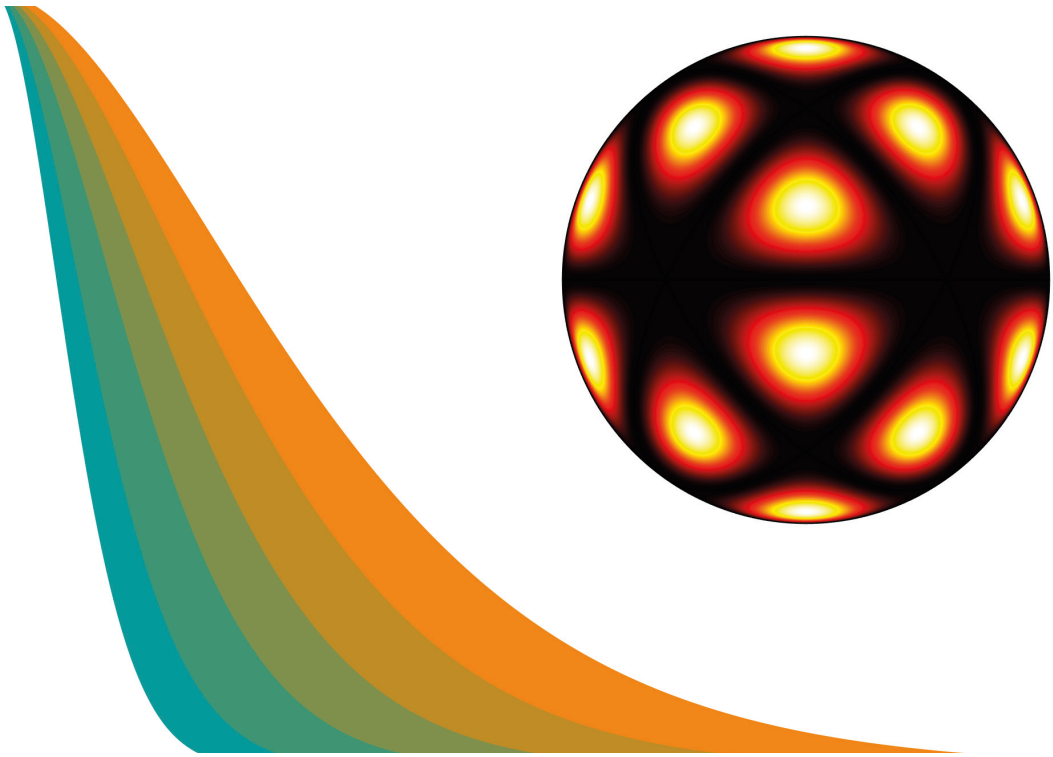
Mälardalen University Doctoral Dissertation 327

Extreme points of the Vandermonde determinant in numerical approximation, random matrix theory and financial mathematics

Asaph Keikara Muhumuza



MÄLARDALEN UNIVERSITY
SWEDEN

Mälardalen University Press Dissertations
No. 327

EXTREME POINTS OF THE VANDERMONDE DETERMINANT IN NUMERICAL APPROXIMATION, RANDOM MATRIX THEORY AND FINANCIAL MATHEMATICS

Asaph Keikara Muhumuza

2020

School of Education, Culture and Communication

ISBN 978-91-7485-484-8
ISSN 1651-4238
Printed by E-Print AB, Stockholm, Sweden

Mälardalen University Press Dissertations
No. 327

EXTREME POINTS OF THE VANDERMONDE DETERMINANT IN NUMERICAL APPROXIMATION, RANDOM MATRIX THEORY AND FINANCIAL MATHEMATICS

Asaph Keikara Muhumuza

Akademisk avhandling

som för avläggande av filosofie doktorsexamen i matematik/tillämpad matematik vid Akademin för utbildning, kultur och kommunikation kommer att offentligen försvaras måndagen den 14 december 2020, 15.15 i Lambda +(digitalt Zoom), Mälardalens Högskola, Västerås.

Fakultetsopponent: Docent Olga Liivapuu, Estonian University of Life Sciences

Akademin för utbildning, kultur och kommunikation

Abstract

This thesis discusses the extreme points of the Vandermonde determinant on various surfaces, their applications in numerical approximation, random matrix theory and financial mathematics. Some mathematical models that employ these extreme points such as curve fitting, data smoothing, experimental design, electrostatics, risk control in finance and method for finding the extreme points on certain surfaces are demonstrated.

The first chapter introduces the theoretical background necessary for later chapters. We review the historical background of the Vandermonde matrix and its determinant, some of its properties that make it more applicable to symmetric polynomials, classical orthogonal polynomials and random matrices.

The second chapter discusses the construction of the generalized Vandermonde interpolation polynomial based on divided differences. We explore further, the concept of weighted Fekete points and their connection to zeros of the classical orthogonal polynomials as stable interpolation points.

The third chapter discusses some extended results on optimizing the Vandermonde determinant on a few different surfaces defined by univariate polynomials. The coordinates of the extreme points are shown to be given as roots of univariate polynomials.

The fourth chapter describes the symmetric group properties of the extreme points of Vandermonde and Schur polynomials as well as application of these extreme points in curve fitting.

The fifth chapter discusses the extreme points of Vandermonde determinant to number of mathematical models in random matrix theory where the joint eigenvalue probability density distribution of a Wishart matrix when optimized over surfaces implicitly defined by univariate polynomials.

The sixth chapter examines some properties of the extreme points of the joint eigenvalue probability density distribution of the Wishart matrix and application of such in computation of the condition numbers of the Vandermonde and Wishart matrices.

The seventh chapter establishes a connection between the extreme points of Vandermonde determinants and minimizing risk measures in financial mathematics. We illustrate this with an application to optimal portfolio selection.

The eighth chapter discusses the extension of the Wishart probability distributions in higher dimension based on the symmetric cones in Jordan algebras. The symmetric cones form a basis for the construction of the degenerate and non-degenerate Wishart distributions.

The ninth chapter demonstrates the connection between the extreme points of the Vandermonde determinant and Wishart joint eigenvalue probability distributions in higher dimension based on the boundary points of the symmetric cones in Jordan algebras that occur in both the discrete and continuous part of the Gindikin set.

ISBN 978-91-7485-484-8
ISSN 1651-4238

Acknowledgements

First and foremost, I would like in a special way to thank my supervisor Professor Sergei Silvestrov who accepted to work with me as his PhD student under the Sida Sweden–Uganda bilateral program 2015–2020. Thank you Professor Sergei for introducing me to this interesting area of research that I have come to love most and for the tireless effort to enable me succeed throughout my PhD studies. In addition, thank you for your persistent guidance, enthusiastic encouragement and constructive critiques during the development of this work.

In addition, I would like to express my sincere appreciation to my co–supervisors Professor Anatoliy Malyarenko, and Dr. Karl Lundengård for their sacrifice in terms of time, efforts and constructive suggestions throughout the various academic discussions we had. I truly learned a lot in all these discussions that I will take with me to wherever I go all over this globe. On the whole you exhibited a special substance of a genius having convincingly guided and encouraged me to undertake and accomplish the right thing even when the road seemed tough. Without your persistent guidance, the goal of this research project would not have been achieved. In the same spirit I wholeheartedly appreciate my other co-supervisors who included Dr. Milica Rančić, Assoc. Prof. John Magero Mango and Dr. Godwin Kakuba for your great advice, compassionate and tireless academic guidance, constructive academic engagements throughout this study.

I am also grateful to Professor emeritus Dmitrii Silvestrov for his insightful comments and useful suggestions for improvements of the PhD thesis and presentation in preparation for PhD defense, as well as for excellent and useful PhD courses lectures. Also, I appreciate the efforts of Professor Predrag Rajkovic for taking time to read my PhD thesis and his positive comments that helped to improve my work.

I would like in a special way to express my deep appreciation to my loving family, my wonderful wife Rebecca Nalule Muhumuza, my dear son Tumwebaze Austine Muhumuza and all the children under my care and mentorship including Jonan Tugume (BSc.Ed), Katusiime Peace (Secretarial), Tamisha Namboira (BA.Admin.), Mayi Namumera (SWASA), Farida Nambote (Dip.Ed.), Gift Niwasiima, Promise Nakayi, Mercy Atim, Bayern Aikiriza and Joan Nakku. Thank for enduring all those several months whenever papa would

be away in Sweden pursuing PhD studies. It has been comforting to know that I could count on all your spiritual and moral support throughout all this time. I would like to pay my special regards to my aging parents. My dad, Mr. Emmanuel Keikara and my mum Mrs. Jane Keikara for the good caring hands and parenting to make me the man I am today. I thank all my siblings especially Wilberforce Bamwiine, Bernard Keikara Mugume, Nelson Keikara Mwebaze, Fred Keikara Mukundane, Frank Gumisiriza, Byaruhanga Hannington, Emmanuel Hama, Yorokamu Kashaija, Jovent Komushomo and Jovlet Arinaitwe for their constant love, friendship and encouragement.

I would like in a special way to express my very great appreciation to the Swedish international development cooperation agency (Sida), International Science Program (ISP) and International Science Programme in Mathematics (IPMS) for all the financial support. In a special way, I thank my Project Coordinators Assoc. Prof. John Magero Mango at Makerere University, Uganda, Assoc. Prof. Bengt-Ove Turresson at Linköping University, Sweden, and Dr. Leif Abrahamsson at ISP, Uppsala for always providing quick answers and ensuring a comfortable stay in Sweden. Special thanks also go to Therese Rwatankoko and Chris Fabian Bengtsson the ISP administrator in charge of Sida Bilateral Sweden–Uganda programme as well as Theresa Lagali Hensen and Josephine Ataro the Maths project 316 administrators at Linköping university and Makerere University respectively.

I wish to express my deepest gratitude to the staff at the School of Education, Culture and Communication, (UKK) Mälardalens University for providing such a wonderful and conducive academic and research environment in Mathematics and Applied Mathematics (MAM). I am particularly greatful to Kristina Konpan (outgoing) who was always ready to attend to our administrative needs. Special thanks to various people who have in one way or the other made my stay in Sweden quite memorable.

I would like to recognize my employers of Busitema University management under the leadership of former Vice Chancellor Professor Mary Okwakor and the current Vice Chancellor Professor Paul Waako for the recommendation, invaluable assistance and support that you all provided throughout my study. More special thanks go to my colleagues at the Faculty of Science Education and department of Mathematics members including Dr. Fulgensia Kamugisha Mbabazi, Dr. Hasifa Nampala, Dr. Richard Awichi, Mr. Jackson Okiring, Mr. Stephen Kadedetsya, Mr. Abubakar Mwasa, Ms. Annet Kyomuhangi, Ms. Rebecca Nalule Muhumuza, Ms. Josephine Nanyondo, and Ms. Topista Nabirye. Thank you for being such a wonderful family.

I would like to thank my fellow PhD students since I enrolled at MDH both under ISP and Sida-Bilateral programs including Benard Abola, Pitos Biganda, Tin Nwe Aye, Samya Suleiman, Elvice Ongonga, Djinja Domingos, those who have already graduated include Dr. Alex Behakanira Tumwesigye, Dr. Betuel Jesus Canhanga, Dr. Carolyne Ogutu, Dr. Jean-Paul Murara, Dr. John Musonda and all other PhD students in Mathematics and

Applied Mathematics, Mälardalens university for the nice moments we shared.

I would like to thank also my fellow PhD students in mathematics who include Benard Abola, Abubakar Mwasa, Innocent Ndikubwayo, Yasin Kikabi, Nalule Rebecca Muhumuza, Annet Kyomuhangi, Olivia Nabawanda, Loy Nankinga, Mary Nanfuka, Josephine Nanyondo, Collins Anguzu, Sam Canpwonyi, Pilly Kimuli, Silas Liliro Kito, Felix Wamuno, Joseph Okello, Martin Arop, Wycliff Ssebunjo, Herbert Mukalazi and Edson Mayanja.

I would like in a special way express my sincere appreciation to my dear spiritual mentors, Pastor Imelda Namutebi Kula and her husband Pastor Tom Kula for their constant love, prayers, counsel and mentorship to stay focused on my academic studies. "*The fear of God is the beginning of wisdom.*" Proverbs 9:10

Västerås, November, 2020
Asaph Keikara Muhumuza

Sida
SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY

Sponsor

UPPSALA UNIVERSITET

International Science Programme (ISP)

MÄLARDALENS HÖGSKOLA
ESKILSTUNA VÄSTERÅS

Host University

MAKERERE UNIVERSITY

Sida Sweden-Uganda Bilateral Programme Coordinating University Uganda

LINKÖPINGS UNIVERSITET

Sida Math Project 316 Coordinating University Sweden

BUSITEMA UNIVERSITY
Pursuing excellence

Employer

This thesis is dedicated to my beloved parents
Mr. Emmanuel Keikara and Mrs. Jane Keikara

Popular Science Summary

Mathematics, naturals sciences and technology are strongly interrelated both in theory and practice. Mathematical theories like analysis, geometry and algebra are all crucial components of mathematical models in many applications. Mathematical models are mainly applied in natural sciences that include physics, biology, earth-science and chemistry, and in technological disciplines including computer science and telecommunication engineering, electrical, mechanical and chemical engineering, as well as in the social-economic science disciplines that include economics, finance, operations research, psychology, sociology and political sciences. The most important thing to note is that a wide variety mathematical models whether linear or non-linear, static or dynamic, explicit or implicit, discrete or continuous, deterministic or stochastic (or probabilistic), strategic or non-strategic and deductive or inductive as used in various science disciplines can all be constructed based on the concept of matrix theory.

In this thesis a special matrix called the Vandermonde matrix is our main focus in studying certain mathematical models in numerical analysis, random matrix theory and random field based on optimization the Vandermonde determinant. Here, mathematical optimization a mathematical programming principle mainly refers to the systematic criteria of selection of a best optimal (or extreme) elements, from some set of available large field of alternative points represented in a matrix form and such elements should maximize or minimize the determinant of the same matrix.

Most mathematical models are characterized by the phenomenon of well–posedness whereby, for example, according to Jacques Hadamard a mathematical model of physical phenomenon is said to be well-posed problem if it has the properties that the solution exists, the solution is unique and the solution's behaviour changes continuously with the initial conditions. In continuum models that must often require to be discretized in order to obtain a numerical solution, whereas the solutions may be continuous with respect to the initial conditions, they may suffer from numerical instability when solved with finite precision, or with errors in the data. Much as the problem may be well–posed, it may still suffer to be ill–conditioned,due to the fact that a small error in the initial data can result in even much larger errors in the final solution. This fact of stability of solutions inspired our study of the Vandermonde matrix and optimization of its determinant a technique that is highly employed in error control for ill-conditioned problem and also indicated by a large condition number.

The study of extreme points of Vandermonde determinant and conditioning inspired us to extend the results to investigate such systems including Coulomb's system and energy level spacing for heavy nuclear atoms which are characterised by joint eigenvalue distribution also called ensembles that occur mainly in random matrix theory and random

fields. These extreme points of the Vandermonde determinant are seen to play a significant role in both physical and biological science based on the zeros of the classical orthogonal polynomials, the Gaussian ensembles and the Wishart ensembles in symmetric cones of Jordan algebras.

Populärvetenskaplig Sammanfattning

Matematik, naturvetenskap och teknologi är starkt sammankopplade både i teori och praktik. Matematiska områden såsom analys, geometri och algebra är kritiska komponenter i konstruktionen av matematiska modeller inom många tillämpningsområden. Matematiska modeller används främst i naturvetenskaper såsom fysik, biologi, geovetenskap och kemi, och inom teknologiska områden såsom datorvetenskap, telekommunikation, elektroteknik, mekanik och kemiteknik, men även inom social-ekonomiska områden såsom ekonomi, finans, operationsanalys, psykologi, sociologi och statsvetenskap. Det som är viktigast att ha i åtanke är att de flesta matematiska modeller, oavsett om de är linjära eller icke-linjära, statiska eller dynamiska, explicita eller implicita, diskreta eller kontinuerliga, deterministiska eller stokastiska (slumpmässiga), strategiska eller ostrategiska, baserade på deduktion eller induktion, kan alla konstrueras baserat på begrepp från matristeori.

I denna avhandling är en speciell slumpmässig matris som kallas för Vandermondematrisen vårt huvudfokus, vi kommer att studera vissa matematiska modeller från numerisk analys, teorin om slumpmässiga matriser och slumpmässiga kroppar baserat på optimering av Vandermonde determinanten. Med matematisk optimering menar vi här systematiskt urval av de mest optimala (eller mest extrema) element från någon stor kropp av möjliga punkter som representeras i matrisform på så sätt att dessa element maximerar eller minimerar determinanten av samma matris.

De flesta modeller ger problem som kan sägas vara väl-ställda, med detta menas, enligt t.ex. Jaques Hadamard, att en matematiska modell av ett fysikaliskt fenomen get välställda problem om problemets lösning existerar, lösningen är entydig och lösningens beteende ändras kontinuerligt om problemets initialvillkor ändras. I kontinuerliga modeller som behöver diskretiseras för att kunna behandlas med numeriska metoder, så kan det vara så att medan lösningen ändras kontinuerligt med avseende på initialvillkoren, så introducerar begränsningar i numerisk precision instabilitet i lösningen. På liknande sätt kan fel i data introducera instabilitet. Stabiliteten av lösningar inspirerade vår undersökning av Vandermondematrisen och metoder för optimering av dess determinant då detta är relevant för felkontroll för dålig ställda problem på grund av kopplingar mellan determinanten och matrisen konditionstal.

Studien av extrempunkter hos Vandermondedeterminanten och kondition insperade vidare undersökning av systems såsom Coulombs system och avstånd mellan energinivåer för tunga kärnpartiklar vilka beskrivs av egenvärdena för en typ av multivariat distribution som kallas för en ensemble och som ofta dyker upp i teorin för slumpmässiga matriser och slumpmässiga kroppar. Extrempunkterna för Vandermondedeterminanten kan beskrivas med hjälp av nollställena till klassiska ortogonala polynom för den Gauss-ensemblen, Wishart-esemblen samt ensembler i den symmetriska konen av Jordan-algebror.

List of Papers

The chapters 2 through to 9 in this thesis are based, respectively, on the following list of papers:

Paper A. Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. *The Generalized Vandermonde Interpolation Polynomial Based on Divided Differences*, SMTDA2018 Conference Proceedings, **ISAST2018**, 443–456, 2018.

Paper B. Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Extreme points of the Vandermonde determinant on surfaces implicitly determined by a univariate polynomial*. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317, 791–818, 2020.
https://doi.org/10.1007/978-3-030-41850-2-33.

Paper C. Muhumuza Asaph K., Silvestrov Sergei, (2019). *Symmetric Group Properties of Extreme Points of Vandermonde Determinant and Schur polynomials*. Accepted for publication in: Sergei Silvestrov, Anatoliy Malyalenko, Milica Rančić M., (Eds.), SPAS 2019: Algebraic Structures and Applications.

Paper D. Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Optimization of the Wishart Joint Eigenvalue Probability Density Distribution Based on the Vandermonde Determinant*. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317, 819–838, 2020.
https://doi.org/10.1007/978-3-030-41850-2-34.

Paper E. Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Properties of the Extreme Points of the Joint Eigenvalue Probability Density Function of the Wishart Matrix*. In ASMDA2019, 18th Applied Stochastic Models and Data Analysis International Conference, ISAST: International Society for the Advancement of Science and Technology. (pp. 559–571), 2019.

Paper F. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Connections Between the Extreme Points of Vandermonde determinants and minimizing risk measure in financial mathematics.* Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), (Eds.), SPAS2019. Algebraic, stochastic and analysis structures for networks, data classification and optimization, Springer Proceedings in Mathematics and Statistics, Springer International Publishing, 2020.

Paper G. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *The Wishart Distribution on Symmetric Cones.* Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), SPAS2019. Algebraic Structures and Applications. SPAS 2019. Springer Proceedings in Mathematics & Statistics, 2020.

Paper H. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *Extreme Points of the Vandermonde Determinant and Wishart Ensembles on Symmetric Cones.* Accepted in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Springer International Publishing, 2020.

The other co-authored paper(s) include:

Paper I. Muhumuza Asaph K., Malyarenko Anatoliy, Silvestrov Sergei, (2017). Lie symmetries of the Black–Scholes type equations in financial mathematics, ASMDA2017 Conference Proceedings, **ISAST2017**, 723–740, 2017.

Parts of this thesis have been presented at the following international conferences and workshops:

- ASMDA2017: The 17th conference of the Applied Stochastic Models and data Analysis International Society and demographic 2017 Workshop, 11-14 June 2019, London, United Kingdom, 6th – 9th June 2017.
- SMTDA2018: The 5th Stochastic Modeling Techniques and Data Analysis International Conference, Chania, Crete, Greece, 12th – 15th June 2018.
- IWAP2018: The 9th International Workshop On Applied Probability, Budapest, Hungary, 18th - 21st June 2018.
- SPAS2017: International Conference on Stochastic Processes and Algebraic Structures-From Theory Towards Applications, Västerås/Stockholm, Sweden, 4th – 6th October, 2017.
- ASMDA2019: The 18th conference of the Applied Stochastic Models and data Analysis International Society and demographic 2019 Workshop, Florence, Italy, 11th – 14th June 2019.
- SPAS2019: International Conference on Stochastic Processes and Algebraic Structures-From Theory Towards Applications, Västerås, Sweden, 30th September – 2nd October 2019.

Notations

The following notations will be used throughout the Thesis unless defined otherwise.

$\mathbb{N},\mathbb{Z},\mathbb{R},\mathbb{C}$	–	The set of Natural, Integers, Real and Complex numbers.
$\mathbf{x},\mathbf{v}$	–	Bold, roman lower letters denote vectors.
$\mathbf{X},\mathbf{V},\mathbf{M}$	–	Bold, uppercase letters denote matrices.
δ	–	The index $\delta = (n-1,\ldots,2,1,0)$, unless stated otherwise.
λ	–	The partition $\lambda = (\lambda_1,\ldots,\lambda_n)$, unless stated otherwise.
$\mathbf{C}_{i,j},\mathbf{M}_{i,j}$	–	Element on the i–th row and j–th column of $\mathbf{M}$.
$\mathbf{M}_{\cdot,j}$	–	Column vector of all elements from the j–th column of $\mathbf{M}$.
$\mathbf{M}_{i,\cdot}$	–	Row vector of all elements from the i–th row of $\mathbf{M}$.
$[a_{ij}]^{nm}$	–	$ij - n\times m$ matrix with element a_{ij} in the i–th row and j–th column.
$\mathbf{V}_{nm}(\mathbf{x})$	–	$n\times m$ – Vandermonde matrix with respect to $\mathbf{x}\in\mathbb{R}^n$.
$\mathbf{V}(\mathbf{x}) = \mathbf{V}_{nn}(\mathbf{x})$	–	n– square Vandermonde matrix with respect to $\mathbf{x}\in\mathbb{R}^n$.
$\det\big(\mathbf{V}(\mathbf{x})\big) = v_n(\mathbf{x})$	–	Determinant of the n– square Vandermonde matrix.
$\mathbf{V}_\delta(\mathbf{x}) = \mathbf{V}_n(\mathbf{x})$	–	Vandermonde matrix with respect to index δ and $\mathbf{x}\in\mathbb{R}^n$.
$\mathbf{V}_{\lambda+\delta}(\mathbf{x})$	–	Vandermonde matrix with respect to partition λ and $\mathbf{x}\in\mathbb{R}^n$.
$\det\big(\mathbf{V}_\delta(\mathbf{x})\big) = a_\lambda(\mathbf{x})$	–	Determinant of the Vandermonde matrix with respect to index δ.
$\det\big(\mathbf{V}_{\lambda+\delta}(\mathbf{x})\big) = a_{\delta+\lambda}(\mathbf{x})$	–	Determinant of the Vandermonde matrix with respect to partition λ.
$s_\lambda(\mathbf{x}) = a_{\delta+\lambda}(\mathbf{x})/a_{\delta+\lambda}(\mathbf{x})$	–	The Schur polynomial with respect to partition λ.
$C^k[\mathbb{K}]$	–	The continuous functions with k–th derivative on the field $\mathbb{K}$.
$\|\mathbf{x}\|_p = \left(\sum_{k=1}^n \|x_k\|^p\right)^{\frac{1}{p}}$	–	The p–norm of $\mathbf{x}\in\mathbb{R}^n$, where $p=2$ is the Euclidean norm.
S_n^p	–	The n–dimension p–sphere, $S_n^p(r) = \left\{\mathbf{x}\in\mathbb{R}^{n+1} \,:\, \sum_{k=1}^n \|x_k\|^{p+1} = r^{p+1}\right\}$.
$\|\cdot\|_{\mathrm{F}}$	–	The Frobenius–norm where $\|\mathbf{X}\|_{\mathrm{F}} = \left(\sum_{i=1}^m\sum_{j=1}^n \|x_{ij}\|^2\right)^{\frac{1}{2}} = \sqrt{\mathrm{tr}(\mathbf{A}^\top\mathbf{A})}$.
$\kappa(\mathbf{X}) = \|\mathbf{X}^{-1}\|\|\mathbf{X}\|$	–	The condition number of $\mathbf{X}$, where $\mathbf{X}^{-1}$ is inverse of $\mathbf{X}$.
$H_n(\cdot), P_n^{(\alpha,\beta)}(\cdot), L_n(x), P_n(\cdot)$	–	The Hermite, Jacobi, Legendre and Laguerre orthogonal polynomials.
$\Gamma(x),\beta(\cdot)$	–	The Gamma and Beta functions, $\Gamma(\alpha) = (\alpha-1)!$, $\beta(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$.
${}_2F_2(a,b;c;x)$	–	The hypergeometric function.
$\dfrac{d^k f}{dx^k} = f^{(k)}(x)$	–	The k–th derivative of the function f with respect to x.
$\dfrac{\partial^n f}{\partial x^n} = f^{(n)}(x)$	–	The n–th partial derivative of the function f with respect to x.
$\mathbb{P}(A)$	–	The probability of event A.
$\mathbb{E}(X), \mathrm{Var}(x) = V(X)$	–	The expectation and variance of random variable X respectively.

Contents

1 Introduction **23**

1.1 Historic Background 27

1.1.1 Vandermonde Matrix 37

1.1.2 Vandermonde Determinant 37

1.1.3 Generalized Vandermonde Matrix 40

1.1.4 Properties of Vandermonde Determinant 41

1.1.5 Relationship with other determinants 43

1.1.6 The Alternant Matrix 43

1.1.7 Calculus of the Vandermonde matrix and its Determinant 45

1.2 Vandermonde Determinant and Symmetric Polynomials 45

1.2.1 Symmetric Polynomials 45

1.2.2 LDU Decomposition of Vandermonde Matrix Using Symmetric Polynomials 51

1.2.3 General Properties of Vandermonde Determinant Based on Symmetric Polynomials . 55

1.2.4 Schur Polynomials 57

1.2.5 Properties of Schur Polynomials 57

1.3 Orthogonal Polynomials 62

1.3.1 Determinantal Representation of Orthogonal Polynomials 62

1.3.2 Vandermonde Determinant and the Christoffel–Darboux Formula 64

1.3.3 Basic Theory of Orthogonal Polynomials 65

1.4 Applications and Occurrences of the Vandermonde Matrix and its Determinant 71

1.4.1 Polynomial Interpolation 72

1.4.2 Fekete points 76
1.4.3 Divided Differences 80
1.4.4 Least Squares Fitting 83
1.4.5 Regression Analysis and Data Smoothing 84
1.4.6 D-Optimal Experimental Design 86
1.5 Random Matrix Theory 89
1.5.1 Overview of Random Matrix Theory 90
1.5.2 Univariate and Multivariate Normal Distribution 91
1.5.3 Wishart Distribution 93
1.5.4 Classical Random Matrix Ensembles 94
1.5.5 Gaussian ensembles 96
1.5.6 Distribution of Level Spacings 100
1.5.7 The Vandermonde determinant in systems with Coulombian interactions 101
1.6 Symmetric Cones and Jordan Algebras 104
1.6.1 Euclidean Jordan Algebras 105
1.6.2 The Cone of Positive Definite Symmetric Matrices 106
1.6.3 Properties and Examples of Jordan Algebras 108
1.6.4 Classification of Irreducible Symmetric Cones 110
1.6.5 Additional Properties 113
1.6.6 Trace, Determinant and Minimal Polynomials 113
1.6.7 Special Functions Defined on Symmetric Cones 114
1.6.8 Gaussian, Chi-Square and Wishart Distributions on Symmetric Cones 118
1.7 Vandermonde Matrix and Determinant in Financial Mathematics 122
1.7.1 Money Market Account 123
1.7.2 Derivatives and Arbitrage Pricing 125
1.7.3 Pricing Derivatives 126
1.7.4 Options 128
1.7.5 Optimization Model in Finance 130
1.8 Summaries of Chapters 135

CONTENTS

2 The Generalized Vandermonde Interpolation Polynomial Based on Divided Differences 141

2.1 Generalized Divided Differences and Vandermonde Determinant 141

2.2 Weighted Fekete Points and Fekete Polynomials 144

2.3 Weighted Lebegue Constant and Lebegue Function 146

2.3.1 Mean Convergence 147

2.3.2 Lebegue Function and Pointwise Convergence 147

2.4 The Optimization of Gaussian Ensembles as Weighted Fekete Points 148

2.5 Fitting Interpolating Polynomial to Experimental Data 148

3 Extreme Points of the Vandermonde Determinant on Surfaces Implicitly Determined by a Univariate Polynomial 153

3.1 Extreme points of the Vandermonde determinant on surfaces defined by a low degree univariate polynomial 153

3.1.1 Critical points on surfaces given by a first degree univariate polynomial 155

3.1.2 Critical points on surfaces given by a second degree univariate polynomial 155

3.2 Critical points on the sphere defined by a p-norm 158

3.2.1 The case p = 4 and n = 4 159

3.2.2 Some results for even n and p 162

3.3 Some results for cubes and intersections of planes 169

4 Symmetric Group Properties of Extreme Points of Vandermonde Determinant and Schur polynomials 177

4.1 Symmetric Group Properties of Vandermonde Matrix and its Determinant 177

4.2 Derivatives, Extreme Points of Vandermonde Determinants and Schur Polynomials 186

4.3 The extreme points of Schur Polynomial on certain surfaces 193

4.3.1 Weighted Schur Polynomials and their extreme points 197

4.4 The Extreme Points of Vandermonde Determinant, Schur Polynomial and Maximum Szegö Limit Theorem 204

4.5 Interpolation with Extreme Points of Schur polynomial 207

5 Optimization of the Wishart Joint Eigenvalue Probability Density Distribution Based on the Vandermonde Determinant 217

5.1 The Vandermonde Determinant and Joint Eigenvalue Probability Densities for Random Matrices . . . 217

5.2 Optimising the joint eigenvalue probability density function . . . 221

6 Properties of the Extreme Points of the Joint Eigenvalue Probability Density Function of the Wishart Type Matrix 229

6.1 Polynomial Factorization of the Vandermonde matrix and Wishart Matrix . . . 229

6.2 Matrix Norm of the Vandermonde and Wishart Matrices . . . 232

6.3 Condition Number of the Vandermonde and Wishart Matrix . . . 235

7 Connections Between the Extreme Points of Vandermonde determinants and minimizing risk measure in financial mathematics 241

7.1 Pricing with Extreme Points Vandermonde Determinant . . . 241

7.2 Optimum Value of Generalized Variance $\mathbb{V}[\hat{\boldsymbol{\beta}}]$ with Extreme Points of Vandermonde Determinant . . . 245

8 The Wishart Distribution on Symmetric Cones 257

8.1 The Wishart Ensembles on Symmetric Cones . . . 257

8.2 Lassalle Measure on Symmetric Cones and Probability Distribution . . . 261

8.3 Degenerate Wishart Ensembles on Symmetric Cones . . . 267

9 Extreme Points of the Vandermonde Determinant and Wishart Ensembles on Symmetric Cones 275

9.1 The Gindikin Set and Wishart Joint Eigenvalue Distribution . . . 275

9.2 A quick jump into Wishart distribution on symmetric cones . . . 276

9.3 Extreme Points of the Degenerate Wishart Distribution and Vandermonde Determinant . . . 279

List of Figures 335

List of Tables 338

List of Definitions 339

List of Theorems 342

List of Lemmas 345

Introduction

This chapter is based on Paper **A**, Paper **B**, Paper **C**, Paper **D**, Paper **F** and Paper **G**, and gives the general overview of the contents of Chapters 2, 3, 4, 5, 6 7, 8, and 9.

Paper A. Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. *The Generalized Vandermonde Interpolation Polynomial Based on Divided Differences*, SMTDA2018 Conference Proceedings, **ISAST2018**, 443–456, 2018.

Paper B. Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *Extreme points of the Vandermonde determinant on surfaces implicitly determined by a univariate polynomial*. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317, 791–818, 2020.
https://doi.org/10.1007/978-3-030-41850-2-33.

Paper C. Muhumuza Asaph K., Silvestrov Sergei, (2019). *Symmetric Group Properties of Extreme Points of Vandermonde Determinant and Schur polynomials*. Accepted for publication in: Sergei Silvestrov, Anatoliy Malyalenko, Milica Rančić M., (Eds.), Algebraic Structures and Applications, SPAS 2019. Springer Proceedings in Mathematics & Statistics 2019.

Paper D. Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Optimization of the Wishart Joint Eigenvalue Probability Density Distribution Based on the Vandermonde Determinant*. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317, 819–838, 2020.
https://doi.org/10.1007/978-3-030-41850-2-34.

Paper E. Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. *Properties of the Extreme Points of the Joint Eigenvalue Probability Density Function of the Wishart Matrix*. In ASMDA2019, 18th Applied Stochastic Models and Data Analysis International Conference. ISAST: International Society for the Advancement of Science and Technology (pp. 559–571), 2019.

Paper F. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *Connections Between the Extreme Points of Vandermonde determinants and minimizing risk measure in financial mathematics.* Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), (Eds.), SPAS2019. Algebraic, stochastic and analysis structures for networks, data classification and optimization, Springer Proceedings in Mathematics and Statistics, Springer International Publishing, 2020.

Paper G. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *The Wishart Distribution on Symmetric Cones.* Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), SPAS2019. Algebraic Structures and Applications, 2020.

Paper H. Muhumuza Asaph K., Lundengård Karl, Malyarenko Anatoliy, Silvestrov Sergei, Mango John M., Kakuba Godwin, (2019). *Extreme Points of the Vandermonde Determinant and Wishart Ensembles on Symmetric Cones.* Accepted in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Springer International Publishing, 2020.

Chapter 1

Introduction

The main fields discussed in this thesis include finding the extreme points of the Vandermonde determinant and their applications in both numerical computations, approximation and random fields. Several of the techniques and approaches that are explored are also applied to multivariate interpolation, random matrix theory and construction of Wishart ensembles using Symmetric cones in Jordan algebras. An overview of the major relations between the different parts of the thesis are illustrated in Figure 1.1. The relations are of many forms including important definitions, dependent results, and conceptual connections. In addition, similarities in proofs and mathematical techniques based on problem formulations will be discussed.

This thesis is based on the eight papers listed on pages 13 and 14, that is, Paper *A*, Paper *B*, Paper *C*, Paper *D*, Paper *E*, Paper *F*, Paper *G* and Paper *H*. The contents of the papers especially the introductory parts have been transferred to the main introduction, rearranged and/or in some cases parts have been omitted to avoid repetition and improve coherence. Generally, the original text and structure of the papers have been preserved.

Chapter 1 gives the general introduction of the major concepts to be used in the later chapters. The historic background of the Vandermonde matrix and its determinant, the structure and definition of Vandermonde matrix, Vandermonde determinant, generalized Vandermonde matrix, general properties of Vandermonde matrix and its determinant. These properties makes both the Vandermonde matrix and Vandermonde determinant more applicable in both scientific and mathematical computations. Section 1.1 outlines some relationships with other determinants while Section 1.2 is devoted to the group properties of the matrix and its Vandermonde determinant based on the symmetric polynomial, Schur polynomials and orthogonal polynomials, and how such are helpful in decomposition of Vandermonde matrix, Cauchy matrix, Hankel matrix and Toeplitz matrix.

Section 1.3 is concerned with the determinantal representation of orthogonal polynomials, the Vandermonde determinant and Christofell–Darboux formula. The classical orthogonal polynomials which include the Hermite, Laguerre, Jacobi and Legendre polynomials.

Section 1.4 introduces some of the occurrences and applications of the Vandermonde matrix and its determinant which include polynomials curve fitting, divided differences, regression analy-

sis, *D*-optimal experiment design and solutions of differential equation.

Section 1.5 discusses the applications of the Vandermonde matrix and its determinant in Random Matrix Theory (RMT), Gaussian ensembles, Wishart ensembles, distribution level spacing and the significance of the Vandermonde determinant in characterisation of certain systems including Coulombian interaction, distribution of electrical charges, sphere packing and Coulomb gas.

Section 1.6 introduces the Vandermonde determinant into Jordan algebras, Symmetric cones and random fields thus gives the general overview on Euclidean Jordan Algebras, the cone of positive definite symmetric matrices, properties and examples of Jordan algebras, classification of irreducible symmetric cones, trace, determinant and minimum polynomials on symmetric cones and the Laplace transform, special functions defined on Symmetric cones and the Wishart distribution on symmetric cones.

Section 1.7 highlights the occurrence of the Vandermonde matrix and its determinant in financial models that include money market accounts, derivatives and arbitrage pricing, derivative securities, options and many other optimization models in finance.

Chapter 2 discusses the general expression of divided differences using the generalized Vandermonde determinant. Section 2.1 describes the generalized divided differences, Section 2.2 introduces the weighted Fekete points and Section 2.3 gives the weighted Lebesgue constant and Lebesgue function as applies to error analysis in interpolation. Section 2.4 applies these concepts in establishing the connects the Gaussian orthogonal ensembles while Section 2.5 gives curve fitting criteria using the zeros of the Jacobi, Laguerre and Hermite polynomials.

Chapter 3 explores the extreme points of the Vandermonde determinant in higher dimension motivated by results of optimization of Vandermonde determinant by Lagrange multiplier explained in [305]. Section 3.1 discusses extreme points of the Vandermonde determinant on surfaces defined by a low degree univariate polynomials, Section 3.1.1 critical points on surfaces given by a first degree univariate polynomial, Section 3.1.2 critical points on surfaces given by a second degree univariate polynomial, Section 3.2 critical points on the sphere defined by a p-norm in which higher degree univariate polynomials are computed, and Section 3.3 extends the results of extreme points of Vandermonde determinants to cubes and intersections of planes.

Chapter 4 studies symmetric group properties of extreme points of Vandermonde determinant and Schur polynomials based on symmetric group properties of polynomial rings. Section 4.1 highlights the symmetric group properties of Vandermonde Matrix and its determinant, Section 4.2 gives the derivatives, extreme points of Vandermonde determinants and Schur polynomials, Section 4.3 discusses the extreme points of Schur Polynomial on certain surfaces and based on classical orthogonal polynomial. Section 4.4 extends the extreme points of generalized Vandermonde determinant and Schur polynomial to the Szegö Limit Theorems, and Section 4.5 gives the application of extreme points of Vandermonde determinant in interpolation with symmetric polynomials and Schur polynomials.

Chapter 5 discusses the optimization of the Wishart joint eigenvalue probability density distribution based on the Vandermonde determinant by use of the classical Lagrange multiplier. Section 5.1 describes optimization of the Vandermonde determinant and joint eigenvalue probability densities on certain surfaces. Section 5.2 gives an illustration of extremes the joint eigenvalue probability

density function on a p-sphere.

Chapter 6 studies the properties of the extreme points of the Joint Eigenvalue Probability Density Function of the Wishart type matrices based on the condition number of both the Vandermonde matrix and Wishart type matrices. Section 6.1 outlines polynomial factorization of the Vandermonde and Wishart Matrix, Section 6.2 discusses the concept of matrix norm of the Vandermonde and Wishart Matrices, and Section 6.3 illustrates evaluation of condition number of the Vandermonde and Wishart Matrix based on eigenvalues.

Chapter 7 investigates the connections between the extreme points of Vandermonde determinant and minimizing risk measure in financial mathematics. Section 7.1 derives the relevant expression of financial models while 7.2 demonstrates the concepts of pricing with extreme points Vandermonde Determinant on an an efficient frontier defined by certain surfaces.

Chapter 8 explores the group properties of the Wishart distribution on symmetric cones in Jordan algebras based on Vandermonde determinant. Section 8.1 outlines the Wishart Ensembles on symmetric cones, Section 8.2 gives the Lassalle measure on symmetric cones and probability distribution, and in Section 8.3 we construct the degenerate and non-degenerate Wishart ensembles on symmetric cones.

Chapter 9 demonstrates the extreme points of the Vandermonde determinant and Wishart ensembles on symmetric cones. Section 9.1 gives the Gindikin set and Wishart joint eigenvalue distribution, Section 9.2 discusses the Wishart distribution on symmetric cones, Section 9.3 illustrates the extreme points of the degenerate and non-degenerate Wishart distribution based on the Vandermonde determinant.

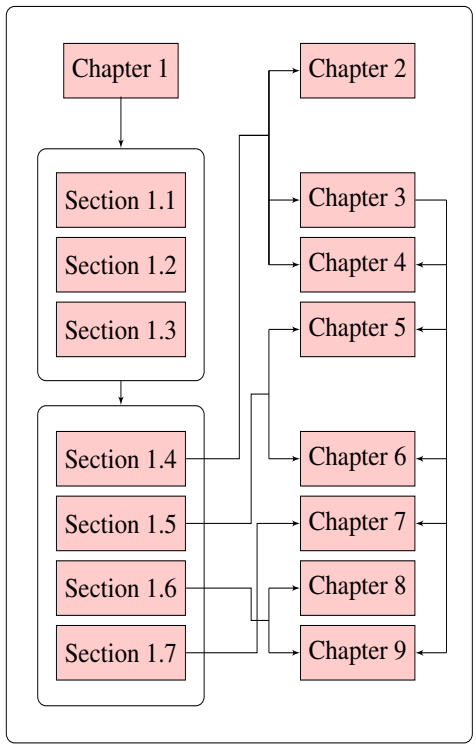


Figure 1.1: An illustration of the relationship between thesis sections and chapters.

1.1 Historic Background

The Vandermonde matrix and its determinant are considered to be a typical example of Stingler's law of eponymy [445], "No scientific discovery is named after its original discoverer." Henri Lebesgue (1875-1941), who gave a conference presentation, titled: "L'œuvre mathématique de Vandermonde." During his conference presentation he made the following assertion as stated in [293]:

> What could have been personal, is the Vandermonde determinant? Yet it is not there, nor anywhere else in Vandermonde's work. Why then was Vandermonde's name given to that determinant?

Lebesgue's conjecture downplays his other memoirs, for instance, he states in [293]:

> Thus, the Vandermonde determinant is not due to Vandermonde; his theory of determinants is not very original, his notations of factorial is unimportant; his study of situation geometry is somewhat childish, what is left? What is left is his first Memoir, about which Cauchy stated ...

The memoir that was referred to as "childish" was the memoir on combinatorics which contained more than just a notation for factorials, that is, the identity

$$\binom{n}{k} = \sum_{j=1}^{k} \binom{m}{j}\binom{n-m}{k-j} \tag{1.1}$$

is presently still referred to as "Vandermonde Theorem or Identity" in most probability and combinatorial textbooks, for example, see [412]. Even though referred to as "childish", the memoir on situational geometry made Vandermonde to be regarded as a foundation of knot theory as stated in [384]. Furthermore, it is known that the Vandermonde identity (1.1) was already known in 1303 by the Chinese mathematician Zhu Shijie, [15], and is expressed as

$$\binom{n_1+\cdots+n_p}{m} = \sum_{k_1+\cdots+k_p=m} \binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_p}{k_p}. \tag{1.2}$$

where (1.2) is a generalization of (1.1).

The matrix is named after Alexandre Théophile Vandermonde (1735-1796) whose life is characterised by his engagements in the French revolution, he had a varied career interests including music, mechanic and political economy, and his brief career in mathematics can be obtained in various literature including [154, 198, 222, 293, 448]. Vandermonde's career began with law studies and work as a concert violinist, transitioned into work in both science and mathematics in the early 1770s. He gradually turned into administrative and leadership positions at various Parisian institutions as well as work in politics and economics in the end of the 1780s [305]. By the end of 1980s, the name Vandermonde matrix and determinant had appeared in various research articles

and textbooks, for example, [3, 20, 40, 61, 66, 87, 110, 111, 119, 123, 141, 157, 179, 203, 210, 216, 221, 229, 251, 278, 282, 283, 284, 289, 316, 334, 354, 355, 364, 366, 373, 411, 431, 438, 445, 447, 451, 462, 487, 488, 490] to list but a few.

The Vandermonde notation for the determinant was based on a linear system of equations in the variables ζi in the j equation by $\overset{j}{i}$ so that the system [470],

$$\begin{aligned}
\overset{1}{1}\zeta 1+\overset{1}{2}\zeta 2+\overset{1}{3}\zeta 3+\overset{1}{4}=0,\\
\overset{2}{1}\zeta 1+\overset{2}{2}\zeta 2+\overset{2}{3}\zeta 3+\overset{2}{4}=0,\\
\overset{3}{1}\zeta 1+\overset{3}{2}\zeta 2+\overset{3}{3}\zeta 3+\overset{3}{4}=0,
\end{aligned}$$

will have the determinant such as

$$\begin{vmatrix}
\overset{1}{1} & \overset{1}{2} & \overset{1}{3}\\
\overset{2}{1} & \overset{2}{2} & \overset{2}{3}\\
\overset{3}{1} & \overset{3}{2} & \overset{3}{3}
\end{vmatrix}$$

where the upper indices represents the exponents, which would translate to the usual notation Vandermonde matrix if the upper indices are taken as a superscripts. The translation of exponents into indices gave birth to the foundation of Cauchy's theory of determinants [78, 79]. The same concept of switching the indices into superscripts had helped Vandermonde himself to the observation that changing one of the indices of the general determinant into an exponent leads to an alternating function [470]. This could have been one the key remarks that could have interested Cauchy and Jacobi in their works, and most probably that could have concretized the naming to Vandermonde determinant.

After thorough analysis of Cauchy's two memoirs [78, 79] that were read to the institute on November 30, 1812 and later published in 1815, T. Muir in his conclusion describes the respective roles Vandermonde and Cauchy as follows [355]:

> On looking back, however, at Cauchy's memoir as a whole, one cannot but be struck with admiration both at the quality and the quantity of its contents. Supposing that none of its theorems had been new, and that it had not even presented a single old theorem in a fresh light, the memoir would have been most valuable, furnishing, as it did, to the mathematicians of the time, an almost exhaustive treatise on the theory of general determinants. It is not too much to say, although it may come to many as a surprise, that the ordinary text-books of determinants supplied to university students of the present day do not contain much more of the general theory than is to be found in Cauchy's memoir of about eight years ago. One apparently trivial instrument, which Cauchy had not received from his predecessors and which he did not make for himself, viz, a notation for determinants whose elements had special values, is at the foundation of the whole difference between his treatise and those at present

> employed. When this want came to be supplied later on, the functions crept steadily into every use, and a fresh impetus was consequently given to the study of them. But if from the work of the said eight years all researches regarding special forms of determinants be left out, and all investigations which ended in mere rediscoveries or in rehabilitations of old ideas, there is a surprisingly small proportion left. If one bears this in mind, and recalls the fact, temporarily set aside, that Cauchy, instead of being a compiler, presented the subject from a perfectly new point of view, added many results previously not thought of, and opened up a whole avenue of fresh investigation, one cannot but assign to him the place of honour among all the workers from 1693 to 1812. It is, no doubt, impossible to call him, as some have done, the formal founder of the theory. This honour is certainly due to Vandermonde, who however, erected on the foundation comparatively little of a superstructure. Those who followed Vandermonde contributed, knowingly or unknowingly, only a stone or two, larger or smaller, to the building. Cauchy, relaid the foundation, rebuilt the whole, and initiated new enlargements; the result being an edifice which the architects of to-day may still admire and find worthy of study.

In [81] the concept of determinant has been described using the method of induction, which in most recent times is referred to as the Laplace's expansion or formula, [220]. Cauchy's method of defining determinant was quite different from the previous methods especially due to Laplace. According to Cauchy [78]:

> Let $a_1, a_2, \ldots, a_n$ be several different quantities in number equal to n. It has been shown that by multiplying the product of the quantities,
>
> $$a_1 a_2 a_3 \cdots a_n$$
>
> by the product of their respective differences,
>
> $$(a_2 - a_1)(a_3 - a_1)\cdots(a_n - a_1)(a_3 - a_2)\cdots(a_n - a_2)\cdots(a_n - a_{n-1}).$$
>
> One obtains the alternating symmetric polynomial
>
> $$S(\pm a_1 a_2^2 \cdots a_n^n),$$
>
> which, as a consequence, happens to be equal to the product
>
> $$a_1 a_2 a_3 \cdots a_n (a_2 - a_1)\cdots(a_n - a_1)(a_3 - a_2)\cdots(a_n - a_2)\cdots(a_n - a_{n-1}).$$
>
> Let us suppose now that one develops this later product and that, in each term of the development, one replaces the exponent of each letter by a second index equal to the exponent: by writing, for instance, $a_{r,i}$ instead of a_r^i and $a_{i,r}$ instead of a_i^r. One will

> obtain as a result a new alternating symmetric polynomial which, instead of being represented by
>
> $$S(\pm a_1^1 a_2^2 \cdots a_n^n),$$
>
> will be represented by
>
> $$S(\pm a_{1,1} a_{2,2} \cdots a_{n,n}).$$
>
> The sign S being relative to the first indices of each letter. Such is most general form of the functions that I shall designate in what follows under the same name of determinant.

To get a better understanding of Cauchy's argument, there is need to keep in mind that his major focus was on the functions of n variables defined in [78] which came as a result of his first memoir [79] where he discussed the functions of n variables that assume less than $n!$ different values whenever the variables are permuted. He referred to "symmetric alternating functions" as those functions assuming only two opposite values which also can be simply called alternating function. Among these polynomials of n variables included the "product of differences" which is also referred to as difference-product as stated in [355]. This difference product generates a sum of monomials with alternating signs that mainly depend on the permutations of the variables and their exponents. The "rule of signs" had already been described by Cauchy before defining determinants [75, 76, 77, 78, 79, 278].

There are three important polynomial functions of n variables say $a_1, a_2, \ldots, a_n$ that are algebraically equal but distinct in their formulation as described in detail in [78, 79, 355] these include

- difference-product: $\prod_{1 \le i < j \le n} (a_j - a_i)$;
- alternating polynomial: $\sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} \prod_{i=1}^{n} a_i^{\sigma(i)-1}$;
- Vandermonde determinant: $\det \left(a_i^j \right)_{1 \le i < j \le n-1}$.

These three functions are expressed in modern notations where S_n is the group of permutations of the finite set $\{a_1, a_2, \ldots, a_n\}$ onto itself and $\varepsilon(\sigma)$ is the signature or parity given by the number of inversions of the permutation σ. It should be noted that the group permutations and the signature as a homomorphism. In the development of the difference product it has been recognized that the same rule of signs as that of a general determinant allows one to express the idea using [78]

$$\prod_{i=1}^{n} a_i \prod_{1 \le i < j \le n} (a_j - a_i) = \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} \prod_{i=1}^{n} a_i^{\sigma(i)-1} \tag{1.3}$$

as a general definition by mutating the exponent of each variable into a second index.

The year 1841 was marked by unprecedented development in theory of determinant as clearly stated in [355]:

> Like in the year 1812 the year 1841 merits a chapter to itself, and in 1841 as in 1812, it is the work of only two authors that concern us. Strange to say, however, the two notable years has an author in common, the writers of 1812 being Binet and Cauchy, and those of 1841 being Cauchy and Jacobi. In 1841 Jacobi's contribution constituted a comprehensive monograph similar to that produced by Cauchy in 1812 and Cauchy's in 1841, as was to be expected, were more nature of the aftermath.

This general theory of determinant supplemented with the publication in Crelle's journal of Jacobi's monograph, that was split into three papers in which he rebuilds the whole theory and turning upside down Cauchy's approach. A brief recount is as exactly outlined in [355]:

> While Jacobi was aware, as we have already partly seen, of the labours of Cramer, Benzout, Vandermonde, Laplace, Gauss, and Binet, his main source of inspiration is Cauchy. Of all the writers since Cauchy's time, indeed, he is the first who gives evidence of having read and mastered the famous memoir of 1812. It scarcely needs to be said, however, that his own individuality and powerful grasp are manifest throughout the whole exposition.
>
> At the outset, there is reversal of former orders of things; Cramer's rule of signs for a permutation and Cauchy's rule being led up by series of propositions instead of one them being made a convention or definition. This implies, of course, that a new definition of a signed permutation is adopted, and that conversely this definition must have appeared as a deduced theorem in any exposition having either of this rules as its starting point.

In brief, Cauchy used the approach of difference-products to define the determinant in which he transformed the exponents used in the Vandermonde determinant into indices, whereas Jacobi used the approach of first defining positive and negative permutations, and then defined the determinant as a polynomial carrying the $\pm$ coefficient assigned according to the sign of the corresponding permutation. This Jacobi's approach to defining the determinant picked up and suppressed the Cauchy's approach. On the whole, Cauchy's approach addressed both pedagogical and mathematical concepts as expressed in his famous 1821 memoir, "Cours d'Analyse", [77], in which he recommended the difference-product approach to determinant as a general and suitable method for solving linear system of equations. He immediately applied his recommendations to Lagrange's interpolation problem as described in [77, 78, 79, 137].

In his third memoir published in Crelle's Journal, Jacobi deals mainly with alternating functions [251]. In response to this, Cauchy responded in 1841, in which he treated the quotients of alternation functions by difference-product as described in [76]. As a matter of fact, when comparing Cauchy's method of calculating $\det\left(\frac{1}{a_i+b_j}\right)_{1\leq i,j\leq n}$ described in [76] to the modern method

of calculating the same determinant, it can be confirmed that Cauchy had a point. However, the name "Cauchy's determinant" for that particular example is hardly heard of or used outside France, whereas for the particular case in which $a_i = i, b_j = j - 1$ is a universally known as "Hilbert matrix".

Importantly to note also is that one year before 1841, the approach of difference-product to defining determinant had just been established by James Joseph Sylvester, who without reference to Cauchy called "zeta-ic multiplication" Cauchy's operation of transforming the exponents into indices in a polynomial [451]. Based on this, Muir's remark on this is quite fascinating [355]:

> This early paper, one cannot but observe, has all the characteristics afterwards so familiar to readers of Sylvester's writings, fervid imagination, vigorous originality, bold exuberance of diction, hasty if not contemptuous disregard of historical research, the outstripping of demonstration of enunciation, and an infective enthusiasm as to the vistas opened by his own work.

Thus, based on the above brief enlisting, the name Vandermonde matrix or determinant does not appear in any of the Vandermonde's research published works, which is not surprising considering that the modern matrix concept did not really take a definite shape until almost a hundred years later following the works of Sylvester and Cayley [452]. It is therefore strange that the Vandermonde matrix was named after him, a thorough discussion on this is found in [2, 355], but a possible reason is the simple formula for the determinant that Vandermonde briefly discusses in his fourth paper that can be generalized to a Vandermonde matrix of any size.

Alexandre Théophile Vandermonde (1735-1796) published only four papers in his entire but rather short mathematical career. Whereas his papers can be said to contain very important ideas of mathematical concepts, they do not qualify any of them to great scientific maturity and therefore the very reason he is considered as minor scientist and mathematician, if compared to the well-known scientists and mathematicians namely Étienne Bézout (1730-1783) and Pirre-Simon de Laplace (1749-1827) as well as chemist Antoine Lavoisier (1743-1794) that he mainly worked with for some time after his mathematical career [395].

Vandermonde's writings can be strongly linked to the Vandermonde matrix or determinant more especially his "memoir on elimination", in which in his own words says [470]:

> This memoir was read to the Academy for the first time on the 20$^{\text{th}}$ of January 1771. It contained different things that I have suppressed here because they have been published since by other Geometers.

Here it is presumed "other geometers" definitely include Laplace's memoir though posterior [290], was published in the same volume as Vandermonde. Thus, what Vandermonde exactly suppressed remains conjectural.

In a similar way like Cauchy in 1812, Vandermonde wrote determinants being defined as a by-product of symmetric polynomials, especially his memoir on elimination is a sequel to the memoir on the solution of equations. It should be noted however, the publication dates, 1774 and 1776, are quite misleading: The first memoir [470] was read to the academy "sometime in November 1770",

that is, just only two months before the second [470]. Presumably, Vandermonde undoubtedly had the first memoir in mind when he wrote the second, thereby it is important to examine the two as a whole. Here is a summary of Vandermonde's four publications.

The first paper [469], " Mémoire surla résolution de équations," mainly discusses properties of the roots of polynomial equations, more in particular the formulas for the sum of the roots and a sum of symmetric polynomials of the powers of the roots. This paper is considered very important since it contains some of the fundamental ideas of group theory though this work generally this work is overshadowed by the works of the contemporary Joseph Louis Lagrange (1736-1813) [285, 286]. As part of the results in his first paper, he notices the most important equality

$$a^2b+b^2c+ac^2-a^2c-ab^2-bc^2=(a-b)(a-c)(b-c),$$

which is a special case of the formula for determinant of the Vandermonde matrix and the corresponding squares as

$$\begin{aligned}&a^4b^2+a^4c^2+b^4c^2+c^4a^2+c^4b^2-2(a^4bc+b^4ac+c^4ab)-2(a^3b^3+a^3c^3+b^3c^3)\\&+2(a^3b^2c+a^3c^2b+b^3a^2c+c^3a^2b+c^3b^2a)-6a^2b^2c^2=(a-b)^2(a-c)^2(b-c)^2.\end{aligned}$$

It seems that Vandermonde did not understand the significance of these expressions as pointed out by Stoy Reed Irving [395].

The second paper [468], "Remarques sur des problémes de situation," mainly discusses the problem of the knight's tour (what sequence of moves allows a knight to visit all squares on a chessboard exactly once). This paper is considered the first mathematical paper that uses the basic ideas of what is now called *knot theory* as stated in [59, 352, 354, 355, 356] where it is stated [355]:

> Those acquainted with the abbreviated symbols that I have named *partial types of combination*, in *Memoir on the resolution of equations*, will recognize here the formation of the *partial type* depending on the second degree, for any number of letters; they will easily see that, by taking our $\alpha,\beta,\gamma,\delta,\&c.$, for instance, as exponents, all terms with equal signs in the development of one of our abbreviations, which also be the development of the *partial type* depending on the second degree, & formed with an equal number of letters.

The third paper [467], "Mémoire sur des irrationnelles de différent ordres avecune application all cercles," is a paper on combinatorics and the most well - known result from the paper is the Zhu–Vandermonde identity

$$\sum_{k=1}^{n}\left[\left(\prod_{j=1}^{k}\frac{r+1-j}{j}\right)\left(\prod_{j=1}^{n-k}\frac{s+1-j}{j}\right)\right]=\left(\prod_{j=1}^{n-k}\frac{r+s+1-j}{j}\right)$$

where $r,s\in\mathbb{R}$ and $n\in\mathbb{Z}$. This same identity was first discovered by Zhu Shijie in 1303 and was later rediscovered by Vandermonde [15, 362].

In the fourth paper [466], "Mémoire sur l'élimination", in which Vandermonde discusses a general method for solving linear system using alternating functions, which has strong relation to determinants. He also notices that exchanging exponents for indices in a class of expressions from his first paper will give a class of expressions that he discusses in his fourth paper. This relation is mirrored in the relationship between the determinant of the Vandermonde matrix and the determinant of a general matrix [505].

In these four main papers of Vandermonde that outlines key of his scientific and mathematical work can be seen to contain many important ideas but these do not bring any of them to maturity and he is therefore usually considered a minor scientist and mathematician compared to well known contemporary mathematicians such as Étienne Bézout (1730–1783)[6] and Pierre-Simon de Laplace (1749–1827) [97] or scientists such as the chemist Antoine Lavoisier (1743–1794)[124] that he worked with for some time after his mathematical career. The Vandermonde matrix does not appear in any of Vandermonde's published works, which is not surprising considering that the modern matrix concept did not really take shape until almost a hundred years later in the works of James Joseph Sylvester (1814-1897) [452], Arthur Cayley (1821-1895) [80] and William Rowan Hamilton (1805-1865)[400]. It is therefore strange and remains a very big point of contention that the Vandermonde matrix was named after him, a detailed discussion on this can be found in [352, 354, 355, 356, 505], but a possible reason is the simple formula for the determinant that Vandermonde briefly discusses in his fourth paper can be generalized to a Vandermonde matrix of any size. One of the main reasons that the Vandermonde matrix has become known is that it has an exceptionally simple expression for its determinant that in turn has a surprisingly fundamental relation to the determinant of a general matrix [305].

The first direct and historic application of the difference-product which is mainly attributed to Cauchy, other than being a natural way of combining say n variables, it mainly appears in the Lagrange interpolation problem. The main historical background of interpolation can be got from [171]. If for instance, $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ are the usual Cartesian coordinates representing the points to be interpolated with $Q = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$ as the unknown polynomial to be fitted. Then, the coefficients $a_0, a_1, a_2, \ldots, a_{n-1}$ satisfy the system [101]:

$$\begin{aligned} a_0 + a_1x_1 + a_2x_1^2 + &\quad \ldots \quad + a_{n-1}x_1^{n-1} = y_1 \\ a_0 + a_1x_2 + a_2x_2^2 + &\quad \ldots \quad + a_{n-1}x_2^{n-1} = y_2 \\ &\quad \ldots \\ a_0 + a_1x_n + a_2x_n^2 + &\quad \ldots \quad + a_{n-1}x_n^{n-1} = y_n \end{aligned} \tag{1.4}$$

If x_i's are assumed to be distinct, the solution is given by the Lagrange interpolation polynomial [101]:

$$Q(x) = \sum_{i=1}^{n} y_i \prod_{i \neq j} \frac{x - x_j}{x_j - x_i}. \tag{1.5}$$

It would be really fair that whoever first wrote the linear system (1.4) should have gotten the credit of discovering the Vandermonde matrix. For instance, the name "Lagrange interpolation" came

from one of the famous lessons that Joseph Louis Lagrange (1736-1813) gave at the École Normale in Paris in 1795 [286]. Indeed, in his lecture, Lagrange did not pretend to expose his research as he stated:

> Newton is the first one who has posed that problem. Here is the solution he gives.

The fact remains that in the *Principia Mathematica*, Isaac Newton (1642-1727) had described a method to determine "a curved line of a parabolic which passes through any number of points" as stated in [366], as what is famously known as Newton's divided differences method. Isaac Newton's major breakthrough came in June 1661, when he was admitted to Trinity College, Cambridge as a sizar, a sort of work-study role [489]. At that time, the college's teachings were based on those of Aristotle, whom Newton supplemented with modern philosophers such as Descartes and astronomers such as Copernicus, Galileo, and Kepler. In 1665, he discovered the generalised binomial theorem and began to develop a mathematical theory that later became infinitesimal calculus. Soon after Newton had obtained his degree in August 1665, the University closed down as a precaution against the outbreak of a Great Plague called Bubonic plague in England. Although he had been undistinguished as a Cambridge student [333], Newton's private studies at his home in Woolsthorpe Manor over the next two years saw the development of his theories on calculus, optics and the law of gravitation. For instance, while sitting at home in the garden there one day, he saw an apple fall from a tree, providing him with the inspiration to eventually formulate his law of universal gravitation. Newton later relayed the apple story to William Stukeley, who included it in a book, "Memoir of Sir Isaac Newton's Life," published in 1752. In 1667 he returned to Cambridge as a fellow of Trinity [365].

According to the *Principia*, Isaac Newton did not explicitly write the system (1.4), however, in a famous letter to to Oldenburg dated October 24, 1676, he mentions a manuscript, *"Methodus differentialis"*, that appear in only after the *Principia*, in 1711. The system (1.4) is explicitly written in [171], where the *"Methodus differentialis"* is reproduced and translated, even though the explicit solution of the system (1.4) also is not given. It may be assumed that the system (1.4) might have been useless and even misleading to Newton. Probably, he was aware that his method was both much faster and numerically stable than the direct application of the system (1.4). It was later established that the first to explicitly write the system (1.4) was Newton's friend Abraham de Moivre (1667-1754), in 1730, for details see "de Moivre's relationship with Newton" in [33]. Unlike Lagrange whose interest in the system (1.4) was interpolation as in (1.5), de Moivre's main motivation was to calculated the coefficients in a linear combination of geometric series, when that linear combination is supposedly equals to another series, the coefficients of which turn out to be the solution of a system equivalent to (1.4). In *Miscellanea analytica*, de Moivre explicitly writes a general system with power coefficient, and gives a solution, thus being the one to prove the non-singularity of the Vandermonde matrix by computing its inverse [110]. It is also know that de Moivre had already published particular cases of the same result in the first edition of his "Doctrine of chances" [111].

The main reasons that the Vandermonde matrix has become very much known and applicable is that it has an exceptionally simple and/or user friendly expression, referred to as difference-product

[78], for its determinant that in turn has a surprisingly fundamental relation to the determinant of a general matrix. This qualifies it to be one the major numerical applications in the evaluation of the Lebesgue constant which is a valuable numerical instrument for linear interpolation and approximation because it provides a measure of how close the interpolant of a function is to the best polynomial approximant of the function [101]. Moreover, if the interpolant is computed by using the Lagrange basis, then the Lebesgue constant also expresses the conditioning of the interpolation problem. In addition, many publications have been devoted to the search for optimal interpolation points in the sense that these points lead to a minimal Lebesgue constant for the interpolation problems on the interval $[-1, 1]$ as for instance contained in [402].

The Vandermonde matrix and its corresponding determinant occurs mainly in many applications both in mathematics and science which include polynomial interpolation [402], least square regression [59], optimal experiment design [395], calculation of discrete Fourier transform [30], solving systems of differential equations with constant coefficients [341], random matrix theory [281, 327], modelling network of synaptic connections between neurons in the brain as applies to neural networks or neuroscience that can help to construct dynamical models based on random connectivity matrix [435], and in recent times financial modelling especially risk models and time series [19].

The extreme points of Vandermonde determinant especially if optimized over various quadratic surface have been proven to be of great significance in modelling scientific phenomena [454]. The optimization of the Vandermonde determinant over various surfaces in a finite–dimensional has been extensively investigated, for instance, see [305, 308]. For this particular case, the extreme points of the Vandermonde determinant are found to be equivalent to roots of the rescaled classical Hermite polynomials and explicit expressions are given for dimensions three up to seven. These extreme points can visualized in three to seven dimensions by using symmetries of the results by projecting all the extreme points onto a two-dimensional plane [308]. These results can be extended to the values of the Vandermonde determinant optimized over such surfaces like the ellipsoid and cylinder in three or higher dimensions. This can be achieved using the method of Lagrange multipliers to find a system of polynomial equations which give the local extreme points as its solutions. Also, using Gröbner basis and other techniques, the extreme points can be either explicitly obtained.

Lagrange interpolation is a classical method for approximating a continuous function by a polynomial that agrees with the function at a number of chosen points (the "nodes"). However, the accuracy of the approximation is greatly influenced by the location of these nodes. Now, a useful way to measure a given set of nodes to determine whether its Lagrange polynomials are likely to provide good approximations is by means of the Lebesgue constant [381, 454]. A brief survey of methods and results for the calculation of Lebesgue constants for some particular node systems is presented mainly based on classical polynomials is also contained in [433]. These ideas were then discussed in the context of Hermite–Fejér interpolation and a weighted interpolation method where the nodes are zeros of Chebyshev polynomials of the second kind.

1.1.1 Vandermonde Matrix

The Vandermonde matrix is such a famous matrix and assumes different structures in various circumstances as will be explained in many examples which may include polynomial interpolation, least squares curve fitting, optimal experiment design, construction of error-detecting and error-correcting codes as discussed in [50, 234, 248, 394]. Other fields of occurrence include financial mathematics in determining if a market with a finite set of traded assets is complete [318], calculating the discrete Fourier transform [391] and related transforms such as the fractional discrete Fourier transform [231], the quantum Fourier transform [117], and the Vandermonde transform [29, 30], solving systems of differential equations with constant coefficients [232], various problems in mathematical physics [2, 472], nuclear physics [90, 403], random matrix theory [18, 117, 167, 327], description of properties of the Fisher information matrix of stationary stochastic processes [276] and many others field as discussed in detail in [305, 308].

The Vandermonde matrix is generally defined as follows [308]:

Definition 1.1.1. *A Vandermonde matrix is an $m \times n$ matrix of the form,*

$$\mathbf{V}_{mn}(\mathbf{x}_n) = [x_j^{i-1}]_{i,j=1}^{m-1,n} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{bmatrix} \tag{1.6}$$

where $x_i \in \mathbb{R}(or\ \mathbb{C}), i = 1, \cdots, n$. If the matrix is square, $m = n$, the notation $\mathbf{V}_n = \mathbf{V}_{mn}$ is always adopted for square $n \times n$ Vandermonde matrix.

1.1.2 Vandermonde Determinant

It is important to note that quite often it is not the Vandermonde matrix itself that is very useful, instead it is the multivariate polynomial given by its determinant that is examined and used especially when expressed as difference-product [78]. The determinant of the matrix (1.6) is referred to as the Vandermonde determinant, Vandermonde polynomial, or simply Vandermondian as can be stated in [2, 454, 472] and can be written using an exceptionally simple formula given by [308].

Before we state definition of Vandermonde determinant, we first give the formal definition of determinant [16, 289, 353]:

Definition 1.1.2. *Let $\mathbb{F}$ be a field, the determinant is a function,* det $: M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ *of a square matrix* $\mathbf{M}$ *such that*

$$\det(\mathbf{M}) = \det(\mathbf{M}_{\cdot,1}, \mathbf{M}_{\cdot,2}, \ldots, \mathbf{M}_{\cdot,n})$$

where $\mathbf{M}_{\cdot,j}, j = 1, \ldots, n$ are the column vectors of all elements of the j-th colum of M.

The determinant function has the following properties:

- The determinant must be multi-linear

$$\det(\mathbf{M}_{\cdot,1},\ldots,a\mathbf{M}_{\cdot,k}+b\mathbf{N}_{\cdot,k},\ldots,\mathbf{M}_{\cdot,n}) \\ = a\det(\mathbf{M}_{\cdot,1},\ldots,a\mathbf{M}_{\cdot,k},\ldots,\mathbf{M}_{\cdot,n})+b\det(\mathbf{M}_{\cdot,1},\ldots,a\mathbf{N}_{\cdot,k},\ldots,\mathbf{M}_{\cdot,n}).$$

- The determinant must be alternating, that is, if $\mathbf{M}_{\cdot,i}=\mathbf{M}_{\cdot,j}$ for some $i\neq j$, then $\det(\mathbf{M})=0$.
- If $\mathbf{I}$ is the identity matrix, then $\det(\mathbf{I})=1$.

Theorem 1.1.1 ([296] **Leibnitz formula for determinants**). *A standard result from linear algebra says that the determinant is unique and that it is given by the following formula*

$$\det(\mathbf{M})=\sum_{\sigma\in S_n}(-1)^{\mathrm{sgn}(\sigma)}\prod_{i=1}^{n}m_{i,\sigma(i)} \tag{1.7}$$

where S_n is the set of permutations of the set $\{1,2,\ldots,n\}$, that is lists that contain the number $1,2,\ldots,n$ exactly once, and if σ is a permutation, then $\sigma(i)$ is the element of that permutation.

Remark 1.1.2. *It is important to note here that formula* (1.7) *is often used as the definition of determinant of a matrix as discussed in [16]. The formula* (1.7) *is mainly attributed to Gottfried Wilhem Leibniz (1646-1716) established in 1693 when he described a method of solving a system of linear equations by method closely related to that of Cramer's rule [352, 356], the particular letter of this result was published as [296] and its translation can be found in [430, 432].*

Thus using the ideas of Cauchy [78] expressed in (1.3) and Leibnitz [296] given in (1.7) we can give the Vandermonde determinant formula which is also given in [454]:

Theorem 1.1.3. *The Vandermonde determinant, $v_n(x_1,\ldots,x_n)$, is given by*

$$v_n(x_1,\ldots,x_n)=\prod_{1\leq i<j\leq n}(x_j-x_i). \tag{1.8}$$

The expression on the right hand side of (1.8) *is referred to as a difference-product and contains $\sum_{i=0}^{n-1} i=\frac{1}{2}n(n-1)$ factors.*

There are different versions of the proof of this theorem and can be obtained by both method of elementary row (or column) operation and combinatorial techniques, for instance, for detailed discussion of these, see [33, 36, 63, 75, 238, 247, 279, 337, 386, 454], but here we provide a more simplified proof based on direct decomposition by deletion of a given row. Here under we give a detailed outline of proof of Theorem 1.1.3 based mainly on induction with a traditional matrix of the same size [472]:

Proof. Considering two matrices $\mathbf{V}^\top$ and $\mathbf{T}$, such that

$$\mathbf{V}^\top = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 1 & -x_n & 0 & \dots & 0 & 0 \\ 0 & 1 & -x_n & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \ddots & -x_n & 0 \\ 0 & 0 & \dots & \dots & 1 & -x_n \end{bmatrix}.$$

Then,

$$\mathbf{VT} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \dots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} 1 & -x_n & 0 & \dots & 0 & 0 \\ 0 & 1 & -x_n & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \ddots & -x_n & 0 \\ 0 & 0 & \dots & \dots & 1 & -x_n \end{bmatrix},$$

it follows that

$$\begin{aligned} \mathbf{VT} &= \begin{bmatrix} 1 & x_1 - x_n & x_1^2 - x_1 x_n & \dots & x_1^{n-1} - x_1^{n-2} x_n \\ 1 & x_2 - x_n & x_2^2 - x_2 x_n & \dots & x_2^{n-1} - x_2^{n-2} x_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n-1} - x_n & x_{n-1}^2 - x_{n-1} x_n & \dots & x_{n-1}^{n-1} - x_{n-1}^{n-2} x_n \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & x_1 - x_n & x_1(x_1 - x_n) & \dots & x_1^{n-2}(x_1 - x_n) \\ 1 & x_2 - x_n & x_2(x_2 - x_n) & \dots & x_2^{n-2}(x_1 - x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n-1} - x_n & x_{n-1}(x_{n-1} - x_n) & \dots & x_{n-1}^{n-2}(x_{n-1} - x_n) \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n-1} & \mathbf{Y} \\ 1 & \mathbf{0} \end{bmatrix}, \end{aligned}$$

where

$$\mathbf{Y} = \begin{bmatrix} x_1 - x_n & x_1(x_1 - x_n) & \dots & x_1^{n-2}(x_1 - x_n) \\ x_2 - x_n & x_2(x_2 - x_n) & \dots & x_2^{n-2}(x_1 - x_n) \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} - x_n & x_{n-1}(x_{n-1} - x_n) & \dots & x_{n-1}^{n-2}(x_{n-1} - x_n) \end{bmatrix}.$$

The matrix $\mathbf{Y}$ can further be expressed as a product of a diagonal matrix $\mathbf{D}$ and submatrix $\mathbf{W} = \mathbf{V}_{n-1}$ of $\mathbf{V}_n$, that is

$$\mathbf{Y} = \begin{bmatrix} x_1 - x_n & 0 & \dots & 0 \\ 0 & x_2 - x_n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_{n-1} - x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 & \dots & x_1^{n-2} \\ 1 & x_2 & \dots & x_2^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n-1} & \dots & x_{n-1}^{n-2} \end{bmatrix} = \mathbf{DW}.$$

Applying the properties of matrix determinant, that is, the property of determinant of block diagonal matrix and the product of matrices. Thus, it follows that

$$\begin{aligned}\det(\mathbf{V}) = |\mathbf{V}| &= \begin{vmatrix} \mathbf{1}_{n-1} & \mathbf{Y} \\ 1 & \mathbf{0} \end{vmatrix} = (-1)^{n-1}|\mathbf{Y}| = (-1)^{n-1}|\mathbf{D}||\mathbf{W}| = |-\mathbf{D}||\mathbf{W}| \\ &= (x_n - x_1)(x_n - x_2)\dots(x_n - x_{n-2})|\mathbf{W}|. \end{aligned} \tag{1.9}$$

The expression in (1.9) relates the determinant of an $n \times n$ Vandermonde matrix to that of an $(n-1) \times (n-1)$ Vandermonde matrix. By repeating the above decomposition procedure, allows us to completely evaluate the determinant of any Vandermonde matrix. □

For example, it can be noticed that when $n = 2$, that is, $\mathbf{V}$ is a 2×2 Vandermonde matrix, then gives $\det(\mathbf{V}) = x_2 - x_1$; when $n = 3$, that is, $\mathbf{V}$ is a 3×3 Vandermonde matrix, then gives $\det(\mathbf{V}) = (x_3 - x_1)(x_3 - x_2)(x_2 - x_1)$; and in general,

$$\begin{aligned}\det(\mathbf{V}) = \prod_{i<j}(x_j - x_i) &= (x_n - x_1)(x_n - x_2)\cdots(x_n - x_{n-1}) \\ &\cdot (x_{n-1} - x_1)(x_{n-1} - x_2)\cdots(x_{n-1} - x_{n-2})\cdots(x_3 - x_1)(x_3 - x_2)(x_2 - x_1). \end{aligned} \tag{1.10}$$

The final expression (1.10) can easily be verified by simple mathematical induction argument based on the decomposition relation in (1.9).

Its also clear from (1.10) that $\det(\mathbf{V}) = |\mathbf{V}| \neq 0$ if and only if $x_j \neq x_i$ for all $i < j, i, j = 1, 2, \dots n$. Thus $\mathbf{V}$ is a non-singular matrix if and only if the n scalars $x_1, x_2, \dots, x_n$ are distinct.

Related to Theorem 1.1.3 above, if $\mathbf{V}_n(\mathbf{x})$ is the generalized Vandermonde matrix of the structure

$$\tilde{\mathbf{V}}_n(\mathbf{x}_n) = [x_j^{i-1}]_{i,j=1}^{m,n-1} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{bmatrix} \tag{1.11}$$

then, the determinant of $\tilde{\mathbf{V}}_n(\mathbf{x})$ is given by

$$\prod_{1 \leq j \leq n} x_j \prod_{1 \leq i < j \leq n} (x_j - x_i) \tag{1.12}$$

1.1.3 Generalized Vandermonde Matrix

Many generalizations of the Vandermonde matrix have been proposed and studied in literature [69, 102, 165, 257, 259, 299, 309, 434, 472]. One of the typical examples is the confluent Vandermonde matrix, also referred to as generalized Vandermonde matrix. By definition, the generalized Vandermonde determinant can be expressed as:

Definition 1.1.3. *A generalized Vandermonde matrix is an $m \times n$ matrix of the form*

$$G_{mn}(\mathbf{x}_n) = \left[x_j^{\alpha_i}\right]_{i,j}^{m,n} = \begin{bmatrix} x_1^{\alpha_1} & \cdots & x_n^{\alpha_1} \\ x_1^{\alpha_2} & \cdots & x_n^{\alpha_2} \\ \vdots & \ddots & \vdots \\ x_1^{\alpha_m} & \cdots & x_n^{\alpha_m} \end{bmatrix} \tag{1.13}$$

where $x_i \in \mathbb{C}, \alpha_i \in \mathbb{C}, i = 1, \cdots, n$.

The generalized Vandermonde matrices [224] and their determinants are directly connected to the Schur polynomials [483], and these texts and literature [107, 108, 109, 152, 173, 275, 414, 504] are devoted to the study of the same.

1.1.4 Properties of Vandermonde Determinant

In this section, we highlight some algebraic varieties of the Vandermonde determinant. Algebraic varieties are the central objects of study in algebraic geometry. In principle, an algebraic variety is defined as the set of solutions or zero locus of a system of polynomial equations over the real or complex numbers [98, 219]. Since the extreme points of the Vandermonde determinant are closely related to the zeros of polynomials as will be presented. Thus, it is important to look at some properties of the Vandermonde determinant based on geometric curves and surfaces which are determined by a system of polynomial equations such as sphere, ellipsoid, paraboloid, and hyperboloid. These properties will help us to explore a variety of applications Vandermonde determinant and its extreme points.

Singularity

If $m < n$, the matrix $\mathbf{V}_n(\mathbf{x})$ has rank (m) if and only if all m_j are distinct. A square Vandermonde matrix is thus invertible if and only if the m_i are distinct [305]. An explicit formula for the inverse is known to be equivalent to the columns of the Lagrange basis polynomial [314]. The inverse of the Vandermonde matrix is of great importance especially in solving linear systems. According to [45, 145, 231, 232, 369, 370, 396, 454, 460], the inverse of the Vandermonde matrix can be expressed as follows:

Theorem 1.1.4. *The elements of the inverse of an $n-$dimensional Vandermonde matrix $\mathbf{V}_n(\mathbf{x})$ can be calculated by*

$$(\mathbf{V}_n^{-1})_{ij} = \frac{(-1)^{j-1}\sigma_{n-j,i}}{\prod\limits_{\substack{i=1 \\ k \neq i}}^{n}(x_k - x_i)}, \tag{1.14}$$

where $\sigma_{j,i}$ is the j-th elementary symmetric polynomial with variable x_i set to zero, that is,

$$\sigma_{j,i} = \sum_{1\le m_1\le m_2<\cdots<m_j\le n}\prod_{k=1}^{j} x_{m_k}(1-\delta_{m_k,i}),\ \ \delta_{a,b} = \begin{cases} 1, & a=b \\ 0, & a\neq b. \end{cases} \tag{1.15}$$

There are various versions of the proof of this theorem especially one proposed by [305, 308]. Alternatively, the inverse of the Vandermonde matrix can also be expressed by [145]:

Theorem 1.1.5. *Let $\mathbf{V}_n(\mathbf{x})$ be a Vandermonde matrix and $\mathbf{W}_n(\mathbf{x})$ be its inverse. Then, the generic element $w_n(i,j)$ of $\mathbf{W}_n(\mathbf{x})$ is:*

$$w_n(i,j) = \phi(n,j)\psi(n,i,j),\ \ i,j = 1,2,\ldots,n, \tag{1.16}$$

where $\phi(n,j)$ are functions recursively defined as

$$\phi(n+1,j) = \frac{\phi(n,j)}{x_{n+1}-x_j},\ \ j=1,2,\ldots,n,$$
$$\phi(n+1,n+1) = \prod_{k=1}^{n}\frac{1}{x_{n+1}-x_j},$$

$\psi(n,i,j)$ is expressed in as

$$\psi(n,i,j) = (-1)^{i+1}\sum_{r=0}^{n-i}(-1)^r x_j^r \sigma(n,n-i-r),\ \ i,j=1,2,\ldots,n,$$

for all

$$\begin{cases} \sigma(n,j) &= 0,\ \ if\ \ (j<0)\vee(n<0)\vee(j>n), \\ \sigma(n,0) &= 1,\ \ if\ \ n=0,1,2,3,\ldots, \\ \sigma(n,j) &= \sigma(n-1,j)+x_n\sigma(m-1,s-1),\ \ n,j\ \ are\ integers. \end{cases}$$

We also note that $\sigma(n,j)$ is the jth order elementary symmetric polynomial associated to the set $x_1,x_2,\ldots,x_n$, that is, the sum of all products of j distinct nodes chosen from space X of dimension n and these can be expressed as

$$\sum_{r=0}^{n}(-1)^r x_i\sigma(n,n-r) = 0,\ \ i=1,2,\ldots,n,$$
$$\sum_{r=0}^{n}(-1)^r x_{n+1}\sigma(n,n-r) = (-1)^n\prod_{j=1}^{n}(x_{n+1}-x_j).$$

1.1.5 Relationship with other determinants

The Leibniz formula, see Theorem 1.1.1, for the determinant which is given by

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^{\operatorname{sgn}(\sigma)} \prod_{i=1}^{n} m_{i,\sigma(i)}, \tag{1.17}$$

where $\prod_{i=1}^{n} m_{i,\sigma(i)}$ is an elementary product of the entries in the ith row and jth column of $\mathbf{M}$ and is of the form $a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$, S_n is the set of permutations over the set $\{1,2,\ldots,n\}$, that is, all lists that contains the numbers $1,2,\cdots,n$ exactly once, and if σ is a permutation, then $\sigma(i)$ is the element of that permutation. Considering the case of a square Vandermonde matrix $\mathbf{A}$, it follows that

$$\det\big(\mathbf{V}_n(\mathbf{x})\big) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} x_i^{\sigma(i)-1} \tag{1.18}$$

where S_n denotes the set of permutations of $\{1,\cdots,n\}$ and $\operatorname{sgn}(\sigma)$ denotes the signature of the permutations σ. This determines the factors as

$$\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} x_i^{\sigma(i)-1} = \prod_{1 \le i < j \le n} (x_j - x_i). \tag{1.19}$$

This generalizes to the main result of [78] given in (1.3). The property is very important is important is investigating the extreme points of generalized Vandermonde determinant.

1.1.6 The Alternant Matrix

The defining property of the Vandermonde polynomial is that it is alternating in the entries, meaning that permuting the X_i by an odd permutation changes the sign, while permuting them by an even permutation does not change the value of the polynomial, in fact, it is the basic alternating polynomial [352, 463].

Definition 1.1.4. *An alternant matrix is a matrix of the form*

$$\mathbf{A}_{mn}(\mathbf{f}_m, \mathbf{x}_n) = [f_i(x_j)]_{i,j}^{m,n} = \begin{bmatrix} f_1(x_1) & f_1(x_2) & \dots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \dots & f_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(x_1) & f_m(x_2) & \dots & f_m(x_n) \end{bmatrix}$$

where $f_i : \mathbb{F} \to \mathbb{F}$ and $\mathbb{F}$ is a field. If the matrix is square, then $m = n$, so that the notation becomes $\mathbf{A}_n = \mathbf{A}_{mn}$.

Thus the matrix depends on the order, and is zero if two variables are equal. This also follows from the formula, but is also consequence of being alternating: if two variables are equal, then switching them both does not change the value and changes sign of the value, yielding $\mathbf{V}_n = -\mathbf{V}_n$, and thus $\mathbf{V}_n = 0$ (assuming the characteristic of $\mathbb{F}$ is not 2, otherwise being alternating is equivalent to being symmetric) [226, 499].

Conversely, the Vandermonde polynomial is a factor of every alternating polynomial: as shown above, an alternating polynomial vanishes if any two variables are equal, and thus one must have $(X_i - X_j)$ as a factor for all $i \neq j$.

Some examples of alternant matrices include:

Jacobian matrices: This matrix can be defined as follows [404]:

Definition 1.1.5. *Let $f : \mathbb{F}^n \to \mathbb{F}^n$ be a vector-valued function that is n times differentiable with respect to each variable, then the Jacobian matrix $\mathbf{J}$ is given by*

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & \ddots & \cdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n}. \end{bmatrix}$$

There are numerous applications of the Jacobian matrix based on the absolute value of its determinant, in particular in how elemental volumes are deformed when changing variables over general surfaces in multivariate calculus.

Wronskian matrix The Wronskian matrix is commonly used to test if a set of solutions are linearly independent as well as finding solutions to ordinary differential equations [51, 52, 53, 60, 181, 236, 311, 376, 377, 497, 498].

Definition 1.1.6. *If $\mathbf{f}_n = (f_1, f_2, \ldots, f_n)$ where $f_i = \dfrac{d^{i-1}}{dx^{i-1}}$, $\mathbf{g}_n = (g_1, g_2, \ldots, g_n)$ where $g_i \in \mathbf{C}^{n-1}[\mathbb{C}]$, then the alternant matrix $\mathbf{A}_n(\mathbf{f}_n; \mathbf{g}_n)$ will be called the Wronskian matrix.*

The classical application of the Wronskian is confirming if the set of solutions to a linear differential equation are independent. This is determined from the determinant in which if the determinant in non-zero, then the solutions are linearly independent.

There are other examples of alternant matrices which includes the Bell matrix which can be used to convert function composition into matrix multiplication. If $\mathbf{D}_i = \dfrac{d^{i-1}}{dx^{i-1}}$ and $g_j(x) = (f(x))^j$, where f is continuous smooth function, that is, infinitely differentiable, then the alternant matrix $\mathbf{B}[f] = \mathbf{A}_n(\mathbf{D}_n, \mathbf{g}_n)$ is called a Bell matrix and its transpose is called Carleman matrix. Detailed discussion of this can be got in [250, 279, 280].

The other other famous example of alternant matrix is called the Moore matrix. This matrix can be constructed by taking an alternant matrix say Vandermonde or Wronskian matrix where by the rows are given by the power of the Frobenius automorphism, $F(\omega) = \omega^p$ where $p < \infty$. More detailed discussion on the Moore matrix can be got in [120, 305, 308, 340, 358, 368].

The alternant property of Vandermonde matrix and its determinant is very useful in the study of symmetric group properties details of which are discussed in Section 1.2.

1.1.7 Calculus of the Vandermonde matrix and its Determinant

In mathematical calculus, a differentiable function of one or more real variables is a function whose derivative exists at each point in its domain. Thus, the graph of a differentiable function must have a (non-vertical) tangent line at each interior point. Thus Vandermonde determinant is differentiable since it is a continuous polynomial. Its derivative can be expressed as [2, 305, 308, 454]:

Theorem 1.1.6. *For any* $1 \leq k \leq n$ *and* $\mathbf{x} = (x_1, \ldots, x_n)$*, then*

$$\frac{\partial v_n(\mathbf{x})}{\partial x_k} = \sum_{\substack{i=1 \\ k \neq i}}^{n} \frac{v_n(\mathbf{x})}{(x_k - x_j)}. \tag{1.20}$$

where $v_n(\mathbf{x})$ *is the Vandermonde determinant given in* (1.8)*.*

This property verifies continuity and differentiability of Vandermonde determinant and is very helpful in the optimization of the Vandermonde determinant by the method of the Lagrange multiplier [272, 273, 406]. The detailed discussion and illustration of this property and related applications can also be obtained in [305, 308, 342, 343, 344, 345, 346, 347, 348, 349, 350].

1.2 Vandermonde Determinant and Symmetric Polynomials

In this section we give key concepts of symmetric polynomials (functions) which are important in characterization of the properties of the Vandermonde matrix and its determinant especially regarding calculus, inverse and decomposition. These concepts will be applied mainly in Chapter 2, Chapter 3, Chapter 4, Chapter 5, Chapter 6, Chapter 7, Chapter 8 and Chapter 9.

1.2.1 Symmetric Polynomials

Symmetric polynomials occur naturally especially in the study of the relation between the roots of a polynomial in one variable and its coefficients, whereby the coefficients can be expressed by polynomial expressions in terms of its roots [312, 440, 472].

Definition 1.2.1. *A* weak composition*,* α*, of a non-negative integer n is a sequence of non-negative integers* $\alpha = (\alpha_1, \alpha_2, \ldots)$ *such that* $\sum_i \alpha_i = n$*.*

Definition 1.2.2 ([440]). *Let* $\mathbf{x} = (x_1, \ldots)$ *be indeterminates and let* $n \in \mathbb{N}$. *A homogeneous symmetric polynomial of degree n over a commutative ring* R *is a formal power series*

$$f(\mathbf{x}) = \sum_{\alpha} c_\alpha x^\alpha \tag{1.21}$$

where

(a) α *ranges over all weak compositions* $\alpha = (\alpha_1, \alpha_2, \ldots)$ *of n (of infinite length),*

(b) $c_\alpha \in \mathrm{R}$,

(c) x^α *stands for the monomial* $x_1^{\alpha_1} x_2^{\alpha_2} \cdots$,

(d) $f(x_{\sigma(1)}, x_{\sigma(2)} \ldots) = f(x_1, x_2, \ldots)$ *for a every permutation* σ *of the positive integers* $\mathbb{P}$.

Considering the ring $\mathbb{Z}[x_1, x_2, \ldots, x_n]$ of polynomials in n independent variables $x_1, x_2, \ldots, x_n$ with integers coefficients. The symmetric group S_n acts on this ring by permuting the variables and a polynomial is symmetric if it is invariant under this action. The symmetric polynomials form a subring [312]

$$\Lambda = \Lambda_n(\mathbf{x}) = \mathbb{Z}[x_1, x_2, \ldots, x_n]^{\mathrm{S}_n}.$$

Λ_n is graded ring and we have

$$\Lambda_n = \bigoplus_{k \geq 0} \Lambda_n^k$$

where Λ_n^k consists of the *homogeneous symmetric polynomials* of degree k, together with the zero polynomial [35, 312, 410].

Definition 1.2.3. *A partition of a positive integer n is a finite non-negative non-increasing sequence of positive integers* $\lambda_1, \lambda_2, \ldots, \lambda_n$, *such that* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ *and* $|\lambda| = \sum_{i=1}^{n} \lambda_i = n$. *It is denoted by* $\lambda \vdash n$.

Definition 1.2.4 ([410]). , *Let* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{N}^n$ *be a partition of n and* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *be indeterminates. Then, the sum of monomials, corresponding to* λ *is given by*

$$m_\lambda(\mathbf{x}) = \sum_{\lambda} x^\lambda = \sum_{|\lambda|=n} x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_n}^{\lambda_n} = \sum_{|\lambda|=n} \prod_{k=1}^{n} x_{i_k}^{\lambda_k} \tag{1.22}$$

where the sum is over all distinct monomials $x^\lambda = x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}$ *and* m_λ *is referred to as monomial symmetric polynomial.*

For example, with $\lambda = (2, 0, \ldots, 0)$, then $m_\lambda = x_1^2 + x_2^2 + x_3^2 \ldots$ and for $\lambda = (2, 1, 0, \ldots, 0)$, gives $m_\lambda = x_1^2 x_2 + x_1 x_2^2 + x_1^2 x_3 + x_1 x_2^2 + x_2^2 x_3 + x_2 x_3^2 + \ldots$.

Definition 1.2.5. *Let $f(x)$ be the polynomial and e_k the polynomials in the n variables x_i, $1 \le i \le n$, be given by*

$$f(x) = \prod_{i=1}^{n}(x - x_i) = \sum_{k=0}^{n}(-1)^k e_k x^{n-k}. \tag{1.23}$$

Then, for each integer $k \ge 0$, the kth elementary symmetric polynomial e_k is the sum of all products of k distinct variables x_i, so that $e_0 = 1$ and

$$e_k = \sum_{i_1 < i_2 < \ldots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k} = m_{1^k} \ \ (\textit{with} \ \ e_0 = m_\emptyset = 1,) \ \ k \ge 1, \tag{1.24}$$

$e_\lambda = e_{\lambda_1} e_{\lambda_2} \cdots$, if $\lambda = (\lambda_1, \lambda_2, \ldots)$ and $\emptyset = (0, \ldots, 0)$ is for zero partition.

Example 1.2.1. $e_1 = \sum_{k=1}^{n} x_k$, $e_2 = \sum_{1 \le k < l \le n} x_k x_l$, $e_3 = \sum_{1 \le k < l < m \le n} x_k x_l x_m$, $\ldots$, $e_n = x_1 x_2 x_3 \ldots x_n$.

The generating function for the e_k is

$$\mathrm{E}[t] = \sum_{r \ge 0} e_r t^r = \prod_{i \ge 1}(1 + x_i t) \tag{1.25}$$

where t is another variable as can be seen by multiplying out the product on the right.

Related to the elementary symmetric polynomials, are polynomials $g_{nk}(x)$ that can be expressed as stated in [472].

Definition 1.2.6. *Let $g_{nk}(x)$ be a polynomial such that the polynomials e_{kl} in the $(n-1)$ variables $x_i, 1 \le i \le n, i \ne k$, be given by:*

$$g_{nk}(x) = \frac{f_n(x)}{x - x_k} = \sum_{l=0}^{n-1}(-1)^l e_{kl} x^{n-1-l} \tag{1.26}$$

$$g_{nn}(x) = f_{n-1}(x) \tag{1.27}$$

for all values of x. Then,

$$e_{nl}^{(n)} = e_l^{n-1}. \tag{1.28}$$

Example 1.2.2. $e_2^{(3)} = x_1 x_2 + x_1 x_3 + x_2 x_3$, $e_{k0}^{(n)} = 1, 1 \le k \le n$, $e_{21}^{(3)} = x_1 + x_3$, $e_{22}^{(3)} = x_1 x_3$, $e_{31}^{(4)} = x_1 + x_2 + x_3$, $e_{32}^{(4)} = x_1 x_2 + x_1 x_4 + x_2 x_4$, $e_{33}^{(4)} = x_1 x_2 x_4$.

The above content can be summed up in the lemma as stated in [472].

Lemma 1.2.3 ([472])**.** $e_{rl}^{(n)} = \sum_{k=0}^{l} e_k^n (-x_r)^{l-k}$.

Proposition 1.2.1 ([472])**.**

$$\prod_{i,j}(1+x_i y_j)=\sum_{\lambda} m_\lambda(\mathbf{x})e_\lambda(\mathbf{y}). \tag{1.29}$$

Definition 1.2.7. *For each $r \geq 0$, the rth complete symmetric polynomial h_r is the sum of all monomials of total degree r in the variables $x_1, x_2, \ldots$, so that*

$$h_n=\sum_{|\lambda|=n} m_\lambda = \sum_{\substack{i_1<i_2<\ldots<i_n \\ \lambda_1+\lambda_2+\ldots+\lambda_n=n}} x_{i_1}^{\lambda_1}x_{i_2}^{\lambda_2}\cdots x_{i_n}^{\lambda_n},\ (h_0=m_{1^0}=1, h_1=e_1), \tag{1.30}$$

and $h_\lambda = h_{\lambda_1}h_{\lambda_2}\cdots$ with $\lambda=(\lambda_1,\lambda_2,\ldots)$ and $\emptyset=(0,\ldots,0)$ is for zero partition. It is convenient to define h_n and e_n to be zero for $n<0$.

The generating function for h_r is given by

$$\mathrm{H}(t)=\sum_{r\geq 0} h_r t^r = \prod_{i\geq 1}(1-x_i t)^{-1}. \tag{1.31}$$

This can easily be seen from the identity $(1-x_i t)^{-1}=\sum_{k\geq 0} x_i^k t^k$, and multiplying these geometric series together.

It should be noted from (1.25) and (1.31) that

$$\mathrm{H}(t)\mathrm{E}(-t)=1 \tag{1.32}$$

or equivalently,

$$\sum_{r=0}^{n}(-1)^r e_r h_{n-r}=0 \tag{1.33}$$

for all $n \geq 1$.

The equation (1.32) leads to a determinant identity, that is, by taking N to be a positive integer and considering the matrices $\mathbf{H}$ and $\mathbf{E}$ of $N+1$ rows and columns given by

$$\mathbf{H}=(h_{i-j})_{0\leq i,j\leq N}, \qquad \mathbf{E}=\left((-1)^{i-j}e_{i-j}\right)_{0\leq i,j\leq N}. \tag{1.34}$$

Applying the standard convention mentioned earlier that $h_r=e_r=0$ for $r<0$. Both $\mathbf{H}$ and $\mathbf{E}$ are lower triangular, with 1's down the diagonal, then the following identity holds

$$\det\big(\mathbf{H}\big)=\det\big(\mathbf{E}\big)=1.$$

The relation (1.34) also shows that the matrices $\mathbf{H}$ and $\mathbf{E}$ are inverses of each other. It follows that each minor of $\mathbf{H}$ is equal to the complementary cofactor of $\mathbf{E}^\top$, the transpose of $\mathbf{E}$.

Proposition 1.2.2 ([440])**.**

$$\prod_{i,j}(1-x_i y_j)^{-1}=\sum_{\lambda} m_\lambda(\mathbf{x})h_\lambda(\mathbf{y}) \tag{1.35}$$

Definition 1.2.8. *Consider a polynomial $P(x)$ with roots x_i, $i = 1,\ldots,n$. For any given $r \geq 1$, the rth power sum is given by*

$$p_r = \sum_{i=1}^{n} x_i^r = m_r, \;\; r \geq 1, \;\; (\textit{with } p_0 = m_\emptyset = 1). \tag{1.36}$$

We also define the power sum with respect to a partition $\lambda = (\lambda_1, \lambda_2, \ldots)$ as $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots$.

The generating function for the p_r is expressed as

$$\mathrm{P}(t) = \sum_{r\geq 1} p_r t^{r-1} = \sum_{i\geq 1}\sum_{r\geq 1} x_i^r t^{r-1} = \sum_{i\geq 1} \frac{x_i}{1-x_i t} = \sum_{i\geq 1} \frac{d}{dt} \log \frac{1}{x_i t}$$

so that

$$\mathrm{P}(t) = \frac{d}{dt} \log \prod_{i\geq 1} (1-x_i t)^{-1} = \frac{d}{dt} \log \mathrm{H}(t) = \mathrm{H}^{'}(t)/\mathrm{H}(t). \tag{1.37}$$

Likewise we have

$$\mathrm{P}(-t) = \frac{d}{dt} \log \mathrm{E}(t) = \mathrm{E}^{'}(t)/\mathrm{E}(t). \tag{1.38}$$

For example, for $n = 3$

$$\begin{aligned} e_3 &= x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + \ldots, \\ h_3 &= x_1^3 + x_2^3 + x_3^3 + \ldots + x_1^2x_2 + x_1x_2^2 + \ldots + x_1x_2x_3 + x_1x_2x_4 + \ldots \\ p_3 &= x_1^3 + x_2^3 + x_3^3 + \ldots. \end{aligned}$$

Lemma 1.2.4 ([440])**.** *The power sums in* (1.36) *can be expressed in terms of the coefficients of the polynomial*

$$\mathrm{P}(x) = x^n + \sum_{i=0}^{r-1} e_i x^i = \prod_{j=1}^{n} (x - x_j), \tag{1.39}$$

as follows

$$\begin{aligned} &\sum_{i=1}^{k-1} e_{n-i} p_{k-i} + k e_{n-k} = 0, \;\; k = 1,2,\ldots,n \\ &\sum_{i=1}^{k-1} e_{n-i} p_{n+k-i} = 0, \;\; k = 1,2,\ldots,m-n, \end{aligned} \tag{1.40}$$

for any $m > n$

The simple proof of the Lemma 1.2.4 follows from (1.37) and (1.38) such that

$$n h_n = \sum_{r=1}^{n} p_r h_{n-r}, \tag{1.41}$$

$$n e_n = \sum_{r=1}^{n} (-1)^{r-1} p_r r_{n-r}, \tag{1.42}$$

for $n \geq 1$, and these equations enable us to express the hs and the es in terms of the ps and vice versa. The Equation (1.42) is due to Isaac Newton, and are famously known as Newton's formulas. From Equation (1.41) it is clear that $h_n \in \mathbb{Q}[p_1,\ldots,p_n]$ and $p_n \in \mathbb{Z}[h_1,\ldots,h_n]$ and hence that $\mathbb{Q}[p_1,\ldots,p_n] = \mathbb{Q}[h_1,\ldots,h_n]$.

Example 1.2.5 ([440])**.** *Evaluating e_n from Equations* (1.33) *we obtain $e_n = \det\left(h_{1-i+j}\right)_{1\leq i,j\leq n}$, and dually $h_n = \det\left(e_{1-i+j}\right)_{1\leq i,j\leq n}$. Similarly from Equation* (1.41)*, we obtain the determinant formulas*

$$p_n = \begin{vmatrix} e_1 & 1 & 0 & \ldots & 0 & 0 \\ 2e_2 & e_1 & 1 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ (n-2)e_{n-2} & e_{n-3} & \ldots & e_1 & 1 & 0 \\ (n-1)e_{n-1} & e_{n-2} & e_{n-3} & \ldots & e_1 & 1 \\ ne_n & e_{n-1} & e_{n-2} & \ldots & e_2 & e_1 \end{vmatrix}, \tag{1.43}$$

and

$$n!e_n = \begin{vmatrix} p_1 & 1 & 0 & \ldots & 0 & 0 \\ p_2 & p_1 & 2 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ p_{n-2} & p_{n-3} & \ldots & p_1 & n-1 & 0 \\ p_{n-1} & p_{n-2} & p_{n-3} & \ldots & p_1 & n-1 \\ p_n & p_{n-1} & p_{n-2} & \ldots & p_2 & p_1 \end{vmatrix} \tag{1.44}$$

where $|\cdot|$ represents determinant of a matrix and dually

$$(-1)^{n-1}p_n = \begin{vmatrix} h_1 & 1 & 0 & \ldots & 0 & 0 \\ 2h_2 & h_1 & 1 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ (n-1)h_{n-2} & h_{n-3} & \ldots & h_1 & 1 & 0 \\ (n-1)h_{n-1} & h_{n-2} & h_{n-3} & \ldots & h_1 & 1 \\ nh_n & h_{n-1} & h_{n-2} & \ldots & h_2 & h_1 \end{vmatrix}, \tag{1.45}$$

and

$$n!h_n = \begin{vmatrix} p_1 & -1 & 0 & \ldots & 0 & 0 \\ p_2 & p_1 & -2 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ p_{n-2} & p_{n-3} & \ldots & p_1 & -n+1 & 0 \\ p_{n-1} & p_{n-2} & p_{n-3} & \ldots & p_1 & -n+1 \\ p_n & p_{n-1} & p_{n-2} & \ldots & p_2 & p_1 \end{vmatrix}. \tag{1.46}$$

Lemma 1.2.6. *From the matrix determinants given in* (1.43), (1.44), (1.45) *and* (1.46) *we can write*

$$\begin{vmatrix} p_0 & p_1 & \cdots & p_{k-1} & p_k \\ p_1 & p_2 & \ddots & p_k & p_{k+1} \\ p_2 & p_3 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & p_{n-1} \\ p_{n-k} & p_{n-k+1} & \cdots & p_{n-1} & p_n \end{vmatrix} = \begin{vmatrix} 1 & 1 & \dots & 1 & 1 \\ x_1 & x_2 & \dots & x_{n-1} & x_n \\ x_1^2 & x_2^2 & \dots & x_{n-1}^2 & x_n^2 \\ x_1^3 & x_2^3 & \dots & x_{n-1}^3 & x_n^3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^n & x_2^n & \dots & x_{n-1}^n & x_n^n \end{vmatrix}^2 = \det\big(\mathbf{V}_{\delta'}(\mathbf{x})\big)^2 \tag{1.47}$$

where $\mathbf{V}_n(\mathbf{x})$ *is the Vandermonde matrix with* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$*, Definition 4.1.1 and Equation* (4.2) *and also in Table 1.1 as well as decomposition* (1.53)*.*

Following the fundamental theorem of symmetric polynomials [98, 312], then it can be shown that the determinant of even powers of Vandermonde matrices are symmetric polynomials.

Theorem 1.2.7 ([472] **Newton's Theorem).** *Any symmetric polynomial in the ring* $\mathbb{K}[x_1, x_2, \ldots, x_n]$ *can be written as a polynomial in elementary symmetric polynomials* $e_1, e_2, \ldots, e_n$ *with coefficients in the field* $\mathbb{K}$ *and this polynomial is unique.*

Proposition 1.2.3 ([440])**.**

$$\prod_{i,j}(1 - x_i y_j)^{-1} = \exp\left(\sum_{n\geq 1}\frac{1}{n}p_n(\mathbf{x})p_n(\mathbf{y})\right) = \sum_{\lambda} z_\lambda^{-1} p_\lambda(\mathbf{x})p_\lambda(\mathbf{y}) \tag{1.48}$$

$$\prod_{i,j}(1 + x_i y_j) = \exp\left(\sum_{n\geq 1}(-1)^{n-1}\frac{1}{n}p_n(\mathbf{x})p_n(\mathbf{y})\right) = \sum_{\lambda} z_\lambda^{-1}\varepsilon_\lambda p_\lambda(\mathbf{x})p_\lambda(\mathbf{y})$$

where for a partition $\lambda = \langle 1^{m_1} 2^{m_2} \cdots \rangle$*,* $z_\lambda = 1^{m_1} m_1! 2^{m_2} m_2! \cdots$ *and* $\varepsilon_\lambda = (-1)^{m_2+m_4+\cdots} = (-1)^{n-\ell(\lambda)}$*.*

Proposition 1.2.4 ([440])**.**

$$h_n(\mathbf{x}) = \sum_{\lambda} z_\lambda^{-1} p_\lambda(\mathbf{x}), \;\; e_n(\mathbf{x}) = \sum_{\lambda} \varepsilon_\lambda z_\lambda^{-1} p_\lambda(\mathbf{x}). \tag{1.49}$$

The above concepts on symmetric polynomials will be helpful in detailed study of Schur polynomials.

1.2.2 LDU Decomposition of Vandermonde Matrix Using Symmetric Polynomials

The Vandermonde matrix can be decomposed by **LDU** factorization [379]. Denoting the Vandermonde matrix with $\mathbf{V} = \mathbf{V}_{\delta'}(\mathbf{x})$ $\big($or $\mathbf{V}_{\delta}(\mathbf{x}), \mathbf{V}_{\lambda'+\delta'}(\mathbf{x}), \mathbf{V}_{\lambda+\delta}(\mathbf{x})\big)$, then $\mathbf{V}$ can be expressed in terms of the factors $\mathbf{L}$ a lower triangular matrix, $\mathbf{D}$ a diagonal matrix and $\mathbf{U}$ an upper triangular matrix.

This decomposition procedure can also help to explain properties of generalized Vandermonde matrices, Cauchy matrices, Hankel matrices, and Toeplitz matrices listed in Table 1.1.

According to [30], the Vandermonde matrix can be used to decompose both the Hankel and Toeplitz matrices whereby for the Hankel matrices

$$\mathbf{H} = \mathbf{V}^{\top}\mathbf{D}\mathbf{V} \tag{1.50}$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{V}^{\top}$ is the transpose of the transpose of the matrix. In the case for the Toeplitz matrices factorization by Vandermonde determinant we have

$$\mathbf{T} = \mathbf{V}_{\mathrm{x}}^{\top^{\star}}\mathbf{D}\mathbf{V} \tag{1.51}$$

where the Vandermonde matrix $\mathbf{V}_{\mathrm{x}}$ is defined by the scalars $(x_k^{\top^{\star}})^{-1}$, the complex conjugate of the scalars, and $\mathbf{V}^{\top^{\star}}$ to indicate the complex conjugate transposition for matrices.

Remark 1.2.8. *The Toeplitz matrix* $\mathbf{T}$ *with entries* $t_{-i} = t_{n-i},\ i = 1,2,\ldots n-1$ *is referred to as a circulant matrix if while the matrix* $\mathbf{H}$ *with entries* $h_{n+i} = h_i, i = 1,2,\ldots,n-1$ *is said to be Hankel circulant. It is also important to note that the Hankel matrix* $\mathbf{H}$ *can be converted into the Toeplitz matrix* $\mathbf{T}$ *by multiplication with permutation matrix* $\mathbf{P}_n = \begin{bmatrix} e_{n-1} & \ldots e_1 & e_0 \end{bmatrix}^{\top}$ *so that* $\mathbf{P}_n\mathbf{T} = \mathbf{H}$ *and* $\mathbf{P}_n^2 = \mathbf{I}_n$.

The powers of the the Vandermonde matrix can be obtained directly from the the **LDU** decomposition described in (1.52) where by $\mathbf{V}^n = \mathbf{L}\mathbf{D}^n\mathbf{U}$ where the elements of $\mathbf{L}$, $\mathbf{D}$ and $\mathbf{U}$ are as given in (1.53). The same applies to the Cauchy matrix, Hankel matrix and Toeplitz matrix.

Since the evaluation of the determinant follows directly from its diagonal matrix, then the powers of the Vandermonde determinant are expressed as $\det(\mathbf{V})^n = \det(\mathbf{V}^n) = \det(\mathbf{D})^n$ for $n \geq 2$.

Let $\ell_i^r(x) = (x-x_i)(x-x_{i+1})\ldots(x-x_r),\ i \leq r \leq n$ where the coefficients of x^n in the expansion of $\ell_i^r(x)$ can be expressed in terms of elementary symmetric polynomials as described in the previous section. Since $\ell_i^r(x)$ is continuous function, then its derivative with respect to x exists thus, for a fixed $x = x_j$ we have

$$\ell'^r_i(x_j) = \left.\frac{\partial \ell_i^r(x)}{\partial x}\right|_{x=x_j} = \prod_{\substack{i=1 \\ i\neq j}}^{r}(x_j - x_i),\ \ i \leq r \leq n.$$

The following formulas for the **LDU** decomposition of Vandermonde matrix $\mathbf{V} = \mathbf{V}_{\lambda'+\delta'}(\mathbf{x})$ are based on the **LU** factors in [379].

Theorem 1.2.9 ([370])**.** *Let* $\mathbf{V} = \left\{x_i^{j-1}\right\}_{j=1}^{n}$. *Then,*

$$\mathbf{V} = \mathbf{L}\mathbf{D}\mathbf{U} \ \textit{and}\ \ \mathbf{V}^{-1} = \mathbf{U}^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}, \tag{1.52}$$

where their respective elements are given by (1.53) *and* (1.54)*:*

$$
\begin{aligned}
d_{kk} &= \prod_{j=0}^{k-1}(x_k - x_j),\ k = 1,2,\ldots,n;\\
l_{kj} &= \sum_{\substack{i_0+\ldots+i_j=k-j\\ i_0,\ldots,i_j\geq 0}} x_0^{i_0}x_1^{i_1}\ldots x_j^{i_j},\ 0 \leq j < k \leq n;\\
u_{kj} &= \prod_{i=0}^{k-1}\frac{x_j - x_i}{x_k - x_i},\ 0 < k < j \leq n;\\
u_{0j} &= 1,\ \ j = 1,\ldots,n;
\end{aligned}
\tag{1.53}
$$

$$
\begin{aligned}
d_{kk}^{-1} &= \prod_{j=0}^{k-1}(x_k - x_j)^{-1},\ k = 1,2,\ldots,n;\\
l_{kj}^{-1} &= (-1)^{k+j}\sum_{0\leq i_0<i_1<i_{k-j-}<k} a_{i_0}a_{i_1}\cdots a_{i_{k-j-1}},\ 0 \leq j < k \leq n;\\
u_{kj}^{-1} &= d_{jj}\prod_{\substack{i=0\\ i\neq k}}^{j}\frac{1}{x_k - x_i},\ 0 \leq k \leq j \leq n.
\end{aligned}
\tag{1.54}
$$

Theorem 1.2.10 ([202, 321]). *Let* $\mathbf{C}$ *be a Cauchy matrix* $\mathbf{C} = \left(\dfrac{1}{t_i - s_j}\right)_{i,j=1}^{n}$. *Then,*

$$
\mathbf{C} = \mathbf{LDU} \ \text{ and } \ \mathbf{C}^{-1} = \mathbf{U}^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1},
\tag{1.55}
$$

where their respective elements are given by (1.56) *and* (1.57)*:*

$$
\begin{aligned}
d_{kk} &= (t_k - s_k)^{-1}\prod_{m=0}^{k-1}\frac{(s_k - s_m)(t_k - t_m)}{(s_k - t_m)(t_k - s_m)},\ k = 1,2,\ldots,n;\\
l_{kj} &= \frac{(t_k - s_k)}{(t_k - s_j)}\prod_{m=1}^{k-1}\frac{(t_k - s_m)(t_j - t_m)}{(t_j - s_m)(t_k - t_m)},\ 1 \leq k < j \leq n;\\
u_{kj} &= \frac{(t_k - s_k)}{(t_k - s_j)}\prod_{m=1}^{k-1}\frac{(s_k - t_m)(s_j - s_m)}{(s_j - t_m)(s_k - s_m)},\ 1 \leq k < j \leq n;\\
d_{kk}^{-1} &= (t_k - s_k)\prod_{\substack{m=1\\ m\neq j}}^{k-1}\frac{(s_k - t_m)(t_k - s_m)}{(s_k - s_m)(t_k - t_m)},\ k = 1,2,\ldots,n;
\end{aligned}
\tag{1.56}
$$

$$
\begin{aligned}
l_{kj}^{-1} &= \frac{(t_j - s_j)}{(t_k - s_j)}\prod_{m=1}^{k-1}\frac{(t_j - s_m)(t_k - t_m)}{(t_j - s_m)(t_k - t_m)},\ 1 \leq k < j \leq n;\\
u_{kj}^{-1} &= \frac{(t_j - s_j)}{(t_j - s_k)}\prod_{\substack{m=1\\ m\neq j}}^{k-1}\frac{(s_j - t_m)(s_k - s_m)}{(s_j - s_m)(s_k - t_m)},\ 1 \leq k < j \leq n.
\end{aligned}
\tag{1.57}
$$

Vandermonde matrices $\mathbf{V}_{\delta'} = \left\{x_i^{j-1}\right\}_{i,j=1}^{n-1}$ $\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \\ x_1^2 & x_2^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_n^{n-1} & x_n^{n-1} & \dots & x_n^{n-1} \end{bmatrix}$	Generalized $\mathbf{V}_\lambda(\mathbf{x}) = \left\{x_i^{\lambda_j}\right\}_{i,j=1}^{n}$ $\begin{bmatrix} x_1^{\lambda_1} & x_2^{\lambda_1} & x_3^{\lambda_1} & \dots & x_n^{\lambda_1} \\ x_1^{\lambda_2} & x_2^{\lambda_2} & x_3^{\lambda_2} & \dots & x_n^{\lambda_2} \\ x_1^{\lambda_3} & x_2^{\lambda_3} & x_3^{\lambda_3} & \dots & x_n^{\lambda_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{\lambda_n} & x_2^{\lambda_n} & x_3^{\lambda_n} & \dots & x_n^{\lambda_n} \end{bmatrix}$
Toeplitz matrices $\mathbf{T} = \left(x_{i-j}\right)_{i,j=0}^{n-1}$ $\begin{bmatrix} t_0 & t_{-1} & x_{-2} & \dots & t_{1-n} \\ t_1 & t_0 & x_{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & t_{-1} \\ t_{n-1} & t_{n-2} & \ddots & x_1 & t_0 \end{bmatrix}$	Hankel matrices $\mathbf{H} = \left(h_{i+j}\right)_{i,j=0}^{n-1}$ $\begin{bmatrix} h_0 & h_1 & h_2 & \dots & h_{n-1} \\ h_1 & h_2 & & & \vdots \\ h_2 & & & & \vdots \\ \vdots & & & & h_{2n-3} \\ h_{n-1} & \dots & \dots & h_{2n-3} & h_{2n-2} \end{bmatrix}$
Circulant matrices $\mathbf{C} = \left(c_{i-j}\right)_{i,j=0}^{n-1}$ $\begin{bmatrix} c_0 & c_{n-1} & \dots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & \ddots & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \dots & c_1 & c_0 \end{bmatrix}$	Cauchy matrices $\mathbf{S} = \left(\frac{1}{s_i - t_j}\right)_{i,j=0}^{n-1}$ $\begin{bmatrix} \frac{1}{s_0-t_0} & \frac{1}{s_0-t_1} & \dots & \frac{1}{s_0-t_{n-1}} \\ \frac{1}{s_1-t_0} & \frac{1}{s_1-t_1} & \dots & \frac{1}{s_1-t_{n-1}} \\ \vdots & \ddots & \ddots & \ddots \\ \frac{1}{s_{n-1}-t_0} & \frac{1}{s_{n-1}-t_1} & \dots & \frac{1}{s_{n-1}-t_{n-1}} \end{bmatrix}$

Table 1.1: Vandermonde type matrices.

Considering the close relationship relationship between the Vandermonde matrices, Hankel matrices, Toeplitz matrices, Circulant matrices and Cauchy matrices, whose structures are as shown in the Table 1.1, then the above **LDU** decomposition procedure for **V** can also be used to decompose the Hankel and Toeplitz matrices by combining the relations (1.52) and (1.53) with (1.50) and (1.51).

1.2.3 General Properties of Vandermonde Determinant Based on Symmetric Polynomials

Here we give outline of the general properties of the Vandermonde determinant based on the symmetric polynomials.

(a) **Cofactors of the Vandermonde Determinant:**

Theorem 1.2.11 ([472]). *If* $\mathbf{V}_{\delta'}(\mathbf{x}) = \left\{x_i^{j-1}\right\}_{i,j=1}^{n}$ *is a Vandermonde matrix, then the scaled cofactors of the Vandermonde determinant* $\det\big(\mathbf{V}_{\delta'}(\mathbf{x})\big) = a_{\delta'}(\mathbf{x})$*, are given by given by the quotient formula*

$$[a_{\delta'}(\mathbf{x})]_n^{ij} = \frac{(-1)^{n-j} e_{i,n-j}^{(n)}}{g_{ni}(xi)} \tag{1.58}$$

where $e_{i,n-j}^{(n)}$ *the symmetric polynomial given in* (1.28) *and the function* $g_{ni}(x_i)$ *is as defined in* (1.26)*.*

Theorem 1.2.12 ([472]). *If* $\mathbf{V}_{\delta'}(\mathbf{x}) = \left\{x_i^{j-1}\right\}_{i,j=1}^{n}$ *is Vandermonde matrix, then*

$$[a_{\delta'}(\mathbf{x})]_{nj}^{(n)} = (-1)^{n-j}[a_{\delta'}(\mathbf{x})]_{n-1} e_{n-j}^{(n-1)} \tag{1.59}$$

(b) **A Hybrid Determinant:** Given two $n \times n$ Vandermonde matrices $\mathbf{V}_{\delta'}(\mathbf{x}) = \left\{x_i^{j-1}\right\}_{i,j=1}^{n}$ and $\mathbf{V}_{\delta'}(\mathbf{y}) = \left\{y_i^{j-1}\right\}_{i,j=1}^{n}$. Let $\mathbf{W}_{rs}$ be a hybrid matrix formed by replacing the rth row of $\mathbf{V}_{\delta'}(\mathbf{x})$ by the sth row of $\mathbf{V}_{\delta'}(\mathbf{y})$.

Theorem 1.2.13 ([472]). *The hybrid determinant of the matrix* $\mathbf{W}_{rs}$ *can be expressed as*

$$\frac{\det(\mathbf{W})_{rs}}{a_{\delta'}(\mathbf{x})} = \frac{g_{nr}(y_s)}{g_{nr}(x_r)} = \frac{\prod_{i=1}^{n}(y_s - x_i)}{(y_s - x_r)\prod_{1\le i<r\le n}(x_r - x_i)}. \tag{1.60}$$

Theorem 1.2.14 ([472]). *Given* $\mathbf{A}$ *is an* $n \times n$ *matrix with entries* $e_{i,j-1}^{(m)}$ *as given in* (1.28)*. Then,*

$$\det(\mathbf{A}) = \left|e_{i,j-1}^{(m)}\right|_n = \begin{vmatrix} e_{10}^{(m)} & e_{11}^{(m)} & \dots & e_{1,n-1}^{(m)} \\ e_{20}^{(m)} & e_{21}^{(m)} & \dots & e_{2,n-1}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ e_{n0}^{(m)} & e_{n1}^{(m)} & \dots & e_{n,n-1}^{(m)} \end{vmatrix}_n = (-1)^{\frac{1}{2}n(n-1)} a_{\delta'}(\mathbf{x}). \tag{1.61}$$

The detailed outlines of the proofs to the above two theorems can also be found in [472].

(c) **The Cauchy Double Alternant:** Considering the Cauchy matrix as given in Table 1.1, then the Cauchy double alternant is the determinant $\det(\mathbf{C}) = \det\left(\frac{1}{x_i - y_j}\right)_n$, seen in Table 1.1 as well as decomposition (1.56), which is calculated from the ratio of determinants for the $n \times n$ of two Vandermonde matrices$\mathbf{V}_{\delta'}(\mathbf{x}) = \left\{x_i^{j-1}\right\}_{i,j=1}^{n}$ and $\mathbf{V}_{\delta'}(\mathbf{y}) = \left\{y_i^{j-1}\right\}_{i,j=1}^{n}$. This can be achieved in the following steps as outlined in [472].

Theorem 1.2.15. *Performing the column operation* $\mathbf{C}'_j = \mathbf{C}_j - \mathbf{C}_n,\ \ 1 \le j \le n-1$, *and then removing all the common factors from the elements of rows and columns. The result becomes*

$$\det(\mathbf{C}) = \frac{\prod_{k=1}^{n-1}(y_k - y_n)}{\prod_{k=1}^{n-1}(x_k - x_n)} \cdot \det(\mathbf{B})_n, \tag{1.62}$$

where $\mathbf{B}$ *is the matrix in the last column is* $\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}_n^{\top}$ *and all other columns are identical with the corresponding columns of* $\mathbf{C}$.

Next performing the row operations $\mathbf{R}'_i = \mathbf{R}_i - \mathbf{R}_n,\ \ 1 \le j \le n-1$, *on* $\mathbf{B}$, *which will decompose into a matrix of order* $n-1$. *After extracting all the common factors from the elements of rows and columns, leads to*

$$\det(\mathbf{B})_n = \frac{\prod_{k=1}^{n-1}(y_k - y_n)}{\prod_{k=1}^{n-1}(x_k - x_n)} \cdot \det(\mathbf{C})_{n-1}. \tag{1.63}$$

Substituting for $\det(\mathbf{B})_n$ *in* (1.62) *yields a reduction formula for* $\det(\mathbf{C})_n$, *which, when applied, results in the required formula*

$$\det(\mathbf{B})_n = \frac{(-1)^{\frac{1}{2}n(n-1)}[a_{\delta'}(\mathbf{x})]_n[a_{\delta'}(\mathbf{y})]_n}{\prod_{1 \le r,s \le n1}(x_r - y_s)}, \tag{1.64}$$

where $[a_{\delta'}(\mathbf{x})]_n$ *and* $[a_{\delta'}(\mathbf{y})]_n$ *are determinants of Vandermonde matrices of size n.*

(d) **Determinant Related to the Vandermonde Determinant:** Let $P_k(x)$ be a polynomial defined as $P_k(x) = \sum_{s=1}^{k} a_{sk}x^{s-1},\ \ k \ge 1$. That is, $P_k(x)$ is not a monomial since the leading coefficient $a_{1,k}$ which is not necessarily equal to one.

Theorem 1.2.16 ([472])**.** *Let* $\mathbf{V}_{\delta'}(\mathbf{x})$ *be a Vandermonde matrix with entries* $x_{ij} = x_i^{j-1}$, *then*

$$\det\big(P_i(x_j)\big)_n = \big(a_{11}a_{22}\cdots a_{nn}\big) \cdot [a_{\delta'}(\mathbf{x})]_n.$$

1.2.4 Schur Polynomials

The Schur polynomials named after Issai Schur (1875–1941) are a family of symmetric polynomials in n variables, indexed by partitions, that generalized the elementary symmetric polynomials [63, 251, 318, 338, 339, 418]. The Schur polynomials in representation theory are the characters of polynomial irreducible representations of the general linear group [175, 312].

Definition 1.2.9. *Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be a partition of a positive integer n such that $\lambda_1 > \lambda_n > \ldots > \lambda_n \geq 0$, then the quotient homogeneous polynomial s_λ when $a_{\lambda+\delta}$ is divided by a_δ in $\mathbb{Z}[x_1, \ldots, x_n]$, that is,*

$$s_\lambda(x_1, \ldots, x_n) = \frac{a_{\lambda+\delta}}{a_\delta} = \frac{a_{(\lambda_1+n-1, \lambda_2+n-2, \ldots, \lambda_n)}(x_1, \ldots, x_n)}{a_{(n-1, n-2, \ldots, 0)}(x_1, \ldots, x_n)} \tag{1.65}$$

is symmetric, that is, it is in the subring with basis $m_\lambda, e_\lambda, h_\lambda, p_\lambda$ of symmetric polynomials over field $\mathbb{K}$. This set of functions is called the Schur functions in the variables $x_1, x_2, \ldots, x_n$, corresponding to the partition λ (where $l(\lambda) \leq n$), and is homogeneous of degree $|\lambda|$.

The Schur polynomials are a generalization of the quotient of determinants of two alternant matrices:

Proposition 1.2.5 ([193])**.** *Given a non-negative integer n, for each partition $\lambda = (\lambda_1, \ldots, \lambda_n)$ of non-negative integers such that $\lambda_1 > \lambda_n > \ldots > \lambda_n \geq 0$, the homogeneous polynomial*

$$a_\lambda = \det\left(\left\{x_i^{\lambda_j}\right\}_{i,j=1}^{n}\right) \tag{1.66}$$

is divisible by the principle Vandermonde determinant

$$a_\delta = \det\left(\left\{x_i^{j-1}\right\}_{i,j=1}^{n}\right) = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$

in $\mathbb{K}[x_1, \ldots, x_n]$.

1.2.5 Properties of Schur Polynomials

The degree d-Schur polynomial in n-variables is a linear basis for the space of homogeneous d-degree symmetric polynomials in n-variables [213, 214, 215].

Theorem 1.2.17 ([312])**.** *For a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, the Schur polynomial is a sum of monomials (complete symmetric polynomials)*

$$s_\lambda(x_1, \ldots, x_n) = \sum_T x^T = \sum_T x_1^{t_1} x_2^{t_2} \cdots x_n^{t_n} \tag{1.67}$$

where the summation is over all semi-standard Young Tableaux T of shape λ. The exponents $t_1, t_2, \ldots, t_n$ give the weight T, that is, each t_i counts the occurrences of the number i in T.

Lemma 1.2.18 ([410]). *Schur polynomials can be expressed as linear combinations of elementary symmetric polynomials m_μ with non-negative integer coefficients $K_\lambda\mu$ called Kostka number such that*

$$s_\lambda = \sum_\mu K_\lambda\mu m_\mu \tag{1.68}$$

where the Kostka numbers $K_\lambda\mu$ are given the number of semi-standard Young Tableaux of shape λ and weight μ.

Jacobi–Trudi Identities ([312]): The Jacobi–Trudi formulas expresses the Schur polynomial as a determinant in term of complete homogeneous symmetric polynomials and elementary symmetric polynomials such that as summarized in the following lemma.

Lemma 1.2.19 ([440]). *The Schur function s_λ can be expressed as a polynomial in the elementary symmetric polynomials e_r, and as a polynomial in the complete symmetric polynomials h_r. The formulas are respectively,*

$$s_\lambda = \det\left(h_{\lambda_i+j-i}\right)_{i,j=1}^{l(\lambda)} = \det\begin{bmatrix} h_{\lambda_1} & h_{\lambda_1+1} & \cdots & h_{\lambda_1+n-1} \\ h_{\lambda_2-1} & h_{\lambda_2} & \cdots & h_{\lambda_2+n-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{\lambda_n-n+1} & h_{\lambda_n-n+2} & \cdots & h_{\lambda_n} \end{bmatrix} \tag{1.69}$$

where $n \geq l(\lambda)$ and $h_i = s_{(i)}$, and

$$s_\lambda = \det\left(e_{\lambda'_i+j-i}\right)_{i,j=1}^{l(\lambda')} = \det\begin{bmatrix} e_{\lambda'_1} & e_{\lambda'_1+1} & \cdots & e_{\lambda'_1+n-1} \\ e_{\lambda'_2-1} & e_{\lambda'_2} & \cdots & e_{\lambda'_2+n-2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{\lambda'_n-n+1} & e_{\lambda'_n-n+2} & \cdots & e_{\lambda'_n} \end{bmatrix} \tag{1.70}$$

where $m \geq l(\lambda')$, $e_i = s_{(1^n)}$ and λ' is the conjugate partition to λ.

Giambelli Identity: This determinantal identity expresses the Schur function for an arbitrary partition in terms of those for the hook partitions contained within the Young diagram [194, 313, 441]. In Frobenius' notation, the partition is denoted by $(a_1,\ldots,a_r|b_1,\ldots,b_n)$ where for each diagonal element in position ii, a_i denotes the number of boxes to the right in the row and b_i denotes the number of boxes beneath in the same column.

Theorem 1.2.20 ([174, 312]). *Giambelli identity expresses the Schur function corresponding to the partition as the determinant*

$$s_{(a_1,\ldots,a_r|b_1,\ldots,b_n)} = \det\left(s_{(a_i|b_j)}\right) \tag{1.71}$$

of those for hook partitions.

Theorem 1.2.21 ([312] **Cauchy Identity**). *The Cauchy identity for Schur functions in infinitely many variables and its dual state that*

$$\sum_{\lambda} s_{\lambda}(x)s_{\lambda'}(y) = \sum_{\lambda} m_{\lambda}(x)e_{\lambda}(y) = \prod_{i,j}(1+x_iy_j), \tag{1.72}$$

where the sum is taken over all partitions λ, and $h_{\lambda}(x), e_{\lambda}(x)$ denote the complete homogeneous symmetric polynomials and elementary symmetric polynomials respectively.

If the sum is taken over the products of Schur polynomials in n variables $(x_1,\ldots,x_n)$, the sum includes only partitions of length $l(\lambda) \leq n$ since, otherwise the Schur polynomial vanish.

Further Identities

Lemma 1.2.22 ([312]). *The Schur polynomials can also be computed via a specialization of a formula for Hall–Littlewood polynomials*

$$s_{\lambda}(x_1,\ldots,x_n) = \sum_{\omega\in S_n/S_n^{\lambda}} \omega\left(x^{\lambda}\prod_{\lambda_i>\lambda_j}\frac{x_i}{x_i-x_j}\right) \tag{1.73}$$

where S_n^{λ} is the subgroup of permutations such that $\lambda\omega(i) = \lambda_i$ for all i, and ω acts on variables by permuting indices.

Lemma 1.2.23 ([312], **The Murnagham–Nakayama Rule**). *The Murnaghan–Nakayama rule expresses a product of a power-sum symmetric polynomial with a Schur polynomial, in terms of Schur polynomials:*

$$p_r \cdot s_{\lambda} = \sum_{\mu}(-1)^{ht(\mu/\lambda)+1}s_{\mu} \tag{1.74}$$

where the sum is over all partitions μ such that μ/λ is a rim-hook of size r and $ht(\mu/\lambda)$ is the number of rows in the diagram μ/λ.

Remark 1.2.24. *Let $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k)$ be a partition of $n = \lambda_1 + \cdots + \lambda_k$. It is customary to interpret λ graphically as a Young diagram, namely a left-justified array of square cells with k rows of lengths $\lambda_1,\ldots,\lambda_k$. A (standard) Young tableau of shape λ is a filling of the n cells of the Young diagram with all the integers $\{1,\ldots,n\}$, with no repetition, such that each row and each column form increasing sequences. For the cell in position (i,j), in the ith row and jth column, the hook $H_{\lambda}(i,j)$ is the set of cells (a,b) such that $a = i$ and $b \geq j$ or $a \geq i$ and $b = j$. The hook length $h_{\lambda}(i,j)$ is the number of cells in $H_{\lambda}(i,j)$.*

Definition 1.2.10. *[410] The hook length formula expresses the number of standard Young tableaux of shape λ, denoted by f^{λ} or d_{λ}, as*

$$f^{\lambda} = \frac{n!}{\prod h_{\lambda}(i,j)},$$

where the product is over all cells (i,j) of the Young diagram.

Lemma 1.2.25 ([339] **The Littlewood–Richardson rule and Pieri's formula**). *The Littlewood–Richardson coefficients depend on three partitions, say λ,μ,ν, of which λ and μ describe the Schur functions being multiplied, and ν gives the Schur function of which this is the coefficient in the linear combination; in other words they are the coefficients $c^{\nu}_{\lambda,\mu}$ such that*

$$s_\lambda s_\mu = \sum_\nu c^{\nu}_{\lambda,\mu} s_\nu. \tag{1.75}$$

The Littlewood–Richardson rule states that $c^{\nu}_{\lambda,\mu}$ is equal to the number of Littlewood–Richardson tableaux of skew shape ν/λ and of weight μ, [41, 194, 313].

Pieri's formula is a special case of the Littlewood–Richardson rule, which expresses the product $h_r s_\lambda$ in terms of Schur polynomials. The dual version expresses $e_r s_\lambda$ in terms of Schur polynomials.

Lemma 1.2.26 ([440] **Specializations**). *Evaluating the Schur polynomial s_λ in $(1,1,\ldots,1)$ gives the number of semi-standard Young tableaux of shape λ with entries in $1,2,\ldots,n$. One can show, by using the Weyl character formula for example, that*

$$s_\lambda(1,1,\ldots,1) = \prod_{1\le i<j\le n} \frac{\lambda_i-\lambda_j+j-i}{j-i}. \tag{1.76}$$

In this formula, λ, the tuple indicating the width of each row of the Young diagram, is implicitly extended with zeros until it has length n. The sum of the elements λ_i is d. See also the Hook length formula which computes the same quantity for fixed λ.

Proof. Let $x_1,\ldots x_n$ be indeterminates and using the Schur determinantal formula (1.65):

$$s_\lambda(x_1,\ldots,x_n) = \frac{\det\left(x_i^{\lambda_j+n-j}\right)_{i,j=1}^{n}}{\det\left(x_i^{n-j}\right)_{i,j=1}^{n}} \tag{1.77}$$

First and foremost, it is not possible to evaluate $s_\lambda(\mathbf{x})$ since as $x_i\to 1$ in (1.77) since it leads to the form $0/0$.

Fixing x to be a new indeterminate, arbitrarily, and setting $x_i = x^{i-1}$, then it follows from (1.77) that we have

$$s_\lambda(1,x,\ldots,x^{n-1}) = \frac{\det\left(x^{i(\lambda_j+n-j)}\right)_{i,j=1}^{n}}{\det\left(x^{i(n-j)}\right)_{i,j=1}^{n}}. \tag{1.78}$$

Now, these two determinants are Vandermonde so that (1.78) gives

$$\begin{aligned} s_\lambda(1,x,\ldots,x^{n-1}) &= \frac{\prod_{1\le i<j\le n}\left(x^{(\lambda_i+n-i)}-x^{(\lambda_j+n-j)}\right)}{\prod_{1\le i<j\le n}\left(x^{(n-i)}-x^{(n-j)}\right)} \\ &= \prod_{1\le i<j\le n} \frac{x^{\lambda_j+n-j}}{x^{n-j}}.\frac{\left(x^{((\lambda_i-\lambda_j)-(j-i))}-1\right)}{\left(x^{(j-i)}-1\right)}. \end{aligned}$$

Applying l'Hôpital's rule where $\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}$ provided the derivatives exist, then it follows that $s_\lambda(1,1,\ldots,1)=\lim_{x\to 1}s_\lambda(1,x,\ldots,x^{n-1})$

$$
\begin{aligned}
&= \prod_{1\le i<j\le n}\left[\frac{(\lambda_j+n-j)\times x^{\lambda_j+n-j-1}\left(x^{((\lambda_i-\lambda_j)-(j-i))}-1\right)}{(n-j)x^{n-j-1}\times(x^{(j-i)}-1)}\right|_{x=1}\\
&\quad+\frac{((\lambda_i-\lambda_j)-(j-i))\times x^{\lambda_j+n-j}\times x^{((\lambda_i-\lambda_j)-(j-i))-1}}{(j-i)\times x^{n-j}\times x^{j-i-1}}\Bigg]\Bigg|_{x=1}\\
&= \prod_{1\le i<j\le n}\left[\frac{(\lambda_j+n-j)\times 1\times(1-1)+((\lambda_i-\lambda_j)-(j-i))\times 1\times 1}{(n-j)\times 1\times(1-1)+(j-i)\times 1}\right]\\
&= \prod_{1\le i<j\le n}\left[\frac{(\lambda_i-\lambda_j)-(j-i)}{(j-i)}\right].
\end{aligned}
$$

□

Example 1.2.27 ([440])**.** *Consider the case $n=3,d=4$. Using Ferrers diagrams or some other method, we find that there are just four partitions of 4 into at most three parts. We have*

$$
\begin{aligned}
s_{(2,1,1)}(x_1,x_2,x_3)&=\frac{1}{\Delta}\det\begin{bmatrix}x_1^4 & x_2^4 & x_3^4\\ x_1^2 & x_2^2 & x_3^2\\ x_1 & x_2 & x_3\end{bmatrix}=x_1x_2x_3\,(x_1+x_2+x_3)\\
s_{(2,2,0)}(x_1,x_2,x_3)&=\frac{1}{\Delta}\det\begin{bmatrix}x_1^4 & x_2^4 & x_3^4\\ x_1^3 & x_2^3 & x_3^3\\ 1 & 1 & 1\end{bmatrix}\\
&=x_1^2x_2^2+x_1^2x_3^2+x_2^2x_3^2+x_1^2x_2x_3+x_1x_2^2x_3+x_1x_2x_3^2
\end{aligned}
$$

and so on, where Δ is the Vandermonde determinant $a_{(2,1,0)}(x_1,x_2,x_3)$. Summarizing:

$$
\begin{aligned}
s_{(2,1,1)}&=e_1e_3\\
s_{(2,2,0)}&=e_2^2-e_1e_3\\
s_{(3,1,0)}&=e_1^2e_2-e_2^2-e_1e_3\\
s_{(4,0,0)}&=e_1^4-3e_1^2e_2+2e_1e_3+e_2^2.
\end{aligned}
$$

Every homogeneous degree-four symmetric polynomial in three variables can be expressed as a unique linear combination of these four Schur polynomials, and this combination can again be found using a Gröbner basis for an appropriate elimination order. For example,

$$\phi(x_1,x_2,x_3)=x_1^4+x_2^4+x_3^4$$

is obviously a symmetric polynomial which is homogeneous of degree four, and we have

$$\phi=s_{(2,1,1)}-s_{(3,1,0)}+s_{(4,0,0)}.\phi=s_{(2,1,1)}-s_{(3,1,0)}+s_{(4,0,0)}.$$

1.3 Orthogonal Polynomials

In this section we briefly highlight the occurrence of Vandermonde matrix and its determinant as basis for derivation of classical orthogonal polynomials.

There is a very close link between random matrix theory (RMT) and the classical theory of orthogonal polynomials [2, 12, 15, 130, 281, 454] via the Vandermonde determinant which is expressed by

$$v_n(\mathbf{x}) = \prod_{1\leq i<j\leq n}(x_i - x_j) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix} \tag{1.79}$$

It should be noted that the factor product in (1.79) occurs to the power, $\beta = 1,2,4$, in the joint eigenvalue probability density (1.153) representing each of the basic Gaussian random matrix ensembles as being orthogonal, unitary and symplectic, respectively. These ensembles, their applications cannot be overemphasised as already briefly explain in the previous section. The Vandermonde determinant in (1.79) implies that the joint eigenvalues density (1.153) can be expressed as generalised determinant or a product of determinants. The usefulness of the Vandermonde determinant and the joint eigenvalue density can be more enriched by re-expressing the determinant in (1.79) and the corresponding entries of the Vandermonde matrix using other polynomial bases the special case being

$$\pi_k(x) = x^k + \text{lower order terms} \tag{1.80}$$

for $k = 0,1,2,\ldots$. This is a monic polynomials with basis monomials of the form $x^k, k = 0,1,2,\ldots$. Thus, the elementary column operations on the Vandermonde determinant in (1.79) can equally be expressed using monomials in (1.80) as

$$\prod_{1\leq i<j\leq n}(x_i - x_j) = \begin{vmatrix} \pi_0(x_1) & \pi_0(x_2) & \dots & \pi_0(x_n) \\ \pi_1(x_1) & \pi_1(x_2) & \dots & \pi_1(x_n) \\ \pi_2(x_1) & \pi_2(x_2) & \dots & \pi_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{n-1}(x_1) & \pi_{n-1}(x_2) & \dots & \pi_{n-1}(x_n) \end{vmatrix}. \tag{1.81}$$

where $\pi_i(x_j), i = 0,1,2,\ldots,n-1; j = 1,2,\ldots,n$ form a suitable polynomial basis. Thus in studying the Gaussian ensembles, it turns out that the monic orthogonal polynomials with respect to the Gaussian weight $\pi_i(x_j) = \exp\left(-x_j^2\right), i = 0,1,2,\ldots,n-1; j = 1,2,\ldots,n$ and with measure on $\mathbb{R}$, the Hermite polynomials are very useful in this case.

1.3.1 Determinantal Representation of Orthogonal Polynomials

Let μ be a measure on the real field $\mathbb{R}$ with a moment generating function in the neighbourhood of 0. Assuming that μ is not fully supported by any finite subset of $\mathbb{R}$, then the space $\mathrm{L}^2(\mu)$ is

infinite-dimensional, and thus the monomials of the basis $\mathscr{B}(x) = \{1, x, x^2, x^3, x^4, \ldots\}$ are linearly independent as elements of $\mathrm{L}^2(\mu)$. Thus, for any such measure, it follows that the monomial basis $\mathscr{B}(x) = \{1, x, x^2, x^3, x^4, \ldots\}$ can be orthogonalized by the Gram–Schmidt algorithm to produce orthogonal sets $\{\pi_n(x)\}_{n\geq 0}$ of $\mathrm{L}^2(\mu)$ consisting of polynomials $\pi_n(x)$ of degree n with the real coefficients such that

$$\langle \pi_n, \pi_m \rangle = \int_{\mathbb{R}} \pi_m(x)\pi_n(x)d\mu(x) = \delta_{m,n}. \tag{1.82}$$

Up to signs, this condition uniquely specifies the polynomials $\pi_n(x)$ and it is conventional to choose the signs so that the leading coefficients γ_n are positive so that the resulting polynomials are the normalized orthogonal polynomials associated with the measure μ. Thus we can write

$$p_n(x) = \gamma_n \pi_n(x), \qquad \pi_n(x) = x^n + \text{lower order terms}. \tag{1.83}$$

Proposition 1.3.1. *The orthogonal polynomials $p_n(x)$ and $\pi_n(x)$ are expressed by determinantal formulas, valid for $n \geq 1$*

$$\begin{aligned} \pi_n(x) &= \mathrm{D}_n(x)/\mathrm{D}_{n-1}, \\ p_n(x) &= \mathrm{D}_n(x)/\sqrt{\mathrm{D}_n\mathrm{D}_{n-1}}, \\ \gamma_n &= \sqrt{\mathrm{D}_{n-1}/\mathrm{D}_n}, \end{aligned}$$

where

$$\mathrm{D}_n(x) = \det \begin{bmatrix} m_0 & m_1 & m_2 & \ldots & m_{n-1} & m_n \\ m_1 & m_2 & m_3 & \ldots & m_n & m_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \ldots & m_{2n-2} & m_{2n-1} \\ 1 & x & x^2 & \ldots & x^{n-1} & x^n \end{bmatrix}, \tag{1.84}$$

$$\mathrm{D}_n = \det \begin{bmatrix} m_0 & m_1 & m_2 & \ldots & m_{n-1} \\ m_1 & m_2 & m_3 & \ldots & m_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \ldots & m_{2n-2} \end{bmatrix}, \tag{1.85}$$

and m_n is the nth moment of the measure μ given by

$$m_n = \int_{\mathbb{R}} x^n d\mu(x). \tag{1.86}$$

Consequently, the polynomials $p_n(x)$ and $\pi_n(x)$ have real coefficients. Moreover,

$$\mathrm{D}_n = \prod_{k=1}^{n} \gamma_k^{-2} = \prod_{k=0}^{n} \gamma_k^{-2}. \tag{1.87}$$

We notice that the matrices $(a_{i,j})$ that occur in (1.85) for which $a_{i,j} = a_{i+j}$, are called Hankel matrices while matrices that are constant down diagonals such that $a_{i,j} = a_{j-i}$ are called Toeplitz matrices.

The monic orthogonal polynomials $\pi_n(x)$ satisfy a simple recursive system of linear equations called the three-term recurrence. Thus multiplying the monic polynomial $\pi_n(x)$ by x, the resulting polynomial $x\pi_n(x)$ is again monic and is of degree $n+1$ and it must equal $\pi_{n+1}(x)+$ lower order polynomials which only involve $\pi_n(x)$ and $\pi_{n-1}(x)$.

Proposition 1.3.2. *There exists real terms a_n and b_n so that*

$$x\pi_n(x) = \pi_{n+1}(x) + a_n\pi_n(x) + b_n\pi_{n-1}(x). \tag{1.88}$$

Moreover, the coefficients b_n obey

$$b_n = \frac{\gamma_{n-1}^2}{\gamma_n^2} \implies \gamma_n^2 = \prod_{i=1}^{n} b_n^{-1}. \tag{1.89}$$

1.3.2 Vandermonde Determinant and the Christoffel–Darboux Formula

As will be discussed in the next section, the Hermite polynomials are both monic and L^2-norm 1 relative to the Gaussian measure on $\mathbb{R}$. Some other classical orthogonal polynomial systems have a similar property, and for such systems it is more convenient to work with normalized orthogonal polynomials $p_n(x)$ than the monic polynomials $pi_n(x)$ given by (1.83). We notice that from (1.79) which contains the Vandermonde determinant term $\prod_{1\le i<j\le n}(x_j - x_i)$ as a the determinant of the matrix of monic polynomials, it can easily be converted to a formula involving the normalized orthogonal polynomials $\pi_n(x)$. This can be easily achieved by replacing $\pi_k(x)$ by $p_k(x)$ in the determinant by just multiplying the kth row by the factor γ_k^{-1}. Thus

$$\begin{aligned}
\Delta_n &= \left(\prod_{k=0}^{n-1}\gamma_k^{-1}\right)\det\begin{bmatrix} p_0(x_1) & p_0(x_2) & \dots & p_0(x_n)\\ p_1(x_1) & p_1(x_2) & \dots & p_1(x_n)\\ p_2(x_1) & p_2(x_2) & \dots & p_2(x_n)\\ \vdots & \vdots & \ddots & \vdots\\ p_n(x_1) & p_n(x_2) & \dots & p_n(x_n)\end{bmatrix} \\
&= \sqrt{\mathrm{D}_{n-1}}\det\begin{bmatrix} p_0(x_1) & p_0(x_2) & \dots & p_0(x_n)\\ p_1(x_1) & p_1(x_2) & \dots & p_1(x_n)\\ p_2(x_1) & p_2(x_2) & \dots & p_2(x_n)\\ \vdots & \vdots & \ddots & \vdots\\ p_n(x_1) & p_n(x_2) & \dots & p_n(x_n)\end{bmatrix}
\end{aligned} \tag{1.90}$$

This leads to the useful formula of the square Vandermonde determinant Δ_n^2 which appears especially in the joint eigenvalue density of the Gaussian unitary ensembles, GUE, and the complex Wishart ensembles.

Lemma 1.3.1. *The square Vandermonde determinant Δ_n^2 can be expressed by*

$$\Delta_n^2 = \mathrm{D}_{n-1} \det\left(\mathrm{K}_n(x_i,x_j)\right)_{1\le i,j\le n} \tag{1.91}$$

where

$$\mathrm{K}_n(x_i,x_j) = \sum_{k=0}^{n} p_k(x_i)p_k(x_j) \tag{1.92}$$

This leads to the import result of the Christoffel–Darboux formula that can be stated in the following lemma

Lemma 1.3.2. *The Christoffel–Darboux formula can be expressed by*

$$\mathrm{K}_n(x,y) = \frac{\pi_n(x)\pi_{n-1}(y) - \pi_n(y)\pi_{n-1}(x)}{x-y}. \tag{1.93}$$

The proof of this Lemma follows directly from the routine calculation of the 3-term recurrence.

1.3.3 Basic Theory of Orthogonal Polynomials

Definition 1.3.1. *A sequence of polynomials $\{\pi_n(x)\}_{n=0}^{\infty}$ with degree$[\pi_n(x)] = n$ for each n is called orthogonal with respect to the weight function $\omega(x)$ on the interval (a,b) with $a < b$ if*

$$\int_a^b \omega(x)\pi_m(x)\pi_n(x)dx = h_n\delta_{mn} \;\; with \;\; \delta_{mn} = \begin{cases} 0, & if \;\; m \neq n \\ 1, & if \;\; m = n. \end{cases}$$

The weight function $\omega(x)$ should be continuous and positive on (a,b) such that the moments

$$\mu_n = \int_a^b \omega(x)x^n dx, \;\; n = 0,1,2,\ldots$$

exist. Then the integral

$$\langle f,g\rangle = \int_a^b \omega(x)f(x)g(x)dx$$

denotes an inner product of the polynomials f and g. The interval (a,b) is called the interval of orthogonality. The interval (a,b) can be finite or infinite.

If $h_n = 1$ for each $n \in \{0,1,2,\ldots\}$ the sequence of polynomials is called orthonormal, and if

$$\pi_n = k_n x^n + \text{lower order terms}$$

and with with $k_n = 1$ for each $n \in \{0,1,2,\ldots\}$ gives the monic polynomials as in (1.80).

Example 1.3.3. *Taking $\omega(x) = 1$ and $(a,b) = (0,1)$ and using Gram–Schmidt orthogonalisation we can generate the orthogonal polynomials as follows. Starting with the monomial basis $\{1,x,x^2,x^3,x^4,\ldots\}$ and choosing $\pi_0(x) = 1$. Then we have*

$$\pi_1(x) = x - \frac{\langle x,\pi_0\rangle}{\langle \pi_0,\pi_0\rangle}\pi_0(x) = x - \frac{\langle x,1\rangle}{\langle 1,1\rangle} = x - \frac{1}{2},$$

since $\langle 1,1\rangle = \int_0^1 dx = 1$ *and* $\langle x,1\rangle = \int_0^1 x dx = x - \frac{1}{2}$.

Further we have

$$\begin{aligned}\pi_2(x) &= x^2 - \frac{\langle x^2,\pi_0\rangle}{\langle \pi_0,\pi_0\rangle}\pi_0 - \frac{\langle x^2,\pi_1(x)\rangle}{\langle \pi_1,\pi_1\rangle}\pi_1(x)\\ &= x^2 - \frac{\langle x^2,1\rangle}{\langle 1,1\rangle} - \frac{\langle x^2,x-\frac{1}{2}\rangle}{\langle x-\frac{1}{2},x-\frac{1}{2}\rangle}(x-\frac{1}{2}) = x^2 - x + \frac{1}{6},\end{aligned}$$

since $\langle x^2,1\rangle = \int_0^1 x^2 dx = \frac{1}{3}$, $\langle x^2,x-\frac{1}{2}\rangle = \int_0^1 x^2(x-\frac{1}{2})dx = \frac{1}{12}$ *and* $\langle x-\frac{1}{2},x-\frac{1}{2}\rangle = \int_0^1 (x-\frac{1}{2})^2 dx = \frac{1}{12}$.

Repeating this process for higher degree terms with respect to the weight $\omega(x) = 1$ we obtain the following sequence of orthogonal polynomials in the interval $(0,1)$:

$$\begin{aligned}&\pi_0(x) = 1, \quad \pi_1(x) = x - \frac{1}{2}, \quad \pi_2(x) = x^2 - x + \frac{1}{6},\\ &\pi_3(x) = x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}, \quad \pi_4(x) = x^4 - 2x^3 + \frac{9}{7}x^2 - \frac{2}{7}x + \frac{1}{70},\\ &\pi_5(x) = x^5 - \frac{5}{2}x^4 + \frac{20}{9}x^3 - \frac{5}{6}x^2 + \frac{5}{42}x - \frac{1}{252},\ldots\end{aligned}$$

The Classical Orthogonal Polynomials

The main classical orthogonal polynomials are named after the mathematicians Charles Hermite (1822–1901), Edmand Nicolas Laguerre (1834–1866), Andrien–Marie Legendre (1752–1833) and Carl Gustav Jacobi (104-1857) [454], as illustrated in the table below:

These classical orthogonal polynomials satisfy an orthogonality relation, a tree term recurrence relation, a second order differential equation and the famous Rodriguez formula [2].

Hermite Orthogonal Polynomials

The Hermite polynomials are orthogonal on the interval $(-\infty,\infty)$ with respect to the Gaussian weight $\omega(x) = e^{-\frac{1}{2}x^2}$. The polynomials can be defined based on their Rodriguez formula:

$$\mathrm{H}_n(x) = \frac{(-1)^n}{\omega(x)}\mathrm{D}^n\omega(x) = (-1)^n e^{x^2}\mathrm{D}^n e^{-\frac{1}{2}x^2}, \ \ n = 0,1,2,\ldots \tag{1.94}$$

Name	$p_n(x)$	weight, $\omega(x)$	(a,b)
Hermite	$\mathrm{H}_n(x)$	e^{-x^2}	$(-\infty,\infty)$
Laguerre	$\mathrm{L}_n^{(\alpha)}(x)$	$x^\alpha e^{-x}$	$(0,\infty)$
Jacobi	$\mathrm{J}_n^{(\alpha,\beta)}(x)$	$(1-x)^\alpha(1+x)^\beta$	$(-1,1)$
Legendre	$\mathrm{P}_n^{(\alpha)}(x)$	1	$(-1,1)$

Table 1.2: Classical orthogonal polynomials.

where the differentiation operator D is differential operator defined by $\mathrm{D} = \dfrac{d}{dx}$ from which we have the Leibniz's rule

$$\mathrm{D}^n[f(x)g(x)] = \sum_{k=0}^{n} \binom{n}{k} \mathrm{D}^k f(x) \mathrm{D}^{n-k} g(x), \quad n = 0,1,2,\ldots \tag{1.95}$$

which is the generalization of the product rule. Since $\mathrm{D}^{n+1} = \mathrm{D}\mathrm{D}^n$, it follows that

$$\begin{aligned}\mathrm{D}^{n+1}\omega(x) &= (-1)^n \mathrm{D}[\omega(x)\mathrm{H}_n(x)] = (-1)^n[\omega'(x)\mathrm{H}_n(x) + \omega(x)\mathrm{H}_n'(x)] \\ &= (-1)^{n+1}\omega(x)[2x\mathrm{H}_n(x) - \mathrm{H}_n'(x)], \quad n = 0,1,2,\ldots,\end{aligned}$$

which gives the relation

$$\mathrm{H}_{n+1}(x) = 2x\mathrm{H}_n(x) - \mathrm{H}_n'(x), \quad n = 0,1,2,\ldots \tag{1.96}$$

From Equation (1.94), it implies $\mathrm{H}_0(x) = 1$. Thus, by induction on Equation (1.96) it follows that $\mathrm{H}_n(x)$ is a polynomial of degree n.

To find the three term recurrence relation we proceed from

$$\omega(x) = e^{-x^2} \implies \omega'(x) = -2x\omega(x).$$

Applying Leibniz's rule (1.95) we obtain

$$\mathrm{D}^{n+1}\omega(x) = \mathrm{D}^n\omega'(x) = \mathrm{D}^n[-2x\omega(x)] = -2xD^n\omega(x) - 2n\mathrm{D}^{n-1}\omega(x),$$

which leads to

$$\mathrm{H}_{n+1}(x) = 2x\mathrm{H}_n(x) - 2n\mathrm{H}_{n-1}(x), \quad n = 1,2,\ldots \tag{1.97}$$

Combining Equations (1.96) and (1.97) we find that

$$\mathrm{H}_n'(x) = 2n\mathrm{H}_{n-1}(x), \quad n = 1,2,3,\ldots \tag{1.98}$$

Differentiating Equation (1.96) gives

$$\mathrm{H}_{n+1}'(x) = 2x\mathrm{H}_n'(x) + 2\mathrm{H}_n(x) - \mathrm{H}_n''(x), \quad n = 0,1,2,\ldots \tag{1.99}$$

Now, using (1.98) and (1.99) to obtain

$$2(n+1)\mathrm{H}_n(x) = 2x\mathrm{H}_n'(x) + 2\mathrm{H}_n(x) - \mathrm{H}_n''(x), \;\; n = 0,1,2,\ldots \tag{1.100}$$

This implies that the Hermite polynomial $\mathrm{H}_n(x)$ satisfies the ordinary second order linear differential equation given by

$$y''(x) - 2xy'(x) + 2ny(x) = 0, \;\; n \in \{0,1,2,3,\ldots\} \tag{1.101}$$

The Hermite polynomials has the explicit expression given by

$$H_n(x) = \begin{cases} n! \sum\limits_{l=0}^{\frac{n}{2}} \dfrac{(-1)^{\frac{n}{2}-l}}{(2l)!\left(\frac{n}{2}-l\right)!}(2x)^{2l} & \text{for even } n, \\ n! \sum\limits_{l=0}^{\frac{n-1}{2}} \dfrac{(-1)^{\frac{n-1}{2}-l}}{(2l+1)!\left(\frac{n-1}{2}-l\right)!}(2x)^{2l+1} & \text{for odd } n. \end{cases} \tag{1.102}$$

These two equations may be combined into one using the floor function:

$$H_n(x) = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)!}(2x)^{n-2m}. \tag{1.103}$$

The Hermite polynomials $\mathrm{H}_n x$ defined in (1.94) satisfy the orthogonality condition such that

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2} \mathrm{H}_m(x)\mathrm{H}_n(x)dx = 2^n n! \delta_{mn}, \;\; m,n \in \{0,1,2,\ldots\} \tag{1.104}$$

The Hermite polynomials (1.94) have the generating function given by

$$e^{2xt-t^2} = \sum_{n=0}^{\infty} \frac{\mathrm{H}_n(x)}{n!} t^n \tag{1.105}$$

Laguerre Orthogonal Polynomials

The Laguerre polynomials are orthogonal on the interval $(0,\infty)$ with weight $\omega(x) = x^\alpha e^{-x}$. The Laguerre polynomials can be defined as:

$$\mathrm{L}_n^{(\alpha)} = \frac{1}{n!}\frac{1}{\omega(x)}\mathrm{D}^n[\omega(x)x^n] = \frac{1}{n!}e^x x^{-\alpha}\mathrm{D}^n[e^{-x}x^{n+\alpha}], \;\; n = 0,1,2,3,\ldots \tag{1.106}$$

Employing the Leibniz' rule in (1.95) to (1.106) gives

$$\begin{aligned} \mathrm{D}^n[e^{-x}x^{x+\alpha}] &= \sum_{k=0}^{n} \binom{n}{k} \mathrm{D}^k e^{-x} \mathrm{D}^{n-k} x^{n+\alpha} \\ &= e^{-x}x^\alpha \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{\Gamma(n+\alpha+1)}{\Gamma(k+\alpha+1)} x^k. \end{aligned}$$

This simplifies to

$$\mathrm{L}_n^{(\alpha)}(x) = \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!}, \;\; n = 0,1,2,3,\ldots \tag{1.107}$$

where

$$\binom{n+\alpha}{n-k} = \frac{\Gamma(n+\alpha+1)}{(n-1)!\Gamma(k+\alpha+1)} = \frac{(k+\alpha+1)_{n-k}}{(n-k)!}, \;\; k = 0,1,2,3,\ldots,n.$$

This confirms the fact that $\mathrm{L}_n^{((\alpha)}(x)$ is a polynomial of degree n.

The orthogonality relation for Laguerre polynomials is given by

$$\int_0^\infty e^{-x} x^\alpha \mathrm{L}_m^{(\alpha)}(x) \mathrm{L}_n^{(\alpha)}(x) dx = \frac{\Gamma(n+\alpha+1)}{n!} \delta_{mn}, \;\; \alpha > -1 \tag{1.108}$$

for $m,n \in \{0,1,2,\ldots\}$.

The moment generating function of the Laguerre polynomials (1.106) is defined by

$$(1-t)^{-\alpha-1} \exp\left(-\frac{xt}{1-t}\right) = \sum_{n=0}^{\infty} \mathrm{L}_n^{(\alpha)}(x) t^n. \tag{1.109}$$

The three term recurrence relationship for the Laguerre polynomials (1.106) is given by

$$(n+1)\mathrm{L}_{n+1}^{(\alpha)}(x) + (x-2n-\alpha-1)\mathrm{L}_n^{(\alpha)}(x) + (n+\alpha)\mathrm{L}_{n-1}^{(\alpha)}(x) = 0, \;\; n = 1,2,3,\ldots \tag{1.110}$$

The Laguerre polynomial $\mathrm{L}_n^{(\alpha)}(x)$ satisfies the ordinary second order linear differential equation

$$xy''(x) + (\alpha+1-x)y'(x) + ny(x) = 0, \;\; n = 0,1,2,\ldots \tag{1.111}$$

Jacobi Orthogonal Polynomials

The Jacobi polynomials are orthogonal on the interval $(-1,1)$ with respect to the Beta distribution with weight $\omega(x) = (1-x)^\alpha (1+x)^\beta$. The Jacobi polynomials can be expressed as:

$$\begin{aligned} \mathrm{P}_n^{(\alpha,\beta)}(x) &= \frac{(-1)^n}{2^n n!} \frac{1}{\omega(x)} \mathrm{D}^n[\omega(x)(1-x^2)^n] \\ &= \frac{(-1)^n}{2^n n!} (1-x)^{-\alpha}(1+x)^{-\beta} \mathrm{D}^n\left[(1-x)^{n+\alpha}(1+x)^{n+\beta}\right], \end{aligned} \tag{1.112}$$

for $n = 0,1,2,3,\ldots$ Applying the Leibniz' rule in (1.95) to (1.112) gives

$$\begin{aligned} \mathrm{D}^n\left[(1-x)^{n+\alpha}(1+x)^{n+\beta}\right] &= \sum_{k=0}^{n} \binom{n}{k} \mathrm{D}^k (1-x)^{n+\alpha} \mathrm{D}^{n-k}(1+x)^{n+\beta} \\ &= n! \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{k} \binom{n+\beta}{n-k} (1-x)^{n+\alpha-k}(1+x)^{\beta+k}. \end{aligned}$$

for $n = 0, 1, 2, \ldots$ It follows that

$$\mathrm{P}_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n} \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{k} \binom{n+\beta}{n-k} (1-x)^{n-k} (1+x)^k, \;\; n = 0, 1, 2, 3, \ldots \tag{1.113}$$

This justifies the fact that $\mathrm{P}_n^{(\alpha,\beta)}(x)$ is a polynomial of degree n.

The orthogonality relation for Jacobi polynomials is given by

$$\int_{-1}^{1} (1-x)^{\alpha} (1+x)^{\beta} \mathrm{P}_m^{(\alpha,\beta)}(x) \mathrm{P}_n^{(\alpha,\beta)}(x) dx = \frac{2^{\alpha+\beta+1} \Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{(2n+\alpha+\beta+1) \Gamma(n+\alpha+\beta+1) n!} \delta_{mn}, \tag{1.114}$$

for $\alpha > -1, \beta > -1$ $m, n \in \{0, 1, 2, \ldots\}$.

The three term recurrence relationship for the Jacobi polynomials (1.112) is given by

$$\begin{aligned} & 2n(n+\alpha+\beta)(2n+\alpha+\beta-2) \mathrm{P}_n^{\alpha,\beta}(x) \\ =& (2n+\alpha+\beta-1) \left\{ (2n+\alpha+\beta)(2n+\alpha+\beta-2)x + \alpha^2 - \beta^2 \right\} \mathrm{P}_{n-1}^{\alpha,\beta}(x) \\ & - 2(n+\alpha-1)(n+\beta-1)(2n+\alpha+\beta) \mathrm{P}_{n-2}^{\alpha,\beta}(x), \quad n = 2, 3, 4, \ldots \end{aligned} \tag{1.115}$$

The Jacobi polynomial $\mathrm{P}_n^{(\alpha,\beta)}(x)$ satisfies the ordinary second order linear differential equation

$$(1-x^2) y''(x) + [\beta - \alpha - (\alpha+\beta+2)x] y'(x) + n(n+\alpha+\beta+1) y(x) = 0, \;\; n = 0, 1, 2, \ldots \tag{1.116}$$

The moment generating function of the Jacobi polynomials (1.112) is defined by

$$\frac{2^{\alpha+\beta}}{R(1+R-t)^{\alpha}(1+R+t)^{\beta}} = \sum_{n=0}^{\infty} \mathrm{P}_n^{(\alpha,\beta)}(x) t^n, \;\; R = \sqrt{1-2xt+t^2}. \tag{1.117}$$

Legendre Orthogonal Polynomials

The Legendre polynomials are a special case of Jacobi polynomials with $(\alpha = 0 = \beta)$ with respect to the weight $\omega(x) = 1$. The polynomials can be written using their corresponding Rodriguez formula:

$$\mathrm{P}_n(x) = \frac{(-1)^n}{2^n n!} \frac{1}{\omega(x)} \mathrm{D}^n [\omega(x)(1-x^2)^n] = \frac{(-1)^n}{2^n n!} (1-x^2)^n, \;\; n = 0, 1, 2, 3, \ldots \tag{1.118}$$

The orthogonality relation for Legendre polynomials is given by

$$\int_{-1}^{1} \mathrm{P}_m(x) \mathrm{P}_n(x) dx = \frac{2}{2n+1} \delta_{mn}, \;\; m, n \in \{0, 1, 2, \ldots\} \tag{1.119}$$

The moment generating function of the Jacobi polynomials (1.118) is defined by

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{\infty} \mathrm{P}_n(x) t^n. \tag{1.120}$$

The three term recurrence relationship for the Legendre polynomials (1.118) is given by

$$(n+1)\mathrm{P}_{n+1}(x) - (2n+1)x\mathrm{P}_n(x) + n\mathrm{P}_{n-1}(x) = 0, \;\; n = 1,2,3,\ldots \tag{1.121}$$

The Legendre polynomial $\mathrm{P}_n(x)$ satisfies the ordinary second order linear differential equation

$$(1-x^2)y''(x) - 2xy'(x) + n(n+1)y(x), \;\; n = 0,1,2,\ldots \tag{1.122}$$

1.4 Applications and Occurrences of the Vandermonde Matrix and its Determinant

In this section we give a brief overview of the occurrences and applications of the Vandermonde matrix and its determinant in real life as applies to Mathematical analysis and computation.

The Vandermonde matrix and its corresponding determinant has numerous numerical applications. These applications are mainly due to the systematic structure of the matrix and simplicity of the expression of the determinant. Here we will discuss applications where we want to optimize the Vandermonde determinant over various surfaces. We will construct a generalized Vandermonde interpolating polynomial using divided differences interpolation polynomial based on the generalized Vandermonde determinant approach. Some results regarding the appropriateness for this method for curve-fitting and approximation will be discussed.

Polynomial interpolation is mostly used when the data set we wish to interpolate is small. The main reason for this is the instability of the interpolation method. One example of this is Runge's phenomenon that shows that when certain functions are approximated by polynomial interpolation fitted to equidistantly sampled points will sometimes lose precision when the number of interpolating points is increased. One way to predict this instability of polynomial interpolation is that the conditional number of the Vandermonde matrix can be very large for equidistant points [184, 187, 188, 189, 190]. There are different ways to mitigate the issue of stability, for example choosing data points that minimize the conditional number of the relevant matrix [184] or by choosing a polynomial basis that is more stable for the given set of data points such as Bernstein polynomials in the case of equidistant points [361]. Other polynomial schemes can also be considered, for instance by interpolating with different basis functions in different intervals, for example, using polynomial splines [416].

While the instability of polynomial interpolation does not prevent it from being useful for analytical examinations it is generally considered impractical when there is noise present or when calculations are performed with limited precision. Often interpolating polynomials are not constructed by inverting the Vandermonde matrix or calculating the Lagrange basis polynomials, instead a more computationally efficient method such as Newton interpolation or Neville's algorithm are used [383]. There are some variants of Lagrange interpolation, such as barycentric Lagrange interpolation, that have good computational performance [39].

In applications where the data is noisy it is often suitable to use least squares fitting, instead of interpolation.

Since most our research study is mainly centred at the extreme points of the Vandermonde determinant as will be discussed in Chapters 2 to 9. It is very important at this moment to give a brief highly of the major occurrences and applications Vandermonde matrix and its determinant in mathematical context. These applications are mainly due to the systematic structure of the matrix and simplicity of the expression of the determinant as discussed in the previous sections. In most of the applications the Vandermonde determinant takes the central advantage due to the fact that it can be optimized globally, on various surfaces or boundaries. This makes the extreme points of the Vandermonde determinant of great significance in all the major applications that will be outlined in the following subsections.

1.4.1 Polynomial Interpolation

The process of fitting $n-1$ degree polynomial to the points $(x_1,y_1),(x_2,y_2),\cdots,(x_n,x_n)$ is usually referred to as polynomial interpolation [257]. If the polynomial is

$$p(x)=c_0+c_1x+c_2x^2+\cdots+c_{n-1}x^{n-1}, \tag{1.123}$$

then, the coefficients c_i may be determined by solving the equation $p(x_k)=y_k,k=1,2,\cdots,n$. The Vandermonde matrix can be used to describe this type of interpolation problem [24, 25] simply by writing the equation system given by $p(x_k)=y_k$ as a matrix equation

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \tag{1.124}$$

We observe that the system matrix is indeed a transposed Vandermonde matrix in (1.11) and that its determinant is connected to the solvability of the system (1.124). By the Vandermonde determinant identity (1.8), the determinant is non-zero if the values of x_i are distinct, in which case the coefficients of q_i are uniquely determined. In fact, $q(x)$ can be explicitly formulated in this case as follows. Let

$$Q(x)=\det\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & \cdots & x_n & x \\ x_1^2 & x_2^2 & \cdots & x_n & x^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_{n-1} & x^{n-1} \\ y_1 & y_2 & \cdots & y_n & 0 \end{bmatrix}$$

Then, $Q(x_i)$ may be found by replacing x and x_i in the last column of the matrix. Subtracting column i from the last column now produces entries of 0, except for the last entry which is $-y_i$.

Consequently, $Q(x_i) = (-\det(\mathbf{V}_n))y_i$ and so

$$p(x) = \frac{-1}{\det(\mathbf{V})}Q(x) \tag{1.125}$$

Here, the Vandermonde determinant is of obvious significance. We notice that if x_i are not distinct, then $Q(x_i) = 0$ for all i and the system (1.124) has a singular matrix. Moreover, polynomial interpolation is not meaningful if the x_i are not distinct.

There are several ways to construct the interpolating polynomial without explicitly inverting the Vandermonde matrix, for instance in [484].

The idea behind Lagrange interpolation is simple, construct a set of n polynomials $\{p_1, \cdots, p_n\}$ such that

$$p_i(x_j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

and then construct the final interpolating polynomial by the sum of these p_i weighted by the corresponding y_i.

The p_i polynomial are called *Lagrange basis* polynomial and can easily be constructed by placing the roots appropriately and then normalizing the result such that $p_i(x) = 1$, this gives the expression

$$l_i(x) = \frac{(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_i-x_1)\cdots(x_i-x_{k-1})(x_i-x_{k+1})\cdots(x_i-x_n)} \tag{1.126}$$

The explicit formula for the full interpolating polynomial is

$$p(x) = \sum_{k=1}^{n} y_k \frac{(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_i-x_1)\cdots(x_i-x_{k-1})(x_i-x_{k+1})\cdots(x_i-x_n)} \tag{1.127}$$

and from this formula the expression for the inverse of the Vandermonde matrix can be found by noting that the jth row of the inverse will consist of the coefficients of p_j, the resulting expression for the elements given in above.

The equation (1.126) is considered to give the error approximation based on Lagrange interpolating polynomial [285, 484].

Theorem 1.4.1. *The interpolating polynomial through the points* $(x_i, f(x_i))$, $i = 1, 2, \ldots, n$ *is given by*

$$P(x) = \sum_{i=1}^{n} \frac{\omega(x)}{(1-x_i)\omega'(x_i)} f(x_i) = \sum_{i=1}^{n} f(x_i) \prod_{j=1}^{n_i} \frac{x-x_j}{x_i-x_j} \equiv \mathbf{L}^\top \mathbf{U}, \tag{1.128}$$

where prime means the derivative with respect to x,

$$\mathbf{L} = \left[\prod_{j=1}^{n_i} \frac{x-x_j}{x_i-x_j} \right]^\top, \text{ and } \omega(x) = \prod_{j=1}^{n} (t-t_j).$$

By decomposition of the Vandermonde matrix in the form [277]:

$$\mathbf{V}^\top \mathbf{V}\mathbf{C} = \mathbf{V}^\top \mathbf{U} \text{ where } \mathbf{V} = [\mathbf{T}_1 \mathbf{T}_2 \dots \mathbf{T}_n]^\top,$$

and $\mathbf{T}_i$ is the vector given by $\mathbf{T}_i = [1, t_i, t_i^2, \dots, t_i^{n-1}]^\top$ and $\mathbf{U} = [f_i] = [f(x_i)]$ for simplicity. If we seek the polynomial of degree $(n-1)$ which yields the least squares fit to $\{f_i\}_{i=1}^n$, necessitates an n dimensional constant vector $\mathbf{C}$ such that $\mathbf{C}$ attains

$$\min \mathbf{C} \sum_{i=1}^{n} (f_i - \mathbf{C}^\top \mathbf{T}_i)^2.$$

Thus, we can obtain the matrix expression for (1.128) as

$$P(x) = \left[\mathbf{V}^{-1}\mathbf{U}\mathbf{T}\right] = \mathbf{T}^\top \mathrm{V}^{-1}\mathbf{U} \tag{1.129}$$

where $\mathrm{L}^\top = \mathbf{T}^\top \mathbf{V}^{-1}$ which can be directly obtained from Equations (1.14) and (1.16).

The Error in Polynomial Interpolation

Our goal here is to provide estimates on the "error" we make when interpolating data that is taken from sampling an underlying function $f(x)$. While the interpolant and the function agree with each other at the interpolation points, there is, in general, no reason to expect them to be close to each other elsewhere. Nevertheless, we can estimate the difference between them, a difference which we refer to as the interpolation error, for more details see [21, 150, 151, 230, 308, 382, 453].

We let Π_n denote the space of polynomials of degree $\leq n$, and let $C^{n+1}[a,b]$ denote the space of functions that have $n+1$ continuous derivatives on the interval $[a,b]$.

Theorem 1.4.2. *Let $f(x) \in C^{n+1}[a,b]$ and $Q_n(x) \in \Pi_n$ such that it interpolates $f(x)$ at the $n+1$ distinct points $x_0, \cdots, x_n \in [a,b]$. Then for all $x \in [a,b]$, there exists $\xi_n \in (a,b)$ such that*

$$f(x) - Q(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_n) \prod_{j=0}^{n} (x - x_j). \tag{1.130}$$

This is a well-known theorem, its proof can be found in [101] and many other texts.

Lebesgue Function and Lebesgue Constants

Given $X = \{x_j,\ j = 0, 1, 2, \cdots, n;\ \in \mathbb{N}\}$ be a set of $n+1$ distinct interpolation points (or nodes) on the real interval $[-1,1]$ such that $-1 \leq x_0 < x_1 < \cdots < x_n \leq 1$. We define a $f \in C[-1,1]$ such that when approximating f from a finite-dimensional $\mathscr{V}_n = \text{span}\{\phi_0, \phi_1, \cdots, \phi_n\}$ with $\phi_i \in C[-1,1]$ for $0 \leq i \leq n$, then there exists at least one element $p_n^* \in \mathscr{V}_n$ that is close to f. When using the $\|f\|\infty = \max\limits_{x \in [-1,1]} |f(x)|$ approximation criteria [335, 379], the element p_n^* is the closest one if the $\phi_0, \phi_1, \cdots, \phi_n$ are a Tchebychev system.

We are interested in the interpolation points x_j that make the interpolation error [101]

$$\left\| f(x) - \sum_{i=0}^{n} \alpha_i \phi_i(x_i) \right\| = \max_{x\in[-1,1]} \left| f(x) - \sum_{i=0}^{n} \alpha_i \phi_i(x_i) \right|$$

as small as possible. In other words, there is an interest in using the interpolating polynomials that are near best approximants [416].

When for instance, we consider the monomials $\phi_i(x) = x^i$ and f to be sufficiently differentiable say $f \in C^{n+1}[-1,1]$, then the interpolant $p_n(x) = \sum_{i=0}^{n} \alpha_i x^j$ satisfying $p_n(x_j) = f(x_j)$, $0 \le j \le n$, the error $\|f - p_n\|$ is bounded such that from (1.130), we have

$$\|f - p_n\|_\infty \le \max_{x\in[-1,1]} \left(\frac{|f^{n+1}(x)|}{(n+1)!} \right) \max_{x\in[-1,1]} \prod_{j=0}^{n} |x - x_j|. \tag{1.131}$$

Throughout this study, we will refer to the inequality (1.131) the first interpolating error formula [101, 416]. It is well known that $\|(x-x_0)(x-x_1)\cdots(x-x_n)\|_\infty$ is minimal on $[-1,1]$ if the x_j are the zeroes of the $(n+1)$th-degree Tchebychev polynomial $T_{n+1}(x) = \cos((n+1)\arccos x)$.

The operator that associates with f its interpolant p_n is linear and given by

$$P_n[x_0,\cdots,x_n] : C([-1,1]) \to \mathscr{V}_n : f(x) \to p_n(x) = \sum_{i=0}^{n} f(x_i) l_i(x), \tag{1.132}$$

where the basic Lagrange polynomial [449, 450, 475, 476] expressed in (1.126). So another bound for the interpolation error is given by

$$\|f - p_n\|_\infty \le (1 + \|P_n\|) \|f - p_n^*\|_\infty, \;\; \|P_n\| = \max_{x\in[-1,1]} \sum_{i=0}^{n} |l_i(x)|$$

where $P_n := P_n[x_0,\cdots,x_n]$ is the linear operator defined by (1.132), and p_n^* is the best uniform polynomial approximation to f.

Definition 1.4.1. *For a fixed $n \in \mathbb{N}$ and a given $x_0,\cdots,x_n$, the Lebesgue function [67, 68] is defined by:*

$$L_n(x) = L_n(x_0,\cdots,x_n;x) = \sum_{i=1}^{n} |l_i(x)|,$$

and the Lebesgue constant [211, 212] is defined by:

$$\Lambda_n = \Lambda_n(x_0,\cdots,x_n) = \max_{-1\le x\le 1} \sum_{i=1}^{n} |l_i(x)|.$$

It can be clearly seen that both $L_n(x)$ and Λ_n depend on the location of the interpolation points x_j as well as on the degree n but not necessarily on the function values $f(x_i)$. It should also be noted that the operator norm of P_n defined in (1.132) is equal to the infinity norm of the Lebesgue function [160, 335, 397]:

$$\|P_n\|_\infty = \lambda_n = \max_{-1\le x\le 1} L_n(x).$$

In the general setting, let $\Omega \subset \mathbb{R}^d$(or $\mathbb{C}^d$) be a compact set, for example, $\Omega = [-1,1]^2 \subset \mathbb{R}^2$ or $\Omega = [-1,1]^3 \subset \mathbb{R}^3$, as in the case of triangle and spheres respectively [385]. Let $\xi_1, \cdots, \xi_N$ in Ω be the interpolating points. With the monomials $\phi_i = \phi(\xi_i) = \phi(\xi_1, \cdots, \xi_N)$ which are given by $\phi_i = \xi_1^{\alpha_1(i)}, \cdots, \xi_N^{\alpha_n(i)} = x^{\alpha(i)}$, $N = \dim(\Phi_n^N(\Omega))$ in a certain order $(\boldsymbol{\xi}, \boldsymbol{\alpha} \in \mathbb{R}^n(\text{or } \mathbb{C}^d))$. We construct the Vandermonde matrix $V(\boldsymbol{\xi}, \boldsymbol{\phi}) = [\phi_j(\xi_i)]$ and $\det(V) \neq 0$. We define the determinantal Lagrange formula [73, 177, 294, 485], by $\Pi_n f(x) = \sum_{j=1}^{N} f(\xi_j) l_j(x)$, where

$$l_j(x_j) = \frac{\det[V(\xi_1, \cdots, \xi_{j-1}, x, \xi_{j+1}, \cdots, \xi_N)]}{\det[V(\xi_1, \cdots, \xi_{j-1}, \xi_j, \xi_{j+1}, \cdots, \xi_N)]}, \;\; l_j(\xi_i) = \sigma_{ij}.$$

1.4.2 Fekete points

When a function is approximated by a polynomial using interpolation the approximation error depends on the chosen interpolation points. The Fekete points is a set of points that provide an almost optimal choice of interpolation points [409] and they are given by maximizing the Vandermonde determinant. Fekete points deeply discussed in [13, 83, 272, 273, 387, 388] and by definition:

Definition 1.4.2. *The Fekete Points are classical optimal interpolating points that maximize the absolute value of the Vandermonde determinant, that is,* $\max |v_n(x_1, \cdots, x_n)|$ *over the compact subset or manifold* $\Omega \subset \mathbb{R}^d$ *(or* $\mathbb{C}^d$*), where* $x_1, \cdots, x_n$*,* $v_n(x_1, \cdots, x_n)$ *and* $|\cdot|$ *is the absolute value.*

Taking the logarithm of the expression for the Vandermonde determinant given in (1.8) gives

$$\log\big(v_n(x_1, \ldots, x_n)\big) = \sum_{1\le i<j\le n} \log(x_j - x_i)$$

and thus $-\frac{1}{2}\log\big(v_n(x_1, \ldots, x_n)\big)$ gives the same as setting $g(x) = \log(x)$ and $\mathrm{V}(x) \equiv 0$ in

$$\mathscr{H}_N(x_1, \ldots, x_N) = \frac{1}{2}\sum_{i\neq j} g(x_i - x_j) + N\sum_{i=1}^{N} \mathrm{V}(x_i). \tag{1.133}$$

the Coulombs law which states that the force between two charged particles is proportional to the product of the charges and the inverse of the square of the distance between the two charges [96].

Finding the Fekete points are also of interest in complexity theory and would help with finding an appropriate starting polynomial for a homotopy algorithm for realizing the Fundamental Theorem of Algebra [426, 428].

Considering a compact domain $\Omega \subset \mathbb{R}^d$(or $\mathbb{C}^d$), it should be noted that the determinant of the Vandermonde matrix constructed with respect to the polynomial basis $\{\phi_k\}$ is key in determining a set for good interpolation points called the Fekete points deeply discussed in [13, 83, 272, 273, 387, 388].

To compute the Lebesgue constant [475, 476], we discretize the subset $\Omega \in \mathbb{R}^n$ with a "grid" $G \subset \Omega$ containing K well spaced points such that

$$\Lambda = \max_{x\in\Omega} \sum_{i=1}^{N} |l_i(x)| \;\; \rightarrow \;\; \Lambda \approx \max_{x\in G} \sum_{i=1}^{N} |l_i(x)|$$

The above results can be summarized in the following theorem:

Theorem 1.4.3 ([457])**.** *Let $x_0, x_1, \cdots, x_n$ be $n+1$ distinct points on $[a,b]$. The linear polynomial P_n which in is $C[a,b]$ associates with any function f the polynomial $P_n f \in P_n$ interpolates f between the $x_k's$ has the form*

$$\|P_n\| = \Lambda_n = \max_{x\in[a,b]} \sum_{k=0}^{n} |l_k(x)|.$$

A common generalisation of the Fekete points is the case where multivariate polynomials are used, see for example [57, 64, 322]. The case where and points in $\mathbb{C}^d$ are interpolated have also been examined, an example of a recent significant results is [38] and a review can be found in [48].

Polynomial interpolation is mostly used when the data set we wish to interpolate is small. The main reason for this is the instability of the interpolation method. One example of this is Runge's phenomenon that shows that when certain functions are approximated by polynomial interpolation fitted to equidistantly sampled points will sometimes lose precision when the number of interpolating points is increased.

One way to predict this instability of polynomial interpolation is that the conditional number of the Vandermonde matrix can be very large for equidistant points [184, 187, 188, 189, 190].

There are different ways to mitigate the issue of stability, for example choosing data points that minimize the conditional number of the relevant matrix [187, 188, 189, 190] or by choosing a polynomial basis that is more stable for the given set of data points such as Bernstein polynomials in the case of equidistant points [361]. Other polynomial schemes can also be considered, for instance by interpolating with different basis functions in different intervals, for example using polynomial splines.

While the instability of polynomial interpolation does not prevent it from being useful for analytical examinations it is generally considered impractical when there is noise present or when calculations are performed with limited precision. Often interpolating polynomials are not constructed by inverting the Vandermonde matrix or calculating the Lagrange basis polynomials, instead a more computationally efficient method such as Newton interpolation or Neville's algorithm

are used [383]. There are some variants of Lagrange interpolation, such as barycentric Lagrange interpolation, that have good computational performance [39].

In applications where the data is noisy it is often suitable to use least squares fitting, instead of interpolation.

B-Spline Fitting

The most popular application of the divided differences operator is the construction of so called *B*-splines, piecewise polynomial functions that allow for very efficient storage and computation of a variety of shapes. The concept of (mathematical) splines first appeared in the 1940s [414, 415] and *B*-splines were developed in the 1960s and 1970s [42, 104, 105, 106]. We can define a *B*-spline using the divided differences as follows.

Definition 1.4.3. *Given a sequence, $\cdots \leq t_{-1} \leq t_0 \leq t_1 \leq t_2 \leq \cdots$ we can define the kth B-spline of order m as*

$$B_{k,m}(x) = \begin{cases} (-1)^m [t_k, \ldots, t_{k+m}] g_k(x,t), & \text{if } t_k \leq x \leq t_{k+1}, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$g_k(x,t) = \begin{cases} (x-t)^{k-1}, & \text{if } x \geq t, \\ 0, & \text{otherwise}. \end{cases}$$

and the divided difference operator acts with respect to t.

Remark 1.4.4. *There are several different ways to define B-splines, above we followed the definition in [416]. In modern literature it is more common that B-splines and their computation are described from the perspective of so-called blossoms [178, 389, 390] rather that the divided difference description.*

B-splines can be used for many things, for example approximation theory [324], geometric modelling [88] and wavelets construction [178].

If we want to do linear interpolation and use some other set of basis functions other than the monomials, then we need to define a generalized version of the divided difference operator.

Definition 1.4.4. *Given a set of m linearly independent functions, $\mathcal{G} = \{g_i\}$, and n values, $x_0, \ldots, x_n$,*

then the generalized divided differences operator that acts on a function $f(x)$ is defined as

$$[x_0,x_1]_G f(x) = \frac{\begin{vmatrix} g_1(x_1) & \dots & g_{n-1}(x_1) & f(x_1) \\ g_1(x_2) & \dots & g_{n-1}(x_2) & f(x_2) \\ \vdots & \ddots & \vdots & \vdots \\ g_1(x_n) & \dots & g_{n-1}(x_n) & f(x_n) \end{vmatrix}}{\begin{vmatrix} g_1(x_1) & \dots & g_n(x_1) \\ g_1(x_2) & \dots & g_n(x_2) \\ \vdots & \ddots & \vdots \\ g_1(x_n) & \dots & g_n(x_n) \end{vmatrix}}$$

Remark 1.4.5. *We mentioned previously that the divided difference operator can be used to construct B-splines and using the generalized divided difference operator similar tools can be constructed using other sets functions than polynomials as a basis, see for example [310].*

Finding the Fekete points are also of interest in complexity theory and would help with finding an appropriate starting polynomial for a homotopy algorithm for realizing the Fundamental Theorem of Algebra [426, 428].

In Chapter 3 we will discuss how to find the maximum points of the Vandermonde determinant for certain special cases. A common generalisation of the Fekete points is the case where multivariate polynomials are used, see for example [57, 64, 322]. The case where and points in $\mathbb{C}^d$ are interpolated have also been examined, an example of a recent significant results is [38] and a review can be found in [48].

Integration of an exponential function over a unitary group

If we let $U(n)$ be the n-dimensional unitary group and dU the Haar measure normalised to 1 then the Harish-Chandra–Itzykson–Zuber integral formula [217, 249], says that if $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices with eigenvalues $\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \ldots \leq \lambda_n(\mathbf{A})$ and $\lambda_1(\mathbf{B}) \leq \lambda_2(\mathbf{B}) \leq \ldots \leq \lambda_n(\mathbf{B})$, then

$$\int_{U(n)} \exp\left(t\operatorname{tr}\left(\mathbf{A}\mathrm{U}\mathbf{B}\mathrm{U}^{\star}\right)\right) d\mathrm{U} = \frac{\det\left(\left[\exp\left(t\lambda_j(\mathbf{A})\lambda_k(\mathbf{B})\right)\right]_{j,k}^{n,n}\right)}{t^{\frac{n(n-1)}{2}} v_n\left(\lambda(\mathbf{A})\right) v_n\left(\lambda(\mathbf{B})\right)} \prod_{i=1}^{n-1} i! \tag{1.134}$$

where $v_n(\cdot)$ is the determinant of the Vandermonde matrix. If $t = 1$ and $\mathbf{A}$ and $\mathbf{B}$ are chosen as diagonal matrices such that

$$\mathbf{A}_{ij} = \begin{cases} a_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \qquad \mathbf{B}_{ij} = \begin{cases} b_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

then formula (1.134) reduces to an expression involving determinants of a generalized Vandermonde matrix and two Vandermonde matrices,

$$\int_{U(n)} (t\mathrm{tr}(\mathbf{AUBU}^{\star}))\, d\mathrm{U} = \frac{\begin{vmatrix} e^{a_1 b_1} & e^{a_1 b_2} & \dots & e^{a_1 b_n} \\ e^{a_2 b_1} & e^{a_2 b_2} & \dots & e^{a_2 b_n} \\ \vdots & \vdots & \dots & \vdots \\ e^{a_n b_1} & e^{a_n b_2} & \dots & e^{a_n b_n} \end{vmatrix}}{v_n(a_1,\dots,a_n) v_n(b_1,\dots,b_n)}.$$

1.4.3 Divided Differences

The coefficients of an interpolating polynomial could be computed by inverting the Vandermonde matrix or using the Lagrange basis polynomials. Another method for the coefficients of the polynomials is based on a computation called divided differences.

Definition 1.4.5. *Let $x_0,\dots,x_n$ then the divided differences operator that acts on a function $f(x)$ is defined as*

$$[x_0,\dots,x_n]f(x) = \begin{cases} f(x_0), & \text{if } n = 0, \\ \dfrac{[x_1,\dots,x_n]f(x) - [x_0,\dots,x_{n-1}]f(x)}{x_n - x_1}, & \text{if } n > 0. \end{cases}$$

The reason that the divided difference operator is interesting in polynomial interpolation is that if we apply it to two distinct points, x_0 and x_1, and a function $f(x)$ then the result is the slope of a line that passes through the two points $(x_0, f(x_0))$ and $(x_1, f(x_1))$,

$$[x_0,x_1]f(x) = \frac{[x_1]f(x) - [x_0]f(x)}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

A line that passes through the two points can then be constructed like this

$$p(x) = f(x_0) + (x - x_0)[x_0,x_1]f(x).$$

It can similarly be shown that a polynomial that interpolates a set of points

$$(x_0, f(x_0)), \dots, (x_n, f(x_n))$$

can be written

$$\begin{aligned} p(x) &= f(x_0) + (x - x_0)[x_0,x_1]f(x) + (x - x_0)(x - x_1)[x_0,x_1,x_2]f(x) + \dots \\ &= +(x - x_0)\dots(x - x_{n-1})[x_0,\dots,x_n]f(x) \end{aligned}$$

This method for interpolation is usually referred to as Newton interpolation and is probably the most well-known application of divided differences. In some literature, for example, see [103, 104, 105], this property is even used as a definition for divided differences.

Since we expect to find the same polynomial whether we use the Lagrange interpolation method described in the previous or the Newton interpolation method described above we also expect there to be some relation between the divided difference operator and the Vandermonde determinant. Turns out there is a fairly simple relation, see [416] for details.

Lemma 1.4.6. *The divided difference operator defined in Definition 1.4.5 can also be written as*

$$[x_0,x_1]f(x) = \frac{\begin{vmatrix} 1 & x_0 & \dots & x_0^{n-1} & f(x_0) \\ 1 & x_1 & \dots & x_1^{n-1} & f(x_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \dots & x_n^{n-1} & f(x_n) \\ 1 & x & \dots & x^{n-1} & f(x) \end{vmatrix}}{v_n(x_0,x_1,\dots,x_n,x)} \prod_{i=0}^{n}(x-x_i). \tag{1.135}$$

where $v_n(x_0,x_1,\dots,x_n,x)$ denotes the Vandermonde determinant.

The divided differences operator can also be used to describe the error that one gets when a function is approximated by interpolating with a polynomial, the following lemma is from [274].

Lemma 1.4.7. *Let $p(x)$ be a polynomial of degree smaller than or equal to n that interpolates the points $\{f(x_i,f(x_i)),\ i=0,\dots,n\}$. For any $x \neq x_i,\ i=0,\dots,n$ the error $f(x)-p(x)$ is given by*

$$f(x)-p(x) = [x_0,\dots,x_n,x]f(x)\prod_{i=0}^{n}(x-x_i).$$

Combining Lemma 1.4.6 and Lemma 1.4.7 gives

$$f(x)-p(x) = \frac{\begin{vmatrix} 1 & x_0 & \dots & x_0^{n-1} & f(x_0) \\ 1 & x_1 & \dots & x_1^{n-1} & f(x_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \dots & x_n^{n-1} & f(x_n) \\ 1 & x & \dots & x^{n-1} & f(x) \end{vmatrix}}{v_n(x_0,x_1,\dots,x_n,x)} \prod_{i=0}^{n}(x-x_i).$$

which gives some insight to why the value of the Vandermonde determinant is important when choosing interpolation points.

Another popular application of the divided differences operator is the construction of so called *B*-splines, piecewise polynomial functions that allow for very efficient storage and computation of a variety of shapes. The concept of (mathematical) splines first appeared in the 1940s [414, 415] and *B*-splines were developed in the 1960s and 1970s [42, 104, 105]. We can define a *B*-spline using the divided differences as follows.

Definition 1.4.6. *Given a sequence, $\cdots \leq t_{-1} \leq t_0 \leq t_1 \leq t_2 \leq \cdots$ we can define the kth B-spline of order m as*

$$B_{k,m}(x) = \begin{cases} (-1)^m [t_k, \ldots, t_{k+m}] g_k(x,t), & \text{if } t_k \leq x \leq t_{k+1}, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$g_k(x,t) = \begin{cases} (x-t)^{k-1}, & \text{if } x \geq t, \\ 0, & \text{otherwise}. \end{cases}$$

and the divided difference operator acts with respect to t.

Remark 1.4.8. *There are several different ways to define B-splines, above we followed the definition in [416]. In modern literature it is more common that B-splines and their computation are described from the perspective of so-called blossoms [178, 389, 390] rather that the divided difference description.*

B-splines can be used for many things, for example approximation theory [324], geometric modelling [88] and wavelets construction [178].

If we want to do linear interpolation and use some other set of basis functions other than the monomials, then we need to define a generalized version of the divided difference operator.

Definition 1.4.7. *Given a set of m linearly independent functions, $\mathrm{G} = \{g_i\}$, and n values, $x_0, \ldots, x_n$, then the generalized divided differences operator that acts on a function $f(x)$ is defined as*

$$[x_0, x_1]_{\mathrm{G}} f(x) = \frac{\begin{vmatrix} g_1(x_1) & g_2(x_1) & \ldots & g_{n-1}(x_1) & f(x_1) \\ g_1(x_2) & g_2(x_2) & \ldots & g_{n-1}(x_2) & f(x_2) \\ \vdots & \vdots & \ddots & \vdots & vdots \\ g_1(x_n) & g_2(x_n) & \ldots & g_{n-1}(x_n) & f(x_n) \end{vmatrix}}{\begin{vmatrix} g_1(x_1) & g_2(x_1) & \ldots & g_n(x_1) \\ g_1(x_2) & g_2(x_2) & \ldots & g_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(x_n) & g_2(x_n) & \ldots & g_n(x_n) \end{vmatrix}}$$

Remark 1.4.9. *We mentioned previously that the divided difference operator can be used to construct B-splines and using the generalized divided difference operator similar tools can be constructed using other sets functions than polynomials as a basis, see for example [310].*

1.4.4 Least Squares Fitting

Suppose we want to find a function

$$f(x) = \sum_{i=1}^{N} \beta_i g_i(x) \tag{1.136}$$

that fits as well as possible in the least squares sense to the data points (x_i, y_i), $i = 1, \ldots, n$. We then get a curve or polynomial fitting problem described by the linear equation system $\mathbf{A}\boldsymbol{\beta} = \mathbf{y}$ where $\boldsymbol{\beta}$ are the coefficients of f, $\mathbf{y}$ is the vector of data values and $\mathbf{A}$ is the appropriate alternant matrix, and these can be expressed as

$$\mathbf{A} = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \ldots & g_N(x_1) \\ g_1(x_2) & g_2(x_2) & \ldots & g_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(x_n) & g_2(x_n) & \ldots & g_N(x_n) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

This is an overdetermined version of the linear interpolation described in the previous section.

How can we actually find the coefficients that minimize the sum of the squares of the residuals? First we can define the square of the length of the residual vector, $\mathbf{e} = \mathbf{A}\boldsymbol{\beta}\mathbf{y}$, as a function

$$S(\mathbf{e}) = \mathbf{e}^\top \mathbf{e} = \sum_{i=1}^{n} |e_i|^2 = (\mathbf{A}\boldsymbol{\beta})^\top (\mathbf{A}\boldsymbol{\beta}).$$

where $(\cdot)^\top$ is the transpose of the matrix. This kind of function is a positive second degree polynomial with no mixed terms and thus has a global minima where $\dfrac{\partial S(\mathbf{e})}{\partial e_i} = 0$ for all $1 \leq i \leq n$. We can find the global minima by looking at the derivative of the function, e_i is determined by β_i and

$$\frac{\partial e_i}{\partial \beta_i} = \mathbf{A}_{i,j}$$

thus

$$\frac{\partial S(\mathbf{e})}{\partial \beta_i} = \sum_{i=1}^{n} 2e_i \frac{\partial e_i}{\partial \beta_i} = \sum_{i=1}^{n} 2(\mathbf{A}_{i,}\boldsymbol{\beta} - y_i)\mathbf{A}_{i,j} = 0 \;\Leftrightarrow\; \mathbf{A}^\top \mathbf{A}\boldsymbol{\beta} = \mathbf{A}^\top \mathbf{y}.$$

This gives

$$\boldsymbol{\beta} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}$$

and by the Gauss-Markov theorem of [182, 183, 319, 406] for a more modern description, if $(\mathbf{A}^\top \mathbf{A})^{-1}$ exists then (19) gives the linear, unbiased estimator that gives the lowest variance possible for any linear, unbiased estimator. The matrix given by $(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top$ is sometimes referred to as the Moore-Penrose pseudo-inverse of $\mathbf{A}$.

Clearly a linear curve fitting model with $g_i(x) = x^{i-1}$ gives an equation system described by a rectangular Vandermonde matrix.

If it is not necessary to exactly reproduce the series of data points a commonly applied alternative to interpolation is least squares fitting. A least squares fitting of a mathematical model to a set of data points $\{f(x_i, y_i),\ i = 1, \ldots, n\}$ is the choice of parameters of the model, here denoted $\boldsymbol{\beta}$, chosen such that the sum of the squares of the residuals

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} (y_i - f(\boldsymbol{\beta}; x_i))$$

is minimized. This choice is appropriate if data series is affected by independent and normally distributed noise.

The most wide-spread form of least squares fitting is linear least squares fitting where, analogously to linear interpolation, the function $f(\boldsymbol{\beta}x)$ depends linearly on $\boldsymbol{\beta}$. This case has a unique solution that is simple to find. It is commonly known as the least squares method and we describe it in detail in the next section. With a non-linear $f(\boldsymbol{\beta}; x)$ it is usually much more difficult to find the least squares fitting and often numerical methods are used, for example the Marquardt least squares method.

Non-linear Least Squares Fitting

So far we have only considered models that are linear with respect to the parameters that specify them. If we relax the linearity condition and simply consider fitting a function with m parameters, $f(\beta_1, \ldots, \beta_n; x)$, to n data points in the least squares sense it is usually referred to as a non-linear least squares fitting.

There is no general analogue to the Gauss-Markov theorem for non-linear least squares fitting and therefore finding the appropriate estimator requires more knowledge about the specifics of the model. In practice non-linear least squares fittings are often found using some numerical method for non-linear optimization of which there are many, see for instance [406] for details.

1.4.5 Regression Analysis and Data Smoothing

In this thesis we will discuss several ways to construct mathematical models. With several mathematical models available it is needed to have some method for comparing the methods and choose the most suitable one. When the model in constructed with a certain application in mind there is often a set of required or desired properties given by the application and choosing the best model is a matter of seeing which model matches the requirements the best. In many cases this process is not straightforward and often there is not one model that is better than the other candidate models in all aspects, a common example is the trade-off between accuracy and complexity of the model. It is often easy to improve the model by increasing its complexity (either by introducing more general and flexible mathematical concepts that are more difficult to analyse or less well understood or by extending the model in a way that increases the cost of computations and simulations using the model), but finding the best compromise between accuracy and complexity can be difficult. In this

section we will discuss how to compare models primarily with respect to accuracy and the number of required parameters.

Regression is similar to interpolation except that the presence of noise in the data is taken into consideration [83, 178, 274, 280, 387, 414, 416, 419, 454, 460]. The typical regression problem assumes that the data points $\{f(x_i,y_i);\ i=1,\ldots,n\}$ are sample from a stochastic variable of the form

$$Y_i = f(\boldsymbol{\beta};x_i)+\varepsilon_i$$

where $f(\boldsymbol{\beta};x)$ is a given function with a fixed number of undetermined parameters $\boldsymbol{\beta}\in\mathscr{B}$ and ε_i for $i=1,\ldots,n$ are samples of a random variable with expected value zero, called the errors or the noise for the data set.

There are many different classes of regression problems and they are defined by the type of function $f(\beta_1,\ldots,\beta_m;x)$ and the distribution of errors.

Here we will only consider the situation when the ε_i variables are independent and normally distributed with identical variance and that the parameter space $\mathscr{B}$ is a compact subset of $\mathbb{R}^k$ and that for all x_i the function $f(\boldsymbol{\beta};x_i)$ is a continuous function of $\boldsymbol{\beta}\in\mathscr{B}$.

Suppose we want to choose the appropriate set of parameters for f based on some set of observed data points. A common approach to this is so called maximum likelihood estimation.

Definition 1.4.8. *The likelihood function, L is the function that gives us the probability that a certain observation, x, of a stochastic variable X is made given a certain set of parameters,* $\boldsymbol{\beta}$*,*

$$\mathrm{L}_x(\boldsymbol{\beta}) = \mathbb{P}(X=x|\boldsymbol{\beta}).$$

Thus choosing parameters that maximize the likelihood function gives the set of parameters that seem to be most likely based on available information. Typically these parameters cannot be calculated exactly and must be estimated, this estimation is called the Maximum Likelihood Estimation (MLE).

To find the MLE we need to find the maximum of the likelihood function. Note that here we will only consider the case where the noise variables, ε_i, are independent and normally distributed with mean zero.

Lemma 1.4.10. *For the stochastic variables* $Y_i = f(\boldsymbol{\beta};x_i)+\varepsilon_i$ *where* $f(\boldsymbol{\beta};x)$ *is a given function with a fixed number of undetermined parameters and* ε_i *for* $i=1,\ldots,n$ *are independent random variables with expected value zero* $\boldsymbol{\beta}\in\mathscr{B}$ *and standard deviation* σ *the likelihood function is given by the joint probability density function for the noise,*

$$L_y(\boldsymbol{\beta}) = (2\pi)^{\frac{n}{2}}\sigma^n\prod_{i=1}^{n}\exp\left(-\frac{(y_i-f(\boldsymbol{\beta};x_i))^2}{\sigma^2}\right)$$

Proof. Since each ε_i is normally distributed with mean zero and standard deviation σ the difference between the observed value and the given function, $y_i-f(\boldsymbol{\beta};x_i)$ is normally distributed with mean

zero and standard deviation σ. Since all the errors are independent the joint probability density function is just the product of n probability density functions of the form

$$p_i(\boldsymbol{\beta};(x_i,y_i)) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i - f(\boldsymbol{\beta};x_i))^2}{\sigma^2}\right)$$

for $i = 1,\ldots,n$. □

For the MLE we only care about what parameters give the maximum of the likelihood function, not the actual value of the likelihood function so we can ignore the constant factor and in practice it also often simple to consider the maximum of the logarithm of the likelihood function. This leads to the following lemma.

Lemma 1.4.11. *Let the set of data points $\{(x_i,y_i), i = 1,\ldots,n\}$ describe a regression problem and the stochastic variables $Y_i = f(\boldsymbol{\beta};x_i) + \varepsilon_i$ where $f(\boldsymbol{\beta};x)$ is a given function with a fixed number of undetermined parameters $\boldsymbol{\beta} \in \mathscr{B}$ and ε_i for $i = 1,\ldots,n$ are independent normally distributed random variables with expected value zero and standard deviation σ. The MLE for the parameters $\boldsymbol{\beta}$ will minimize the sum of the squares of the residuals,*

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta};x_i))^2.$$

Proof. Since the natural logarithm is a monotonically increasing function $-\ln(L_y(\boldsymbol{\beta}))$ will have a minimum point where $L_y(\boldsymbol{\beta})$ has a maximum point. By Lemma 1.4.10 we obtain

$$\begin{aligned}
-\ln(L_y(\boldsymbol{\beta})) &= -\ln\left((2\pi)^{\frac{n}{2}}\sigma^n\prod_{i=1}^{n}\exp\left(-\frac{(y_i - f(\boldsymbol{\beta};x_i))^2}{\sigma^2}\right)\right)\\
&\quad -\ln\left((2\pi)^{\frac{n}{2}}\sigma^n\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta};x_i))^2\\
&\quad -\ln\left((2\pi)^{\frac{n}{2}}\sigma^n\right) + \frac{1}{\sigma^2}S(\boldsymbol{\beta}).
\end{aligned}$$

Since the first term and the factor in front of $S(\boldsymbol{\beta})$ does not depend on $\boldsymbol{\beta}$ the minimum point of $S(\boldsymbol{\beta})$ will coincide with the maximum of the likelihood function. □

Here, it can be seen that finding the MLE is equivalent to using the curve fitting technique describes in Section 1.4.4.

1.4.6 D-Optimal Experimental Design

For the class of linear non-weighted regression problems described in the previous Section, minimizing the square of the sum of residuals gives the maximum likelihood estimation of the parameters that specify the fitted function. This estimation naturally has a variance as well and minimizing

this variance can be interpreted as improving the reliability of the fitted function by minimizing its sensitivity to noise in measurements. This minimization is usually done by choosing where to sample the data carefully, in other words, given the regression problem defined by

$$y_i = f(\boldsymbol{\beta};x_i) + \varepsilon_i$$

for $i = 1,\ldots,n$ with the same conditions on $f(\boldsymbol{\beta};x)$ and ε_i as in the previous Section. We want to choose a design $\{x_i \; i = 1,\ldots,n\}$ that minimizes the variance of the values predicted by the regression model. This is usually referred to as G-optimality.

To give a proper definition of G-optimality we will need the concept of the Fisher information matrix. When motivating the expression for the Akaike Information Criteria, (AIC), described in detail in [4, 5, 62, 70, 118, 131, 176, 280, 305, 308, 419] the matrix

$$\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln\left(g(\hat{\boldsymbol{\beta}};X)\right)\right]_{i=1,j=1}^{n,n}$$

In that context we were interested how much information was lost when the model g was used instead of the data. If the model g is the true distribution, twice differentiable, and has only one parameter, $\boldsymbol{\beta}$, it is possible to describe how information about the model that is contained in the parameter using the Fisher information

$$I(\boldsymbol{\beta}) = -\mathrm{E}_X\left[\frac{\partial^2\log\left(g(\boldsymbol{\beta};X)\right)}{\partial\boldsymbol{\beta}^2}\right].$$

Essentially this expression measures the probability of a particular outcome being observed for a known value of $\boldsymbol{\beta}$, so if the Fischer information is only large in near a certain points it is easy to tell which parameter value is the true parameter value and if the Fisher information does not have a clear pea it is difficult to estimate the correct value of $\boldsymbol{\beta}$. When the model has several parameters the Fisher information is replaced by the Fischer information matrix.

Definition 1.4.9. *For a finite design* $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{R}^n$ *the Fisher information matrix ,* $\mathbf{M}$*, is the matrix defined by*

$$\mathbf{M}(\boldsymbol{\beta}) = -\mathrm{E}_X\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln\left(g(\hat{\boldsymbol{\beta}};X)\right)\right]_{i=1,j=1}^{n,n}$$

Thus, the concept of information in the AIC and the concept of information here are two different but related concepts, for a detailed discussion of this relation see [70].

There is a lot of literature on the Fisher information matrix and but in the context of the least squares problems discussed here we have a fairly simple expression for its elements, see [328] for details.

Lemma 1.4.12. *For a finite design* $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{R}^n$ *the Fisher information matrix for the type of least squares fitting problem considered in this section can be computed by*

$$\mathbf{M}(\mathbf{x}) = \sum_{i=1}^{n}\mathbf{f}(x_i)\mathbf{f}(x_i)^\top$$

where $\mathbf{f}(x_i) = \begin{bmatrix} f_1(x) & f_2(x) & \dots & f_n(x) \end{bmatrix}^\top$.

Definition 1.4.10. The G-optimality criterion: *A design, $\boldsymbol{\xi} = \{x_1, \dots, x_n\}$, is said to be G-optimal if it minimizes the maximum variance of any predicted value*

$$\operatorname{Var}\big(\mathbf{y}(\boldsymbol{\xi})\big) = \min_{x_i, i=1,\dots,n} \max_{x \in \mathscr{X}} \operatorname{Var}\big(\mathbf{y}(\mathbf{x})\big) = \min_{\mathbf{z} \in \mathscr{X}} \max_{\mathbf{x} \in \mathscr{X}} \mathbf{f}(\mathbf{x})^\top \mathbf{M}(\mathbf{z}) \mathbf{f}(\mathbf{x}).$$

The *G*-optimality condition was first introduced in [432] (the name *G*- optimality comes from later work by Kiefer and Wolfowitz where they describe several different types of optimal design using alphabetical letters [269, 270]) and is an example of a minimax criterion, since it minimizes the maximum variance of the values given by the regression model [328].

There are many kinds of optimality conditions related to *G*-optimality. One which is suitable for us to consider is *D*-optimality. This type of optimality was first introduced in [481] and instead of focusing on the variance of the predicted values of the model it instead minimizes the volume of the confidence ellipsoid for the parameters (for a given confidence level).

Definition 1.4.11. The D-optimality criterion: *A design $\boldsymbol{\xi}$ is said to be D-optimal if it maximizes the determinant of the Fisher information matrix*

$$\det\big(\mathbf{M}(\boldsymbol{\xi})\big) = \max_{\mathbf{x} \in \mathscr{X}} \det\big(\mathbf{M}(\mathbf{x})\big).$$

The *D*-optimal designs are often good design with respect to other types of criterion, see for example [206, 207] for a brief discussion on this and is often practical to consider due to being invariant with respect to linear transformations of the design matrix. A well-known theorem called the Kiefer–Wolfowitz equivalence theorem shows that under certain conditions G-optimality is equivalent to D-optimality.

Theorem 1.4.13. Kiefer–Wolfowitz equivalence theorem *For any linear regression model with independent, uncorrelated errors and continuous and linearly independent basis functions $f_i(x)$ defined on a fixed compact topological space $\mathbf{X}$ there exists a D-optimal design and any D-optimal design is also G-optimal.*

This equivalence theorem was originally proven in [268] but the formulation above is taken from [328]. Thus maximizing the determinant of the Fisher information matrix corresponds to minimizing the variance of the estimated $\boldsymbol{\beta}$. Interpolation can be considered a special case of regression when the sum of the square of the residuals can be reduced to zero. Thus we can speak of D-optimal design for interpolation as well, in fact optimal experiment design is often used to find the minimum number of points needed for a certain model. For a linear interpolation problem defined by the alternant matrix $\mathbf{A}(\mathbf{f}, \mathbf{x})$ the Fisher information matrix is $\mathbf{M}(\mathbf{x}) = \mathbf{A}(\mathbf{f}; \mathbf{x})^\top \mathbf{A}(\mathbf{f}; \mathbf{x})$ and since $\mathbf{A}(\mathbf{f}; \mathbf{x})$ is an $n \times n$ matrix $\det\big(\mathbf{A}(\mathbf{f}; \mathbf{x})^\top\big) \det\big(\mathbf{A}(\mathbf{f}; \mathbf{x})\big) = \det\big((\mathbf{f}; \mathbf{x})\big)$. Thus the maximization of the determinant of the Fisher information matrix is equivalent to finding the extreme points of the determinant of an alternant matrix in some volume given by the set of possible designs.

A standard case of this is polynomial interpolation where the x-values are in a limited interval, for instance $-1 \leq x_i \leq 1$ for $i = 1, 2, \ldots, n$. In this case the regression problem can be written as $\mathbf{V}_n(\mathbf{x})^\top \boldsymbol{\beta} = \mathbf{y}$ where $\mathbf{V}_n(\mathbf{x})$ is a Vandermonde matrix as defined in equation 1.6 and the constraints on the elements of $\boldsymbol{\beta}$ means that the volume we want to optimize over is a cube in n dimensions. There is a number of classical results that describe how to find the D-optimal designs for weighted univariate polynomials with various efficiency functions, for example see [158], and [305] where they demonstrate one way to optimize the Vandermonde determinant over a cube.

The shape of the volume to optimize the determinant in is given by constraints on the data points. For example, if there is a cost associated with each data point that increases quadratically with x and there is a total budget, C, for the experiment that cannot be exceeded the constraint on the x-values becomes $x_1^2 + x_2^2 + \ldots + x_n^2 \leq C$ and the determinant needs to be optimized in a ball. In Chapter 2 we examine the optimization of the Vandermonde determinant over several different surfaces in several dimensions.

The D-optimal design is mainly used to improve the stability of an interpolation problem as an alternative to the non-linear fitting. It should be noted that while choosing a D-optimal design can give an approximation method that is more stable since it minimizes the variance of the parameters, the approximating function can still be highly sensitive to changes in parameters (the variance of the predicted values can be minimized but still high) so it does necessarily maximize stability or stop instability phenomenons similar to Runge's phenomenon for polynomial interpolation [305].

1.5 Random Matrix Theory

In this section we give an overview of the random matrix theory (RMT) and some applications. Considering the close relationship between the Vandermonde determinant and determinantal probability distribution, then it is important that we give a brief overview of RMT in connection to Gaussian ensembles and Wishart ensembles.

In probability theory and mathematical physics, a random matrix is a matrix-valued random variable, that is, a matrix in which some or all elements are random variables. Many important properties of physical systems can be represented mathematically as matrix problems. For example, the thermal conductivity of a lattice can be computed from the dynamical matrix of the particle-particle interactions within the lattice.

We take advantage of this fact of Vandermonde determinant to establish the relationship between the product of Vandermonde matrices and joint eigenvalue probability density functions for large random matrices that occur in various areas of both classical mechanics, mathematics, statistics and many other areas of science. We also illustrate the optimization of these densities based of the extreme points of Vandermonde determinant.

The extreme points of the Vandermonde determinant appears in random matrix theory, for example to compute the limiting value of the so called Stieltjes transform using the method sometimes called the 'Coulomb gas analogy' [327]. This is also closely related to many problems in quantum mechanics and statistical mechanics. For an overview of some other applications of the extreme

points see [348].

In the next section we give a brief overview of random matrix theory (RMT).

1.5.1 Overview of Random Matrix Theory

Random matrices were first introduced in mathematical statistics in the late 1920s [496] and today the joint probability density function of eigenvalues of random matrices play a significant role both in probability theory, mathematical physics and quantum mechanics [199]. A random matrix, in simple terms can be defined as any matrix whose real or complex valued entries are random variables.

Random matrix theory primarily discusses the properties large or complex matrices with random variables as entries by utilizing the existing probability laws, in particularly, Gaussian distributions [8, 9]. The main motivational question in the probabilistic approach to random matrices is: what can be said about the probabilities of a few or if not all of its eigenvalues and eigenvectors? This question is significant in many areas of science including particle physics, mathematics, statistics and finance as highlighted here under.

In nuclear physics random matrices were applied in the modelling of the nuclei of heavy atoms [317, 256, 374, 375, 493]. The main idea was to investigate the spacing between the lines in the electromagnetic spectrum of a heavy atom nucleus, for example. Uranium 238, which resembles the separation between the eigenvalues of a random matrix [208, 327]. These random matrices have also been employed in solid-state physics to model the chaotic behaviour of large disordered Hamiltonians in terms of mean field approximation [116]. Random matrices have also been applied in quantum chaos to characterise the spectral statistics of quantum systems [49, 95].

Random unitary matrix transformations has also appears in theoretical physics, for example, the boson sampling model [1] has been applied in quantum optics to describe the advantages of quantum computation over classical computation. Random unitary transformations can also be directly implemented in an optical circuit, by mapping their parameters to optical circuit components [405].

Other applications in theoretical physics include, analysing the chiral Dirac operator [266, 473] quantum chromodynamics, quantum gravity in two dimensions [170], in mesoscopic physics random matrices are used to characterise materials of intermediate length [413], spin-transfer torque [407], the fractional quantum Hall effect [71], Anderson localization [254], quantum dots [511] and superconductors [18], electrodynamic properties of structural materials [509], describing electrical conduction properties of disordered organic and inorganic materials [507], quantum gravity [121] and string theory [47].

In mathematics some application include the distribution of the zeros of the Riemann zeta function [265], enumeration of permutations having certain particularities in which the random matrices can help to derive polynomials permutation patterns [367], counting of certain knots and links as applies to folding and colouring [47].

In multivariate statistics random matrices were introduced for statistical analysis of large samples in estimation of covariance matrices [138, 142, 143, 320, 444, 496]. More significant results

have proven that to extend the classical scalar inequalities for improved analysis of a structured dimension reduction based on largest eigenvalues of finite sums of random Hermitian matrices [461].

Random matrices have also been applied to financial modelling especially risk models and time series [19, 218, 471, 496] since their discovery by Wishart in 1928.

Random matrices also are increasingly used to model the network of synaptic connections between neurons in the brain as applies to neural networks or neuroscience. Neuronal networks can help to construct dynamical models based on random connectivity matrix [435]. This has also helped to establish the link relating the statistical properties of the spectrum of biologically inspired random matrix models to the dynamical behaviour of randomly connected neural networks [180, 260, 351, 459, 480].

In optimal control theory random matrices appear as coefficients in the state equation of linear evolution. In most problems the values of the parameters in these matrices are not known with certainty, in which case there are random matrices in the state equation and the problem is known as one of stochastic control [86, 464, 465].

1.5.2 Univariate and Multivariate Normal Distribution

Definition 1.5.1. *The univariate normal probability density function (Gaussian normal density) for a random variable X, which is the basis for construction of many multivariate distributions that occur in statistics, can be expressed as [8]:*

$$\mathbb{P}_X(x) = k \cdot \exp\left\{-\frac{1}{2}\alpha(x-\beta)^2\right\} \equiv k \cdot \exp\left\{-\frac{1}{2}(x-\beta)\alpha(x-\beta)\right\} \tag{1.137}$$

where α and k is chosen so that the integral of (1.137) *over the entire $x-$axis is unity and β is equal to the expectation of X, that is, $\mathbb{E}[X] = \beta$. It is then said that X follows a normal probability density function with parameters α and β, also expressed as $X \sim \mathcal{N}(\alpha, \beta)$.*

The density function of the multivariate normal distribution of random variables say $X_1, \cdots, X_p$ is defined analogously. If the scalar variable x in (1.137) is replaced by the vector $\mathbf{X} = (X_1, \cdots, X_p)^\top$, the scalar constant β is replaced by a vector $\mathbf{b} = (b_1, \cdots, b_p)^\top$ and the expression

$$\alpha(x-\beta)^2 = (x-\beta)\alpha(x-\beta)$$

is replaced by the quadratic form

$$(\mathbf{X}-\mathbf{b})^\top \mathbf{A}(\mathbf{X}-\mathbf{b}) = \sum_{i,j=1}^{p} a_{ij}(x_i - b_i)(x_j - b_j). \tag{1.138}$$

where $\mathbf{A}$ is the positive definite matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}, \tag{1.139}$$

then the density of the p-variate normal distribution becomes

$$\mathbb{P}(\mathbf{X}) = K \cdot \exp\left\{\frac{1}{2}(\mathbf{X}-\mathbf{b})^{\top}\mathbf{A}(\mathbf{X}-\mathbf{b})\right\} \tag{1.140}$$

where $\top$ denotes transpose and $K > 0$ is chosen so that the integral over the entire p-dimensional Euclidean space $x_1, \cdots, x_p$ is unity.

Theorem 1.5.1. *If the density of a p-dimensional random vector* $\mathbf{X}$ *is*

$$\sqrt{|\mathbf{A}|}(2\pi)^{-\frac{1}{2}p}\exp\left\{-\frac{1}{2}(\mathbf{X}-\mathbf{b})^{\top}\mathbf{A}(\mathbf{X}-\mathbf{b})\right\},$$

then the expected value of $\mathbf{X}$ *is* $\mathbf{b}$ *and the covariance matrix is* $\mathbf{A}^{-1}$*, for details see* [8]. Conversely, given a vector $\boldsymbol{\mu}$ and a positive definite matrix $\boldsymbol{\Sigma}$, there is a multivariate normal density

$$\mathbb{P}(\mathbf{X}) = (2\pi)^{-\frac{1}{2}p}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{(\mathbf{X}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\} \tag{1.141}$$

such that the expected value of the density is μ and the covariance matrix is $\boldsymbol{\Sigma}$.

The density (1.141) is often denoted as $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

For example, the diagonal elements of the covariance matrix, $\boldsymbol{\Sigma}_{ii}$, is the variance of the ith component of $\mathbf{X}$, which may sometimes be denoted by σ_i^2. The correlation between X_i and X_j is defined as

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} = \frac{\sigma_{ij}}{\sigma_i\sigma_j}$$

where σ_k denotes the standard deviation of X_k and $\sigma_{ij} = \boldsymbol{\Sigma}_{ij}$. This measure of association is symmetric in X_i and X_j such that $\rho_{ij} = \rho_{ji}$. Since

$$\begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{pmatrix} = \begin{pmatrix} \sigma_i^2 & \sigma_i\sigma_j\rho_{ij} \\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2 \end{pmatrix}$$

is positive-definite, the determinant

$$\begin{vmatrix} \sigma_i^2 & \sigma_i\sigma_j\rho_{ij} \\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2 \end{vmatrix} = \sigma_i^1\sigma_j^2(1-\rho_{ij}^2)$$

is positive. Therefore $-1 < \rho_{ij} < 1$.

1.5.3 Wishart Distribution

The matrix distribution that is now known as a Wishart distribution, was first derived by Wishart in the late 1920s [496]. It is usually regarded as a multivariate extension of the χ^2-distribution.

Theorem 1.5.2. *The sum of squares, $\boldsymbol{\chi}^2 = Z_1^2 + \cdots + Z_n^2$ of n- independent standard normal variables Z_i of mean 0 and variance 1, that is, distributed as $\mathcal{N}(0,1)$ has a χ^2-distribution defined by:*

$$\mathbb{P}_{\boldsymbol{\chi}^2}(x) = \frac{1}{2^{\frac{1}{2}n}\Gamma\left(\frac{1}{2}n\right)} e^{-\frac{1}{2}x^2}(\boldsymbol{\chi}^2)^{\frac{1}{2}n-1}. \tag{1.142}$$

where $\Gamma(\cdot)$ is the Gamma function [378].

Definition 1.5.2. *Let $\mathbf{X} = (X_1, \cdots, X_n)$, where $X_i \sim \mathcal{N}(\mu_i, \boldsymbol{\Sigma})$ and $\mathbf{X}_i$ is independent of $\mathbf{X}_j$, where $i \neq j$. The matrix $\mathbf{W} : p \times p$ is said to be Wishart distributed [496] if and only if $\mathbf{W} = \mathbf{X}\mathbf{X}^\top$ for some matrix $\mathbf{X}$ in a family of Gaussian matrices $\mathbf{G}_{m\times n}, m \leq n$, that is, $\mathbf{X} \sim \mathcal{N}_{m,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I})$ where $\boldsymbol{\Sigma} \geq 0$. If $\boldsymbol{\mu} = 0$ we have a central Wishart distribution which will be denoted by $\mathbf{W} \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n)$, and if $\boldsymbol{\mu} \neq 0$ we have a non-central Wishart distribution which will be denoted $\mathbf{W} \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n, \Delta)$, where $\Delta = \boldsymbol{\mu}\boldsymbol{\mu}^\top$ and n is the number of degrees of freedom.*

In our study, we shall mainly focus on the central Wishart distribution for which $\boldsymbol{\mu} = 0$ and $\mathbf{X} \sim \mathcal{N}_{m,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I})$.

Theorem 1.5.3 ([8])**.** *Given a random matrix $\mathbf{W}$ which can be expressed as $\mathbf{W} = \mathbf{X}\mathbf{X}^\top$ where $\mathbf{X}_1, \cdots, \mathbf{X}_n$, $(n \geq p)$ are independent, each with the distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, the distribution of $\mathbf{W} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n)$. If $\boldsymbol{\Sigma} > 0$, then the random matrix $\mathbf{W}$ has a joint density functions:*

$$\mathbb{P}(\mathbf{W}) = \begin{cases} \dfrac{1}{2^{\frac{np}{2}}\Gamma_p\left(\frac{n}{2}\right)} |\mathbf{W}|^{\frac{n-p-1}{2}} \exp\left(-\dfrac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{W}\right)\right), & \text{if } \mathbf{W} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1.143}$$

where the multivariate Gamma function is given by

$$\Gamma_p(n/2) = \pi^{\frac{p(n-1)}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(n+1-i)\right). \tag{1.144}$$

If $p = 1, \boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{1}$, then the Wishart matrix is identical to a central $\boldsymbol{\chi}^2$-variable with n degrees of freedom as defined in (1.142).

Theorem 1.5.4 ([253, 372])**.** *If $\mathbf{X}$ is distributed as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the probability density distribution of the eigenvalues of $\mathbf{X}\mathbf{X}^\top$, denoted $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$, is given by:*

$$\mathbb{P}(\boldsymbol{\lambda}) = \frac{\pi^{-\frac{1}{2}n}\det(\boldsymbol{\Sigma})^{-\frac{1}{2}n}\det(\mathbf{D})^{\frac{1}{2}(n-p-1)}}{2^{\frac{1}{2}np}\Gamma_p\left(\frac{1}{2}n\right)\Gamma_p\left(\frac{1}{2}p\right)} \prod_{i<j}(\lambda_i - \lambda_j)\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{D}\right)\right) \tag{1.145}$$

where $\mathbf{D} = \mathrm{diag}(\lambda_i)$.

It will prove useful that (1.145) contains the term $\prod_{i<j}(\lambda_i - \lambda_j)$ which is the determinant of a Vandermonde matrix [454]. A Vandermonde matrix is a well-known type of matrix that appears in many different applications both in mathematics, physics and recently in multivariate statistics, most famously curve-fitting using polynomials, for details see [454].

In the next section, we will discuss some well-known ensembles that that appear in the mathematical study of random matrices.

1.5.4 Classical Random Matrix Ensembles

The Gaussian ensembles of random matrices have been extensively investigated and dates back to the works in statistical distribution of the widths and spacings of nuclear resonance levels [494] and statistical theory of energy levels of complex systems [135]. Thus random matrix theory as thoroughly discussed in [167, 281, 327] has proved to be pivotal in high dimensional and/or multivariate statistical analysis plus many other applications based Wishart matrix [8] as well as orthogonal polynomials [454]. Therefore, we attempt to investigate the properties of the extreme points of the joint eigenvalue probability density function of the random Wishart matrix optimized over the unit p-sphere [348, 349]. We also apply the techniques of Vandermonde polynomial optimization [308], matrix factorisation [371] and eigenvalue optimization [204], matrix norms [114] and condition number [140].

The key famously known classical ensembles include the Gaussian Orthogonal Ensembles (G.O.E), the Gaussian Unitary Ensembles (G.U.E), the Gaussian Symplectic Ensembles (G.S.E), the Wishart Ensembles (W.E), the MANOVA Ensembles (M.E) and the Circular Ensembles (C.E). These can be derived from the multivariate Gaussian matrix, $\mathbf{G}_\beta, \beta = 1,2,4$. Since, the multivariate Gaussian possesses an inherent orthogonal property from the standard normal distribution, that is, they remain invariant under orthogonal transformations. More detailed discussions on these ensembles can be found in [8, 11, 327, 357, 492, 496].

Definition 1.5.3 ([281])**.** *The Gaussian Orthogonal Ensembles (G.O.E) are characterised by the symmetric matrix* $\mathbf{X} = \mathbf{G}_1(N,N)$ *obtained as* $\left(\mathbf{X}+\mathbf{X}^\top\right)/2$*. The diagonal entries of* $\mathbf{X}$ *are independent and identically distributes (i.i.d) with a standard normal distribution* $\mathcal{N}(0,1)$ *while the off-diagonal entries are i.i.d with a standard normal distribution* $\mathcal{N}_1(0,1/2)$*. That is, a random matrix* $\mathbf{X}$ *is called the Gaussian Orthogonal Ensemble (GOE), if it is symmetric and real-valued* $(X_{ij} = X_{ji})$ *and has*

$$X_{-ij} = \begin{cases} \sqrt{2}\xi_{ii} \sim \mathcal{N}_1(0,1), & \text{if } i = j \\ \xi_{ij} \sim \mathcal{N}_1(0,1/2), & i < j. \end{cases} \tag{1.146}$$

Definition 1.5.4 ([129, 281])**.** *The Gaussian Unitary Ensembles (G.U.E), are characterised by the Hermitian complex-valued matrix* $\mathbf{H} = \mathbf{G}_2(N,N)$ *obtained as* $\left(\mathbf{H}+\mathbf{H}^{\top^*}\right)/2$ *where* $\top^*$ *is the Hermitian transpose, that is, the Hermitian or conjugate transpose of* $\mathbf{H}$*, that can be expressed as* $(\mathbf{H}^{\top^*})_{ij} = \overline{\mathbf{H}}_{ji}$*. The diagonal entries of* $\mathbf{H}$ *are independent and identically distributes (i.i.d) with*

a standard normal distribution $\mathcal{N}(0,1)$ while the off-diagonal entries are i.i.d with a standard normal distribution $\mathcal{N}_2(0,1/2)$. That is, random matrix $\mathbf{H}$ is called a Gaussian Unitary Ensemble (GUE), if it is complex-valued, Hermitian ($\mathbf{H}_{ij}^{\top^*} = \overline{\mathbf{H}}_{ji}$), and the entries satisfy

$$\mathbf{H}_{ij} = \begin{cases} \sqrt{2}\xi_{ii} \sim \mathcal{N}_2(0,1), & \text{if } i = j \\ \frac{1}{\sqrt{2}}(\xi_{ij} + \sqrt{-1}\eta_{ij}) \sim \mathcal{N}_2(0,1/2), & i < j. \end{cases} \tag{1.147}$$

Definition 1.5.5 ([19, 281])**.** *The Gaussian Symplectic Ensembles (G.S.E), are characterised by the self-dual matrix $\mathbf{S} = \mathbf{G}_4(N,N)$ obtained as $\left(\mathbf{S} + \mathbf{S}^{\top^*}\right)/2$ where $(\cdot)^{\top^*}$ represents the conjugate transpose of a quaternion matrix. The diagonal entries $\mathbf{H}$ are independent and identically distributes (i.i.d) with a standard normal distribution $\mathcal{N}(0,1)$ while the off-diagonal entries are i.i.d with a standard normal distribution $\mathcal{N}_4(0,1/2)$.*

Definition 1.5.6. *[19, 281] The Wishart Ensembles (W.E), $\mathcal{W}_\beta(m,n), m \geq n$, are characterised by the symmetric, Hermitian or self-dual matrix $\mathbf{W} = \mathbf{W}_\beta(N,N)$ obtained as $\mathbf{W} = \mathbf{A}\mathbf{A}^\top, \mathbf{W} = \mathbf{H}\mathbf{H}^\top$, or $\mathbf{W} = \mathbf{S}\mathbf{S}^\top$ where $(\cdot)^\top$ represents the usual transposes of defined in G.O.E, G.U.E and G.S.E above respectively.*

Definition 1.5.7 ([19, 281])**.** *The MANOVA Ensembles (M.E), $\mathrm{J}_\beta(m_1,m_2,n), m_1,m_2 \geq n$, are characterised by the symmetric, Hermitian or self-dual matrix $\mathbf{A}/(\mathbf{A}+\mathbf{B})$ where $\mathbf{A}$ and $\mathbf{B}$ are $\mathbf{W}_\beta(m_1,n)$ and $\mathbf{W}_\beta(m_2,n)$ respectively.*

Definition 1.5.8 ([129, 281])**.** *The Circular Ensembles (C.E), are characterised by the special matrix $\mathbf{U}\mathbf{U}^\top$ where $\mathbf{U}_\beta, \beta = 1,2$ is a uniformly distributed unitary matrix.*

Lemma 1.5.5 ([281])**.** *From the Gaussian normal distribution with mean μ and variance σ^2, that is, $\mathbf{X} \sim \mathcal{N}(\mu,\sigma^2)$, given by* (1.137) *and the multivariate normal distribution with mean vector μ and the covariance matrix is $\mathbf{\Sigma}$, $\mathcal{N}_N(\boldsymbol{\mu},\mathbf{\Sigma})$ given in* (1.141)*, then it can be verified that the joint density of $\mathbf{A}$ is written as:*

$$\mathbb{P}_{\mathbf{X}}(\mathbf{A}) = \frac{1}{2^{n/2}} \frac{1}{\pi^{n(n+1)/4}} \exp\left(-\|\mathbf{A}\|_F^2/2\right)$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix.

Definition 1.5.9. *The Frobenius norm is in principle an extension of the usual Euclidean norm for vectors:*

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{tr(\mathbf{A}^*\mathbf{A})}$$

where tr is the matrix track and $\mathbf{A}^*$ denotes the hermite conjugate of $\mathbf{A}$.

Theorem 1.5.6 ([8, 129])**.** *If we let $\mathbf{X}$ be an $N \times N$ random matrix with entries that are independently identically distributed as $\mathcal{N}(0,1)$, then the joint density distribution of the Gaussian ensembles is given by:*

$$\textit{Gaussian:} \quad \left| \begin{array}{cc} \textit{Orthogonal} & \beta = 1 \\ \textit{Unitary} & \beta = 2 \\ \textit{Symplectic} & \beta = 4 \end{array} \right| \qquad \mathbb{P}_\beta(\mathbf{A}) = \frac{1}{2^{n/2}} \frac{1}{\pi^{n(n+1)\ \beta/4}} \exp\left(-\frac{1}{2}\|\mathbf{A}\|_F^2\right).$$

Theorem 1.5.7 ([129, 327])**.** *Considering a Wishart matrix* $\mathbf{W}_\beta(m,n) = \mathbf{X}\mathbf{X}^\top$ *where* $\mathbf{X} = \mathbf{G}_\beta(m,n)$ *is a multivariate Gaussian matrix. Then, the joint elements of* $\mathbf{W}_\beta(m,n)$ *can be computed in two steps, first writing* $\mathbf{W} = \mathbf{Q}\mathbf{R}$ *and then integrating out* $\mathbf{Q}$ *leaving* $\mathbf{R}$*. Secondly applying the transformation* $\mathbf{W} = \mathbf{R}\mathbf{R}^\top$*, which is the famous Cholesky factorization of matrices in numerical analysis. Then the joint density distribution for Wishart ensembles of* $\mathbf{W}$ *is given by:*

$$\textit{Wishart:} \quad \left| \begin{array}{cc} \textit{Orthogonal} & \beta = 1 \\ \textit{Unitary} & \beta = 2 \\ \textit{Symplectic} & \beta = 4 \end{array} \right| \qquad \mathbb{P}_\beta(W) = \frac{\exp\left(-tr(\mathbf{W}/2)\right)\left(\det \mathbf{W}\right)^{\beta(m-n+1)/2-1}}{2^{mn\beta/2}\Gamma_n^\beta(m\beta/2)}.$$

Here we notice that the density distribution for both the Gaussian and Wishart ensembles are made up of determinant term and exponential trace term. This generalizes the fact that indeed the determinant term is actually the Vandermonde determinant in (1.8) for the case of the joint eigenvalue density functions. This concept further explained in the next section.

1.5.5 Gaussian ensembles

The most studied random matrix ensembles are the Gaussian ensembles. For example, taking an $N \times N$ matrix whose entries are independently sampled from the Gaussian probability density function with mean 0 and variance 1. For $N = 6$ we may generate the following non-symmetric matrix

$$\mathbf{X} = \begin{pmatrix} 1.2448 & 0.0561 & -0.8778 & 1.1058 & 1.1759 & 0.7339 \\ -0.1854 & 0.7819 & -1.3124 & 0.8786 & 0.3965 & -0.3138 \\ -0.4925 & -0.6234 & 0.0307 & 0.8448 & -0.2629 & 0.7013 \\ 0.1933 & -1.5660 & 2.3387 & 0.4320 & -0.0535 & 0.2294 \\ -1.0143 & -0.7578 & 0.3923 & 0.3935 & -0.4883 & -2.7609 \\ 1.8839 & 0.4546 & -0.4495 & 0.0972 & -2.6562 & 1.3405 \end{pmatrix}$$

whose entries are positive, some negative, and all not far from zero. Since the matrix $\mathbf{X}$ is non-symmetric, that is, $\mathbf{X}_{ij} \neq \mathbf{X}_{ji}$, then $\mathbf{X}$ has complex eigenvalues. To obtain real distinct eigenvalues, we symmetrize $\mathbf{X}$ so that into real Hermitian form so that

$$\mathbf{H} = \left(\mathbf{X} + \mathbf{X}^\top\right)/2 \tag{1.148}$$

where $(\cdot)^\top$ denotes the transpose of the defined matrix. It follows that the symmetric sample matrix becomes

$$\mathbf{H} = \begin{pmatrix} 1.2448 & -0.0646 & -0.6852 & 0.6496 & 0.0807 & -0.5750 \\ -0.0646 & 0.7819 & -0.9679 & -0.3436 & -0.1806 & 0.0704 \\ -0.6852 & -0.9679 & 0.0307 & 1.5917 & 0.0647 & 0.1258 \\ 0.6496 & -0.3436 & 1.5917 & 0.4320 & 0.1700 & 0.1633 \\ 0.0807 & -0.1806 & 0.0647 & 0.1700 & -0.4883 & -2.7085 \\ -0.5750 & 0.0704 & 0.1258 & 0.1633 & -2.7085 & 1.3405 \end{pmatrix} \tag{1.149}$$

whose eigenvalues are given as

$$\lambda = \{-2.4932, -1.7534, 0.3307, 1.4459, 2.3823, 3.4294\}.$$

The Gaussian density distribution of such values for different samples is as shown in Figure 1.2. Before we define all the cases of ensembles, it is important to first define Gaussian matrices, Hermitian matrices and quaternion:

Definition 1.5.10. *The real $N \times N$ symmetric matrix $\mathbf{X} = (\mathbf{X}_{ij})$ whose entries are such that $\mathbf{X}_{ij}$ are independent and identically distributed as normal distribution that has mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a Gaussian (random) matrix. The density of the entries $\mathbf{X}_{ii}$ on the major diagonal has the density $f(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right)$ and the off-diagonal entries $\mathbf{X}_{ij}$ have the density $f(x_{ij}) = \frac{1}{\sqrt{\pi}} \exp\left(-x_{ij}^2\right)$.*

Definition 1.5.11. *The Hermitian matrices is that they have real elements on the major diagonal, and complex conjugates for off-diagonal entries. The Hermitian matrix is also a self-adjoint matrix, that is, a complex square matrix that is equal to its own conjugate transpose so that the element in the i-th row and j-th column is equal to the complex conjugate of the element in the j-th row and i-th column, for all indices i and j. Thus $x_{ij} = \overline{x_{ji}}$ or $\mathbf{H} = \overline{\mathbf{H}^\top}$.*

Definition 1.5.12. *The quaternion self-dual matrices are $2N \times 2N$ matrices constructed as*

$$\mathbf{H} = [\mathbf{X} \;\; \mathbf{Y}; -\text{conj}(\mathbf{Y}) \; \text{conj}(\mathbf{X})];$$

where $\mathbf{H}$ is as expressed in (1.148)*, $\mathbf{X}$ and $\mathbf{Y}$ are complex matrices, while* conj *denotes complex conjugation of entries. The Quaternions are generally a number system that extends the complex numbers and are represented in the form:*

$$x = a + bi + cj + dk$$

where a, b, c, and d are real numbers, and i, j, and k are the fundamental quaternion units. These can be combined under multiplication as in Table 1.3

$\times$	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

Table 1.3: Quaternion multiplication and equivalently $i^2 = j^2 = k^2 = ijk = -1$.

The above ideas of random matrices can be extended to an arbitrary number of N of variables. For a given joint probability density function $f(x_1,\ldots,x_N)$, the quantity $f(x_1,\ldots,x_N)dx_1\ldots dx_N$ is the probability that one finds the first variable x_1 in the interval $[x_1, x_1+dx_1]$, the second variable x_2 in the interval $[x_2, x_2+dx_2]$ and so on. The marginal probability density function $f(x_1)$ is the probability that the first variable will be in the interval $[x_1, x_1+dx_1]$ can be computed as

$$f(x) = \int_{x_2\in\mathbb{R}} \ldots \int_{x_N\in\mathbb{R}} f(x_2,\ldots,x_N)dx_N\ldots dx_2.$$

If for a given set of random variables that is a function of another one say, $x_i = x_i(t)$, then there is a relation between the joint probability density function of the two sets given by

$$f(x_1,\ldots,x_N)dx_1\ldots dx_N = \underbrace{f(x_1(t),\ldots,x_N(t))|\,J(x\to t)|}_{\hat{f}(t_1,\ldots,t_N)} dt_1\ldots dt_N$$

where $J(x\to t)$ is their Jacobian of transformation given by $J(x\to t) = \det\left(\dfrac{\partial x_i}{\partial t_i}\right)$ and $|\cdot|$ is that absolute value of the given term.

Now, from the Gaussian random matrix (1.149), the joint probability density function $f(\mathbf{H})$ of the N^2 entries $\{x_{11},\ldots,x_{NN}\}$ of the matrix $\mathbf{H}$ that are independent Gaussian random variables is given by:

1. The Gaussian unitary ensemble GUE(n) is described by the Gaussian measure with density

$$f(\mathbf{H}) = f(x_{11},\ldots,x_{NN}) = \prod_{i,j=1}^{N}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_{ij}^2\right)\right] = \frac{1}{\mathrm{Z}_{\mathrm{GUE}(N)}}e^{-\frac{N}{2}\mathrm{tr}\left(\mathbf{H}^2\right)} \qquad (1.150)$$

   on the space of $N\times N$ Hermitian matrices $\mathbf{H} = (\mathrm{H}_{ij})_{i,j=1}^{N}$. Here $\mathrm{Z}_{\mathrm{GUE}(N)} = 2^{N/2}\pi^{N^2/2}$ is a normalization constant, chosen so that the integral of the density is equal to one. The term unitary refers to the fact that the distribution is invariant under unitary conjugation.

2. The Gaussian orthogonal ensemble GOE(N) is described by the Gaussian measure with density

$$f(\mathbf{H}) = f(x_{11},\ldots,x_{NN}) = \prod_{i,j=1}^{N}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{N}{4}x_{ij}^2\right)\right] = \frac{1}{\mathrm{Z}_{\mathrm{GUE}(N)}}e^{-\frac{N}{4}\mathrm{tr}\left(\mathbf{H}^2\right)} \qquad (1.151)$$

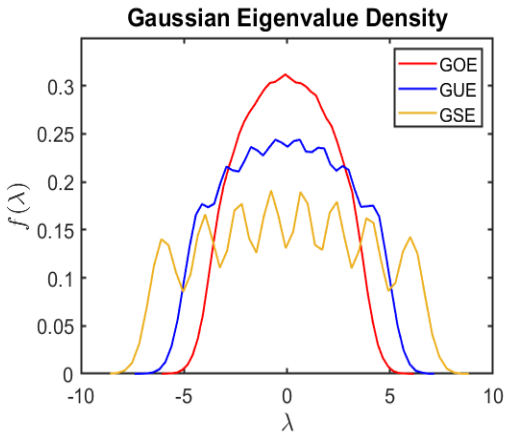


Figure 1.2: The Gaussian Eigenvalue Densities for GOE, $\beta = 1$, GUE $\beta = 2$, and GSE $\beta = 4$ for $N = 8$ and number of samples $T = 50000$.

on the space of $n \times n$ real symmetric matrices $\mathbf{H} = (H_{ij})_{i,j=1}^{n}$. Its distribution is invariant under orthogonal conjugation.

3. The Gaussian symplectic ensemble GSE(N) is described by the Gaussian measure with density

$$f(\mathbf{H}) = f(x_{11},\ldots,x_{NN}) = \prod_{i,j=1}^{N}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-Nx_{ij}^2\right)\right] = \frac{1}{\mathrm{Z}_{\mathrm{GUE}(N)}}e^{-N\mathrm{tr}\left(\mathbf{H}^2\right)} \tag{1.152}$$

on the space of $N \times N$ Hermitian quaternionic matrices, for example, symmetric square matrices composed of quaternions, $\mathbf{H} = (H_{ij})_{i,j=1}^{N}$. Its distribution is invariant under conjugation by the symplectic group, and it models Hamiltonians with time-reversal symmetry but no rotational symmetry.

The joint probability density for the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of GUE/GOE/GSE is given by

$$\frac{1}{\mathrm{Z}_{\beta,N}}\prod_{i<j}\left|\lambda_j - \lambda_i\right|^{\beta}\prod_{k=1}^{n}e^{-\frac{\beta}{4}\lambda_k^2}, \tag{1.153}$$

where the Dyson index, $\beta = 1$ for GOE, $\beta = 2$ for GUE, and $\beta = 4$ for GSE, counts the number of real components per matrix element;

$$\mathrm{Z}_{\beta,N} = (2\pi)^{N/2}\prod_{j=1}^{N}\frac{\Gamma(1+j\beta/2)}{(1+\beta/2)}$$

is a normalization constant which can be explicitly computed, see [420]. In the case of GUE ($\beta = 2$), the formula (1.153) describes a determinantal point process. Eigenvalues repel as the joint probability density has a zero (of β th-order) for coinciding eigenvalues $\lambda_j = \lambda_i$.

The Gaussian factor $\prod_{j=1}^{N} e^{-\frac{\beta}{4}\lambda_j^2} = \exp\left(-\frac{1}{2}\sum_{j=1}^{N}\lambda_j^2\right)$ kills any configuration of eigenvalues $\{\lambda\}$ where λ_j'a are big, that is, as $\lambda_j \to \infty$ the factor vanishes to zero, thus the eigenvalues do not stay too far from the origin. On the other hand, the difference-product factor $\prod_{i<j}|x_{j-x_i}|$ kills the configuration when two eigenvalues are close to each other, since the factor vanishes to zero, that is $\lim_{i\neq j}|x_j - x_i| \to 0$.

The other effect of the the repulsion factor $\prod_{i<j}|x_{j-x_i}|$ is that it makes the eigenvalues strictly and strongly non-independent. In other words, each eigenvalue feels the presence of others, thus the joint eigenvalue probability density function (1.153) does not factorize at all.

For the distribution of the largest eigenvalue for GOE, GUE and Wishart matrices of finite dimensions, for details see [85].

1.5.6 Distribution of Level Spacings

Considering a 2×2 GOE matrix $\mathbf{H} = \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix}$ where $x_1, x_2 \sim \mathcal{N}(0,1)$ and $x_3 \sim \mathcal{N}(0, \frac{1}{2})$, the we aim to evaluate the probability density function of the spacing $c = \lambda_2 - \lambda_1$ between the two eigenvalues $\lambda_2 > \lambda_1$ also called the Wigner surmise [195]. The two eigenvalues are indeed random variables expressed in terms of the entries of the roots of the characteristic polynomial

$$\lambda^2 - \mathrm{tr}(\mathbf{H})\lambda + \det(\mathbf{H})$$

whose zeros are

$$\lambda_{1,2} = \frac{1}{2}\left(x_1 + x_2 \pm \sqrt{(x_2 - x_1)^2 + 4x_3^2}\right).$$

This implies that the spacing between the two roots is given by

$$s = \lambda_2 - \lambda_1 = \sqrt{(x_2 - x_1)^2 + 4x_3^2}.$$

By definition, the density of the spacing of the eigenvalues λ_1 and λ_2

$$p(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}x_1^2}}{\sqrt{2\pi}}\frac{e^{-\frac{1}{2}x_2^2}}{\sqrt{2\pi}}\frac{e^{-x_3^2}}{\sqrt{\pi}}\delta\left(s - \sqrt{(x_2 - x_1)^2 + 4x_3^2}\right) dx_1 dx_2 dx_3. \tag{1.154}$$

By change of variable such that

$$\begin{cases} x_1 - x_2 = r\cos(\theta) \\ 2x_2 = r\sin(\theta) \\ x_1 + x_2 = \phi \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1}{2}(r\cos(\theta) + \phi) \\ x_2 = \frac{1}{2}(\phi - r\sin(\theta)) \\ x_3 = \frac{1}{2}(r\sin(\theta)) \end{cases}.$$

On computing the corresponding Jacobian we obtain

$$J = \det\begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} & \frac{\partial x_1}{\partial \phi} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} & \frac{\partial x_2}{\partial \phi} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \theta} & \frac{\partial x_3}{\partial \phi} \end{pmatrix} = \det\begin{pmatrix} \frac{1}{2}(\cos(\theta)) & -\frac{1}{2}r(\sin(\theta)) & \frac{1}{2} \\ -\frac{1}{2}(\cos(\theta)) & -\frac{1}{2}r(\sin(\theta)) & \frac{1}{2} \\ (\sin(\theta)) & r(\sin(\theta)) & 0 \end{pmatrix} = -\frac{r}{4}.$$

It follows that

$$\begin{aligned} p(s) &= \frac{1}{2^3\pi^{\frac{3}{2}}}\int_0^{2\pi}\int_0^{s}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[\left(\frac{r\cos(\theta)+\phi}{2}\right)^2+\left(\frac{-r\cos(\theta)+\phi}{2}\right)^2\right.\right. \\ &\quad \left.\left.+\left(\frac{r\sin(\theta)+\phi}{\sqrt{2}}\right)^2\right]\right\}\rho r d\phi dr d\theta \qquad (1.155) \\ &= \frac{\sqrt{4\pi}}{2^3\pi^{\frac{3}{2}}}\int_0^{2\pi}\exp\left\{-\frac{1}{2}\left[\frac{s^2\cos^2(\theta)}{2}+\frac{s^2\sin^2(\theta)}{2}\right]\right\} \\ &= \frac{1}{2}se^{-\frac{1}{4}s^2} \qquad (1.156) \end{aligned}$$

where the identity $\cos^2\theta+\sin^2\theta=1$ is used in the simplifications. It can be noticed that the spacing probability density function in (1.155) is actually normalized in that

$$\int_0^{\infty} p(s)ds = 1.$$

For the ordered sequence of eigenvalues $\lambda_1 < \ldots < \lambda_n < \lambda_{n+1} < \ldots$, the spacing probability density function is often rescaled by defining $\bar{p}(s) = \mathbb{E}[s]p(\mathbb{E}[s]s)$ where $\mathbb{E}[s] = \int_0^{\infty} sp(s)ds$ is the mean level spacing and upon scaling it follows that $\int_0^{\infty}\bar{p}(s)ds = \int_0^{\infty}s\bar{p}(s)ds = 1$.

It follows immediately from the rescaled spacing probability density function that from (1.153),

$$\begin{cases} p_\beta(s) = \frac{\pi}{2}se^{-\frac{\pi}{4}s^2}, & \text{for GOE } \beta = 1 \\ p_\beta(s) = \frac{32}{\pi^2}s^2e^{-\frac{4}{\pi}s^2}, & \text{for GUE } \beta = 2 \\ p_\beta(s) = \frac{2^{18}}{3^6\pi^3}s^4e^{-\frac{64}{9\pi}s^2}, & \text{for GSE } \beta = 4 \end{cases}. \qquad (1.157)$$

The eigenvalue spacing probability density functions (1.157) are as illustrated in Figure 1.3.

1.5.7 The Vandermonde determinant in systems with Coulombian interactions

Several interesting mathematical problems that feature Vandermonde matrices and Vandermonde determinant can be described as questions about systems with Coulombian interactions. The name

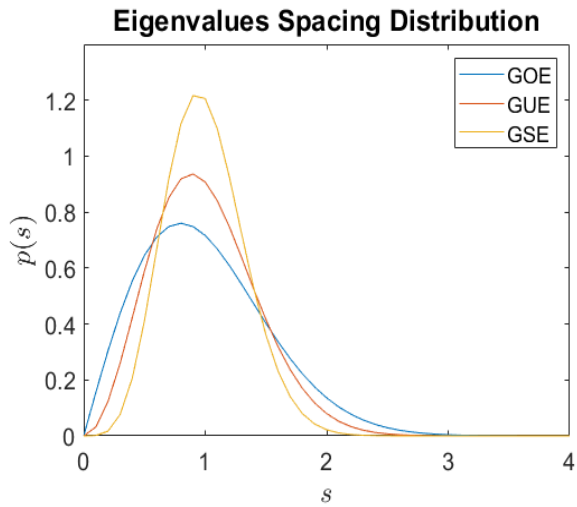


Figure 1.3: The Gaussian Eigenvalue Spacing Distributions for GOE, $\beta = 1$, GUE $\beta = 2$, and GSE $\beta = 4$.

Coulombian interaction come from Charles-Augustin Coulomb (1736-1806) who is probably most well-known for quantifying the force between two charged particles (what is today known as Coulomb's law) in 1785 [96]. Coulombs law states that the force between two charged particles is proportional to the product of the charges and the inverse of the square of the distance between the two charges.

Taking the logarithm of the expression for the Vandermonde determinant given in (1.79) and (1.153) gives

$$\log\big(v_n(x_1,\ldots,x_n)\big) = \sum_{1\le i<j\le n} \log(x_j - x_i)$$

and thus $-\frac{1}{2}\log\big(v_n(x_1,\ldots,x_n)\big)$ gives the same as setting $g(x) = \log(x)$ and $V(x) \equiv 0$ in

$$\mathscr{H}_N(x_1,\ldots,x_N) = \frac{1}{2}\sum_{i\neq j} g(x_i - x_j) + N\sum_{i=1}^{N} V(x_i). \tag{1.158}$$

the Coulombs law which states that the force between two charged particles is proportional to the product of the charges and the inverse of the square of the distance between the two charges [96].

The points x_i usually belong to $\mathbb{R}^d$ (or some subset thereof) but there is also research that involves more general manifolds. A common goal is to minimize this energy or find some other extreme points. There are many areas where this kind of problems, or closely related problems

appear. See the extended version of [421] for a recent review of the field. In this section we will mention a few examples of interesting systems with Coulombian interactions that are connected to the Vandermonde determinant and its applications in electrostatics are discussed in [305] and in other chapters of this thesis.

Distribution of electrical charges

The most classical example of a system with Coulombian interactions is a system of charged particles confined to some volume, even if it was not studied (from a mathematical point of view) until almost a hundred years after Coulomb's law was introduced [223, 443]. The classical mathematical formulation of this problem considers $p+1$ charges fixed at points $a_0,\ldots,a_p \in \mathbb{C}$ with weights $\eta_0,\cdots,\eta_p$ and n moveable charges $x_1,\ldots,x_n$. The questions is then what x-values give the extreme points of $L(x_1,\ldots,x_n)$ given by

$$\mathrm{L}(x_1,\ldots,x_n) = \sum_{k=1}^{n}\sum_{j=0}^{p} \eta_j \log\left(\frac{1}{|a_j - x_k|}\right) + \sum_{1\le i<k\le n} \log\left(\frac{1}{|a_i - x_k|}\right).$$

More background on this type of problem together with a collection of recent results can be found in [173]. If there are no fixed charges the problem becomes equivalent to maximising the absolute value of the Vandermonde determinant similar to finding the Fekete points. The problems discussed in Chapter Chapter 5 and Chapter 9 belongs to the class of equations that are called Schrödinger-like in [173].

Sphere Packing

There are several different interaction kernels apart from the logarithmic interaction kernel characterised by, $g(x) = -\log(x)$, that are interesting in mathematical physics, especially statistical mechanics and quantum mechanics. One important class of interaction kernels are those given by $g(x) = \dfrac{1}{|x|^s}$ where s is a positive integer. When this interaction kernel is used value given by formula (1.158) is called the Riesz s-energy. There is a large body of significant literature, in [421] over 30 references are listed as introduction to different related problems [91, 92, 93, 94].

It is worth noting that $\lim_{s\to\infty}\left(1-\dfrac{1}{|x|^s}\right) = -\log(|x|)$ which connects the minimisation of the Riesz s-energy to the Fekete points.

If we instead $s\to\infty$ the problem of minimising the Riesz s-energy formally corresponds to the optimal sphere-packing problem, that is finding the arrangement of non-overlapping identical spheres that cover as much of a space as possible. This is a classical problem where extensive effort has gone into finding optimal packings but for many years the problem was only fully solved in one, two and three dimensions, until recently when surprisingly simple proofs were found for 8 and 24 dimensions (seemingly without giving any results for any number of dimensions in-between). For a thorough collection of classical results see [58] and for the recent results see [91, 92, 93, 477].

Coulomb Gas

In mathematical physics a system of particles whose energy can be described by (1.158) is often called a Coulomb gas [167, 326, 421]. One of the most wide reaching results in the analysis of Coulomb gases was that many gas systems can be described using random matrices that belongs to a so–called–ensemble which is defined by matrices with random elements. The foundational results were found in the early 1960s and applied to the cases where $\beta = 1, \beta = 2$ and $\beta = 4$, [131, 132, 133, 134, 135, 136]. These cases will be briefly discussed in Section 1.5.5 and describe where the Vandermonde determinant appears the probability density functions for the eigenvalues of the random matrices. If the same theory is extended to other values of it can also be connected to equations similar to the Harish-Chandra–Itzykson–Zuber integral formula described in Equation (1.134), [167].

1.6 Symmetric Cones and Jordan Algebras

This section provides the overview of the structural analysis of symmetric cones and Jordan algebras which are very important for understanding of the generalized concepts such as Gamma functions, Beta function, hypergeometric functions, zonal polynomials, orthogonal polynomials and Wishart Ensembles that will be discussed mainly in Chapter 8 and Chapter 9.

Let $(\mathrm{E}, \langle .,. \rangle)$ be an Euclidean vector space and let $\mathrm{GL}(\mathrm{E})$ be the group of general linear transformations on E. Let Ω a convex cone satisfying $\alpha x + \beta y \in \Omega$ for any $\alpha, \beta > 0$ such that $\alpha + \beta < 1$ and $x, y \in \Omega$. Considering Ω to be a subset of E, then the set Ω has several interesting properties.

To describe them, let $\mathrm{GL}(\mathrm{E})$ be the group of all invertible linear operators in E (GL stands for General Linear). Also, let $\overline{\Omega}$ denote the closure of Ω in E, then the convex cone Ω is proper if $\overline{\Omega} \bigcap (-\overline{\Omega}) = \{0\}$, full if $\Omega - \Omega = \mathrm{E}$, where the set $\Omega - \Omega = \{x - y \,:\, x \in \Omega, y \in \Omega\}$ is the smallest subset E containing Ω.

Definition 1.6.1 ([155]). *The* automorphism group $G(\Omega)$ *of the set* Ω *is defined by*

$$\mathrm{G}(\Omega) = \{\, g \in \mathrm{GL}(\mathrm{E}) \colon g\Omega = \Omega \,\}.$$

An element $g \in g \in \mathrm{GL}(\mathrm{E})$ *belongs to* $\mathrm{G}(\Omega)$ *if and only if* $g\overline{\Omega} = \overline{\Omega}$.

Let Ω be again the set of all $m \times m$ positive-definite matrices. Let g be an invertible $m \times m$ matrix with positive determinant. The linear operator $\rho(g)$ acting in the linear space $\mathrm{Sym}(m, \mathbb{R})$ by

$$\rho(g)x = gxg^{\top}, \qquad x \in \mathrm{Sym}(m, \mathbb{R}),$$

leaves Ω invariant. We proved that the group $\mathrm{G} = \mathrm{GL}^{+}(m, \mathbb{R})$ is a subgroup of $G(\Omega)$ (the upper index, $+$, stands for the positive determinant).

We write down a list of properties of the set Ω.

Theorem 1.6.1. *The set* Ω *has the following properties:*

(i) Ω *is a* cone*:* $x \in \Omega$ *and* $\lambda > 0$ *imply that* $\lambda x \in \Omega$*.*

(ii) Ω *is* convex*:* $x, y \in \Omega$ *and* $\lambda \in [0,1]$ *imply that* $\lambda x + (1-\lambda)y \in \Omega$*.*

(iii) The open convex cone Ω *is* self-dual*: we have* $\Omega^* = \Omega$*, where*

$$\Omega^* = \{\, y \in \mathrm{E} \colon (x|y) > 0 \quad \textit{for all} \quad x \in \overline{\Omega} \setminus \{0\} \,\}$$

is the open dual cone *of* Ω*.*

(iv) The open cone Ω *is* homogeneous*: for all* $x, y \in \Omega$ *there exists* $g \in \mathrm{G}(\Omega)$ *such that* $gx = y$*.*

(v) Finally, the symmetric *(self-dual and homogeneous) cone* Ω *is* irreducible*: there do not exist nontrivial subspaces* E_1, E_2 *and symmetric cones* $\Omega_1 \subset \mathrm{E}_1$, $\Omega_2 \subset \mathrm{E}_2$ *such that* $\mathrm{E} = \mathrm{E}_1 \oplus \mathrm{E}_2$ *and* $\Omega = \Omega_1 \times \Omega_2$*.*

Do there exists symmetric irreducible cones besides the cone $\Omega = \Pi_m(\mathbb{R})$ *of* $m \times m$ *positive-definite matrices with real entries? Can we define the Wishart distributions on all symmetric irreducible cones?*

To answer these questions, turn the linear space $\mathrm{E} = \mathrm{Sym}(m, \mathbb{R})$ into an *algebra*.

1.6.1 Euclidean Jordan Algebras

Definition 1.6.2 ([155])**.** *A real linear space* E *is called an* algebra *if* $(x,y) \mapsto x \circ y$ *is a bilinear mapping from* $\mathrm{E} \times \mathrm{E}$ *to* E *is defined. The above mapping is called a* product*.*

Example 1.6.2. *Let* $\mathrm{E} = \mathrm{Sym}(m, \mathbb{R})$*. Define the product by*

$$x \circ y = \frac{1}{2}(xy + yx).$$

One can easily see that this product is indeed bilinear. Moreover, one can check the following properties of the introduced product.

$$x \circ y = y \circ x, \tag{1.159a}$$

$$x \circ (x^2 \circ y) = x^2 \circ (x \circ y), \tag{1.159b}$$

$$(x \circ u|v) = (u|x \circ v) \tag{1.159c}$$

for all x, y, u, and v in E*.*

Definition 1.6.3. *An algebra* $(\mathrm{E}, \circ)$ *is called a* Jordan algebra *if it satisfies (1.159a) and (1.159b).*

Definition 1.6.4. *A Jordan algebra defined on a linear space E with an inner product is called* Euclidean *if it satisfies (1.159c).*

Jordan algebras were introduced by P. Jordan, J. von Neumann, and E. Wigner in [256, 479].

The Euclidean Jordan algebra $\mathrm{Sym}(m,\mathbb{R})$ has one more important property.

Definition 1.6.5. *An nonempty subset* I *of a commutative (i.e., satisfying (1.159a)) algebra is called an* ideal *if* $x \circ y \in \mathrm{I}$ *as long as* $x \in \mathrm{E}$ *and* $y \in \mathrm{I}$.

It is obvious that any algebra E contains at least two ideals: $\{0\}$ and E. The above ideals are called *trivial*.

Definition 1.6.6. *An algebra* E *is called* simple *if it does not contain nontrivial ideals.*

The Euclidean Jordan algebra $\mathrm{Sym}(m,\mathbb{R})$ is simple, see [155, Theorem V.3.7].

The following result explains why we introduced Jordan algebras. It describes a one-to-one correspondence between irreducible symmetric cones and simple Euclidean Jordan algebras.

Theorem 1.6.3. *In a simple Euclidean Jordan algebra, the set* Ω *of squares of all invertible elements is an irreducible symmetric cone. Conversely, any irreducible symmetric cone is a set of squares of invertible elements of a certain simple Euclidean Jordan algebra.*

1.6.2 The Cone of Positive Definite Symmetric Matrices

Denoting by $\mathrm{Sym}(m,\mathbb{R})$ the space of $m \times m$ real symmetric matrices. Then, an inner product on this space is given by

$$\begin{aligned}(x|y) &= \mathrm{tr}(xy) = \sum_{i,j} x_{ij} y_{ij} \\ &= \sum_i x_{ii} y_{ii} + 2 \sum_{i<j} x_{ij} y_{ij}.\end{aligned}$$

Thus for a symmetric matrix x, one associates the quadratic form

$$Q(\boldsymbol{\xi}) = \sum_{i,j} x_{ij} \xi_i \xi_j.$$

Considering the vector $\boldsymbol{\xi}$ as an $m \times 1$ matrix, then the quadratic form can be written as

$$Q(\boldsymbol{\xi}) = (x|\boldsymbol{\xi}\boldsymbol{\xi}^\top).$$

where $\top$ represents transpose. The quadratic form Q, or the symmetric matrix x, is said to be positive definite [155] if

$$\forall\, \boldsymbol{\xi} \in \mathbb{R}^m,\ \boldsymbol{\xi} \neq \mathbf{0},\ \text{then}\ \ Q(\boldsymbol{\xi}) > 0,$$

and positive semi-definite if

$$\forall\, \boldsymbol{\xi} \in \mathbb{R}^m, Q(\boldsymbol{\xi}) \geq 0.$$

We recall that a positive quadratic form Q is a sum of the squares of k independent linear forms $(0 \leq k \leq m)$ where

$$Q((\xi)) = \sum_{j=1}^{k} \left(\sum_{i=1}^{m} \alpha_{ij} \xi_j^2 \right) \text{ or } x = \sum_{j=1}^{k} \alpha_j \alpha_j^\top = \boldsymbol{\alpha}\boldsymbol{\alpha}^\top,$$

where x is a symmetric matrix corresponding to Q; $\alpha_j = (\alpha_{1j}, \ldots, \alpha_{mj})^\top$ and $\boldsymbol{\alpha} = \alpha_{ij}$ is a real $n \times k$ matrix. The number k is the rank of x. The matrix x is positive definite if and only if $k = m$.

Lemma 1.6.4 ([155]). *Let $\Omega = \Pi_m(\mathbb{R})$ be the set of positive definite symmetric matrices. Then, the set Ω is an open convex cone whose closure $\overline{\Omega}$ is the set of positive symmetric matrices.*

Proof. First, we prove that the the symmetric cone Ω is self-dual. Let y belong to the self-dual $\Omega^\star$. If $\boldsymbol{\xi}$ is any non-zero $m \times 1$ matrix, then the matrix $x = \boldsymbol{\xi}\boldsymbol{\xi}^\top$ belongs to $\overline{\Omega} \setminus \{0\}$, therefore

$$\sum_{ij} y_{ij} \xi_i \xi_j = (x|y) > 0,$$

which directly implies that y belongs to Ω and $\Omega^\star \subset \Omega$. Thus, any element in x in $\overline{\Omega} \setminus \{0\}$ can be written as

$$x = \sum_{j=1}^{k} \alpha_j \alpha_j^\top,$$

where the α_j are independent $m \times 1$ matrices and $k \geq 1$. Therefore, if y belongs to Ω, then

$$(x|y) = \sum_{j=1}^{k} (y|\alpha_j \alpha_j^\top) > 0,$$

which proves that $\Omega \subset \Omega^\star$.

Next, we prove that Ω is homogeneous. For an element g of the group $\mathrm{GL}(m, \mathbb{R})$ and a symmetric matrix x, we set

$$\rho(g)x = gxg^\top,$$

then $\rho(g)$ is a linear transformation of the space $\mathrm{Sym}(m, \mathbb{R})$ which belongs to $G(\Omega)$. If x belongs to Ω, then

$$x = \boldsymbol{\alpha}\boldsymbol{\alpha}^\top,$$

where $\boldsymbol{\alpha}$ is invertible $m \times m$ matrix, or

$$x = \rho(\boldsymbol{\alpha})\mathbf{I}_m.$$

which show that Ω is homogeneous. □

1.6.3 Properties and Examples of Jordan Algebras

Now, for an element x in V, let $\mathrm{L}(x)$ be a bilinear map of V defined by

$$\mathrm{L}(x)y = xy.$$

Then, an algebra V over the field $\mathbb{F}$ is said to be a Jordan algebra if for all elements x and y in V:

$$xy = yx \tag{J_1}$$
$$x(x^2y) = x^2(xy). \tag{J_2}$$

The property (J_2) says that for every x the endomorphisms and $\mathrm{L}(x^2)$ commute. Thus, using the notation

$$[S,T] = ST - TS,$$

where S and T are two endomorphisms of the vector space V, thus the property J_2 can be written as

$$[\mathrm{L}(x),\mathrm{L}(x^2)] = 0. \tag{J_2'}$$

In general, a Jordan algebra is not associative.

Example 1.6.5 ([155])**.** *(1) If* A *is an associative algebra over* $\mathbb{F}$*, we can define on* A *a Jordan algebra structure by defining the new product:*

$$x \circ y = \frac{1}{2}(xy + yx).$$

It is easy to check that with the new product A *is a Jordan algebra, we call the operation* $\circ$ *the Jordan product.*

(2) If V *is a linear subspace of an associative algebra* A *which is square stable, that is, for any* x *in* V, x^2 *belongs to* V*, then* V *is equipped with Jordan product*

$$x \circ y = \frac{1}{2}(xy + yx),$$

is a Jordan algebra. (In fact, we have $x \circ y = \frac{1}{4}\big((x+y)^2 - (x-y)^2\big)$.

(3) A special case of (2) above, take

$$\mathrm{A} = \mathrm{M}(m,\mathbb{R})$$

the algebra of $m \times m$ *matrices with entries in* $\mathbb{R}$*, and* V *the space* $\mathrm{Sym}(m,\mathbb{R})$ *of symmetric matrices.*

(4) Let W *be a vector space over* **F**, *and* B *be a symmetric bilinear form defined on* W. *Then, on the vector space* $\mathrm{V} = \mathbb{F} \times \mathrm{W}$ *we define the product*

$$(\lambda, u)(\mu, v) = (\lambda\mu + \mathrm{B}(u,v), \lambda v + \mu u).$$

If $x = (\lambda, u)$, *then*

$$x^2 = (\lambda^2 + B(u,u), 2\lambda u).$$

If we write $\mathrm{T} = \mathrm{L}(0,u)$, *then*

$$\begin{aligned} \mathrm{L}(x) &= \lambda\mathbf{I} + \mathrm{T} \\ \mathrm{L}(x^2) &= \big(\lambda^2 + \mathrm{B}(u,u)\mathbf{I}\big) + 2\lambda\mathrm{T}, \end{aligned}$$

therefore $\mathrm{L}(x)$ *and* $\mathrm{L}(x^2)$ *commute and* V *is a Jordan algebra.*

Proposition 1.6.1. *Let* V *be a Jordan algebra. Then, the following identities hold:*

(i) $[\mathrm{L}(x), \mathrm{L}(y^2)] + 2[\mathrm{L}(y), \mathrm{L}(x,y)] = 0$

(ii) $[\mathrm{L}(x), \mathrm{L}(yz)] + [\mathrm{L}(y), \mathrm{L}(zx)] + [\mathrm{L}(z), \mathrm{L}(xy)] = 0$

(iii) $\mathrm{L}(x^2y) - \mathrm{L}(x^2)\mathrm{L}(y) = 2\big(\mathrm{L}(xy) - \mathrm{L}(x)\mathrm{L}(y)\big)\mathrm{L}(x) = 0$

Proof. Differentiating (J_2') in the x direction:

$$D_x[\mathrm{L}(x), \mathrm{L}(y^2)] = 0,$$

we obtain (i).
Applying D_z to (i) one obtains (ii).
Applying both sides of (i) to an element z, the resulting identity can be rewritten as

$$\mathrm{L}(x)\mathrm{L}(y^2z) - \mathrm{L}(y^2)(xz) = 2\mathrm{L}(xy)(yz) - 2\mathrm{L}(y)\big((xy)z\big),$$

valid for all x, y and z.
Using commutativity, this can be written as

$$\mathrm{L}(y^2z)x - \mathrm{L}(y^2)\mathrm{L}(z)x = 2\mathrm{L}(yz)\mathrm{L}(y)x - 2\mathrm{L}(y)\mathrm{L}(z)\mathrm{L}(y)x.$$

This again can be written as

$$\mathrm{L}(y^2z) - \mathrm{L}(y^2)\mathrm{L}(z) = 2\big(\mathrm{L}(yz) - \mathrm{L}(y)\mathrm{L}(z)\big)\mathrm{L}(y),$$

which is exactly (iii). □

Relationship between Euclidean Jordan Algebras and Symmetric Cones

For every Euclidean Jordan algebra E the set of squared elements x^2 where $x \in \mathrm{E}$ is a closed cone. The interior of this cone denoted by E_+, is a symmetric cone and is called the cone of positive elements of E. Equivalently, E_+ is the set of all squared x^2 elements, that is, $\det(x) \neq 0$. Conversely, if Ω is a symmetric cone and $\varepsilon \in \Omega$, then one can construct a Euclidean Jordan algebra E such that $\mathrm{E}_+ = \Omega$ and ε is the identity element of E.

1.6.4 Classification of Irreducible Symmetric Cones

The irreducible symmetric cones are in one-to-one correspondence with simple Euclidean algebras, which classified into four families of classical Jordan algebras together with a single exceptional Jordan algebra.

The first three families of classical Jordan algebras are matrix spaces. More specifically, let $\mathbb{D} = \mathbb{R}, \mathbb{C}$ or the quaternions $\mathbb{H}$. Denote by $\bar{x}$ the conjugate of x in $\mathbb{D}$, $\mathrm{Re}\, x$ the real part of x, and $\mathrm{H}_m(\mathbb{D})$ the set of all $m \times m$ Hermitian matrices over $\mathbb{D}$.

Recalling that $\mathbf{X}^\star$, the adjoint of the matrix $\mathbf{X}$ is obtained by taking the conjugate of each of the entries and then transposing the matrix. Then, the matrix $\mathbf{X}$ is Hermitian if it is equal to $\mathbf{X}^\star$.

This space of Hermitian matrices equipped with the Jordan product

$$\mathbf{X} \circ \mathbf{Y} = \frac{1}{2}\big(\mathbf{XY} + \mathbf{YX}\big)$$

and the scalar product

$$\big(\mathbf{X}|\mathbf{Y}\big) = \mathrm{Re}\,\mathrm{tr}\big(\mathbf{XY}\big)$$

is a simple Euclidean Jordan algebra of rank r and corresponds to the irreducible symmetric cone of positive definite matrices over $\mathbb{R}, \mathbb{C}$ or $\mathbb{H}$, respectively.

The dimension of $\mathrm{H}_m(\mathbb{D})$ over $\mathbb{R}$ is $m + m(m-1)d/2$, where d is called the Pierce constant, is equal to the dimension of the space of $\mathbb{D}$ over $\mathbb{R}$. In particular, when $\mathbb{D}$ is the set of real numbers $\mathrm{H}_m(\mathbb{R})$ is indeed $\mathrm{S}_m(\mathbb{R})$.

The fourth class of the Jordan algebras is the Minkowski space $\mathbb{R} \times \mathbb{R}^{n-1}, n > 2$ which is the vector space equipped with the product $(\varsigma, x)(\xi, y) = \varsigma\xi + xy, \varsigma y + \xi x$. This exactly corresponds to the Lorentz cone

$$\mathbb{L}_n = \{(\varsigma, x) \in \mathbb{R} \times \mathbb{R}^{n-1} : \ \varsigma > \|x\|\}.$$

The exceptional Jordan algebra can be described as follows. First notice that $\mathbb{H} \times \mathbb{H}$ with the product

$$(x_1, y_1)(x_2, y_2) = (x_1 x_2 - \bar{y}_2 y_1, y_1 \bar{x}_2 + y_2 x_1)$$

is a non-associative algebra, called octonions, $\mathbb{O}$. Any element in $\mathbb{O}$ can be written as $x + jy$, where $j = (0, 1)$ and $x, y \in \mathbb{H}$. The conjugation in $\mathbb{O}$ can be defined by $\overline{x + jy} = \bar{x} - j\bar{y}$. The exceptional Jordan algebra is $\mathrm{H}_3(\mathbb{O})$.

E	Ω	n	r	d
$\mathbb{R}^1$	$(0,\infty)$	1	1	0
$\mathbb{R}^1 \times \mathbb{R}^{m-1}$	Λ_m	m	2	$m-2$
$\mathrm{Sym}(m,\mathbb{R})$	$\Pi_m(\mathbb{R})$	$m(m+1)/2$	m	1
$\mathrm{Herm}(m,\mathbb{C})$	$\Pi_m(\mathbb{C})$	m^2	m	2
$\mathrm{Herm}(m,\mathbb{H})$	$\Pi_m(\mathbb{H})$	$m(2m-1)$	m	4
$\mathrm{Herm}(3,\mathbb{O})$	$\Pi_3(\mathbb{O})$	27	3	8

Table 1.4: Classification of simple Euclidean Jordan algebras.

In Table 1.4 we introduce simple Euclidean Jordan algebras. This table is compiled by combining information from [155] and [302]. Simple Euclidean Jordan algebras have been classified by [256]. We explain the content of Table 1.4.

In Table 1.4, the symbol m runs over the set of all positive integers ≥ 3. In the first column, the algebra $\mathbb{R}^1 \times \mathbb{R}^{m-1}$ is called the *Lorentz algebra*. The product in this algebra has the form

$$(\lambda,u)\circ(\mu,v) = (\lambda\mu + (u|v), \lambda v + \mu u), \qquad \lambda,\mu \in \mathbb{R}^1, \quad u,v \in \mathbb{R}^{m-1}.$$

The corresponding cone, Λ_m, is called the *Lorentz cone*. It has the form

$$\Lambda_m = \{\,(\lambda,u) \in \mathbb{R}^1 \times \mathbb{R}^{m-1} \colon \lambda^2 - (u|u) > 0, \lambda > 0\,\}.$$

The symbol $\mathrm{Herm}(m,\mathbb{C})$ (resp. $\mathrm{Herm}(m,\mathbb{H})$, resp. $\mathrm{Herm}(3,\mathbb{O})$) denotes the Jordan algebra of all Hermitian matrices of size $m \times m$ with complex entries (resp. with entries in the skew field $\mathbb{H}$ of quaternions, resp. with entries in the algebra $\mathbb{O}$ of octonions). The scalar product in all of the above algebras has the form

$$(x|y) = \mathrm{Re}\,\mathrm{tr}(xy),$$

while the Jordan product is standard: $x \circ y = \frac{1}{2}(xy + yx)$. All the cones Ω are the sets of positive-definite matrices in the corresponding algebras. All algebras in Table 1.4 are pairwise non-isomorphic. In small dimensions, we have the following isomorphisms:

$$\begin{aligned}
\mathrm{Sym}(1,\mathbb{R}) &\sim \mathrm{Herm}(1,\mathbb{C}) \sim \mathrm{Herm}(1,\mathbb{H}) \sim \mathrm{Herm}(1,\mathbb{O}) \sim \mathbb{R}^1,\\
\mathrm{Sym}(2,\mathbb{R}) &\sim \mathbb{R}^1 \oplus \mathbb{R}^2,\\
\mathrm{Herm}(2,\mathbb{C}) &\sim \mathbb{R}^1 \oplus \mathbb{R}^3,\\
\mathrm{Herm}(2,\mathbb{H}) &\sim \mathbb{R}^1 \oplus \mathbb{R}^5,\\
\mathrm{Herm}(2,\mathbb{O}) &\sim \mathbb{R}^1 \oplus \mathbb{R}^9.
\end{aligned} \tag{1.160}$$

In what follows, the symbol n always denotes the dimension of the real linear space E. All simple Euclidean Jordan algebras, except the Lorentz cones, will be called *matrix algebras*.

To explain the fourth column of Table 1.4, denote by e the identity element of the algebra E. For any $x \in \mathrm{E}$, put

$$m(x) = \min\{k > 0 \colon (e, x, x^2, \ldots, x^k) \quad \text{are linearly dependent}\}. \tag{1.161}$$

The number $m(x)$ is bounded from above by n, the dimension of E.

Definition 1.6.7. *The* rank *of a Jordan algebra* E *is given by*

$$r = \max\{m(x) \colon x \in \mathrm{E}\}.$$

To explain the meaning of the last column of Table 1.4, we start from the following result, see [155].

Theorem 1.6.6. *Any simple Jordan algebra contains a* Jordan frame*, that is, the set* $\{c_1, \ldots, c_r\}$ *such that*

- *its elements are* orthogonal*:* $c_i \circ c_j = 0$ *if* $i \neq j$*;*
- *its elements are* idempotents*:* $c_i^2 = c_i$*;*
- *its elements constitute a* resolution of identity*:* $c_1 + \cdots + c_r = e$.

Denote by $\mathscr{L}(c_i)$ the linear operator in E acting by

$$\mathscr{L}(c_i)x = c_i \circ x, \qquad x \in \mathrm{E}.$$

By [155, Lemma IV.1.3], the linear operators $\mathscr{L}(c_i)$ and $\mathscr{L}(c_j)$ commute. Therefore, the admit a simultaneous diagonalization Let $\mathrm{E}_{ii} = \mathrm{E}(c_i, 1)$ be the one-dimensional eigenspace of the linear operator $\mathscr{L}(c_i)$ that corresponds to the eigenvalue 1. Let $\mathrm{E}(c_i, 1/2)$ be the eigenspace that corresponds to the eigenvalue $1/2$, and let

$$\mathrm{E}_{ij} = \mathrm{E}(c_i, 1/2) \cap \mathrm{E}(c_j, 1/2).$$

Theorem 1.6.7 ([155])**.** *The space* E *decomposes into the orthogonal direct sum*

$$\mathrm{E} = \bigoplus_{1 \leq i \leq j \leq r} \mathrm{E}_{ij}. \tag{1.162}$$

The subspaces E_{ij} *with* $i \neq j$ *have the same dimension.*

Denote the above dimension by d. It follows that

$$n = r + d\frac{r(r-1)}{2}. \tag{1.163}$$

The number d is given in the last column of Table 1.4.

How to define a Wishart distribution on an irreducible symmetric cone Ω? First, we define the *determinant* and the *trace* of an element x of the corresponding Euclidean Jordan algebra E.

1.6.5 Additional Properties

Throughout our discussion, Ω will be assumed to be an irreducible symmetric cone of positive elements of a simple Jordan algebra E of dimension n, rank r and the Pierce constant d defined by $n = r + r(r-1)d/2$.

(1) For each x in Ω and $g \in \mathrm{G}$ we have

(a) $\det \mathrm{L}(x) = \det(x)^{n/r}$,

(b) $\det \mathrm{P}(x) = (\det x)^{2n/r}$,

(c) $\det gx = \det(g)^{r/n} \det(x)$ for each $g \in \mathrm{G}$,

(d) $\det \mathrm{P}(y)x = (\det(x))^2 \det x$,

(e) $(gx)^{-1} = g^{\star -1} x^{-1}$, where $g^\star$ is the adjoint of g,

(f) $\mathrm{P}(x)^{-1} = \mathrm{P}(x^{-1})$,

(g) $\mathrm{P}(x)^\star = \mathrm{P}(x)$, that is, $\mathrm{P}(x)$ is Hermitian.

(b) K acts transitively on the set of primitive idempotents and the set of all Jordan frames.

1.6.6 Trace, Determinant and Minimal Polynomials

Let $\mathbb{R}[\mathrm{X}]$ be the algebra of polynomials in one variable with real coefficients. It is well-known that any ideal in $\mathbb{R}[\mathrm{X}]$ is generated by a unique monic polynomial. In particular, for any $x \in \mathrm{E}$, the ideal

$$\mathrm{J}(x) = \{\, p \in \mathbb{R}[\mathrm{X}] \colon\, p(x) = 0 \,\}$$

is generated by a polynomial called the *minimal polynomial* of x. Its degree, $m(x)$, is determined by Equation (1.161). An element x is called *regular* if $m(x) = r$. By [155, Proposition II.2.1], the set of regular elements is open and dense in E. There exist unique polynomials $a_1, a_2, \ldots, a_r$ such that the minimal polynomial of every regular element x is given by

$$f_x(\lambda) = \lambda^r - a_1(x)\lambda^{r-1} + a_2(x)\lambda^{r-2} + \cdots + (-1)^r a_r(x). \tag{1.164}$$

Moreover, the polynomial a_j is homogeneous of degree j.

Definition 1.6.8. *The* trace *of x is* $\mathrm{tr}(x) = a_1(x)$. *The* determinant *of x is* $\det(x) = a_r(x)$.

A Wishart distribution on an irreducible symmetric cone Ω is defined exactly as in Definition 1.6.12, that is, the Laplace transform of the Wishart random variable $\mathbf{Y}$ is defined on the set

$$\boldsymbol{\Sigma} - \Omega = \{\boldsymbol{\Sigma} - x \colon x \in \Omega\}$$

and is given by

$$\mathscr{L}_{\mathbf{Y}}(x) = \left(\det\left(e - \boldsymbol{\Sigma}^{-1}x\right)\right)^{-\lambda}. \tag{1.165}$$

The result by [197] takes the form

Theorem 1.6.8. *The right hand side of Equation* (1.165) *defines the Laplace transform of a random variable if and only if*

$$\lambda \in \Lambda = \left\{0, \frac{d}{2}, d, \ldots, \frac{(r-1)d}{2}\right\} \cup \left(\frac{(r-1)d}{2}, \infty\right). \tag{1.166}$$

1.6.7 Special Functions Defined on Symmetric Cones

The wide structure of the Symmetric cones has inspired many research studies in the fields of harmonic analysis and random fields. In this section we give a brief description of special functions defined on symmetric cones based on [155].

The Gamma Function of a Cone

Fixing a Jordan frame $c_1, c_2, \ldots, c_r$ in E. For each $1 \leq j \leq r$, then $e_j = c_1 + c_2 + \ldots + c_j$ is defined as the idempotent. In general, it can be shown that if c is an idempotent element of E, then the only possible eigenvalues of the linear transformation $\mathrm{L}(c)$ are $0, 1/2, 1$. One can easily see that the the eigenspace, $\mathrm{E}(e_1, 1)$, corresponding to eigenvalues 1 of $\mathrm{L}(e_j)$ is a Jordan algebra with the multiplication inherited from E.

Let Ω_j be the cone of positive elements and $\overset{(j)}{\det}$ the determinant with respect to this Jordan algebra. Let $\mathrm{P}_j : \mathrm{E} \to \mathrm{E}(e_1, 1)$ be the orthogonal projection on $\mathrm{E}(e_1, 1)$. The principal minor, $\Delta_j(x)$ is a homogeneous polynomial of degree j on E defined by $\Delta_j(x) = \overset{(j)}{\det}(P_j(x))$. We extend this definition as follows. For each $\mathrm{s} = (s_1, s_2, \ldots, s_r) \in \mathbb{C}^r$ we get

$$\Delta_s(x) = \Delta_1^{s_1 - s_2}(x)\Delta_2^{s_2 - s_3}(x) \ldots \Delta_r(x)^{s_r}.$$

For each $\mathrm{s} \in \mathbb{C}^r$ the Gamma function is given by

$$\Gamma_\Omega(\mathrm{s}) = \int_\Omega \exp\{-\mathrm{tr}(x)\}\Delta_\mathrm{s}(x)\det(x)^{-\frac{n}{r}}dx. \tag{1.167}$$

This integral is absolutely convergent if $\mathrm{Re}\,\mathrm{s}_j > (j-1)d/2$, for $j = 1, \ldots, r$. Moreover

$$\Gamma_\Omega(z) = (2\pi)^{\frac{n-r}{2}} \prod_{j=1}^{r} \Gamma(\mathrm{s}_j - (j-1)\frac{d}{2})$$

In particular, identifying $z \in \mathbb{C}$ with $(z, \ldots, z) \in \mathbb{C}^r$, we have

$$\Gamma_\Omega(z) = \int_{\mathrm{E}_+} \exp\{-\mathrm{tr}(x)\}\det(x)^{z-\frac{n}{r}}dx. \tag{1.168}$$

Thus we have the following definition for the Gamma function on a symmetric cone:

Definition 1.6.9. *The gamma function determined by the cone Ω is*

$$\Gamma_\Omega(s) = \int_\Omega \exp(-\operatorname{tr}(x))(\det(x))^{s-n/r}\,\mathrm{d}x, \qquad \operatorname{Re} s > n/r - 1.$$

When $\Omega = (0,\infty)$, we recover Equation (1.167), and when $\Omega = \Pi_m(\mathbb{R})$, we recover (1.168). By [155, Corollary VII.1.3, part (i)], we have

$$\Gamma_\Omega(s) = (2\pi)^{(n-r)/2} \prod_{i=0}^{r-1} \Gamma(s - id/2). \tag{1.169}$$

Note that when $\Omega = \Pi_m(\mathbb{R})$, we have

$$\Gamma_\Omega(s) = 2^{(n-r)/2}\Gamma_m(s) \tag{1.170}$$

because of different parametrisations.

When $\lambda \in ((r-1)d/2, \infty)$, the Wishart distribution is supported by Ω and has probability density

$$f_{\mathbf{Y}}(x) = \frac{(\det(\mathbf{\Sigma}))^\lambda}{\mathbf{\Sigma}_{\mathbf{\Sigma}}(\lambda)} \exp(-\operatorname{tr}(\mathbf{\Sigma} \circ x))(\det(x))^{\lambda - n/r} \mathbb{1}_{\mathbf{\Sigma}}(x). \tag{1.171}$$

For the case of $\Omega = \Pi_m(\mathbb{C})$, the Wishart distribution was studied by [205] and [267], for the case of $\Omega = \Pi_m(\mathbb{H})$ by [11], for the case of $\Omega = \Pi_3(\mathbb{O})$ by [166].

Beta Functions

Definition 1.6.10. *The beta function on a symmetric cone Ω is defined by the integral*

$$\mathrm{B}_\omega(a,b) = \int_{\Omega \bigcap (x-\Omega)} \Delta_{a-\frac{n}{r}}(x)\Delta_{b-\frac{n}{r}}(e-x)dx,$$

where $a, b \in \mathbb{C}^r$ and $e - \Omega = \{e - x : x \in \Omega\}$.

This integral converges absolutely if $\operatorname{Re} a_j > (j-1)d/2$ and $\operatorname{Re} b_j > (j-1)d/2$. In this case,

$$\mathrm{B}_\Omega(a,b) = \frac{\Gamma_\Omega(a)\Gamma_\Omega(b)}{\Gamma_\Omega(a+b)}$$

and

$$\int_{\Omega \bigcap (x-\Omega)} \Delta_{a-\frac{n}{r}}(y)\Delta_{b-\frac{n}{r}}(x-y)dy = \mathrm{B}_\Omega(a,b)\Delta_{a+b-\frac{n}{r}}(x).$$

The Space of Polynomials

Definition 1.6.11. *A function* $f:\ \mathrm{E}\to\mathbb{R}$ *is a polynomial on* E *if there is a basis* $\{v_1,v_2,\ldots,v_n\}$ *of* E *and a polynomial* $p\in\mathbb{R}[t_1,\ldots,t_n]$ *such that for any linear combination* $x=\sum_{i=1}^{n}\xi_i v_i$ *then*

$$f(x)=p(\xi_1,\ldots,\xi_n).$$

One can check that this definition is independent of the choice of basis. The set of all polynomials over E is denoted by $\mathscr{P}(\mathrm{E})$. The group action of G on E can be naturally extended to an action on $\mathscr{P}(\mathrm{E})$ by defining $gp(x)=p(g^{-1}x)$. Let $\mathscr{P}_{\boldsymbol{\lambda}}(\mathrm{E})$ be the subspace of $\mathscr{P}(\mathrm{E})$ generated by the polynomials $g\Delta_{\boldsymbol{\lambda}}, g\in\mathrm{G}$. Then, every p in $\mathscr{P}_{\boldsymbol{\lambda}}(\mathrm{E})$ is a homogeneous polynomial of degree $|\boldsymbol{\lambda}|$.

Spherical Polynomials

Recalling that K, the stabilizer of the identity e, is a compact Lie subgroup of G, thus there exists a Haar measure on K. Thus for each partition $\boldsymbol{\lambda}$, the spherical polynomial $\Phi_{\boldsymbol{\lambda}}$ is given by

$$\Phi_{\boldsymbol{\lambda}}(x)=\int_{\mathrm{K}}\Delta_{\boldsymbol{\lambda}}(kx)d\mu_{\mathrm{K}}(k),$$

where μ_{K} is the normalized Haar measure on K. The function $\Phi_{\boldsymbol{\lambda}}$ is indeed a homogeneous polynomial of degree $|\boldsymbol{\lambda}|$ and is invariant under the action of K, that is, $\Phi_{\boldsymbol{\lambda}}(kx)=\Phi_{\boldsymbol{\lambda}}(x)$ for any $k\in\mathrm{K}$ and xE. Moreover, the spherical polynomial $\Phi_{\boldsymbol{\lambda}}$ is up to a constant factor, the only K$-$invariant polynomial in $\Phi_{\boldsymbol{\lambda}}(\mathrm{E})$. More precisely, if p is a K$-$invariant homogeneous polynomial in $\Phi_{\boldsymbol{\lambda}}(\mathrm{E})$, then

$$\int_{\mathrm{K}}p(kx)d\mu_{\mathrm{K}}(k)=p(e)\Phi_{\boldsymbol{\lambda}}(x). \tag{1.172}$$

Consequently, for any $g\in\mathrm{G}$

$$\int_{\mathrm{K}}\Phi_{\boldsymbol{\lambda}}(x)(gkx)d\mu_{\mathrm{K}}(k)=\Phi_{\boldsymbol{\lambda}}(ge)\Phi_{\boldsymbol{\lambda}}(x). \tag{1.173}$$

If $x\in\Omega$ and $\mathrm{Re}\,\gamma>(r-1)d/2$, then for any y in Ω and $g\in\Omega$ we have

$$\int_{\Omega}e^{-\mathrm{tr}(xy)}\Phi_{\boldsymbol{\lambda}}(gx)\det(x)^{\gamma-\frac{n}{r}}dx=\Gamma_{\Omega}(\lambda+\gamma)\Delta^{-\gamma}(y)\Phi_{\boldsymbol{\lambda}}(gx^{-1}). \tag{1.174}$$

Zonal Polynomials

The Pochhammer symbol for Ω is defined by

$$(\mathrm{s})_{\lambda}=\frac{\Gamma_{\Omega}(\mathrm{s}+\lambda)}{\Gamma_{\Omega}(\mathrm{s})},\qquad \mathrm{s}\in\mathbb{C}.$$

This definition generalizes the classical Pochhammer symbol. The Zonal polynomial Z_λ is a homogeneous K−invariant polynomial of E defined by

$$\mathrm{Z}_\lambda(x) = d_\lambda \frac{|\lambda|!}{\left(\frac{n}{r}\right)_\lambda} \Phi_\lambda(x), \tag{1.175}$$

where d_λ is the dimension of the vector space $\mathscr{P}_\lambda(\mathrm{E})$, and $\mathscr{P}_\lambda$ is the spherical polynomial. Zonal polynomials are K−invariant polynomials normalized by the property

$$\mathrm{tr}(x)^k = \sum_{|\lambda|=k} \mathrm{Z}_\lambda(x), \;\; x \in \Omega. \tag{1.176}$$

Notice that the function $p(x) = \mathrm{tr}(x)^k$ on Ω is a K−invariant homogeneous polynomial in $\mathscr{P}_\lambda(\mathrm{E})$. It follows from (1.173) that for each $x \in \mathrm{E}$ and $y \in \omega$

$$\int_{\mathrm{K}} \mathrm{Z}_\lambda(\mathrm{P}(y^{\frac{1}{2}})kx) d\mu_{\mathrm{K}}(x) = \frac{\mathrm{Z}_\lambda(y)\mathrm{Z}_\lambda(x)}{\mathrm{Z}_\lambda(e)}. \tag{1.177}$$

Hypergeometric Polynomials

Let $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ be real numbers with $a_i - (i-1)d/2 \geq 0, b_j - (j-1)d/2 \geq 0$ and $x, y \in \mathrm{E}$. The hypergeometric function ${}_p\mathrm{F}_q$ is defined by

$${}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, x, y) = \sum_{k=1}^{\infty} \sum_{|\lambda|=k} \frac{(a_1)_\lambda \cdots (a_p)_p}{(b_1)_\lambda \cdots (b_p)_p} \frac{\mathrm{Z}_\lambda(x)}{k!} \frac{\mathrm{Z}_\lambda(y)}{\mathrm{Z}_\lambda(e)}.$$

For $y = e$ we obtain ${}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, x) = {}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, x, e)$. This implies that

$${}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, x) = \sum_{k=1}^{\infty} \sum_{|\lambda|=k} \frac{(a_1)_\lambda \cdots (a_p)_p}{(b_1)_\lambda \cdots (b_p)_p} \frac{\mathrm{Z}_\lambda(x)}{k!}.$$

This series converges absolutely if $p \leq q$ and diverges if $p > q$. Furthermore, ${}_0\mathrm{F}_0(x) = e^{\mathrm{tr}(x)}$ and ${}_0\mathrm{F}_1(b, x) = \det(e - x)^{-b}$.

Suppose $x \in \Omega$ and $g \in \mathrm{G}$. By using the integral Equation (1.177) for $p \leq q$ we obtain

$$\int_{\mathrm{K}} {}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, g(kx)) d\mu_{\mathrm{K}}(x) = {}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, x, ge). \tag{1.178}$$

Similarly, if $\mathrm{Re}\,\gamma > (r-1)d/2, y \in \Omega$ and $p < q$, or $y \in e - \Omega$ and $p = q$, then by applying Equation (1.174) we obtain

$$\begin{aligned} &\int_\Omega e^{-\mathrm{tr}(xy)} {}_p\mathrm{F}_q(a_1, \ldots, a_p, b_1, \ldots, b_q, gx, z) \det(x)^{\gamma - \frac{n}{r}} dx \\ = \;& \Gamma_\Omega(\gamma) \det(\gamma)^{-\gamma} {}_{p+1}\mathrm{F}_q(a_1, \ldots, a_p, \gamma, b_1, \ldots, b_q, gy^{-1}, z). \end{aligned} \tag{1.179}$$

We conclude this section with the following important proposition which can be used to derive Beta and Wishart type distributions [155].

Proposition 1.6.2. *Suppose that* $p \leq q+1, \operatorname{Re}\eta_1 > (r-1)d/2$ *and* $\operatorname{Re}\eta_2 > (r-q)d/2$. *If* $g \in G$ *such that* $ge \in \Omega \bigcap (e-\Omega)$, *then*

$$\begin{aligned} & \int_{\Omega\bigcap(e-\Omega)} {}_p\mathrm{F}_q(a_1,\ldots,a_p,b_1,\ldots,b_q,gx)\det(x)^{\eta_1-\frac{n}{r}}\det(e-x)^{\eta_2-\frac{n}{r}}dx \\ = \; & \mathrm{B}_\Omega(\eta_1,\eta_2)_{p+1\,p+1}\mathrm{F}_q(a_1,\ldots,a_p,\eta_1,b_1,\ldots,b_q,\eta_1+\eta_2,ge). \end{aligned} \tag{1.180}$$

1.6.8 Gaussian, Chi-Square and Wishart Distributions on Symmetric Cones

Consider the following classical statistical problem. Let X be a normal random variable with mean μ and variance σ^2. Let $x_1, \ldots, x_N$ be a sample from the normal population distributed like X. It is well known that the maximum likelihood estimates of the parameters μ and σ^2 are

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{\mu})^2.$$

Moreover, the random variable $\mathrm{Y} = \dfrac{N}{\sigma^2}\hat{\sigma}^2$ has the *chi-square distribution with N degrees of freedom*. The probability density of the chi-square distribution with N degrees of freedom was derived by William S. Gosset, a brewer of Guinness beer, in [446]. Gosset published his research under the pen name "Student". According to [510],

> At Guinness the scientific brewers, including Gosset, were allowed by the company to publish research so long as they did not mention (1) beer, (2) Guinness, or (3) their own surname.

The above probability density has the form

$$f_{\mathrm{Y}}(x) = \frac{1}{2^{N/2}\Gamma(N/2)}\exp(-x/2)x^{N/2-1}\mathbb{1}_{(0,\infty)}(x),$$

where Γ is the gamma-function:

$$\Gamma(s) = \int_0^\infty \exp(-x)x^{s-1}\,\mathrm{d}x. \tag{1.181}$$

It follows from (1.181) that the function

$$f_{\mathrm{Y}}(x) = \frac{x^{\lambda-1}\exp(-x/\sigma)}{\sigma^\lambda\Gamma(\lambda)}\mathbb{1}_{(0,\infty)}(x)$$

is a probability density as long as $\lambda > 0$ and $\sigma > 0$. The corresponding probability distribution is the *gamma distribution* with *shape parameter* λ and *scale parameter* σ. For particular values of $\lambda = N/2$ and $\sigma = 2$, we return back to the chi-square distribution with N degrees of freedom.

Laplace Transform and The Wishart Density

Recalling the fact that the *Laplace transform* of a random variable Y is defined as

$$\mathscr{L}_{\mathrm{Y}}(s) = \mathbb{E}[\exp(s\mathrm{Y})]$$

for all complex numbers s such that the expectation exists.

Remark 1.6.9. *In many sources, the above definition contains an opposite sign:*

$$\mathscr{L}_{\mathrm{Y}}(s) = \mathbb{E}[\exp(-s\mathrm{Y})].$$

We choose the convention by [74] and subsequent papers.

In particular, for the gamma distribution we have

$$\mathscr{L}_{\lambda,\sigma}(s) = (1-\sigma s)^{-\lambda}, \qquad \mathrm{Re}\, s < \sigma^{-1}.$$

Observe that

$$\lim_{\lambda \downarrow 0} \mathscr{L}_{\lambda,\sigma}(s) = 1.$$

The right hand side is the Laplace transform of the random variable $\mathrm{Y} = 0$. We say that this random variable has gamma distribution with shape parameter $\lambda = 0$ and an arbitrary scale parameter $\sigma > 0$.

It is easy to generalise the above discussion to the case of *random vectors*. Specifically, let $\mathbf{X}$ be a m-dimensional normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a sample from a normal population distributed like $\mathbf{X}$. The maximum likelihood estimates of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{\top}.$$

Moreover, the random variable $\mathbf{Y} = N\hat{\boldsymbol{\Sigma}}$ has the *classical Wishart distribution with N degrees of freedom and covariance matrix* $\boldsymbol{\Sigma}$, see [357]. Denote this distribution by $\mathscr{W}_m^c(N, \boldsymbol{\Sigma})$. If $N \geq m$, then the probability density of the above distribution has the form

$$f_{\mathbf{Y}}(x) = \frac{1}{2^{Nm/2}\Gamma_m(N/2)(\det\boldsymbol{\Sigma})^{N/2}} \exp\bigl(-\mathrm{tr}\bigl(\boldsymbol{\Sigma}^{-1}x\bigr)/2\bigr)(\det(x))^{(N-m-1)/2}\mathbb{1}_{\Omega}(x), \tag{1.182}$$

where Ω is the set of all symmetric positive-definite $m \times m$ matrices, Γ_m is the *multivariate gamma function* defined for all $s \in \mathbb{C}$ with $\mathrm{Re}\, s > (m-1)/2$ as

$$\Gamma_m(s) = \int_{\Omega} \exp(-\mathrm{tr}(x))(\det(x))^{s-(m+1)/2}\,\mathrm{d}x, \tag{1.183}$$

see [357, Definition 2.1.10], and where tr denotes the trace of a matrix.

For the case of $s = 2$, this density was derived in [164]. The case of an arbitrary m was considered in [496]. The integral in the right hand side of Equation (1.183) was calculated in [246].

Theorem 1.6.10 ([246]). *We have*

$$\int_{\Omega} \exp(-\operatorname{tr} x)(\det x)^{s-(m+1)/2}\,\mathrm{d}x = \pi^{m(m-1)/4}\prod_{i=0}^{m-1}\Gamma(s-i/2).$$

Later, this integral appeared in [427] and became known in the number-theoretical community as the *Siegel integral*.

Let $\mathrm{E} = \mathrm{Sym}(m,\mathbb{R})$ be the linear space of all symmetric $m \times m$ matrices with real entries. Introduce the scalar product on E by $(x|y) = \operatorname{tr}(xy)$. The Laplace transform of a E-valued random matrix $\mathbf{Y}$ is defined by

$$\mathscr{L}_{\mathbf{Y}}(x) = \mathbb{E}\big[\exp\big((x|Y)\big)\big]$$

for all $x \in \mathrm{E}$ for which the expectation exists. In the case of the classical Wishart distribution we obtain

$$\mathscr{L}_{\mathbf{Y}}(x) = \big(\det\big(\mathbf{I} - 2\boldsymbol{\Sigma}x\big)\big)^{-N/2}, \tag{1.184}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix.

It is convenient to change slightly the parametrisation of the classical Wishart distribution. Denote $\mathscr{W}_m(N,\boldsymbol{\Sigma}) = \mathscr{W}_m^c(2N,\boldsymbol{\Sigma}^{-1}/2)$. The probability density (1.182) becomes

$$f_{\mathbf{Y}}(x) = \frac{\big(\det(\boldsymbol{\Sigma})\big)^N}{\Gamma_{\Omega}(N)}\exp(-\operatorname{tr}(\boldsymbol{\Sigma}x))(\det x)^{N-(m+1)/2}\mathbb{1}_{\Omega}(x),$$

while the Laplace transform (1.184) becomes

$$\mathscr{L}_{\mathbf{Y}}(x) = \big(\det\big(\mathbf{I} - \boldsymbol{\Sigma}^{-1}x\big)\big)^{-N}.$$

In particular, when $m = 1$, we obtain an alternative parametrisation of the chi-square distribution:

$$f_{\mathbf{Y}}(x) = \frac{1}{\Gamma(N)}\exp(-x)x^{N-1}\mathbb{1}_{(0,\infty)}(x).$$

Definition 1.6.12. *A $\overline{\Omega}$-valued random matrix $\mathbf{Y}$ has* Wishart distribution with shape parameter λ and scale parameter Σ *if and only if*

$$\mathscr{L}_{\mathbf{Y}}(y) = \big(\det\big(I - \boldsymbol{\Sigma}^{-1}y\big)\big)^{-\lambda}. \tag{1.185}$$

We introduce a family of Wishart distributions as a particular case of the following general construction. Let μ be a measure defined on the Borel σ-field of the Euclidean finite-dimensional space E. Let

$$\mathscr{L}_{\mu}(y) = \int_{\mathrm{E}} \exp((x|y))\,\mathrm{d}\mu(y)$$

be the Laplace transform of μ, and assume that the interior $\mathrm{Y}(\mu)$ of the set of all $y \in \mathrm{E}$ for which $\mathscr{L}_{\mu}(y) < \infty$ is not empty.

Definition 1.6.13. *The set* $\mathrm{F}(\mu) = \{ \mathsf{P}_{y,\mu} : y \in \mathrm{Y}(\mu) \}$ *of probability measures on E defined by*

$$\mathrm{d}\mathsf{P}_{y,\mu}(x) = \frac{1}{\mathscr{L}_\mu(y)} \exp((x|y))\, \mathrm{d}\mu(x)$$

is called the natural exponential family *generated by* μ.

Observe that

$$\begin{aligned} \mathscr{L}_{\mathsf{P}_{y,\mu}}(z) &= \frac{1}{\mathscr{L}_\mu(y)} \int_{\mathrm{E}} \exp((x|z)) \exp((x|y))\, \mathrm{d}\mu(x) \\ &= \frac{1}{\mathscr{L}_\mu(y)} \int_{\mathrm{E}} \exp((x|z+y))\, \mathrm{d}\mu(x) \\ &= \frac{\mathscr{L}_\mu(z+y)}{\mathscr{L}_\mu(y)}. \end{aligned}$$

The standard reference for natural exponential families is [22, 23].

Example 1.6.11. *Define the measure* μ_λ *by*

$$\mathrm{d}\mu_\lambda(x) = \frac{1}{\Gamma_m(\lambda)} (\det x)^{\lambda-(m+1)/2} \mathbb{1}_\Omega(x)\, \mathrm{d}x, \qquad \lambda > \frac{m-1}{2}.$$

The Laplace transform of this measure is

$$\mathscr{L}_{\mu_\lambda}(y) = \big(\det(-y)\big)^{-\lambda}, \qquad y \in -\Omega = \{ -x : x \in \Omega \}.$$

The corresponding natural exponential family is

$$\begin{aligned} \mathrm{d}\mathsf{P}_{\mathbf{\Sigma},\mu_\lambda}(x) &= \frac{1}{\mathscr{L}_{\mu_\lambda}(\mathbf{\Sigma})} \exp((x|\mathbf{\Sigma}))\, \mathrm{d}\mu_\lambda(x) \\ &= \frac{\big(\det(-\mathbf{\Sigma})\big)^\lambda}{\Gamma_m(\lambda)} \exp((x|\mathbf{\Sigma}))(\det x)^{\lambda-(m+1)/2} \mathbb{1}_\Omega(x)\, \mathrm{d}x \end{aligned}$$

for $\mathbf{\Sigma} \in -\Omega$. *We would like to run* $\mathbf{\Sigma}$ *over* Ω. *For that, replace* $\mathbf{\Sigma}$ *with* $-\mathbf{\Sigma}$. *We obtain the distribution*

$$\mathrm{d}\mathsf{P}_{\mathbf{\Sigma},\mu_\lambda}(x) = \frac{\big(\det(\mathbf{\Sigma})\big)^\lambda}{\Gamma_m(\lambda)} \exp(-\operatorname{tr}(\mathbf{\Sigma}x))(\det x)^{\lambda-(m+1)/2} \mathbb{1}_\Omega(x)\, \mathrm{d}x. \tag{1.186}$$

This distribution is called the Wishart distribution with shape parameter $\lambda > (m-1)/2$ and scale parameter $\mathbf{\Sigma} \in \Omega$.

In particular, the Wishart distribution with shape parameter $\lambda = N/2$ *and scale parameter* $2\mathbf{\Sigma}^{-1}$ *is the classical one. When* $m = 1$, *we obtain an alternative parametrisation of the exponential distribution:*

$$\mathrm{d}\mathsf{P}_{\sigma,\mu_\lambda}(x) = \frac{\sigma^\lambda}{\Gamma(\lambda)} \exp(-\sigma x) x^{\lambda-1} \mathbb{1}_{(0,\infty)}(x)\, \mathrm{d}x.$$

The Laplace transform of the Wishart distribution is

$$\mathscr{L}_{\mathrm{P}_{\Sigma,\mu_\lambda}}(y)=\frac{\mathscr{L}_{\mu_\lambda}(y-\mathbf{\Sigma})}{L_{\mu_\lambda}(-\mathbf{\Sigma})}=\frac{\big(\det(\mathbf{\Sigma})\big)^{\lambda}}{\big(\det(\mathbf{\Sigma}-y)\big)^{\lambda}}=\big(\det\big(I-\mathbf{\Sigma}^{-1}y\big)\big)^{-\lambda}$$

Does such a distribution exist for the remaining values of λ? The answer to this question and related ideas are discussed in detailed and illustrated in Chapter 9.

Detailed discussion and construction of the degenerate and non-degenerate Wishart distributions in the cones including $\mathbb{R}^1\times\mathbb{R}^{m-1}$, $\mathrm{Sym}(m,\mathbb{R})$, $\mathrm{Herm}(m,\mathbb{C})$, $\mathrm{Herm}(m,\mathbb{H})$ and $\mathrm{Herm}(3,\mathbb{O})$ follows in Chapter 8 and Chapter 9.

1.7 Vandermonde Matrix and Determinant in Financial Mathematics

In this section we briefly discuss some mathematical concepts commonly in pricing theory and portfolio construction. These and other related ideals will be further discussed in Chapter 7.

According to [125], one of the most influential ideas of modern finance is the Efficient Market Hypothesis (EMH), the notion that prices in financial markets fully reflect all available information and that there are no trading strategies that produce positive, expected, risk-adjusted excess returns. This, still arguable, situation follows from the view that, in an intensely competitive financial market, the response of investors to new information is rapid and rational, bidding prices up or down until they eliminate any advantage to trading on the new information, for more detailed discussion see, [169, 201, 245, 261, 262, 263, 359, 363].

The above market environment can best be described as an Arbitrage-free market which is a market that is characterised by the No Free Lunch with Vanishing Risk, NFLVR. The NFLVR condition achieved by utilizing a sequence of time self-financing portfolios which converge to an arbitrage strategy, whereby an approximate self-financing portfolio, more detailed discussion can be got from [37, 43, 99, 100, 112, 113, 424, 425, 439, 442, 495, 508].

Under assumptions of no arbitrage, there exists a unique probability, called an equivalent martingale measure such that the price of a claim is the expectation of its discounted pay-off as discussed in [287, 292, 300, 399].

The condition of NFLVR is considered as the basis for the fundamental theorem of asset pricing which provides the necessary and sufficient conditions for a market to be arbitrage free and for the market to be complete [240, 242]. An arbitrage opportunity in simple terms is a means of earning or making money without an initial investment and without the possibility of making a loss. In economics and finance [239], a complete market is a market is a market with two conditions:

(1) Negligible transaction costs and therefore also perfect information,

(2) there is a price for every asset in every possible state of the world.

Completeness is a common property of market models as discussed in [161, 329, 330, 331, 422, 423] in relation to modelling derivatives.

A derivative security (or simply a derivative) is a security whose value depends on the values of other more basic securities or assets called underlying assets or variables. The underlying variable or entity can be an asset, stock of goods, a bond, a currency exchange, index or interest rate, or the quotation of commodities such as god, oil, or wheat. Derives can be used for a number of purposes, including

(i) insuring against price movement or the principle of hedging;

(ii) increasing exposure to price movements for speculation, anticipation or getting getting access to otherwise hard-to-trade assets or markets

Some of the most common derivatives include forwards, futures, options, swaps and variations of these such as synthetic collateralized debt obligation and credit default swaps.

In recent years, derivative securities have become more important than ever in financial markets [271]. Futures, options and swaps and many other exotic options are traded outside of exchanges called over-the-counter (OTC) markets, by financial institutions and their corporate clients. Derivatives are one of the three main categories of financial instruments, the other categories being stocks, that is, equities or shares, and debts, that is, bonds and mortgages.

1.7.1 Money Market Account

According to [271], considering a bank deposit with initial principle $F = 1$. The amount of the deposit after time t periods is denoted by B_t. The interest paid for a period t is equal to $B_{t+1} - B_t$. If the interest paid is proportion to the amount B_t, it is called compound interest [271]. That is, the compound interest is such that the amounts of deposit satisfy the relation

$$B_{t+1} - B_t = rB_t, \;\; t = 0,1,2,3,4,\ldots, \tag{1.187}$$

where the multiplier $r > 0$ is called the interest rate. It follows from (1.187) that

$$B_{t+1} = (1+r)B_t, \tag{1.188}$$

Thus using the recursion (1.188) for $t = 0,1,2,3,4,\ldots$ we have,

$$\begin{aligned}
B_1 &= B_0(1+r) \\
B_2 &= B_1(1+r) = B_0(1+r)^2 \\
B_3 &= B_2(1+r) = B_0(1+r)^3 \\
B_4 &= B_3(1+r) = B_0(1+r)^4 \\
\cdots &\quad \cdots \\
B_t &= B_0(1+r)^t = (1+r)^t, \;\; \text{for } B_0 = 1.
\end{aligned} \tag{1.189}$$

The deposit B_t is often called the money-market account.

Suppose now that the annual interest rate is r and the interest is paid n times each year. We divide one year into n equally spaced sub-periods, so that the interest rate for each period is given by r/n. Following the same argument as above, it is readily seen that the amount of deposit after m periods is given by

$$B_m = \left(1+\frac{r}{n}\right)^m, \ \ m = 0,1,2,3,4,\ldots. \tag{1.190}$$

For example, when the interest is semi-annual compounding, the amount deposited after two years is given by

$$B_4 = \left(1+\frac{r}{2}\right)^4.$$

Suppose, $t = m/n$ or $m = nt$ for some integers m and n, and let $B(t)$ denote the amount of deposit at time t for the compounding interest with annual interest r. From (1.190) we have

$$B(t) = \left(1+\frac{r}{n}\right)^{nt}. \tag{1.191}$$

Now, what if we let n tend to infinity? That is, what if the interest is continuous compounding? This can easily be established because, form the common limit theorem

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n = e^1 \equiv 2.718281828459\cdots$$

we have

$$B(t) = \left[\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n\right]^t = e^{rt}, \ \ t \geq 0. \tag{1.192}$$

Consider next the case that the interest rates vary in time. For simplicity, we first assume that the interest rates are a step function of time. That is, the rate at time t is given by

$$r(t) = r_i \ \ \text{if} \ \ t_{i-1} \leq t \leq t_i, \ \ i = 1,2,\ldots,$$

where $t_0 = 0$. Then, form (1.192), we have that

$$B(t_1) = e^{r_1 t_1} \tag{1.193}$$

$$\frac{B(t_2)}{B(t_1)} = e^{r_2(t_2-t_1)}, \ \ \text{and so on.} \tag{1.194}$$

Hence, for a time t such that $t_{n-1} \leq t \leq t_n$ we obtain

$$B(t) = \exp\left\{\sum_{k=1}^{n-1} r_k\delta_k + r_n(t-t_{n-1})\right\}, \ \ \delta_k \equiv t_k - t_{k-1}. \tag{1.195}$$

Recalling that the integral of any (Riemann) integrable function $r(t)$ is the limit of the sum, that is,

$$\int_0^t r(u)du = \lim_{n\to\infty}\left[\sum_{k=1}^{n-1} r_k\delta_k + r_n(t-t_{n-1})\right], \tag{1.196}$$

where the limit is taken over all possible sequences of partitions of the interval $[0,t]$. The next theorem summarizes the concept of interest rate in continuous time.

Theorem 1.7.1. *Suppose that the instantaneous interest rate at time t is $r(t)$. If the interest is continuous compounding, then, the time t money-market account is given by*

$$B(t) = \exp\left\{\int_0^t r(u)du\right\}, \ t \geq 0. \tag{1.197}$$

Remark: We notice that, even though the interest rates $r(t)$ are random, the money market is given by (1.197); in which case, $B(t)$ is also a random variable.

1.7.2 Derivatives and Arbitrage Pricing

In mathematical finance, a risk-neutral measure, also called an equilibrium measure, or equivalent martingale measure, is a probability measure such that each share price is exactly equal to the discounted expectation of the share price under this measure [125, 126, 128, 146, 149, 161]. This is heavily used in the pricing of financial derivatives due to the fundamental theorem of asset pricing, which implies that in a complete market a derivative's price is the discounted expected value of the future payoff under the unique risk-neutral measure. Such a measure exists if and only if the market is arbitrage-free.

In discrete, that is, finite state market, the following hold [271]

1. The First Fundamental Theorem of Asset Pricing: A discrete market, on a discrete probability space $(\Omega, \mathscr{F}, \mathbb{P})$, is arbitrage-free if and only if, there exists at least one neutral probability measure $\mathbb{Q}$ that is equivalent to the original probability measure, $\mathbb{P}$.

2. The Second Fundamental Theorem of Asset Pricing: An arbitrage-free market (S,B) consisting of a collection of stocks S and risk free bond B is complete if and only if there exists a unique risk-neutral measure that is equivalent to $\mathbb{P}$ and has a numeraire B.
   If $S = (S_t)_{t=0}^T$ is a semi-martingale with values in $\mathbb{R}^d$ then S does not allow for a free lunch with vanishing risk if and only if there exists an equivalent martingale measure $\mathbb{Q}$ such that S is a sigma-martingale under $\mathbb{P}$.

Definition 1.7.1. *The term "portfolio" refers to any combination of financial assets such as stocks, bonds and cash. Portfolios may be held by individual investors and/or managed by financial professionals, hedge funds, banks and other financial institutions. It is a generally accepted principle that a portfolio is designed according to the investor's risk tolerance, time frame and investment objectives. The monetary value of each asset may influence the risk/reward ratio of the portfolio.*

In mathematical finance, a replicating portfolio for a given asset or series of cash flows is a portfolio of assets with the same properties (especially cash flows). This is meant in two distinct senses: static replication, where the portfolio has the same cash flows as the reference asset (and no

changes need to be made to maintain this), and dynamic replication, where the portfolio does not have the same cash flows, but has the same "Greeks" as the reference asset, meaning that for small (properly, infinitesimal) changes to underlying market parameters, the price of the asset and the price of the portfolio change in the same way. Dynamic replication requires continual adjustment, as the asset and portfolio are only assumed to behave similarly at a single point (mathematically, their partial derivatives are equal at a single point).

In financial mathematics, a self-financing portfolio is a portfolio having the feature that, if there is no exogenous infusion or withdrawal of money, the purchase of a new asset must be financed by the sale of an old one [43].

Definition 1.7.2. *Let $h_i(t)$ denote the number of shares of stock number 'i' in the portfolio at time t, and $S_i(t)$ the price of stock number 'i' in a frictionless market with trading in continuous time. Let*

$$V(t) = \sum_{i=1}^{N} h_i(t) S_i(t). \tag{1.198}$$

Then the portfolio $(h_1(t), \ldots, h_n(t))$ is self-financing if

$$dV(t) = \sum_{i=1}^{n} h_i(t) dS_i(t). \tag{1.199}$$

1.7.3 Pricing Derivatives

Discount Bonds and Coupon-Bearing Bonds

A discount bond is a bond that is issued for less than its par–or face–value [271, 439]. A financial security that promises to pay a single cash-flow of magnitude F at a future time, called maturity, is called a discount bond, and the amount of F is called the face value. By taking F as a unit of money, we can assume without loss of generality that $F = 1$.

A bond more commonly traded in practice is a security that promises to pay a stream of certain payments, called coupons, at future times as well as the face value at maturity. Such a bond is called a coupon-bearing bond and its cash-flow, if it pays coupon C_i at time t_i as shown in the Figure 1.4. Since each cash-flow C_i is equivalent to a cash-flow at a discount bond with face value C_i, a coupon bearing bond can be thought of as a portfolio of discount bonds with face values C_i.

There are two types of bonds, default-free bonds and corporate bonds. A corporate bond issued by a firm promises to pay a stream of payments; but there is a possibility that the bond defaults before maturity and cannot meet the debt obligations. That is, corporate bonds are exposed to the credit risk. On the other hand, a default-free bond issued by the government such as U.S. involves such no risk.

Suppose that the prices of default-free bonds for all maturities are observed in (or can be imputed from) from the market. Since the prices of default-free discount bonds with the same maturity are the same, the discount bonds can be used as a benchmark of future values. Namely, let

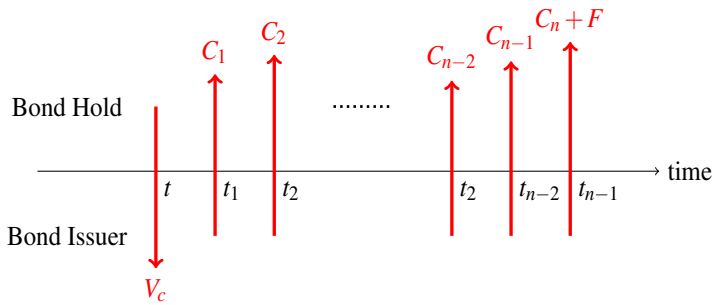


Figure 1.4: Illustration of the cash-flow of a bond coupon holder at different time period before maturity.

$v(t,T), t \le T$, be the time t price of the discount bond that pays 1 (B_0) dollar for sure at maturity T. Then the present value of the certain cash-flow B at future time T is given by $Bv(t,T)$, where t denotes the current time. Of course, $v(t,t) = 1$. The function $v(t,T)$ with respect to T is called the term structure of the default-free discount bonds. It is also called the discount function, since the present value of $1(B_0)$ dollar is paid at future time T is given by $v(t,T)$.

Yield to Maturity

Yield to maturity is the total return anticipated on a bond if the bond is held until maturity time [153, 439]. Suppose that, at time t, an investor purchases a security for $S(t)$ that pays $S(T)$ dollars for sure at maturity T. The rate of return per unit of time, $R(t,T)$ say, from this investment is defined by

$$R(t,T) = \frac{S(T) - S(t)}{(T-t)S(t)}, \quad t \le T. \tag{1.200}$$

It follows that

$$S(T) = S(t)[1 + (T-t)R(t,T)].$$

Next, suppose that the rate of return per unit time is computed in the sense of compounded interests. From (1.190), we then have

$$R(t,T) = S(T) = S(t)\left[1 + \frac{(T-t)R_n(t,T)}{n}\right]^n, \quad n = 1,2,\ldots, \tag{1.201}$$

where the subscript n in $R_n(t,T)$ means that the interests are compounded n times each year. Denoting the rate of return per unit of time in the sense of the continuous compounding by using $Y(t,T) = \lim_{n\to\infty} R_n(t,T)$, it follows from (1.192) that

$$S(T) = S(t)e^{(T-t)Y(t,T)}, \quad t \le T, \tag{1.202}$$

or, equivalently,

$$Y(t,T) = \frac{1}{t-T}\log\left(\frac{S(T)}{S(t)}\right), \ \ t \leq T. \tag{1.203}$$

The rate of return per unit of time in the continuous compounding is called the yield to maturity (or simply the yield). In particular, if the security is the default-free discount bond with maturity T, that is, $S(t) = v(t,T)$ and $S(T) = 1$, the we obtain

$$Y(t,T) = -\frac{\log v(t,T)}{t-T}, \ \ t \leq T. \tag{1.204}$$

In what follows, we use Equation (1.204) as the definition of yield of the discount bond.

1.7.4 Options

Options have have been extensively studied especially after the major breakthrough by Black and Scholes [161], whose work on a new method to determine the value of derivatives won the Nobel prize in 1997 being awarded to Professor Robert C. Merton, Harvard University, Cambridge, USA and Professor Myron S. Scholes, Stanford University, Stanford, USA [46]. More extensive work has been done in modelling options including [65, 99, 100, 113, 126, 196, 240, 241, 242, 380, 422]. These models have also been extended to methods involving Lie groups and Lie Symmetries as studied in [192, 244, 350].

An option is the simplest example of derivative instrument. An option is a contract that gives the right (but not the obligation) to its holder to buy or sell some amount of the underlying asset at a future date, for a specified price. Therefore in an option contract we need to specify:

(i) an underlying asset;

(ii) an exercise price K, the so-called strike price;

(iii) a date T, the so called maturity.

A Call option gives the right to buy, whilst a Put option gives the right to sell. An option is called European if the right to buy or sell can be exercised only at maturity, and it is called American if it can be exercised at any time before maturity.

Let us consider a European call option with strike price K, maturity T and let us denote the underlying asset at maturity by S_T. At the time T we have two possibilities as may be depicted in the Figure 1.5 below.

(i) If $S_T > K$, the pay-off of the options is equal to $S_T - K$, corresponding to the profit obtained by exercising the option (that is, by buying the underlying asset at price K and then selling it at the market price S_T).

(ii) If $S_T < K$, exercising the option is not profitable and the pay-off is zero.

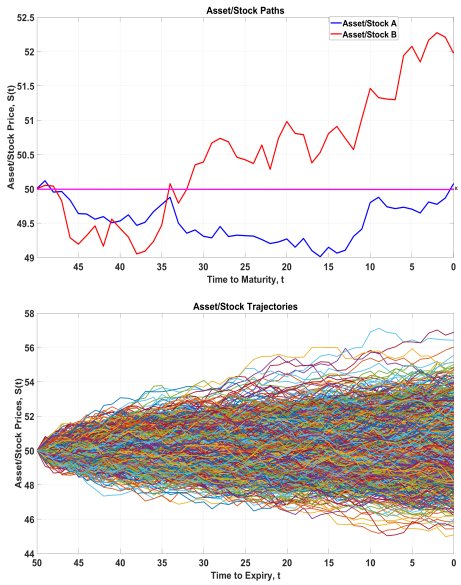


Figure 1.5: Illustration of asset trajectories for two assets A, B and general N with the same maturity time T.

In conclusion, the pay-off of a European option of a European Call option is

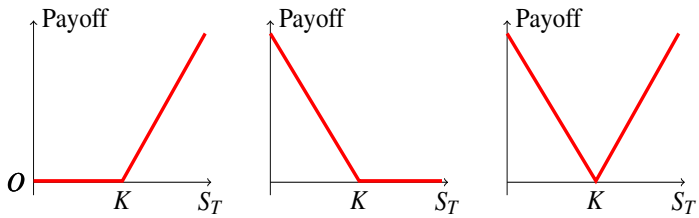
$$(S_T - K)^+ = \max\{S_T - K, 0\}.$$

The graphical representation of the pay-off as a function of S_T for a Call option is as shown in Figure 1.6(a): We notice that the pay-off increases with S_T and gives a potentially unlimited profit. Analogously, we see that the pay-off of a European Put option is

$$(K - S_T)^+ = \max\{K - S_T, 0\}.$$

The graphical representation of the pay-off as a function of S_T for a Put option is as shown in the Figure 1.6(b).

Call and Put options are the basic derivative instruments and for this reason they are often called plain vanilla options. Combining such types of options, it is possible to build new derivatives: for example, by buying a Call and a Put option with the same underlying asset, strike and maturity we obtain a derivative, the so called Straddle, whose pay-off increases the more S_T is far from the strike as shown in Figure 1.6(c). This kind of derivative interesting when one expects a wide movement of the price of the underlying asset without being able to foresee the direction.



(a) Call Option (b) Put Option (c) Straddle

Figure 1.6: Illustration of the pay-off, S_T, for the European: (a) Call option, (b) Put option and (c) Straddle

1.7.5 Optimization Model in Finance

We consider an investment where the investor views the outcome of any investment in probabilistic terms; that is, he thinks of the possible results as some sort of probability distribution with two parameters representing the expectation or returns and the risk factors. Therefore, in assessing the desirability of a particular investment, the investor is willing to take decision on the basis of say two parameters of such a probability distribution, that is, its expected returns $\mathbb{E}[X] = \mu$ value and the risk measure or standard deviation σ, where X is the random variable of the distribution.

The above above ideas can best be described in terms of utility function

$$U = f(E_{\mathrm{B}}, \sigma_{\mathrm{B}}) \tag{1.205}$$

where E_{B} is the expected wealth, σ_{w} the predicted standard deviation of the possible divergence of the future wealth E_{B} from the actual expected wealth E_{B} and B is the total bank deposits, investment or wealth.

Under normal investment environment, every investor prefers a higher expected future wealth to lower ones, leaving other factors constant,

$$\frac{dU}{dE_{\mathrm{B}}} > 0. \tag{1.206}$$

In an attempt to achieve this, however, the investor tends to exhibit the scenario of risk-aversion, that is, opting for an investment offering a lower value of σ_{B} to the one with a greater level independent of the level of E_{B},

$$\frac{dU}{d\sigma_{\mathrm{B}}} < 0. \tag{1.207}$$

These assumptions suggests that the indifference curves relating E_{B} and σ_{B} are upward sloping, see Figure 1.7.5.

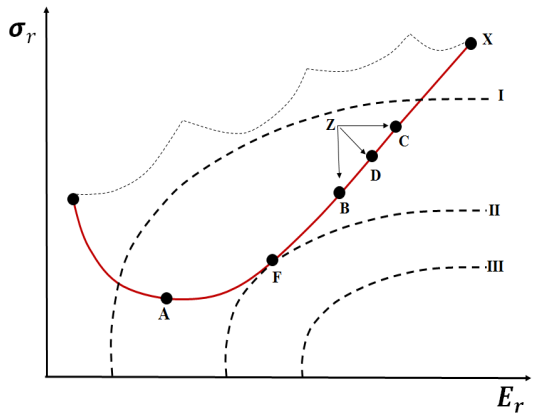


Figure 1.7: AFBDC is the investment opportunity curves.

Assuming the investor decides to commit a given amount B_t of his present wealth to a given investment. Setting B_{t+1} to be the terminal wealth and r the rate of return on his investment, then

$$r = \frac{B_{t+1} - B_t}{B_t}, \tag{1.208}$$

implying that

$$B_{t+1} = rB_t + B_t = (r+1)B_t, \tag{1.209}$$

The above relationship (1.205) makes possible to express the investor's utility in terms of interest rates r, since the terminal wealth B is directly related to the rate of return

$$U = f(E_\mathrm{r}, \sigma_\mathrm{r}). \tag{1.210}$$

Figure 1.7.5 summarizes the mode of investor preferences in a family of indifference curves, that is the investment opportunity curves, which are successive curves that indicate higher levels of utility as one moves down or to the right.

Therefore, the model of investment behaviour considers the investor as aiming at investing in investment opportunities that maximizes his utility in terms of future expected returns. Such investments available to to the investor may be represented in the E_B and σ_B plane. If all such investment plans involve some risk, the area composed of such points will posses an appearance similar to that shown in Figure 1.7.5.

According to Figure 1.7.5, the investor will choose from among all possible plans the investment that places him on the indifference curve representing the highest level of utility in this case point F. The decision can be made in two stages, that is, first find the set of efficient investment plan, and second choose one that is most optimal from among this set.

An investment plan is said to be efficient or optimum if and only if there is no alternative with either

(i) the same E_{B} and a lower σ_{B},

(ii) the same σ_{B} and and a higher E_{B}, or higher E_{B} and a lower σ_{B}

Thus, from Figure 1.7.5 the investment Z is inefficient (not optimum) since investments B,C and D dominate it, among other factors allows arbitrage. The only plans which would be chosen must be along the lower right-hand boundary, AFBDCX-the investment opportunity curve. More detailed discussion of the same principle of investment plan can be got form [423].

Generally, optimization models in finance are constructed based on Markowitz theory, [320], though these models may vary dependent on the different styles of the investors that may include optimal costs, returns and risk on the investment.

The standard mean-variance risk measurement model [320] with $N-$ kinds of assets available for investment options, if the yield of the i-th asset during the $j-$period is s_{ij}, $i = 1,2,\ldots,N$, $j = 1,2,\ldots,M$, the proportion of the total assets of the $i-$th asset is $x_i, i = 1,2,\ldots,N$, the average yield and its variance of the $M-$th period of the $i-$th security are given by

$$\bar{\beta}_i = \frac{1}{M}\sum_{j=1}^{M} s_{ij}, \quad \sigma_i^2 = \sum_{j=1}^{M}(s_{ij}-\bar{s}_i)^2.$$

The covariance of the i-th and j-th assets yield can be expressed as

$$\sigma_{ik} = \frac{1}{M}\sum_{j=1}^{M}(s_{ij}-\bar{s}_i)(s_{ij}-\bar{s}_k).$$

The investment risk can be defined as

$$V = \sum_{i=1}^{N}\sigma_i^2 x_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sigma_{ij}x_i x_j. \tag{1.211}$$

Denoting by $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\top$ as the portfolio vector, then the investment risk V given in (1.211) can be expressed as

$$V = \mathbf{x}^\top \mathbf{Q}\mathbf{x}, \tag{1.212}$$

where

$$\mathbf{Q} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \ldots & \sigma_{NN} \end{bmatrix}, \quad \begin{cases} \sigma_{ij} = \sigma_i^i, & \text{if } i = j \\ \sigma_{ik} = \sigma_{ki}, & \text{if } i \neq k \\ i, j = 1,2,\ldots,N \end{cases}$$

which is a symmetric and semidefinite matrix.

Denoting the yield vector by $\bar{\mathbf{s}} = (\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_N)^\top$, then we can express the investment yield as

$$S = \sum_{i=1}^{N} \bar{s}_i x_i = \mathbf{x}^\top \bar{\mathbf{s}}.$$

In the case of an investment allowing short term trading, if the decision is to minimize the risk, the one constructs the risk model as follows:

$$\text{minimize } V = \mathbf{x}^\top \mathbf{Q}\mathbf{x} \text{ subject to } \sum_{i=1}^{N} x_i = 1. \tag{1.213}$$

The risk minimization model in (1.213) is only based on risk, not considering other factors like returns, interest rate, and cost of investments. In such optimization problems of financial portfolio, if different conditions are considered, then the optimization model can be reformulated differently. For instance, if one considers how to purchase a number of assets from N kinds of risky assets, the variance risk is as small as possible under the premise of the yield to a certain level. Based on the above conditions, we can construct the the following model [26]

$$\begin{cases} \min & V = \mathbf{x}^\top \mathbf{Q}\mathbf{x} \\ \text{s.t} & \sum_{i=1}^{N} x_i = 1 \\ & \sum_{i=1}^{N} \bar{s}_i x_i = \mathbf{s}_p \\ & x_i \geq 0, i = 1, 2, \ldots, N. \end{cases} \tag{1.214}$$

where $\mathbf{s}_p$ is the expected return of investment portfolio.

The set of points that satisfies such an optimization problem are found to lie on a a boundary curve or path also called an efficient frontier [320].

Definition 1.7.3. *The efficient frontier is the set of of optimal portfolios that offer the highest expected returns for a defined level of risk or the lowest risk for a given level of expected returns. Portfolios that lie below the efficient frontier are called sub-optimal because they do not provide enough returns for the level of risk. The portfolios that cluster to the right of the efficient frontier are sub-optimal because they have a higher level of risk for a defined rate of return.*

The Figure 1.7.5 illustrates the structure of an efficient frontier. Each of the points on the surface or boundary line represents the most optimal or extreme points whose portfolio or combination as assets or securities would maximize the returns at any given level of standard deviation, risk or volatility. The points that are inferior to the efficient frontier represent the assets or securities who portfolio or combination would either offer less returns or same returns as the points on the efficient frontier but with high risk or they may offer less returns with the same high risk.

This concept of portfolio optimization and construction of the efficient frontier in asset price is our major motivation in pricing assets or securities using extreme points of the Vandermonde

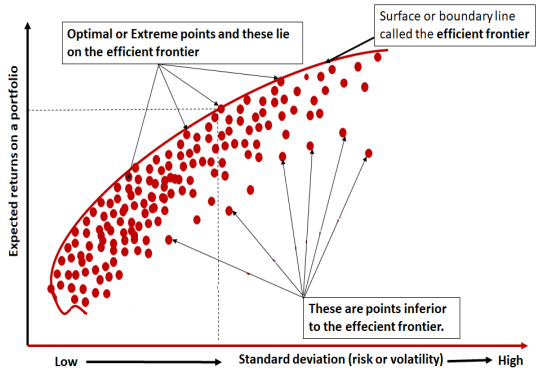


Figure 1.8: Illustration of the portfolio construction using optimal or extreme points lying on the efficient frontier or boundary surface

determinant. In this case efficient frontier represents the boundary of a smooth surface for example the p-sphere, cubes, ellipsoids, and paraboloids that preserve convexity which all form a family of the $p-$norms in $N-$dimension [348]. In the next section we describe the set up of the polynomial based models and pricing with extreme points of the Vandermonde determinant.

1.8 Summaries of Chapters

Chapter 1

This chapter gives the general introduction of the major concepts to be used in the later chapters. The historic background of the Vandermonde matrix and its determinant, the structure and definition of Vandermonde matrix, Vandermonde determinant, generalized Vandermonde matrix, general properties of Vandermonde matrix and its determinant that some properties of the Vandermonde matrix and Vandermonde determinant that make it more applicable in both scientific and mathematical computations, and some of its relationships with other determinants.

Chapter 2

This chapter is based on Paper A [347] and discusses the generalized Vandermonde interpolation polynomial based on divided differences. Some results regarding the appropriateness for this method for curve-fitting and approximation are discussed. The proposed interpolation technique will be tested by construction of approximative models based on joint eigenvalue probability distribution and classical orthogonal polynomials. More details of the are as discussed in Section 2.1, Section 2.2 we define the weighted Fekete points, Section 2.3 the weighted Lebesgue constant and Lebesgue function and these concepts are applied in Section 2.4 that connects the Gaussian orthogonal ensembles with weighted Fekete points which leads to Section 2.5 that gives the possible interpolation polynomial fitting based on different types of weights, that is, the Jacobi, Laguerre and Hermite whose zeros would be the weighted Fekete points for experimental data.

Chapter 3

This chapter is based on Paper B [348] and investigates the extreme points of the Vandermonde determinant on surfaces implicitly determined by a univariate polynomial in higher dimension motivated by results of optimization of Vandermonde determinant by Lagrange multiplier explained in [305]. We derived polynomial expressions that has the coordinates of the extreme points as roots when the surface is a sphere or cube as in Section 3.1.1 critical points on surfaces given by a first degree univariate polynomial, Section 3.1.2 Critical points on surfaces given by a second degree univariate polynomial, Section 3.2 Critical points on the sphere defined by a p-norm in which higher degree univariate polynomials are computed, and Section 3.3 extends the results of extreme points of Vandermonde determinants to cubes and intersections of planes.

Chapter 4

This chapter is based on Paper C [342] and studies the symmetric group properties of extreme points of the Vandermonde determinant and Schur polynomials. We applied these symmetric group properties to investigate the relationship between the extreme points Vandermonde determinant

and Schur polynomials. In Section 4.1 discusses the symmetric group properties of Vandermonde Matrix and its Determinant, Section 4.2 gives the derivatives, extreme points of Vandermonde determinants and Schur polynomials, Section 4.3 explores the extreme points of Schur polynomials on certain surfaces and given by zeros of classical orthogonal polynomial, Section 4.4 extends the extreme points of generalized Vandermonde determinant and Schur polynomial to the Szegö Limit Theorems, and Section 4.5 gives the application of extreme points of Vandermonde determinant in interpolation with symmetric polynomials and Schur polynomials.

Chapter 5

This chapter is based on Paper D [349] and discusses the optimization of the Wishart joint eigenvalue probability density distribution Based on the Vandermonde Determinant we were motivated by the fact that a number of models from mathematics, physics, probability theory and statistics can be described in terms of Wishart matrices and their eigenvalues as in 5.1. The most prominent example being the Gaussian orthogonal ensembles of the spectrum of Wishart type matrix. We aimed to expressing extreme points of the joint eigenvalue probability density distribution of a Wishart matrix using optimisation techniques for the Vandermonde determinant over certain surfaces implicitly defined by univariate polynomials which is illustrated in Section 5.2.

Chapter 6

This chapter is based on Paper E [346] and studies the properties of the extreme points of the joint eigenvalue probability density function of the Wishart matrix we established the usefulness of polynomial decomposition and the properties of the extreme points of the joint eigenvalue probability density function in optimization of condition number of the Vandermonde and Wishart matrices by the technique of maximizing the Vandermonde determinant which is explained in detail in Section 6.1.

We also established that the condition number of the Vandermonde matrix is inversely proportional to the absolute value of its determinant while the condition of the Wishart matrix is inversely proportional to the square of the Vandermonde determinant. Therefore, the extreme points of the joint probability density function of the Wishart matrix that maximize the Vandermonde determinant can be used to minimize the condition number of both the Vandermonde determinant and the Wishart matrix. The points that maximizes the Vandermonde determinant are often referred to as Fekete points as in Section 6.2.

We were also able to illustrate that indeed the extreme points of the Vandermonde determinant are indeed related to the eigenvalues of the Wishart and these extreme points have a joint eigenvalue density function as the Gaussian ensembles. These points which are also zero of classical orthogonal polynomials provide the most stable and economical interpolating points as explained in Section 6.3.

Chapter 7

This chapter is based on Paper F [343] and investigates the connections between the extreme points of Vandermonde determinant and minimizing risk measure in financial mathematics. The extreme points of the Vandermonde determinant optimized over various surfaces were found helpful in approximating the efficient frontier and can play a useful role in asset pricing and portfolio construction. For instance, to determine most appropriate asset allocation, that is, which assets can make the best combination based on their risk measure to be able to maximize the returns by minimizing the risk. Here extreme points of the Vandermonde determinant we found helpful in constructing the most suitable path to represent the risk-returns trade off in which case indicates how the potential returns rises or falls with with increase in risk which is discussed in Section 7.1.

The pricing model also can constructed using the extreme points of Vandermonde determinant once the right asset allocation has been made. This can help to divide the capital equitably between the most appropriate and optimal assets. This model can continuously be assessed by use of suitable portfolio weights depending on the available information as in Section 7.2.

Chapter 8

This chapter is based on Paper G [344] and investigates group properties of the Wishart distribution on symmetric vones in Jordan algebra. The Wishart probability distributions was constructed on the basis of symmetric cones and how this extends to higher dimension as discussed in Section 8.1. This density is mainly characterised by the structure of the Vandermonde determinant and the exponential weight that is dependent on the trace of the given matrix by use of Lassalle measure on symmetric cones and probability distribution in Section 8.2. The symmetric cones especially the Gindikin set form a suitable basis for the construction of the degenerate and non-degenerate Wishart distributions in the field of $\mathrm{Herm}(m,\mathbb{C})$, $\mathrm{Herm}(m,\mathbb{H})$, $\mathrm{Herm}(3,\mathbb{O})$ denotes respectively the Jordan algebra of all Hermitian matrices of size $m \times m$ with complex entries, the skew field $\mathbb{H}$ of quaternions, and the algebra $\mathbb{O}$ of octonions as discussed in Section 8.3.

Chapter 9

This chapter is based on Paper H [345] and explores the symmetric cones especially the Gindikin set that forms a suitable basis for the construction of the degenerate and non-degenerate Wishart distributions in the field of $\mathrm{Herm}(m,\mathbb{C})$, $\mathrm{Herm}(m,\mathbb{H})$, $\mathrm{Herm}(3,\mathbb{O})$ which denotes, respectively, the Jordan algebra of all Hermitian matrices of size $m \times m$ with complex entries, the skew field $\mathbb{H}$ of quaternions, and the algebra $\mathbb{O}$ of octonions.

We also illustrated the extreme points of the Vandermonde determinant and Wishart ensembles on symmetric cones in Section 9.1 which gives the relationship between Gindikin set and Wishart joint eigenvalue distribution. The generalization of the Wishart probability distributions in higher dimension based on the boundary points of the symmetric cones as discussed in Section 9.2. Further, the characterisation of the Wishart density by the general structure of the Vandermonde

Conclusion

The thesis mainly discusses the extreme points of the Vandermonde determinant on various surfaces, their applications in numerical approximation, random matrix theory and financial mathematics. Some mathematical models that employ these extreme points such as curve fitting, data smoothing, experimental design, electrostatics, risk control in finance and method for finding these extreme points on certain surfaces are demonstrated.

In Chapter 1, we explored the theoretical background necessary for later chapters. We review the historical background of the Vandermonde matrix and its determinant, some of its properties based on group theory, symmetric polynomials, classical orthogonal polynomials and random matrix theory.

In Chapter 2, we established the relationship between generalized Vandermonde interpolation polynomial based and divided difference. We extended this relationship to the concepts of weighted Fekete points, joint eigenvalue probability distributions and zeros of the classical orthogonal polynomials as stable interpolation points.

In Chapter 3, we generated univariate polynomials whose zeros are the extreme points of Vandermonde determinant when optimized on p-spheres of finite dimension. We also give the generalization of the extreme points of Vandermonde determinant on various surfaces, which are given as roots of classical orthogonal polynomials.

In Chapter 4, obtained the symmetric group properties of the extreme points of Vandermonde and Schur polynomials. We also generated results on optimization of Schur polynomials on various surfaces which are also given as zeros of classical orthogonal polynomials in relation with generalized Vandermonde determinant. We illustrated the use of these extreme points in computation of Szego limits, interpolation with symmetric polynomials and Schur polynomials.

In Chapter 5, we applied the extreme points of Vandermonde determinant in optimization of the joint eigenvalue probability density distribution of a Wishart matrix on surfaces implicitly defined by univariate polynomials. We illustrated our results on the p-sphere.

In Chapter 6, we applied the extreme points of the vandermonde determinant that maximize the joint eigenvalue probability density distribution in computation of the condition numbers of the Vandermonde and Wishart matrices which is necessary in characterization of stability of Vandermonde matrix.

In Chapter 7, we demonstrated use of the extreme points of Vandermonde determinants as minimizing risk measures in financial mathematics. This was illustrated with an application to

optimal portfolio selection.

In Chapter 8, we constructed the Wishart probability distributions in higher dimension based on the symmetric cones in Jordan algebras. The the non-degenerate Wishart distribution and degenerate Wishart distribution were derived on both the continuous and discrete part of the Gindikin set.

In Chapter 9, we illustrated the extreme points of the Vandermonde determinant which are also zeros of Laguerre orthogonal polynomial in computation of the optimal points of Wishart joint eigenvalue probability distributions based on the boundary points of the Gindikin set.

References

[1] Aaronson Scott, Alex Arkhipov. The computational complexity of linear optics. Theory of Computing, 9: 143–252, 2013.

[2] Abramowitz Milton, Stegun Irene A. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York , 1964.

[3] Aitken Alexander C. Determinants and matrices. Oliver & Boyd, Edinburgh, 1939.

[4] Akaike Harotugu. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723, 1974.

[5] Akaike Hirotugu. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and Frigyes Csáki, editors, 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, 267–281, 1973.

[6] Alfonsi Liliane. Étienne Bézout (1730-1783): Mathématicien des Lumiéres (Histoire des sciences) (French) L'Harmattan, Paperback – 1 Nov. 2011.

[7] Anderson Greg W., Guionnet Alice, Zeitouni Ofer. An introduction to random matrices. Cambridge University Press, 2010.

[8] Anderson Theodore W. An Introduction to Multivariate Statistical Analysis. Wiley Interscience, 2003.

[9] Anderson T. W., Girshick M. A. Some extensions of Wishart distribution. Ann. Math. Statist., 15(4), 345–357, 1944.

[10] Andersson Steen A., Brøns Hans K., Jensen Soren Tolver. Distribution of eigenvalues in multivariate statistical analysis. Ann. Statist. 11(2), 392–415, 1983.

[11] Andersson Steen A. Invariant normal models. Ann. Statist., 3:132–154, 1975. ISSN 0090-5364.

[12] Andrews George E., Askey Richard, Roy Ranjan. Special functions, Encyclopedia of Mathematics and its Applications, vol. 71. Cambridge University Press, Cambridge (1999).

[13] Angelos James R., Kaufman Edwin H., Henry Mark S., Lenker Terry D. Optimal nodes for polynomial interpolation. In: Chui, CK, Schumaker, LL, Ward, JD (eds.) Approximation Theory. Academic Press, New York, VI, 17–20, 1989.

[14] Arazy Jonathan, Upmeier Harald. Boundary measures for symmetric domains and integral formulas for the discrete Wallach points. Integral Equations Operator Theory 47(4), 375–434, 2003.

[15] Askey Richard. Orthogonal Polynomials and Special Functions. Society for Industrial and Applied Mathematics, 1975.

[16] Axler Sheldon. Linear Algebra Done Right. Springer International Publishing, 3rd edition, 2015.

[17] Azad Hassan, Laradji Abdallah, Mustafa Muhammad Tahir. Polynomial solutions of certain differential equations arising in physics. Mathematical Methods in Applied Science, 2012.

[18] Bahcall Safi R. Random Matrix Model for Superconductors in a Magnetic Field. Physical review letters, 77(26), 5276–5279, 1996.

[19] Bai Zhidong, Zhaoben Fang, Ying-Chang Liang. Spectral Theory of Large Dimensional Random Matrices and Its application to Wireless Communiaction and Finance: Random Matrix Theory and Its Applications. World Scientific Publishing Co., Pte., Ltd, 2014.

[20] Baltzer Robert A. Theorie und Anwendung der Determinanten. Hirzel, Leipzig, 1857.

[21] Banach Stefan. Über die Baire'sche Kategorie gewisser Funktionenmengen. Studia Math. 3(1): 174–179, 1931.

[22] Barndorff–Nielsen E. Ole. Information and exponential families in statistical theory. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester (2014).

[23] Barndorff–Nielsen Ole E. Processes of normal inverse Gaussian type. Finance Stock, 2, 41–68, 1998.

[24] Barnsley Michael Fielding. Fractals Everywhere. Academic Press, Inc., 1988.

[25] Barnsley Michael Fielding. Fractal functions and interpolation. Constructive Approximation, 2(1), 303–329, 1986.

[26] Bartholomew-Biggs M. C. Non-Linear Optimization with Financial Applications. Kluwer Academic Publishers, Net Library, Inc., 2005.

[27] Basor Estelle L., Tracy Craig A. The Fisher–Hartwig conjecture and generalizations. Phy. A 177, 167–173, 1991

[28] Basor Estelle L., Widom Harold. On a Toeplitz determinant identity of Borodin and Okounv. Integral Eq. Oper. Th. 37(4), 397–401, 2000.

[29] Bäckström Tom, Fischer Johannes, Boley Daniel. Implementation and evaluation of the Vandermonde transform. In 22nd European Signal Processing Conference (EUSIPCO), 71–75, 2014.

[30] Bäckström Tom. Vandermonde factorization of Toeplitz matrices and applications in filltering and warping. IEEE Transactions on Signal Processing, 61(24), 6257–6263, 2013.

[31] Beckermann Bernhard, Golub Gene H., Labahn George. On the numerical condition of a generalized Hankel eigenvalue problem. Numerische Mathematik, 106(1), 41–68, 2007.

[32] Beckermann Bernhard, Labahn George. Effective computation of rational approximants and interpolants. Reliable Computing, 6(4): 365–390, 2000.

[33] Bellhouse David R., Genest Christian. Maty's biography of Abraham de Moivre, translated, annotated and augmented. Statistical Science, 22(1), 109–136, 2007.

[34] Bellman Richard. Introduction to Matrix Analysis. McGraw–Hill Book Company, New York, 2nd edition, 1970.

[35] Bender Edward A., Knuth Donald E. Enumeration of plane partitions. J. Combin. Theory Ser. A, 13, 40–54, 1972.

[36] Benjamin Arthur T., Dresden Gregory P. A combinatorial proof of Vandermonde's determinant. The American Mathematical Monthly, 114(4), 338–341, April 2007.

[37] Benth Fred Espen. Option Theory with Stochastic Analysis. Springer-Verlag, Berlin, 2004.

[38] Berman Robert, Boucksom Sébastien, Nyström David W. Fekete points and convergence towards equilibrium measures on complex manifolds. Acta Mathematica, 207(1), 1–27, 2011.

[39] Berrut Jean–Paul, Trefethen Lloyd N. Barycentric Lagrange interpolation. SIAM Review, 46(3), 501–517, 2004.

[40] Bertrand Gérald. Trattato di algebra elementare. Le Monnier, Firenze, 1859.

[41] Bessenrodt Christine, Luoto Kurt W., van Willigenburg Stephanie. Skew quasisymmetric Schur functions and noncommutative Schur functions. Adv. Math., 226, 4492–4532, 2011.

[42] Birkhoff Garrett, De Boor Carl. Piecewise polynomial interpolation and approximation. In Henry. L. Garabedian, editor, Approximation of functions-Proceeding of the General Motors Symposium of 1964, 164–190, 1965.

[43] Björck Åke. Arbitrage Theory in Continuous Time. Oxford University, Press, 2000.

[44] Björck Åke, Elfving Tommy. Algorithms for confluent Vandermonde systems. Numer. Math., 21, 130–137, 1973.

[45] Björck Åke, Victor Pereyra. Solution of Vandermonde systems of equations. Mathematics of Computation, 24(112), 893–903, 1970.

[46] Black Fischer, Scholes Myron. The pricing of option and corporate liabilities. Journal of political Economy, 81, 637–654, 1973.

[47] Bleher Pavel M., Bleher Pavel, Its Alexander. Random Matrix Models and Their Applications, eds. Vol. 40. Cambridge University Press, 2001.

[48] Bloom Thomas F., Bos Len P., Calvi Jean-Paul, Levenberg Norman. Polynomial interpolation and approximation in Cd. arXiv:1111.6418.

[49] Bohigas Oriol, Marie-Joya Giannoni, Charles Schmit. Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws. Phys. Rev. Lett., 52(1), 1–4, 1984.

[50] Bonello Nicholas , Sheng Chen, Lajos Hanzo. Construction of regular quasi-cyclic protograph LDPC codes based on Vandermonde matrices. IEEE transactions on vehicular technology, 57(8), 2583–2588, July 2008.

[51] Bôcher Maxime. Certain cases in which the vanishing of the Wronskian is a sufficient condition for linear dependence. Transactions of the American Mathematical Society, 2(2), 139–149, April 1900.

[52] Bôcher Maxime. On linear dependence of functions of one variable. Bulletin of the American Mathematical Society, 120–121, 1900.

[53] Bôcher Maxime. The theory of linear dependence. Annals of Mathematics, 2(1), 81–96, 1900.

[54] Böttcher Albrecht, Silbermann Bernd. Introduction to large truncated Toeplitz matrices. Springer, 1998.

[55] Böttcher Albrecht, Silbermann Bernd. Operator-valued Szegö-Widom limit Theorems. Oper. Theory Adv. Appl., 71, Birkhäuser, 33–53, 1994.

[56] Böttcher Albrecht, Silbermann Bernd. Analysis of Toeplitz operators. Springer, 1990.

[57] Bos Leonard P., De Marchi Stefano, Sommariva Alvise, Vianello Marco. Computing multivariate Fekete and Leja points by numerical linear algebra. SIAM Journal of Numerical Analysis, 48(5), 1984–1999, 2010.

[58] Bos Leonard P., Calvi Jean–Paul, Levenberg Norman, Sommariva Alvise, Vianello Marco. Geometric weakly admissible meshes, discrete least squares approximation and approximate Fekete points. Math. Comp., in press, 2010.

[59] Bose Ray Chandra, Ray-Chaudhuri Dwijendra K. On a class of error correcting binary group codes. Information and Control, 1960.

[60] Bostan Alin, Dumas Phillippe. Wronskians and linear dependence. American Mathematical Monthly, 117(8), 722–727, 2010.

[61] Bourbaki Nicolas. Elements of Mathematics: Algebra I, Springer, New York, 1989.

[62] Bozgodan Hamparsum. Model selection and Akaike's information criterion (AIC): the general theory and analytical extensions. Psychometrika, 52(3), 345–370, 1987.

[63] Bressoud David M. Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture. Spectrum. Cambridge University Press, 1999.

[64] Briani Matteo, Sommariva Alvise, Vianello Marco. Computing Fekete and Lebesgue points: Simplex, square, disk. Journal of Computational and Applied Mathematics, 236(9), 2477–2486, 2012.

[65] Brigo Damiano, Mercurio Fabio. Interest rate models-theory and practice: with smile, inflation and credit. Springer Science & Business Media, 2007.

[66] Brioschi Francesco. La teorica dei determinanti, e le sue pricipali applicazioni. Bizzoni, Pavia, 1854.

[67] Brutman Lev. Lebesgue functions for polynomial interpolation - a survey. Ann. Numer. Math. 4(1–4), 111–127, 1997.

[68] Brutman Lev. On the Lebesgue function for polynomial interpolation. SIAM J. Numer. Anal. 15(4), 694–704, 1978.

[69] Buck Marshall W., Coley Raymond A., Robbins David P. A generalized Vandermonde determinant. Journal of Algebraic Combinatorics, 1, 105–109, 1992.

[70] Burnham Kenneth P., Anderson David R. Model Selection and Multimodel Inference. Springer-Verlag New York Inc., 2002.

[71] Callaway David J. E. Random matrices, fractional statistics, and the quantum Hall effect. Phys. Rev. B., 43(10), 8641–8643, 1991.

[72] Calvi Jean–Paul, Levenberg Norman. Uniform approximation by discrete least squares polynomials. J. Approx. Theory, 152, 82–100, 2008.

[73] Carlitz Leonard. Note on Lebesgue's constants. Proc. Amer. Math. Soc. 12, 932–935, 1961.

[74] Casalis Muriel, Letac Gérard. Characterization of the Jørgensen set in generalized linear models. Test 3(1), 145–162, 1994.

[75] Cauchy Augustin–Louis. Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variables qu'elles renferment. Journal de l'Ecolé Polytechnique, 10(17), 29–112, 1815. Reprinted in Euvres complétes d'Augustin Cauchy series 2, Gauthier–Villars, Paris, Volume 1, 91–161, 1899.

[76] Cauchy Augustin–Louis. Mémoire sur les fonctions alternées et sur les sommes alternées, in Exercices d'Analyse et de Physique Mathématique tome ii. Bachelier, Paris, 151–159, 1841.

[77] Cauchy Augustin–Louis. Cauchy, Cours d'Analyse de l'École. Royale Polytechnique Debure, Paris, 1821.

[78] Cauchy Augustin–Louis. Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variables qu'elles renferment. Journal de l'École Polytechnique, xviie Cahier, Tome x, 91–169, 1815.

[79] Cauchy Augustin–Louis. Mémoire sur le nombre des valeurs qu'une fonction peut acquérir lorsqu'on y permute de toutes les maniéres possibles les quantités qu'elle renferme. Journal de l'École Polytechnique, xviie Cahier, Tome x, 64–90, 1815.

[80] Cayley Arthur. A memoir on the theory of matrices. Philosophical Transactions of the Royal Society of London, 148, 17–37, 1858.

[81] Cayley Arthur. Note sur la methode d'elimination de Bezout. J. Reine Angew. Math., 53, 366–367, 1857.

[82] Chen William Y. C., James Louck D. Interpolation for symmetric functions. Advances in mathematics, 117(1), 147–156, 1996.

[83] Chen Qi, Babuška Ivo. Approximate optimal points for polynomial interpolation of real functions in an interval and in a triangle. Comput. Methods Appl. Mech. Eng. 128(3-4), 405–417, 1995.

[84] Cheney Elliott Ward. Introduction to Approximation Theory. McGraw-Hill, New York, 1966.

[85] Chiani Marco. Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution. Journal of Multivariate Analysis, 129, 69–81, 2074.

[86] Chow Gregory P. Analysis and Control of Dynamic Economic Systems. New York, Wiley, 1976.

[87] Chrystal George. Algebra. Black, London, 1886.

[88] Chui Charles K. An Introduction To Wavelets. Academic Press, 1992.

[89] Ciftci Hakan, Hall Richard L., Saad Nasser, Dogu Ebubekir. Physical applications of second–order linear differential equations that admit polynomial solutions. A: Mathematical and Theoretical, 2010.

[90] Cirafici Michele, Sinkovics Annamaria, Szabo Richard J. Cohomological gauge theory, quiver matrix models and Donaldson–Thomas theory. Nuclear Physics, Section B, 809(3), 452–518, 2009.

[91] Cohn Henry, Kumar Abhinav, Miller Stephen D., Radchenko Danylo, Viazovska Maryna. The sphere packing problem in dimension 24. Annals of Mathematics, 185, 1017–1033, 2017.

[92] Cohn Henry. A conceptual breakthrough in sphere packing. Notices American Mathematical Society, 64(2), 102–15, 2017.

[93] Cohn Henry, Miller Stephen D. Some properties of optimal functions for sphere packing in dimensions 8 and 24. arXiv:1603.04759, 2016.

[94] Conway John Horton, Neil James Alexander S. Sphere Packings, Lattices and Groups, 3rd Ed. Springer-Verlag Berlin Heidelberg New York, 1999.

[95] Cotler Jordan, Hunter–Jones Nicholas, Liu Junyu, Yoshida Beni. Chaos, Complexity and Random Matrices. Journal of High Energy Physics, 2017, 48, 2017.

[96] Coulomb Charles-Augustin. Premier mémoire sur l'électricité et le magnétisme. Histoire de l'Académie royale des sciences avec les mémoires de mathématiques et de physique pour la même année tirés des registres de cette académie. Annáe MDCCLXXXV, 569–577, 1785.

[97] Coulston Gillispie C., Grattan-Guinness Ivor. Pierre-Simon Laplace, 1749–1827, a life in exact science. Princeton University Press, 2000.

[98] Cox David John, Little Don O'Shea. Ideals, Varieties, and Algorithms (second ed.) Springer-Verlag, 1997.

[99] Cox John C., Ross Stephen A., Ross Rubinstein M. Option pricing: A simplified approach. Journal of financial Economics, 7(3), 229–263, 1979.

[100] Dana Rose–Anne, Monique Jeanblanc. Financial markets in continuous time. Springer Finance. Springer-Verlag, Berlin, 2007.

[101] Davis Philip J. Interpolation and Approximation. Blaisdell, New York, 1963.

[102] D'Andrea Carlos, Tabera Luis. Tropicalization and irreducibility of generalized Vandermonde determinants. Proceedings of the American Mathematical Society, 137, 3647–3656, 2009.

[103] De Boor Carl. Divided differences. Surveys in Approximation Theory, 6, 50–62, 2005.

[104] De Boor Carl. A Practical Guide to Splines. Springer Verlag, 1978.

[105] De Boor Carl. On calculating with b-splines. Journal of Approximation Theory, 1, 46–69, 1972.

[106] De Boor Carl, Pinkus Allan. Proof of the conjectures of Bernstein and Erdös concerning the optimal nodes for polynomial interpolation. Bull. J. Approx. Theory 24(4), 289–303, 1978.

[107] De Marchi Stefano, Morandi Cecchi Maria. Polynomials arising in factoring generalized Vandermonde determinants III: Computations and their roots. Neural, Parallel and Scientiffic Computations, 14, 25–38, 2006.

[108] De Marchi Stefano. Polynomials arising in factoring generalized Vandermonde determinants II: A condition for monicity. Applied Mathematics Letters, 15(5), 627–632, 2002.

[109] De Marchi Stefano. Polynomials arising in factoring generalized Vandermonde determinants: an algorithm for computing their coefficients. Mathematical and Computer Modelling, 34(3), 271–281, 2001.

[110] De Moivre Abraham. Miscellanea analytica de seriebus et quadraturis. London, 1730.

[111] De Moivre Abraham. The doctrine of chances, London, 1718.

[112] Delbaen Freddy, Chachermayer Walter. The mathematics of arbitrage. Springer Finance. Springer–Verlag, Berlin, 2006.

[113] Delbaen Freddy, Shirakawa Hiroshi. A note on option pricing for the constant elasticity of variance model. Asia–Pacific Financial Markets, 9, 85–99, 2002.

[114] Demmel James W. Applied numerical linear algebra. SIAM, 56, 1997.

[115] Demmel James W. The probability that a numerical analysis problem is difficult. Mathematics of Computation, 50(182), 449–480, 1988.

[116] Derrida Bernard. Random–energy model: Limit of a family of disordered models. Physical Review Letters, 45(2), 79, 1980.

[117] Desurvire Emmanuel. Classical and Quantum Information Theory. Cambridge University Press, 2009.

[118] Dette Holger, Trampisch Matthias. A general approach to D-optimal designs for weighted univariate polynomial regression models. Journal of the Korean Statistical Society, 39, 1–26, 2010.

[119] Dhombres Jean. Une histoire de l'objectivité scientifique et le concept de postérité, in R. Guesnerie & H. Hartog, Des sciences et des techniques, un débat. EHESS, Paris, 127–148, 1998.

[120] Dickson Leonard Eugene. Linear Groups with an Exposition of the Galois Field Theory. B. G. Teubner, 1901.

[121] Di Francesco Philippe. Quantum Gravity, Matrix Models and Graph Combinatorics. NATO Science Series II: Mathematics, Physics and Chemistry. Springer, Dordrecht, 221, 33–88, 2004.

[122] Dimitrov Dimitar K., Shapiro Boris. Electrostatic Problems with a Rational Constraint and Degenerate Lamé Equations. Potential Anal 52, 645–659, 2020.

[123] Dodgson Charles L. An elementary treatise on determinants. Mac Millan, London, 1867.

[124] Donovan Arthur, Antoine Lavoisier. Science, administration and revolution. Vol. 5. Cambridge University Press, 1996.

[125] Dothan Michael. Efficiency and Arbitrage in Financial Markets. International Research Journal of Finance and Economics, 19.19, 102–106, 2008.

[126] Dothan Michael. Prices in financial markets. The Clarendon Press Oxford University Press, New York, 1990.

[127] Dubiner Moshe. Spectral methods on triangles and other domains. J. Sci. Comput., 6, 345–390, 1991.

[128] Duffie Darrell. Dynamic asset pricing theory, 3rd ed. NJ: Princeton University Press, Princeton, 2001.

[129] Dumitriu Ioana, Edelman Alan. Matrix models for beta ensembles. Journal of Mathematical Physics, 45(11), 5830–5847, 2002.

[130] Dunkl Charles F., Xu Yuan. Orthogonal Polynomials of Several Variables. Cambridge University Press, Cambridge, 2001.

[131] Dyson Freeman John. A meeting with Enrico Fermi. Nature, 427(6972), 297, 2004.

[132] Dyson Freeman John. Statistical theory of energy levels of complex systems IV. Journal of Mathematical Physics, 3(6), 1199–1215, 1962.

[133] Dyson Freeman John. Statistical theory of energy levels of complex systems III. Journal of Mathematical Physics, 3(1), 166–175, 1962.

[134] Dyson Freeman John. Statistical theory of energy levels of complex systems II. Journal of Mathematical Physics, 3(1), 157–165, 1962.

[135] Dyson Freeman John. Statistical theory of energy levels of complex systems I. Journal of Mathematical Physics, 3(1), 140–156, 1962.

[136] Dyson Freeman John. The approximation to algebraic numbers by rationals. Acta Mathematica, 79(1), 225–240, 1947.

[137] Echols Ho W. On Some Forms of Lagrange's Interpolation Formula. Mathematical problems for the next century. The Annals of Mathematics, 8(1/6), 22–24, 1893.

[138] Edelman Alan, Rao N. Raj. Random matrix theory. Acta Numerica., 14, 233–297, 2005.

[139] Edelman Alan. The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. Linear Algebra Appl. 159, 55–80, 1991.

[140] Edelman Alan. Eigenvalues and condition numbers of random matrices. SIAM Journal on Matrix Analysis and Applications, 9(4), 543–560, 1988.

[141] Edwards Harold M. Galois theory. Springer, New York, 1984.

[142] Efron Bradley, Morris Carl N. Stein's paradox in statistics. Scientific American, 236 (5), 119–127, 1977.

[143] Efron Bradley, Morris Carl N. Multivariate empirical Bayes and estimation of covariance matrices. The Annals of Statistics, 4(1), 22–32, 1976.

[144] Ehrhardt Torsten. Toeplitz determinants with several Fisher-Hartwig singularities. PhD Thesis Techn. Univ. Chemnitz, 1997.

[145] Eisinberg Alfredo, Fedele Guiseppe. On the inversion of the Vandermonde matrix. Applied Mathematics and Computation, 174(2), 1384–1397, 2006.

[146] Elliott Robert J., Kopp Ekkehard P. Mathematics of financial markets, 2nd ed. Springer Finance. Springer-Verlag, New York, 2005.

[147] El–Mikkawy Moawwad E. A. Explicit inversion Of A Generalized Vandermonde Matrix. International Journal of Computer Mathematics, 80(6), 2003.

[148] El–Mikkawy Moawwad E. A, El-Desouky Beih. On a connection between symmetric polynomials, generalized Stirling numbers and the Newton general divided difference interpolation polynomial. Appl. Math. Comput., 138, 375–385, 2003.

[149] Epps Thomas Wake. Pricing derivative securities. World Scientific, Singapore, 2000.

[150] Erdös P. (Budapest). Problems and results on the theory of interpolation. I. Acta Math. Acad. Sci. Hung. 9, 381–388, 1958.

[151] Erdös P. (Budapest). Some remarks on the theory of graphs. Bull. Am. Math. Soc. 53, 292–294, 1947.

[152] Ernst Thomas. Generalized Vandermonde determinants. U. U. D. M. Report 6, 2000.

[153] Fabozzi Frank J., Mann Steven V. A handbook of fixed income securuties (Seventh ed.) kvrv, New York, 2012.

[154] Faccarello Gilbert. Du conservatoire à l'École Normale: quelques notes sur A.T. Vandermonde (1735–1796), Cahiers d'Histoire du CNAM, 2/3, 15–57, 1993.

[155] Faraut Jacques, Korányi Adam. Analysis on symmetric cones. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1994.

[156] Faraut Jacques. Formule du binôme généralisée. In: Harmonic analysis (Luxembourg, 1987), Lecture Notes in Math. Springer, Berlin, 1359, 170–180, 1988.

[157] Farnell A. B. A special Vandermondian determinant. Amer. Math. Monthly, 66(1959), 564–569, 1950.

[158] Fedorov Valerii V. Theory of Optimal Experiments. Academic Press, Inc., 1972.

[159] Fejér Leopold. Bestimmung derjenigen Abszissen eines Intervalles, für welche die Quadratsumme der Grundfunktionen der Lagrangeschen Interpolation im Intervalle ein Möglichst kleines Maximum Besitzt Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Série 2, Volume 1(3), 263–276, 1932.

[160] Finch Steven R. Mathematical Constants. Encyclopedia of Mathematics and its Applications, Cambridge University Press, 94, 2003.

[161] Fischer Black, Scholes Myron. The pricing of options and corporate liabilities. Journal of Political Economy, 81, 637–659, 1975.

[162] Fisher Michael E., Hartwig Robert E. Toeplitz determinants: Some applications, theorems and conjectures. Adv. Chem. Phys. 15, 333–353, 1968.

[163] Fisher Ronald A. The sampling distribution of some statistics obtained from non-linear equations. Ann. Eugenics 9, 238–249, 1939.

[164] Fisher Ronald A. Frequency distribution of the values of the correlation coefficients in samples from an infinitely large population. Biometrika 10(4), 507–521, 1915.

[165] Flowe Randolph P., Harris Gary A. A note on generalized Vandermonde determinants. SIAM Journal on Matrix Analysis and Applications, 14(4), 1146–1151, 1993.

[166] Forrester Peter J. Octonions in random matrix theory. Proc. A., 473(2200), 20160800, 11, 2017.

[167] Forrester Peter J. Log-Gases and Random Matrices. Princeton University Press, 2010.

[168] Forrester Peter J., Warnaar Ole S. The importance of the Selberg integral. Bull. Amer. Math. Soc. (N.S.) 45(4), 489–534, 2008.

[169] Föllmer Hans, Schied Alexander. Stochastic Finance. An Introduction in Discrete Time, 2nd ed. Gruyter Studies in Mathematics 27, Walter de Gruyter, New York, 2002.

[170] Franchini Fabio, Vladimir Kravtsov E. Horizon in random matrix theory, the Hawking radiation, and flow of cold atoms. Phys. Rev. Lett., 103 (16), 166401, 2009.

[171] Fraser Douglas C. Newton's interpolation formulas, reprinted from The Journal of the Institute of Acturaries, Layton, London, 51(2), 77–106, 1918, &, 211–232, 1919.

[172] Freud G (Budapest). On Polynomial Approximation with Weights $\exp\{-\frac{1}{2}x^{2k}\}, k \geq 1$. Acta Mathematica Academiae Scientarum Hungariocal Tomus 24(3–4), 363–371, 1973.

[173] Fröberg Ralf, Shapiro Boris. On Vandermonde varieties. Mathematica Scandinavica, 119(1), 73–91, 2016.

[174] Fulton William. Young tableaux, London Mathematical Society Student Texts, 35, Cambridge University Press, 1997.

[175] Fulton William, Harris Joe. Representation Theory: a first course. Springer-Verlag, 1991.

[176] Gaffke Norbert, Krafft Olaf. Exact D-Optimum Designs for Quadratic Regression. Journal of the Royal Statistical Society. Series B (Methodological), 44(3), 394–397, 1982.

[177] Galkin Pavel V. Estimates for Lebesgue constants (in Russian). Trudy Mat. Inst. Steklov. 109, 3–5, 1971; Engl. transl. in Proc. Steklov Inst. Math. 109, 1–4, 1971.

[178] Gallier Jean H. Curves and Surfaces in Geometric Modelling: Theory and Algorithms. Morgan Kaufmann, 1999.

[179] Gantmacher Feliks Ruvimovič. The theory of matrices. Chelsea, New York, 1959.

[180] García del Molino L. C., Pakdaman Khashayar, Touboul Jonathan, Wainrib Gilles. Synchronization in random balanced networks. Physical Review E., 88(4), 2013.

[181] Gatto Letterio, Scherbak Inna. On generalized Wronskians. In P. Pragacz, editor, Contributions to Algebraic Geometry, EMS Congress Series Report, 257–296, 2012.

[182] Gauss Carl Friedrich. Theoria combinationis observationum erroribus minimis obnoxiae, Pars Posterior. Gottingae, 1823.

[183] Gauss Carl Friedrich. Theoria combinationis observationum erroribus minimis obnoxiae, Pars Prior. Gottingae, 1821.

[184] Gautschi Walter. Optimally scaled and optimally conditioned Vandermonde and Vandermonde-like matrices. BIT Numerical Mathematics, 51(1), 103–125, 2011.

[185] Gautschi Walter. How (stable) are Vandermonde systems. Asymptotic and computational analysis 124, 193–210, 1990.

[186] Gautschi Walter. A survey of Gauss–Christoffel quadrature formulae. In EB Christoffel, Birkhauser, Basel, 72-147, 1981.

[187] Gautschi Walter. On inverses of Vandermonde and confluent Vandermonde matrices III. Numerische Mathematik, 29(4), 445–450, 1978.

[188] Gautschi Walter. Optimally conditioned Vandermonde matrices. Numerische Mathematik, 24(1), 1–12, 1975.

[189] Gautschi Walter. On inverses of Vandermonde and confluent Vandermonde matrices II. Numerische Mathematik, 5, 425–430, 1963.

[190] Gautschi Walter. On inverses of Vandermonde and confluent Vandermonde matrices. Numerische Mathematik, 4, 117–123, 1962.

[191] Gautschi Walter, Inglese G. Lower bounds for the condition number of Vandermonde matrices. Numerische Mathematik, 52(3), 241–250, 1987.

[192] Gazizov Rafail K., Ibragimov Nail H. Lie symmetry analysis of differential equations in finance. Nonlinear Dynamics, Kluwer Academic Publishers, Netherlands, 17(4), 387–407. 1998.

[193] Geloun Joseph B., Mahouton Norbert H. Multi–indicial symmetric functions. arXiv preprint arXiv:0906.1252, 2009.

[194] Gessel Ira M. Multipartite P-partitions and inner products of skew Schur functions. Contemp. Math. 34, 289–301, 1984.

[195] Giacomo Livan, Marcel Novaes, Pierpaolo Vivo. Introduction to Random Matrices, Theory and Practice. Springer International Publishing, (2018)

[196] Gibbons Michael R. Multivariate tests of financial model. Journal of Financial Economics, 10(1), 3–27, 1982.

[197] Gindikin Simon G. Invariant generalized functions in homogeneous domains. Funkcional. Anal. i Priložen. 9(1), 56–58, 1975.

[198] Gillispie Charles Coulston. Dictionary of scientific biography. vol. XIII, Scribner, New York, 1976.

[199] Girko Viacheslav Leonidovich. Theory of Random Determinants. Kluwer Academic Publishers, 1990.

[200] Girshick Meyer A. On the sampling theory of roots of determinantal equations. Ann. Math. Statistics 10, 203–224 (1939).

[201] Glasserman Paul. Monte Carlo methods in financial engineering, vol. 53 of Applications of Mathematics (New York). Stochastic Modelling and Applied Probability, Springer-Verlag, New York, 2004.

[202] Gohberg Israel, Koltracht Israel. Triangular factors of Cauchy and Vandermonde matrices. Integral Equations and Operator Theory, 26(1), 46–59, 1996.

[203] Gohierre de Longchamps Gaston. de Longchamps Cours de Mathématiques spéciales, Delagrave, Paris, 1883.

[204] Golub Gene H., Van Loan, Charles F. Matrix Computations, Third edition. Johns Hopkins University Press, Baltimore, 1996.

[205] Goodman Nathaniel R. Statistical analysis based on a certain multivariate complex Gaussian distribution. (An introduction). Ann. Math. Statist., 34, 152–177, 1963.

[206] Goos Peter. The Optimal Design of Blocked and Split-Plot Experiments. Number 164 in Lecture Notes In Statistics. Springer Verlag, New York, Inc., 2002.

[207] Goss David. Basic Structures of Function Field Arithmetic. Springer- Verlag Berlin Heidelberg, 1996.

[208] Guest Philip George. The spacing of observations in polynomial regression. The Annals of Mathematical Statistics, 29(1), 294–299, 1958.

[209] Gustafson Robert A., Milne Stephen C. Schur functions, Good's identity, and hypergeometric series well poised in $SU(n)$. Advances in Mathematics s, 48(2) 177–188, 1983.

[210] Günter Siegmund, Günter Adam. Lehrbuch der Determinanten-theorie für Studirende. Besold, Erlangen 1875.

[211] Günttner Rüdiger. Note on the lower estimate of optimal Lebesgue constants. Acta Math. Hung. 65(4), 313–317, 1994.

[212] Günttner Rüdiger. Evaluation of Lebesgue constants. SIAM J. Numer. Anal. 17(4), 512–520, 1980.

[213] Haglund James, Haiman Mark D., Loehr Nicholas A. A combinatorial formula for MacDonald polynomials. J. Amer. Math. Soc. 18, 735–761 2005.

[214] Haglund James, Luoto Kurt W., Mason Sarah Katherine, van Willigenburg Stephanie J. Refinements of the Littlewood–Richardson rule. Trans. Amer. Math. Soc. 363, 1665–1686, 2011.

[215] Haglund James, Luoto Kurt W., Mason Sarah K., van Willigenburg Stephanie J. Quasisymmetric Schur functions. J. Combin. Theory Ser. A, 118, 463–490, 2011.

[216] Hanus Paul Henry. An elementary treatise on the theory of determinants. Ginn & Co., Boston, 1886.

[217] Harish-Chandra. Differential operators on a semisimple Lie algebra. American Journal of Mathematics, 79(1), 87–120, 1957.

[218] Harnad John (Ed.). Random Matrices, Random Processes and Integral Systems. CRM–Series in Mathematical Physics, Springer Science and Business Media, 2011.

[219] Harris Joe. Algebraic Geometry - A first course. Springer-Verlag, 1992.

[220] Harvey Rose E. Linear Algebra. A Pure Mathematical Approach. Springer, 57–60, 2002.

[221] Hawkes Herbert E. Advanced algebra. Gin & Co., Boston, 1905.

[222] Hecht Jacqueline. Un exemple de multidisciplinarité: Alexandre Vandermonde. Population, 26(4), 641–676, 1971.

[223] Heine Eduard. Handbuch der Kugelfunctionen, Theorie und Anwendungen. G. Reimer, Berlin, 1878.

[224] Heineman Ellis Richard. Generalized Vandermonde determinants. Transactions of the American Mathematical Society, 31(3), 464–476, July 1929.

[225] Helgason Sigurdur. Differential geometry, Lie groups, and symmetric spaces, Pure and Applied Mathematics, vol. 80. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York–London, 1978.

[226] Helgert Hermann J. Alternant codes. Information and Control, 26, 369–380, 1974.

[227] Helwig Karl–Heinz. Jordan–Algebren und symmetrische Räume. I. Math. Z. 115, 315–349, 1970.

[228] Henderson Dean, Plaschko Peter. Stochastic differential equations in science and engineering. World Scientific Publishing Co. Pte Ltd, 2006.

[229] Henwood Mervyn R., Ivan Rival. Eponymy in Mathematical nomenclature: what's in a name, and what should be? Math. Intelligencer, 2(4), 204–205, 1980.

[230] Hesthaven Jan S. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. SIAM J. Numer. Anal. 35(2), 655–676, 1998.

[231] Héctor Manuel Moya–Cessa, Francisco Soto–Eguibar. Discrete fractional Fourier transform: Vandermonde approach. arxiv: 1604.06686v1 [math.GM], 2016.

[232] Héctor Manuel Moya–Cessa, Francisco Soto–Eguibar. Diérential Equations: An Operational Approach. Rinton Press, 2011.

[233] Higham Nicholas J. Accuracy and Stability of Numerical Algorithms, 2nd edition. SIAM, Philadelphia, 2002.

[234] Hocquenghem Alexis. Codes correcteurs d'erreurs. Chiffres, 2, 147–156, 1959.

[235] Hoel Paul G. Efficiency problems in polynomial estimation. The Annals of Mathematical Statistics, 29(4), 1134–1145, 1958.

[236] Hoene-Wroński, Józef Maria. Ráfutation de la Théorie des fonctions analytiques. de Lagrange., 1812.

[237] Hoffman Dean Gunnar. Packing problems and inequalities, in Klarner, David A. (ed.), The Mathematical Gardner, Springer, 212–225, 1981.

[238] Holst Anders, Ufnarovski Victor. Matrix Theory. Studentlitteratur AB, Lund, 2014.

[239] Huang Chi–fu., Litzenberger Robert H. Foundations for financial economics. North-Holland Publishing Co., New York, 1988.

[240] Hull John C. Options, Futures and Other Derivatives. Prentice Hall College Div,, New York, 2000.

[241] Hull John C. Options, Futures and Other Derivative Securities. Prentice Hall, Englewood Cliffs, 1993.

[242] Hull John C. Options, Futures and Other Derivatives. Pearson/Prentice Hall, Ninth edition, 2015.

[243] Hsu Pau Lo. On the distribution of roots of certain determinantal equations. Ann. Eugenics 9, 250–258, 1939.

[244] Ibragimov Nail H. CRC Handbook of lie Group Analysis of Differential Equations. CRC Press, Boca Raton, FL. (ed) Vol.1, 1994, Vol. 2, 1995; Vol.3, 199.

[245] Ingersoll Jonathan E. Theory of Financial Decision Making. Blackwell, Oxford, 1997.

[246] Ingham Albert Edward. An integral which occurs in statistics. Math. Proc. Cambridge Philos. Soc. 29(2), 271–276, 1933.

[247] Ira Gessel. Tournaments and Vandermonde's determinant. Journal of Graph Theory, 3(3), 305–307, 1979.

[248] Irving Stoy Reed, Gustave Solomon. Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics, 8(2), 300–304, 1960.

[249] Itykson Claude, Zuber Jean-Bernard. The planar approximation II. Journal of Mathematical Physics, 21(3), 411–421, 1980.

[250] Jabotinsky Eri. Analytic iteration. Transactions of the American Mathematical Society, 108(3), 457–477, 1963.

[251] Jacobi Carl Gustav Jacob. De fonctionibus alternatibus earumque divisione per productum et differentiis elementorum con atum. J. Reine Angew. Mathematik, xxvii, 360–371, 1841.

[252] James Alan T. Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. 35, 475–501, 1964.

[253] James Alan T. The distribution of latent roots of the covariance matrix. The Annals of Mathematical Statistics, 31(1), 151–158, 1960.

[254] Janssen Martin, Krystian Pracz. Correlated random band matrices: localization-delocalization transitions. Phys. Rev. E., 61 (6 Pt A): 6278–86, 2000.

[255] Johnstone Iain M. On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. 29(2), 295–327, 2001.

[256] Jordan Pascual, von Neumann John, Wigner Eugene P. On an algebraic generalization of the quantum mechanical formalism. Ann. of Math. (2) 35(1), 29–64, 1934.

[257] Kalman Dan. The generalized Vandermonde matrix. Mathematics Magazine, 57(1), 15–21, January 1984.

[258] Kaltofen Erich L., Wen–shin Lee, Zhengfeng Yang. Fast estimates of Hankel matrix condition numbers and numeric sparse interpolation. In Proceedings of the 2011 International Workshop on Symbolic-Numeric Computation, 130–136, 2012.

[259] Kan E. An inversion procedure of the generalized Vandermonde matrix. IEEE Transactions on Automatic Control, 16(5), 492–493, 1971.

[260] Kanaka Rajan, Abbott Larry F. Eigenvalue Spectra of Random Matrices for Neural Networks. Physical Review Letters, 97 (18), 188104, 2006.

[261] Karatzas Ioannis. Lectures on the mathematics of finance, vol. 8 of CRM Monograph Series. American Mathematical Society, Providence RI, 1988.

[262] Karatzas Ioannis, Shreve Steven E. Methods of mathematical finance, vol. 39 of Applications of Mathematics (New York). Springer-Verlag, New York, 1998.

[263] Karatzas Ioannis, Shreve Steven E. Brownian motion and stochastic calculus, second ed., vol. 113 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1991.

[264] Karlin Samuel, Studden William J. Tchebycheff system: With appications in Analysis and Statistics. Interscience Publishers, John Wiley & Sons, 1966.

[265] Keating Jon. The Riemann zeta-function and quantum chaology. Proc. Internat. School of Phys. Enrico Fermi., CXIX, 145–185, 1993.

[266] Kemal Melih Şene. Universality in Random Matrix Models of Quantum Chromodynamics. Doctoral dissertation, 91191 State University of New York, 1999.

[267] Khatri Chinubhai G. Classical statistical analysis based on a certain multivariate complex Gaussian distribution. Ann. Math. Statist., 36, 98–114, 1965.

[268] Kiefer Jack, Wolfowitz Jacob. The equivalence of two extremum problems. Canadian Journal of Mathematics, 12, 363–366, 1960.

[269] Kiefer Jack, Wolfowitz Jacob. Optimum designs in regression problems. The Annals of Mathematical Statistics, 30(2), 271–294, 1959.

[270] Kiefer Jack, Wolfowitz Jacob. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. The Annals of Mathematical Statistics, 29(3), 675–699, 1958.

[271] Kijima Masaaki. Stochastic processes with applications to finance. CRC Press, 2013.

[272] Kilgore Theodore A. A characterization of the Lagrange interpolating projection with minimal Tchebycheff norm. Bull. J. Approx. Theory 24(4), 273–288, 1978.

[273] Kilgore Theodore A. Optimization of the norm of the Lagrange interpolation operator. Bull. Am. Math. Soc. 83(5), 1069–1071, 1977.

[274] Kincaid David, Cheney Ward. Numerical Analysis: Mathematics of Scientific Computing. American Mathematical Society, 3rd edition, 2002.

[275] Kitamoto Takuya. On the computation of the determinant of a generalized Vandermonde matrix. In Computer Algebra in Scientific Computing: 16th International Workshop, CASC 2014, 242–255, 2014.

[276] Klein André. Matrix algebraic properties of the Fisher information matrix of stationary processes. Entropy, 16, 2013–2055, 2014.

[277] Klinger Allen. The Vandermonde Matrix. The American Mathematical Society, 74(5),571–574, 1967.

[278] Knobloch Edgar. Déterminants et élimination chez Leibniz, Revue d'Histoire des Sciences, 54(2), 143–164, 2001.

[279] Knuth Donald Ervin. Convolution polynomials. Mathema Journal, 4, 67–78, 1992.

[280] Konishi Sadanori, Kitagawa Genshiro. Information Criteria and Statistical Modeling. Springer, 2008.

[281] König Wolfgang. Orthogonal polynomial ensembles in probability theory. Probability Surveys, 2, 385–447, 2005.

[282] Kowalewski Gerhard. Determinantentheorie, de Gruyter, Berlin, 1942.

[283] Kronecker Leopold. Vorlesungen über die Theorie der Determinanten Teubner, Leipzig, 1903.

[284] Kronecker Leopold. Vorlesungen über Zahlentheorie Vol. 1, Teubner, Leipzig, 1901.

[285] Lagrange Joseph Louis. Réflexions sur la résolution algébrique des équations. In Œuvres de Lagrange, J. A. Serret, 3, 205–421, 1869.

[286] Lagrange Joseph Louis. Leçon cinquième: sur l'usage des courbes dans la solution des problèmes, in Séances des Écoles Normales recueillies par les sténographes et revues par les professeurs, Reynier, Paris, 1795.

[287] Lamberton Damien, Lapeyre Bernard. Introduction to stochastic calculus applied to finance, 2nd ed. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2008.

[288] Lang Serge. Algebra, Graduate Text in Mathematics, Third Ed. Springer-Verlag, New York, 2002.

[289] Lang Serge. Introduction to linear algebra, Addison-Wesley, Reading, MA, 1970.

[290] Laplace Pierre-Simon. Pierre-Simon Laplace philosophical essay on probabilities: translated from the fifth french edition of 1825 with notes by the translator. Vol. 13. Springer Science & Business Media, 2012.

[291] Lassalle Michel. Algèbre de Jordan et ensemble de Wallach. Invent. Math. 89(2), 375–393, 1987.

[292] Laurence Peter. Quantitative modeling of derivative securities. From theory to practice. Routledge, 2017.

[293] Lebesgue Henri Léon. L'œuvre mathématique de Vandermonde, in Notices d'Histoire des Mathématiques, Université de Genève, 18–39, 1958.

[294] Lebesgue Henri Léon. Sur la représentation trigonométrique approchée des fonctions satisfaisant àune condition de Lipschitz. Bull. Soc. Math. France, 38, 184–210, 1910; also in Oeuvres Scientifiques, v.3, L'Enseignement Math., 363–389, 1972.

[295] Lehmann Erich L., Romano Joseph P. Testing statistical hypotheses, third edn. Springer Texts in Statistics. Springer, New York (2005)

[296] Leibniz Gottfried Wilhelm, VI. Leibniz an de l'Hospital. In Leibniz Gesammelte Werk. A. Asher & Comp., 1850.

[297] Levenberg Kenneth. A method for the solution of certain nonlinear problems in least squares. Quarterly of Applied Mathematics, 2(2), 164–168, 1944.

[298] Lévy Paul E. The arithmetic character of the Wishart distribution. Proc. Cambridge Philos. Soc. 44, 295–297, 1948.

[299] Li Hsuan–Chu, Tan Eng–Tjioe. On a special generalized Vandermonde matrix and its LU factorization. Taiwanese Journal of Mathematics, 12(7), 1651–1666, 2008.

[300] Lipton Alexander. Mathematical methods for foreign exchange. A financial engineer's approach. World Scientific Publishing Co. Inc., River Edge, NJ, 2001.

[301] Ljung Lennart, Pflug Georg Ch., Harro Walk. Applied Stochastic approximation and optimization of random systems. Birkhäuser, Vol. 17, 2012.

[302] Loos Ottmar. Symmetric spaces. II: Compact spaces and classification. W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[303] Louck James D. Theory of Angular Momentum in N–dimensional space. Tech. Rep. LA-2451, Los Alamos Scientific Labaratory, 1960.

[304] Lubinsky Doron S. A Survey of Weighted Polynomial Approximation with Exponential Weights. 1 January, 2007.

[305] Lundengård Karl. Extreme points of the Vandermonde determinant and phenomenological modelling with power exponential functions. (Doctoral dissertation, Mälardalen University), 2019.

[306] Lundengård Karl, Österberg Jonas, Silvestrov Sergei. Extreme points of the Vandermonde determinant on the sphere and some limits involving the generalized Vandermonde determinant. arXiv, eprint arXiv:1312.6193.

[307] Lundengård Karl, Rancic Milica R., Javor Vesna., Silvestrov Sergei On some properties of multi-peaked analytically extended function for approximation of lightning discharge currents. Chapter 10 in Engineering Mathematics I: Electrostatics, Fluid Mechanics, Material Physics and Financial Engineering, Volume 178 of Springer Proceedings in Mathematics & Statistics, 151–176, 2016.

[308] Lundengård Karl, Österberg Jonas, Silvestrov Sergei. Optimization of the determinant of the Vandermonde matrix and related matrices. AIP Conference Proceedings 1637, 627, 2014.

[309] Lupaş Laurenţiu. On the computation of the generalized Vandermonde matrix inverse. IEEE Transactions on automatic control, 20(4), 559–561, 1975.

[310] Lyche Tom, Schumaker Larry L. L-spline wavelets. Wavelet Analysis and Its Applications, 5, 197–212, 1994.

[311] Ma Wen-Xiu. Wronskians: generalized Wronskians and solutions to the Korteweg-de Vries equation, 2004.

[312] MacDonald Ian Grant. Symmetric Functions and Hall polynomials. Oxford University Press, USA, 1995.

[313] MacMahon Percy Alexander. Combinatory analysis. Vol. I, II (bound in one volume). Reprint of An introduction to combinatory analysis (1920) and Combinatory analysis. Vol. I, II (1915, 1916). Dover Publications, Mineola, 2004.

[314] Macon Nathaniel, Spitzbart Abraham. Inverses of Vandermonde matrices. The American Mathematical Monthly, 65(2), 95–100, 1958. Chaos, Solitons and Fractals, 19, 163–170, 2004.

[315] Malyarenko Anatoliy. Invariant random fields on spaces with a group action. pringer Science & Business Media, 2012.

[316] Mansion Paul. Éléments de la théorie des déterminants, avec de nombreux exercices, 1880.

[317] Marčenko Vladimir A., Pastur Leonid A. Distribution of eigenvalues for some sets of random matrices. Ann. Math. Statist., 1(4), 457–483, 1967.

[318] Mark Herbert, Ainsworth Davis. Martingale Representation and All That. Birkhäuser Boston, Boston, MA, 57–68, 2005.

[319] Markov Andrey A. Rasprostranenie predel'nyh teorem ischisleniya veroyatnostej na summu velichin svyazannyh v cep'. Zapiski Akademii Nauk po Fiziko-matematicheskomu otdeleniyu, 25(3), 1908.

[320] Markowitz Harry M. Portfolio selection. The journal of finance, 7(1), 77–91, 1952.

[321] Martínez Jose-Javier, Peña Juan Manuel. Factorization of Cauchy–Vandermonde matrices. Linear Algebra Appl., 284, 229–237, 1998.

[322] Marzo Jordi, Ortega-Cerdà Joaquim. Equidistribution of Fekete points on the sphere. Constructive Approximation, 32(3), 513–521, 2010.

[323] Massam Hélène, Neher Erhard. On transformations and determinants of Wishart variables on symmetric cones. J. Theoret. Probab. 10(4), 867–902, 1997.

[324] Massopust Peter. Interpolation and approximation with splines and fractals. Oxford University Press, 2010.

[325] Matjila D. M. Bounds for Lebesgue Functions for Freud Weights. Journal of Approximation Theory. 79, 385–406, 1994.

[326] Mehta Madan Lal. Random Matrices. Elsevier, 3rd edition, 2004.

[327] Mehta Madan Lal. Random Matrices and the Statistical Theory of Energy Levels. Academic Press, New York, London. 1967.

[328] Melas Viatcheslav B. Functional Approach to Optimal Experimental Design, volume 184 of Lecture Notes in Statistics. Springer Science + Business Media, Inc, 2006.

[329] Merton Robert C. Continuous-time finance. Cambridge, MA: Blackwell., xix, 732 p., 1999.

[330] Merton Robert C. The theory of rational option pricing. Bell J. Econ. Manag. Sci., 4, 141–183, 1973.

[331] Merton Robert C. Optimum consumption and portfolio rules in a continuous time model. Journal of Economics Theory, 3(4), 79–93, 1971.

[332] Mhaskar Hrushikesh N., Saff Edward B. Weighted Analogues of Capacity, Transfinited Diameter and Chebyshev Constant. Springer–Verlag New York, 1992.

[333] Michael Hoskins. Cambridge Illustrated History of Astronomy, p. 159. Cambridge University Press, 1997.

[334] Mikami Yoshio. On the Japanese theory of determinants, Isis, 2(1), 9–36, 1914.

[335] Mills Terence M., Smith Simon J. The Lebesgue constant for Lagrange interpolation on equidistant nodes. Numer. Math. 61(1), 111–115, 1992.

[336] Milne–Thomson, Louis Melville. The Calculus of Finite Differences. MacMillan and Co Limited, London, 1933.

[337] Mirsky Leon. An Introduction To Linear Algebra. Oxford University Press, 1955.

[338] Moh Tzuong Tsieng. Algebra. Series on University Mathematics, 1992.

[339] Molev Alexander I. Littlewood–Richardson polynomials. Journal of Algebra. 321(11), 3450–3468, 2009.

[340] Moore Eliakim Hastings. A two-fold generalization of Fermat's theorem. Bullentin of the Americal Mathematical Society, 2, 189–199, 1896.

[341] Moya-Cessa Héctor Manuel, Soto-Eguibar Francisco. Differential Equations: An Operational Approach. Rinton Press, 2011.

[342] Muhumuza Asaph K., Silvestrov Sergei. Symmetric Group Properties of Extreme Points of Vandermonde Determinant and Schur polynomials. Accepted for publication in: Sergei Silvestrov, Anatoliy Malyalenko, Milica Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2019. Springer Proceedings in Mathematics & Statistics, 2020.

[343] Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. Connections Between the Extreme Points of Vandermonde determinants and minimizing risk measure in financial mathematics. Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), (Eds.), SPAS2019. Algebraic, stochastic and analysis structures for networks, data classification and optimization, Springer Proceedings in Mathematics and Statistics, Springer International Publishing, 2020.

[344] Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. Wishart Distribution on Symmetric Cones. Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), SPAS2019. Algebraic Structures and Applications. SPAS 2019. Springer Proceedings in Mathematics & Statistics, 2020.

[345] Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. Extreme Points of the Vandermonde Determinant and Wishart Ensembles on Symmetric Cones. Accepted for publication in: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Springer International Publishing, 2019.

[346] Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. Properties of the extreme points of the joint eigenvalue probability density function of the random Wishart matrix. In ASMDA2019, 18th Applied Stochastic Models and Data Analysis International Conference, ISAST: International Society for the Advancement of Science and Technology, 559-571, 2019.

[347] Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. The Generalized Vandermonde Interpolation Polynomial Based on Divided Differences, SMTDA2018 Conference Proceedings, ISAST2018, 443–456, 2018.

[348] Muhumuza Asaph K., Lundengård Karl, Österberg Jonas, Silvestrov Sergei, Mango John M., Kakuba Godwin. Extreme points of the Vandermonde determinant on surfaces implicitly determined by a univariate polynomial. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317. Springer, Cham. 19 June 2020.

[349] Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. Optimization of the Wishart Joint Eigenvalue Probability Density Distribution Based on the Vandermonde Determinant. In: Silvestrov S., Malyalenko A., Rančić M., (Eds.), Algebraic Structures and Applications. SPAS 2017. Springer Proceedings in Mathematics & Statistics, vol 317. Springer, Cham. 19 June 2020.

[350] Muhumuza Asaph K., Lundengård Karl, Silvestrov Sergei, Mango John M., Kakuba Godwin. Lie symmetries of the Black-Scholes type equations in financial mathematics, ASMDA2017 Conference Proceedings, ISAST2017, 723–740, 2017.

[351] Muir Dylan R., Thomas Mrsic-Flogel. Eigenspectrum bounds for semirandom matrices with modular and spatial structure for neural networks. Phys. Rev. E., 91(4), 042808, 2015.

[352] Muir Thomas, Metzler William Henry. A Treatise on the Theory of Determinants. Dover Publications Inc., New York, 1966.

[353] Muir Thomas [1933], A treatise on the theory of determinants, Revised and enlarged by William H. Metzler, New York, NY: Dover, 1960.

[354] Muir Thomas. The theory of determinants in the historical order of development, Volume II Mac Millan, London, 1911.

[355] Muir Thomas. The theory of determinants in the historical order of development, Volume I Mac Millan, London, 1906.

[356] Muir Thomas. The Theory of Determinants in the Historical Order of its Development: Part I: Determinants in General, Leibniz (1693) to Cayley (1841). MacMillan and Co. London, 1890.

[357] Muirhead Robb I. Aspect of Multivariate Statistical Theory. Wiley Interscience, 1982.

[358] Mullen Gary L., Panario Daniel. Handbook of Finite Fields. CRC Press, 2013.

[359] Musiela Marek, Rutkowski Marek. Martingale methods in financial modelling, second ed., vol. 36 of Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005.

[360] Nadakuditi Raj Rao, Edelman Alan. Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. IEEE Trans. Signal Process. 56(7, part 1), 2625–2638, 2008.

[361] Natanson Isidor Pavlovich. Constructive Function Theory, Volume 1: Uniform Approximation. Frederick Ungar Publishing Co., Inc., 1964.

[362] Needham Joseph, Wang Ling. Science and Civilisation in China, Volume 3: Mathematics and the Sciences of the Heavens and the Earth. Cambridge University Press, 1959.

[363] Neftci Salih N. Introduction to the mathematics of financial derivatives. 2nd ed. Orlando, FL: Academic Press,xxvii, 527 p., 2000.

[364] Neumann Olaf. From Euler through Vandermonde to Gauss, in R.E. Bradley & C.E. Sandifer eds. Leonhard Euler: life and legacy, Elsevier, Amsterdam, 2007.

[365] Newton Isaac. A Cambridge Alumni Database. University of Cambridge, RY644J.

[366] Newton Isaac. Philosophiæ Naturalis Principia Mathematica, London 1687.

[367] Novak Jonathan I. Topics in Combinatorics and Random Matrix Theory. Kingston, Ontario, Canada, 2009.

[368] Ore Øystein. On a special class of polynomials. Transactions of the American Mathematical Society, 35(3), 559–584, 1933.

[369] Oruç Halil. LU factorization of the Vandermonde matrix and its applications. Applied Mathematics Letters, 20, 982–987, 2007.

[370] Oruç Halil, Hakan K. Akmaz. Symmetric functions and the Vandermonde matrix. Journal of computational and applied mathematics, 172(1) 49–64, 2004.

[371] Oruç Halil, Phillips George M. Explicit factorization of the Vandermonde matrix. Linear Algebra and its Applications, 315, 113–123, 2000.

[372] Parlett Beresford N. The Symmetric Eigenvalue Problem (Vol. 20). The Society for Industrial and Applied Mathematics, (SIAM), 1998.

[373] Pascal Ernesto. Determinanti, teoria ed applicazioni, Hoepli, Milano, 1897.

[374] Pastur Leonid A. A simple approach to the global regime of Gaussian ensembles of random matrices. Ukrainian Math. J., 57(6), 936–966, 2005.

[375] Pastur Leonid A., Shcherbina Mariya. Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. Journal of Statistical Physics, 86(1–2), 109–147, 1997.

[376] Peano Giuseppe. Sur le déterminant Wronskien. Mathesis, 9, 75–76, 1889.

[377] Peano Giuseppe. Sur le Wronskiens. Mathesis, 9, 110–112, 1889.

[378] Pearson Kelly Jeanne. On the criterion that a certain given system of deviations from the probable in the case of correlated system of variables is such that it can reasonably supposed to have arisen from random sampling. Pil. Mg., 50(302), 157–175, 1900.

[379] Philips George M. Interpolation and approximation by polynomials. Springer, New York, 2003.

[380] Pliska Stanley R. Introduction to the mathematics of financial derivatives discrete models. Wiley, 1997.

[381] Pólya George, Szegő Gabor. Problems and theorems in analysis. Springer, New York, 1945.

[382] Powell Michael J. D. On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria. Comput. J., 9, 404–407, 1967.

[383] Press William H., Teukolsky Saul A., Vetterling William T., Flannery Brian P. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, 3rd edition, 2007.

[384] Przytycki Józef Henryk. History of the knot theory from Vandermonde to Jones, in XXIV-th National Congress of the Mexican Mathematical Society, Mexico City, 173–185, 1991.

[385] Qu Chong Kai, Wong Roderick. Szegö's conjecture on Lebesgue constants for Legendre series. Pacific J. Math. 135, 157–188, 1988.

[386] Quinn Jennifer J. Visualizing Vandermonde's determinant through nonintersecting lattice paths. Journal of Statistical Planning and Inference, 140(8), 2346–2350, 2010.

[387] Rack Heinz–Joachim. An example of optimal nodes for interpolation revisited. In: Advances in Applied Mathematics and Approximation Theory. Springer Proc. Math. Stat., Springer, New York, 41, 117–120, 2013.

[388] Rack Heinz–Joachim. An example of optimal nodes for interpolation. Int. J. Math. Educ. Sci. Technol. 15(3), 355–357, 1984.

[389] Ramshaw Lyle. Blossoms are polar forms. Computer Aided Geometric Design, 6(4), 323–358, 1989.

[390] Ramshaw Lyle. Blossoming: A connect-the-dots approach to splines. resreport 19, Digital Systems Research Center, 1987.

[391] Rao Kamisetti Ramamohan, Kim Do Nyeon, Hwang Jae–Jong. Fast Fourier Transform–Algorithms and Applications. Springer, 2011.

[392] Rao Raj N., Mingo James A., Speicher Roland, Edelman Alan. Statistical eigen–inference from large Wishart matrices. Ann. Statist. 36(6), 2850–2885, 2008.

[393] Ratnarajah Tharmalingam, Vaillancourt Rémi, Alvo Mayer. Complex random matrices and the capacity of a Rician channel. Problemy Peredachi Informatsii 41(1), 3–27, 2005.

[394] Ray Chandra Bose, Dwijendra Kumar Ray-Chaudhuri. On a class of error correcting binary group codes. Information and Control, 1960.

[395] Reed Irving S., Gustave Solomon. Polynomial codes over cer-tain finite fields. Journal of the Society for Industrial and Applied Mathematics, 8(2), 300–304, 1960.

[396] Richard Turner L. Inverse of the Vandermonde matrix with applications. Technical report, National Aeronautics and Space Administration, Lewis Research Center, Cleveland Ohio, 1966.

[397] Rivlin Theodore J. The Lebesgue constants for polynomial interpolation, Functional Analysis and Its Application. Int. Conf.; Madras, 1973, H. G. Garnir et al., eds., Springer-Verlag, Berlin, 1974.

[398] Rossi Hugo E., Vergne Michèle. Analytic continuation of the holomorphic discrete series of a semi-simple Lie group. Acta Math. 136(1-2), 1–59, 1976.

[399] Rouge Richard, Nicole El Karoui. Pricing via utility maximization and entropy. Mathematical Finance, 10(2), 259–276, 2000.

[400] Rowan Hamilton William. The Mathematical Papers of Sir William Rowan Hamilton. Vol. 3. CUP Archive, 1931.

[401] Roy Samarendra N. p-Statistics or some generalisations in analysis of variance appropriate to multivariate problems. Sankhya 4(3), 381–396, 1939.

[402] Rubin Stanley G., Khosla Prem K. Polynomial interpolation methods for viscous flow calculation. Journal of Computational Physics, 24(3), 217–244, 1977.

[403] Rubinstein Abraham, Romero Carlos, Paolone Mario, Rachidi Farhad, Rubinstein Marcos, Zweiacker Pierre, Daout Bertrand. Lightning measurement station on mount Säntis in Switzerland. In Proceedings of X International Symposium on Lightning Protection, Curitiba, Brazil, 463–468, 2009.

[404] Rudin Walter. Real and Complex Analysis. WCB/McGraw–Hill Book Company, 3rd edition, 1987.

[405] Russel Nicholas J., Chakhmakhchyan Levon, O'Brien Jeremy L., Laing Anthony. Direct dialling of Haar random unitary matrices. New Journal of Physics, 9(3), 033007, 2017.

[406] Ruszczyński Andrzej. Nonlinear Optimization. Princeton University Press, 2006.

[407] Rychkov Vyacheslav S., Borlenghi Studio, Jaffres Henri, Fert Albert, Waintal Xavier. Spin torque and waviness in magnetic multilayers: a bridge between Valet-Fert theory and quantum approaches. Phys. Rev. Lett., 103 (6), 066602, 2009.

[408] Saff Edward B. Logarithimic Potential Theory with Applications to Approximation Theory. Survey in Approximation Theiory, 5, 165–200, 2010.

[409] Saff Edward B., Totik Vilmos. Logarithmic Potentials with External Fields. Springer-Verlag Berlin Heidelberg, 1997.

[410] Sagan Bruce E. The Symmetric Group Representations, Combinatorial Algorithms, and Symmetric Functions, (Graduate Texts in Mathematics). Springer 2001.

[411] Salmon George. Lessons introductory to the modern higher algebra, Hodges, Smith & Co., Dublin, 1859.

[412] Santos David A. Probability: an introduction, Jones & Bartlett, Sudbury MA, 2011.

[413] Sánchez David., Büttiker Markus. Magnetic-field asymmetry of nonlinear mesoscopic transport. Phys. Rev. Lett., 93(10), 106802, 2004.

[414] Schlickewei Hans P., Viola Carlo. Generalized Vandermonde determinants. Acta Arithmetica, XCV(2), 123–137, 2000.

[415] Schoenberg Isaac J. Contributions to the problem of approximation of equidistant data by analytic functions. part B: On the problem of osculatory interpolation, a second class of analytic approximation formulae. Quart. Appl. Math., 4, 112–141, 1964.

[416] Schumaker Larry L. Spline Functions: Basic Theory. Cambridge University Press, 3rd edition, 2007.

[417] Schur Issai. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. Matematische Zeitschrift, 1(4), 377–402, 1918.

[418] Schur Issai. Ueber eine klasse von matrizen, die sich einer gegebenen matrix zuordnen lassen... Friedrich Wilhelms Universität, Berlin., Springer 1901.

[419] Schwarz Gideon. Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464, 1978.

[420] Selberg Atle. Bemerkninger om et multipelt integral. Norsk Mat. Tidsskr. 26, 71–78, 1944.

[421] Serfaty Sylvia. Systems of points with Coulombian interactions. European Mathematical Society Newsletter, 12(110), 16–21, 2018.

[422] Sharp Keith P. Stochastic differential equations in finance. Applied mathematics and Computation, 38, 207–413, 1990.

[423] Sharp William F. Capital Asset Prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, XIX(3) 425–442, 1990.

[424] Shreve Steven E. Stochastic calculus for finance. II. The binomial asset pricing model. I. Springer Finance. Springer-Verlag, New York, 2004.

[425] Shreve Steven E. Stochastic calculus for finance. I. Continuous-time models. Springer Finance. Springer-Verlag, New York, 2004.

[426] Shub Michael, Smale Steve. Complexity of Bezout's theorem, III: condition number and packing Journal of Complexity, 9, 4–14, 1993.

[427] Siegel Carl Ludwig. Über die analytische Theorie der quadratischen Formen. Ann. of Math. (2) 36(3), 527–606, 1935.

[428] Smale Steve. Mathematical problems for the next century. The Mathematical Intelligencer, 20(2), 7–15, 1998.

[429] Smale Steve. On the efficiency of algorithms of analysis. Bulletin (New Series) of The American Mathematical Society, 13(2), 87–121, 1985.

[430] Smith David Eugene. Leibniz on determinants. In A Source Book in Mathematics, volume 1. Dover Publications Inc., New York, 1959.

[431] Smith Jonathan D. H. In defense of eponymy, Math Intelligencer, 3(2), 89–90, 1980.

[432] Smith Kirstine. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of the observations. Biometrika, 12(1/2), 1–85, 1918.

[433] Smith Simon J. Lebesgue constants in polynomial interpolation. Annales Mathematicae et Informaticae 33 (2006), 109–123, 2005.

[434] Sobczyk Garrett. Generalized Vandermonde determinants and applications. Aportaciones Matematicas, 30, 41–53, 2002.

[435] Sompolinsky Haim, Andrea Crisanti, Hans-Jurgen Sommers. Chaos in Random Neural Networks. Physical Review Letters, 61(3), 259–262, 1988.

[436] Spindler Karlheinz. Abstract Algebra with Applications: Volume 1: Vector Spaces and Groups. CRC Press 1993.

[437] Spivey Michael Z., Zimmer Andrew M. Symmetric Polynomials, Pascal matrices, and Stirling matrices. Linear Algebra Appl., 428(4), 1127–1134, 2008.

[438] Spottiswoode William. Elementary theorems relating to determinants, London, 1851.

[439] Staff Investopedia. Portfolio, Investopedia. Archived from the original on 2018-04-20, Retrieved 2018-04-19, 2003-11-25.

[440] Stanley Richard P. Enumerative combinatorics, Vol.2. Cambridge University Press, Cambridge, MA, 1999.

[441] Stanley Richard P. Ordered structures and partitions. CMem. Amer. Math. Soc. 119, 1972.

[442] Steele John M. Stochastic calculus and financial applications, vol. 45 of Applications of Mathematics (New York). Springer-Verlag, New York, 2001.

[443] Stieltjes Thomas Joannes. Sur certains polynômes qui vériént une équation différentielle linéaire du second ordre et sur la theorie des fonctions de Lamé. Acta Mathematics, 6, 321–326, 1885.

[444] Stein Charles. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Stanford University, United States, 1956.

[445] Stigler Stephen Mack. Statistics on the table: the history of statistical concepts and methods, Harvard University Press, Cambridge MA, 1999.

[446] Student. The probable error of a mean. Biometrika 6(1), 1–25, 1908.

[447] Suárez Fernández Antonio, Gascó Luis Gonzaga. Lecciones de combinatoria con las determinantes y sus principales aplicaciones Alufre. Valencia, 1882.

[448] Sullivan Charles R. The first chair of political economy in France: Alexandre Vandermonde and the Principles of Sir James Steuart at the Ecole normale of the year iii, Franch Historical Studies, 20(4), 635–664, 1997.

[449] Sündermann Burkhard. Lebesgue constants in Lagrangian interpolation at the Fekete points. Mitt. Math. Ges. Hamb. 11(2), 204–211, 1983.

[450] Sündermann Burkhard. On projection constants of polynomial space on the unit ball in several variables. Math. Z. 188(1), 111–117, 1984.

[451] Sylvester James Joseph. On derivation of coexistence: Part I. Being the theory of simultaneous simple homogeneous equations, Philosophical Magazine, xvi(1840), 37–43, 1840.

[452] Sylvester James Joseph. Additions to the articles in the September number of this journal, “on a new class of theorems,” and on Pascal’s theorem. Philosophical Magazine Series 3, 37(251), 363–370, 1850.

[453] Szabados József, Vértesi Péter. Interpolation of Functions. World Scientific, Teaneck, 1990.

[454] Szegő Gábor. Orthogonal Polynomials. Colloquium Publications. XXIII. American Mathematical Society. 1939.

[455] Szegő Gábor. Beitrage zur Theorie der Toeplitzschen Formen. Math. Z. 6, 167–202, 1920.

[456] Tang Wei–Pai, Golub Gene Howard. The block decomposition of a Vandermonde matrix and its applications. BIT, 21, 505–517, 1981.

[457] Taylor Mark A., Wingate Beth A., Rachel Vincent E. An Algorithm for Computing Fekete Points in the Triangle. SIAM Journal of Numerical Analysis 38(5), 1707–1720, 2000.

[458] Téllez Gabriel, Forrester Peter J. Exact finite-size study of the 2D OCP at $\Gamma = 4$ and $\Gamma = 6$. Journal of Statistical Physics, 97(3), 489–521, 1999.

[459] Timme Marc, Wolf Fred, Geisel Theo. Topological Speed Limits to Network Synchronization. Physical Review Letters, 92(7), 074101, 2004.

[460] Traub Joseph F. Associated polynomials and uniform methods for the solution of linear problems. SIAM Review, 8(3), 277–301, 1966.

[461] Tropp Joel A. User-Friendly Tail Bounds for Sums of Random Matrices. Foundations of Computational Mathematics, 12(4), 389–434, 2011.

[462] Trudi Nicola. Teoria de' determinanti e loro applicazioni, Pellerano, Napoli, 1862.

[463] Turnbull Herbert Westren, Aitken Alexander Craig. An Introduction to the Theory of Canonical Matrices. Dover Publications, Inc., 1961.

[464] Turnovsky Stephen. Optimal stabilization policies for stochastic linear systems: The case of correlated multiplicative and additive disturbances. Review of Economic Studies, 43(1), 191–194, 1976.

[465] Turnovsky Stephen. The stability properties of optimal economic policies. American Economic Review, JSTOR 1814888, 64(1), 136–148, 1974.

[466] Vandermonde Alexandre–Théophile. Mémoire sur l'élimination. Histoire de l'Académie royale des sciences avec les mémoires de 206 mathématiques et de physique pour la même année tirés des registres de cette académie. Année MDCCLXXII Seconde Partie, 516–532, 1776.

[467] Vandermonde Alexandre–Théophile. Mémoire sur des irrationnelles de differents ordres avec une application au cercle. Histoire de l'Académie royale des sciences avec les mémoires de mathématiques et de physique pour la même année tirés des registres de cette académie. Année MDCCLXXII Premiére Partie, 489–498, 1775.

[468] Vandermonde Alexandre–Théophile. Remarques sur des problémes de situation. Histoire de l'Académie royale des sciences avec les mémoires de mathématiques et de physique pour la même année tirés des registres de cette académie. Année MDCCLXXI, 566–574, 1774.

[469] Vandermonde Alexandre–Théophile. Mémoire sur la résolution des équations. Histoire de l'Académie royale des sciences avec les mémoires de mathématiques et de physique pour la même année tirés des registres de cette académie. Année MDCCLXXI, 365–416, 1774.

[470] Vandermonde Alexandre-Théophile. Mémoire sur l'élimination, in Histoire de l'Académie royale des sciences avec les mémoires de mathématiques et de physique pour la même année tirés des registres de cette académie. Année mdcclxxi seconde partie, Paris 1776, 516–532, 1770.

[471] Van der Vaart A. W. Asymptotic statistics (Vol. 3). Cambridge university press, 2000.

[472] Vein Robert, Dale Paul. Determinants and Their Applications in Mathematical Physics. Springer-Verlag New York, 1999.

[473] Verbaarschot J. M. Jacobas , Wettig Tilo. Random Matrix Theory and Chiral Symmetry in QCD. Annu. Rev. Nucl. Part. Sci., 50, 343–410, 2000.

[474] Verde–Star Luis. Divided differences and combinatorial identities. Studies Appl. Math., 85, 215–242, 1991.

[475] Vértesi Péter. Optimal Lebesgue constant for Lagrange interpolation. SIAM J. Numer. Anal. 27(5), 1322-1331, 1990.

[476] Vértesi Péter. On the optimal Lebesgue constants for polynomial interpolation. Acta Math. Hung. 47(1-2), 165–178, 1986.

[477] Viazovska Maryna S. The sphere packing problem in dimension 8. Annals of Mathematics, 185(2), 991–1015, 2017.

[478] von Neumann John, Traub Charles H. The Collected Works of John von Neumann: 6-Volume Set. Reader's Digest Young Families, 1963.

[479] Von Neumann John, Herman Goldstine H. Numerical inverting of matrices of high order. Bull. Amer. Math. Soc., 53(11), 1021–1099, 1947.

[480] Wainrib Gilles, Jonathan Touboul. Topological and Dynamical Complexity of Random Neural Networks. Physical Review Letters, 110(11), 118101, 2013.

[481] Wald Abraham. On the efficient design of statistical investigations. The Annals of Mathematical Statistics, 14(2), 134-140, 1943.

[482] Wallach Nolan R. The analytic continuation of the discrete series. I, II. Trans. Amer. Math. Soc. 251(1–17), 19–37, 1979.

[483] Walter Ledermann, Neumann Peter M. The Life of Issai Schur through Letters and Other Documents, in Studies in Memory of Issai Schur: Edited by: Anthony Joseph, Anna Melnikov, Rudolf Rentschler 2003.

[484] Waring Edward. Problems concerning interpolations. Philosophical Transactions of the Royal Society of London, 69, 59–67, 1779.

[485] Watson George Neville. The constants of Landau and Lebesgue. Quart. J. Math. 1, 310–318, 1930.

[486] Watson George Neville. Asymptotic expansions of Hypergeometric series. Proc. London Math. Soc. 4, 4–9, 1919.

[487] Wedderburn Joseph Henry Maclagan. Lectures on matrices. AMS Colloquium publications, vol. 17, 1934.

[488] Weld Laenas Gifford. Determinants. Wiley, New York, 1906.

[489] White Michael, 1959-2018. (20 Benthruary 2012). Isaac Newton: the last sorcerer (Paperback ed.) London. p. 46, 2012.

[490] Whittacker Edmund T., Robinson George. The calculus of observations. Old Bailey, London 1924.

[491] Widom Harold. Asymptotic behavior of Block Toeplitz matrices and determinants II. Adv. Math. 21, 1–29, 1976.

[492] Wigner Eugene Paul. Random Matrices in Physics. SIAM Review, 9(1), 1–23, 1967.

[493] Wigner Eugene Paul. Characteristic vectors of bordered matrices with infinite dimensions. Annals of Mathematics, 62(3), 548–564, 1955.

[494] Wigner Eugene Paul. On the statistical distribution of the widths and spacings of nuclear resonance levels. Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press, 47(4), 1951.

[495] Wilmott Paul, Howison Sam, Dewynne Jeff. The mathematics of financial derivatives: a student introduction. Cambridge University Press, Cambridge, 1995.

[496] Wishart John. The generalised product moment distribution in samples from a normal multivariate population. Biometrika, 20A(1/2), 32-52, 1928.

[497] Wolsson Kenneth. A condition equivalent to linear dependence for functions with vanishing Wronskian. Linear Algebra and its Applications, 116, 1–8, 1989.

[498] Wolsson Kenneth. Linear dependence of a function set of m variables with vanishing generalized Wronskians. Linear Algebra and its Applications, 117, 73–80, 1989.

[499] Xambö–Descamps Sebastian. Block Error-Correcting Codes. Springer–Verlag Berlin Heidelberg, 1st edition, 2003.

[500] Yang Yongzhi. Generalized Leibniz functional matrices and factorization of some well-known matrices. Linear Algebra Appl., 430(1), 511–531, 2009.

[501] Yang Sheng–Liang, Hong You. On a connection between the Pascal, Stirling and Vandermonde matrices. Discrete Appl. Math., 155(2), 2025–2030, 2007.

[502] Yang Sheng–Liang. On the LU factorization of the Vandermonde matrix. Discrete Appl. Math., 146(2), 102–105, 2005.

[503] Yang Sheng Liang, Qiao Zhan Ke. Stirling matrix and its property. Int. J. Appl. Math., 14(2), 145–157, 2003.

[504] Yang Shang-Jun, Wu Hua-Zhang, Zhang Quan-Bing. Generalization of Vandermonde determinants. Linear Algebra and its Applications, 336, 201–204, 2001.

[505] Ycart Bernard. A case of mathematical eponymy: the Vandermonde determinant. Revue d'Histoire des Mathématiques, 9(1), 43–77, 2013.

[506] Zamarashkin Nikolai Leonidovich, Tyrtyshnikov Eugene Evgenévich. Distribution of Eigenvalues and singular numbers of Toeplitz matrices under weakened requirements of generating function. Mat. Sb. 188, 83–92, 1997.

[507] Zanon N., Pichard J.-L. Random matrix theory and universal statistics for disordered quantum conductors with spin-dependent hopping. Journal de Physique, 49(6), 907–920, 1988.

[508] Zhu You Lan, Wu Xiao Nan, Chern I–Liang. Derivative securities and difference methods. Springer Finance. Springer-Verlag, New York, 2004.

[509] Ziegler Klaus. Random Matrix Approach to Light Scattering on Complex Particles. In The Fifth International Kharkov Symposium, Ukraine, June 21–26 2004, on Physics and Engineering of Microwaves, Millimeter, and Submillimeter Waves (IEEE Cat. No. 04EX828), 1, 208–210, 2004.

[510] Ziliak Stephen T. How large are your *G*-values? Try Gosset's Guinnessometrics when a little "*p*" is not enough. Amer. Statist. 73(1), 281–290 (2019).

[511] Zumbühl Dominik M., Miller John Boris, Marcus Charles M., Campman K., Gossard A. C. Spin–orit coupling, antilocalization, and parallel magnetic fields in quantum dots. Phys. Rev. Lett., 89(27) 276803, 2002.