i, p) = 1 \Rightarrow (l_i t_i, p) = (d_2, p) = 1$ we see that $\phi \circ \psi \in A$. It follows that A satisfies the closure property. A being finite, this implies $A \subseteq Aut(G)$ and we are done.

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: 27^{th} November 2015

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. 84, Nos. 3-4, July-December, (2015)

EDITORIAL BOARD

THE INDIAN MATHEMATICAL SOCIETY
Founded in 1907
Registered Office: Center for Advanced Study in Mathematics
Savitribai Phule Pune University, Pune - 411 007

Back volumes of our periodicals, except for a few numbers out of stock, are available.

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(India). Printed in India